

Ana Cláudia Almeida Moreno

Diagnóstico Molecular na era da sequenciação de 3^a geração e da PCR digital

Universidade Fernando Pessoa

Porto, 2013

Ana Cláudia Almeida Moreno

Diagnóstico Molecular na era da sequenciação de 3^a geração e da PCR digital

Universidade Fernando Pessoa

Porto, 2013

© 2013
Ana Cláudia Almeida Moreno
TODOS OS DIREITOS RESERVADOS

Ana Cláudia Almeida Moreno

Diagnóstico Molecular na era da sequenciação de 3ª geração e da PCR digital

Ana Cláudia Almeida Moreno

Trabalho apresentado à Universidade Fernando Pessoa como parte dos requisitos para a obtenção do grau de Mestre em Ciências Farmacêuticas, sob a orientação do Prof. Doutor José Manuel Cabeda

Resumo

O campo científico do diagnóstico molecular vive atualmente tempos áureos de desenvolvimento. A busca de tecnologias revolucionárias capazes de fornecer informação de um modo rápido, preciso e a baixo custo, nunca foi tão intensa.

A descoberta do DNA e a necessidade urgente do seu estudo, impulsionaram o desenvolvimento de novas técnicas de análise, como a reação em cadeia da polimerase (PCR) e a sequenciação. Desde então, estes procedimentos não pararam de evoluir, constituindo hoje um pilar essencial para a investigação e prática clínica.

A PCR, uma técnica de biologia molecular capaz de amplificar sequencias específicas de moléculas de DNA, é atualmente um método indispensável e rotineiro na investigação médica e científica. Desde a primeira técnica de PCR à PCR digital, investigadores em todo o mundo procuraram desenvolver ferramentas cada vez mais poderosas e eficazes, capazes de serem utilizadas em todos os campos científicos.

A demanda por um método que permitisse desvendar a informação codificada no genoma humano impulsionou a evolução das técnicas de sequenciação. A sequenciação pelo método dideoxy, desenvolvido por Sanger, foi a base para a atual sequenciação de terceira geração. O aparecimento destas novas técnicas proporcionou uma inexorável evolução tecnológica marcada por um elevado rendimento num curto espaço de tempo. Análises genómicas que anteriormente eram luxos inalcançáveis, tornaram-se progressivamente mais acessíveis pela diminuição significativa do custo por análise.

No futuro próximo, os diagnósticos moleculares continuarão a desempenhar um papel crítico na qualidade da saúde da população mundial. Inúmeras técnicas de testes moleculares permitem a deteção e caracterização das mais variadas doenças, assim como a monitorização da sua terapêutica.

É importante perceber como cada uma destas técnicas funciona e qual a mais indicada para cada tipo de laboratório. Neste trabalho, procuramos oferecer ao leitor uma visão alargada das vantagens, limitações e aplicações associadas a cada tecnologia atualmente disponível.

Palavras-chave: Diagnóstico Molecular, PCR, PCR digital, sequenciação de última geração, análise de DNA.

Abstract

These are exciting times for scientists in the field of molecular diagnostics. The demand for revolutionary technologies that deliver a fast, inexpensive and accurate genomic information has never been greater. This journey began with the discovery of DNA. The urge to study this vital molecule led to the development of new and revolutionizing techniques, such as the polymerase chain reaction (PCR) and sequencing. Since then, these procedures never stopped evolving and nowadays provide vital information for clinical practice.

The PCR is a scientific technique in molecular biology that is able to amplify a single piece of a particular DNA sequence. It is now a common and often indispensable method used in medical and biological research labs for a variety of applications. Between the first PCR and the digital PCR available nowadays, researchers sought to develop more effective and powerful tools to be used in all scientific fields.

The quest to find a method able to unveil the information behind the human genome pushed the DNA sequencing tools to the next level. The dideoxy terminator sequencing developed by Sanger was the workhorse to develop what we currently call the 3rd generation sequencing. The arrival of these new next generation techniques caused an enormous technological shift marked by dramatic throughput increases, in a less time consuming process. Genomic analysis that were, for most, unreachable luxuries just a few years ago, are being increasingly democratized at a rapid pace due to striking decreases in the cost of each run.

In the upcoming years, molecular diagnostics will continue to be of critical importance to worldwide public health. There are innumerable techniques available for molecular-based tests for detection and characterization of disease, as well as monitoring of drug response. It is important to understand how each of these techniques work and which one would be more suitable in each particular laboratory. In this paper we seek to give the reader a good grasp of the advantages and limitations associated to the technological machinery available today.

Keywords: Molecular diagnostics, PCR, digital PCR, NGS sequencing, DNA analysis

Dedicatória

À minha mãe, Dulez

Agradecimentos

Queria agradecer ao Professor Doutor José Manuel Cabeda pela orientação, ajuda, dedicação, esclarecimento, simpatia e disponibilidade ao longo da execução deste trabalho. Todo o seu incentivo e conhecimento científico foram decisivos para a conclusão e qualidade do mesmo.

À minha mãe por todo o carinho e apoio incondicional, não só durante a execução deste trabalho, mas também ao longo de todo o meu percurso académico. Também à minha irmã, a quem agradeço profundamente a paciência e o incentivo.

Agradeço por último à Universidade Fernando Pessoa e ao seu corpo docente pela qualidade de ensino prestado.

ÍNDICE GERAL

SUMÁRIO

ABSTRACT

DEDICATÓRIA

AGRADECIMENTOS

ÍNDICE GERAL

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

LISTA DE ABREVIATURAS

1. Introdução	1
1.1. Diagnóstico Molecular: Uma perspectiva histórica	1
2. Diagnóstico baseado no PCR	5
2.1. PCR Convencional	5
2.2. PCR Multiplex	11
2.3. Nested PCR	13
2.4. RT-PCR	14
2.5. PCR em Tempo Real (qPCR)	15
2.6. PCR digital	20
3. Diagnóstico baseado na sequenciação	24
3.1. Sequenciação de 1ª geração	24
3.2. Sequenciação de 2ª geração	27
3.3. Sequenciação de 3ª geração	40
4. O Diagnóstico Molecular	49
4.1. Diagnóstico Genético	49
4.2. Diagnóstico oncológico	52
4.3. Farmacogenómica	54
4.4. Diagnóstico de doenças infecciosas	57
4.5. Metagenómica	61
5. Conclusão	64
6. Bibliografia	66

Índice de Figuras

Figura 1 – Demonstração da variação da temperatura ao longo do ciclo de PCR e ao longo das 3 fases	6
Figura 2 – Esquema do PCR	8
Figura 3 – Sistemas de fluorescência	17
Figura 4 – Detecção do produto do PCR em tempo real	18
Figura 5 – Análise por PCR digital para verificar desequilíbrio alélico	23
Figura 6 – Esquema do progresso de sequenciação <i>shotgun de novo</i>	25
Figura 7 – Esquema do processo de sequenciação do Roche 454.....	29
Figura 8 – O processo de sequenciação do Illumina	32
Figura 9 – Princípios do “ <i>2 base encoding</i> ”	35
Figura 10 – Esboço do <i>chip</i> do <i>IonProton™</i>	37
Figura 11 – Sequenciadores comercializados pela Life Technologies.....	38
Figura 12 – Visualização dos dois chips de sequenciação usados pelo <i>IonProton™ Sequencer</i>	39
Figura 13 – Sequenciação executada pelo HeliScope	43
Figura 14 – Sequenciação SMRT pela DNA polimerase.....	44
Figura 15 – Diferença da marcação fluorescente nas duas gerações de sequenciação.....	45
Figura 16 – Moléculas diferentes provocam alterações na corrente elétrica, proporcionando a sua identificação	46
Figura 17 – Leque de aplicações das plataformas GridION e MinION.....	48

Índice de Tabelas

Tabela 1 – Cronologia do diagnóstico por DNA	2
Tabela 2 – Comparação de plataformas de segunda e terceira geração	37

Lista de Abreviaturas

A – Adenina

AD – Hereditariedade Autossômica Dominante

AMV – *Avian Myeloblastoma virus*

AR – Hereditariedade Autossômica Recessiva

Bp – *Base pairs*

C – Citosina

cDNA – DNA complementar

CML – *Cronic Myeloid Leukemia*

CMV – Citomegalovirus

DNA – Ácido desoxirribonucleico

EBV – Vírus de Epstein-Barr

EV – Enterovírus

FAP – *Familiar adenomatosis polyposis*

FDA – *Food and drug administration*

FET – Transístores de efeito de campo

G – Guanina

Gb – Gigabase

HCB – Vírus da hepatite B

HCV – Vírus da hepatite C

HIV – Vírus da imunodeficiência humana

HNPCC – Carcinoma colorretal hereditário não-polipótico

HPV – Vírus do papiloma humano

HSV – Vírus herpes simplex

HTLV – Vírus T-linfotrópico humano

dNTP – desoxinucleotidos

ddNTP – di-desoxinucleótido

dUTP – *Deoxyuridine triphosphate*

Kpb – *kilo base pairs*

M-MuLV – *Moloney murine leukaemia virus*

Mb – Megabases

mRNA – RNA mensageiro

MRSA – *Staphylococcus aureus* resistentes à meticilina

Nm – Nanometro

PA – Picoamperes

PCR – *Polymerase chain reaction*

PGM – *Personal genome machine*

PTP – *Picotiter plate*

qPCR – PCR em tempo real

RNA – Ácido ribonucleico

RT-PCR – *Reverse transcriptase polymerase chain reaction*

RSV – Vírus sincicial respiratório

SBS – Sequenciação por síntese

SIDA – Síndrome da imunodeficiência adquirida

SMRT – *Single molecule real time*

SNP – Polimorfismo de nucleótido único

T – Timina

Tm – Temperatura de fusão

tSMS – *True Single Molecule Sequencing*

U – Uracilo

VZV – Vírus da varicela zóster

ZMW – *Zero Mode Waveguide*

1. Introdução

1.1. Diagnóstico Molecular: Uma perspectiva histórica

O intenso desenvolvimento tecnológico, aliado à compreensão cada vez mais profunda do Genoma Humano, permitiu o rápido desenvolvimento de técnicas cada vez mais rápidas e automatizadas de diagnóstico molecular. A aceleração deste processo foi possível graças aos progressos na capacidade de extração da amostra, na amplificação do alvo, no aumento da sensibilidade dos métodos e dos mecanismos de detecção específica (Bermingham, 2003).

Após a apresentação do modelo estrutural do DNA em 1953 por Watson-Crick, foi possível a elucidação de como o DNA é transcrito e traduzido em proteínas. Deste modo, a criação de protocolos de clonagem, sequenciação e hibridação do DNA permitiu a posterior sequenciação do DNA e o desenvolvimento da técnica de PCR (*Polymerase Chain Reaction*; Tabela 1; Tsongalis & Silverman, 2006).

A *Polymerase Chain Reaction*, ou PCR, foi inventada em 1984 pelo bioquímico americano, Kary Mullis., que recebeu o Prémio Nobel e o Prémio Japonês pelo seu desenvolvimento em 1993. No entanto o princípio básico de replicação de uma peça de DNA usando dois primers já havia sido descrito por Gobind Khorana em 1971, altura em que o progresso neste campo foi limitado devido à dificuldade de síntese de primers e purificação de polimerases. O PCR, utilizado no âmbito da biologia molecular, permite a análise de quantidades infinitesimais de DNA a partir da produção de milhares de milhões de cópias da amostra original. Esta técnica rapidamente se tornou uma das mais usadas em laboratórios por todo o mundo devido à sua simplicidade, rapidez e baixo custo. (Joshi & Deshpande, 2010)

Atualmente, esta técnica possibilita a investigação e o diagnóstico molecular de diversas doenças (Joshi & Deshpande, 2010).

Tabela 1: Cronologia do diagnóstico por DNA (Adaptado de Demidov, 2003)

Data	Principais desenvolvimentos
1953	Descoberta da estrutura em dupla hélice do DNA
1958 - 1959	Descoberta e isolamento de DNA polimerases
1960	Primeiras técnicas de hibridação e detecção eletroquímica de DNA
1965	Síntese enzimática de RNA
1969	Hibridação in situ
1970	Descoberta da transcriptase reversa e das enzimas de restrição
1975	Southern Blot
1977	Métodos para sequenciação de DNA
1983 - 1985	Primeira síntese de olinucleótidos modificados
1985 - 1986	Invenção dos primeiros passos do PCR
1986	Automatização da sequenciação de DNA
1987 - 1988	Descoberta das primeiras sondas de DNA
1990	Aparecimento da nanotecnologia de DNA
1991	Ligase chain reaction
1992	Primeiros métodos para sequenciação do genoma
1992 - 1993	PCR em tempo real
1995	Primeira bactéria (<i>Haemophilus influenzae</i>) sequenciada
1996	Sondas molecular beacon primeiro eucariota é sequenciado
2000	Sequenciação genómica da mosca da fruta (<i>Drosophila melanogaster</i>)
2001	Primeiras versões da sequenciação do genoma humano

Por outro lado, é também utilizado como método padrão em diversos laboratórios de investigação, quer isoladamente quer como método base de outras técnicas moleculares

mais complexas, tornando-se cada vez mais numa tecnologia de uso multidisciplinar (Joshi & Deshpande, 2010).

Nesta perspectiva, e no intuito de alargar o seu espectro de aplicabilidade, foram desenvolvidas outras técnicas derivadas como o RT-PCR (*reverse transcriptase polymerase chain reaction*), que permite avaliar os níveis de RNA, o qPCR, utilizado para a quantificação e genotipagem de DNA em tempo real, o PCR multiplex e o Nested PCR. Ao permitir a deteção e quantificação de mudanças na expressão dos genes, estas técnicas têm providenciado uma maior compreensão dos mecanismos patológicos e servem agora de base para o diagnóstico e investigação científica (Joshi & Deshpande, 2010).

A PCR digital, desenvolvida por Vogelstein e Kinzler, é um exemplo do poder da PCR, fornecendo a oportunidade de um diagnóstico molecular sem precedentes (Pohl e Shih, 2004). Esta tecnologia baseia-se na aplicação de condições ótimas de PCR para a amplificação de uma única molécula de DNA alvo, procedendo-se posteriormente à deteção de sequências específicas obtidas pelo PCR, com o objetivo de contar os alelos obtidos (Pohl e Shih, 2004). Após este processo, várias análises estatísticas podem ser executadas para medir a distribuição dos alelos mutantes. Esta abordagem fornece uma base estatística rigorosa tornando possível uma análise do estado dos alelos em amostras de tecidos e fluidos corporais. Assim, a PCR digital transforma o sinal analógico convencional num sinal linear e digital passível de ser partilhado mundialmente (Pohl & Shih, 2004).

A tecnologia de sequenciação foi descrita pela primeira vez por Sanger em 1977 (Sanger *et al.*, 1977), como um método para a determinação de sequências de

nucleótidos. Esta tecnologia permitiu, ao longo de duas décadas, um grande número de importantes avanços a nível da genética (Shendure & Ji, 2008).

Apesar da utilidade e da evolução desta técnica, as suas limitações tornaram-se cada vez mais evidentes (Shendure & Ji, 2008). Assim, o método clássico de Sanger sofreu diversas modificações e adaptações ao longo dos anos, mais recentemente pela influência do Human Genome Project, o qual impulsionou a automatização e o aumento da rentabilidade da técnica (Moorthie *et al.*, 2011).

Os novos métodos que foram surgindo, foram intitulados como sequenciação de próxima geração ou sequenciação de 2ª e de 3ª geração. A chegada desta última geração de métodos de sequenciação mudou a base da abordagem científica em termos de investigação e clínica, permitindo obter uma enorme quantidade de dados de uma forma bastante económica. A possibilidade de sequenciar o genoma completo de vários organismos impulsionou estudos evolucionários em larga escala, os quais seriam impossíveis de concretizar há apenas alguns anos atrás. A aplicação mais ampla destas técnicas tem sido a investigação sobre o genoma humano, proporcionando um conhecimento cada vez mais vasto de como as diferenças genéticas afetam a saúde e a doença (Metzker, 2010).

2. Diagnóstico baseado no PCR

2.1. PCR Convencional

Poucas tecnologias, na área da ciência, foram tão aclamadas como a PCR. Esta revelou ser uma técnica extremamente útil não só para o desenvolvimento de novas formas de diagnóstico molecular, mas também como a base para o desenvolvimento de inúmeras novas técnicas revolucionárias como a sequenciação e a clonagem (Bermingham, 2003).

A PCR começou a ser desenvolvida na década de 80 por Kary Mullis, o qual recebeu em 1993 o prémio Nobel da química precisamente por esse trabalho. As reações originais eram feitas em tubos contendo os componentes necessários e eram manualmente mudados entre banhos de água maria para se obter as temperaturas adequadas a cada etapa da reação. A enzima que fazia a amplificação, a DNA polimerase da *E. coli* era termossensível o que obrigava à adição de nova enzima no final de cada ciclo. Posteriormente foi desenvolvido equipamento capaz de realizar os ciclos térmicos para automatizar o longo processo manual (Sanders, 2007).

Mais tarde, a descoberta de polimerases termoestáveis revelou ser o maior avanço na tecnologia de automatização do PCR, não havendo a perda de enzima em cada ciclo (Bermingham, 2003).

O PCR apesar de ser uma técnica simples exige uma mão cheia de passos para serem obtidos os resultados pretendidos. Primeiro, a molécula de DNA alvo é submetida a altas temperaturas (90 - 97° C) o que provoca a desnaturação da dupla cadeia (Joshi & Deshpande, 2010). Após a desnaturação, um par de pequenas sequências de DNA sintético, chamadas oligonucleotídeos ou primers, são ligados à molécula de DNA alvo (hibridação, ou hibridização ou annealing). Estes servem como um ponto de partida

para adição de nucleótidos, feita pela DNA polimerase. A cadeia molde original dita a sequência dos nucleótidos que são adicionados, ou seja, há uma complementaridade em relação à sequência de bases que se encontra na molécula de DNA original. *In vitro* todos estes processos podem ser alcançados por alterações cíclicas de temperatura, facilitando a desnaturação do DNA, a hibridação dos primers e a ação da polimerase (Bermingham, 2003).

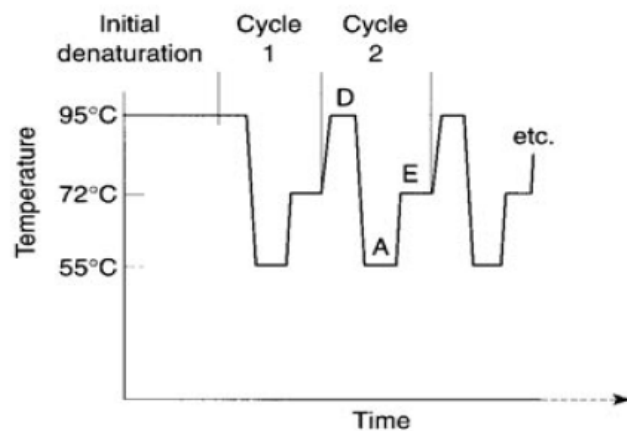


Figura 1: Demonstração da variação da temperatura ao longo do ciclo de PCR e durante as 3 fases: desnaturação (D), *annealing* (A) e extensão (E). (Joshi & Deshpande, 2010)

Mais detalhadamente, primeiro a molécula de DNA alvo é desnaturada, isto é, separada em duas cadeias simples não emparelhadas, pelo aquecimento a 94 - 98° C. A temperatura é depois diminuída até 50 - 72° C para permitir que os primers hibridizem (a temperatura neste passo depende das sequências dos primers). Em seguida, a temperatura aumenta novamente para os 60 - 72° C para permitir que a polimerase estenda os primers, criando uma cópia complementar da cadeia molde de DNA (Figura 1; Joshi & Deshpande, 2010)

O segundo ciclo, começa com novo aquecimento até aos 94 – 98 ° C, esta temperatura desnatura as novas moléculas acabadas de sintetizar. No segundo passo do *annealing*, os *primers* podem-se ligar às moléculas sintetizadas e desnaturadas, permitindo à polimerase atuar até chegar ao fim do primeiro primer. No fim do segundo ciclo novas cadeias foram sintetizadas e obtiveram-se moléculas com o comprimento correto (Figura 2; Bermingham, 2003).

Nos ciclos subsequentes, o número de moléculas sintetizadas ultrapassam amplamente o número de moléculas alvo iniciais e aumentam para o dobro em cada ciclo. Se o PCR fosse 100% eficaz, uma molécula iria originar 2^n moléculas depois de n ciclos. Na prática, 20 a 40 ciclos são normalmente usados, do que resulta uma amplificação média de cerca de $2^{30} \approx 10^9$ vezes o número de moléculas original. (Bermingham, 2003)

Sempre que for possível obter um par de *primers* específicos para uma região de DNA, então essa região pode ser extensamente amplificada por PCR. Na maior parte das vezes, é a sequência e a combinação dos primers que influencia o sucesso do ensaio. Dada a importância dos primers para o sucesso da reação de PCR, o seu desenvolvimento e obtenção deve seguir certas regras. Os primers devem ter um comprimento de 14 a 40 nucleótidos, com um equilíbrio no conteúdo de bases G+C por volta dos 50%, não devem ser complementares uma da outra (para evitar a formação de *primer-dimer*) e não devem formar uma estrutura secundária estável (Bermingham, 2003).

Os *primers* são desenhados de forma a que haja a hibridização em cadeias e zonas opostas da sequência alvo, deste modo são estendidas uma em direção à outra pela adição de nucleótidos a partir da extremidade 3' (Bermingham, 2003).

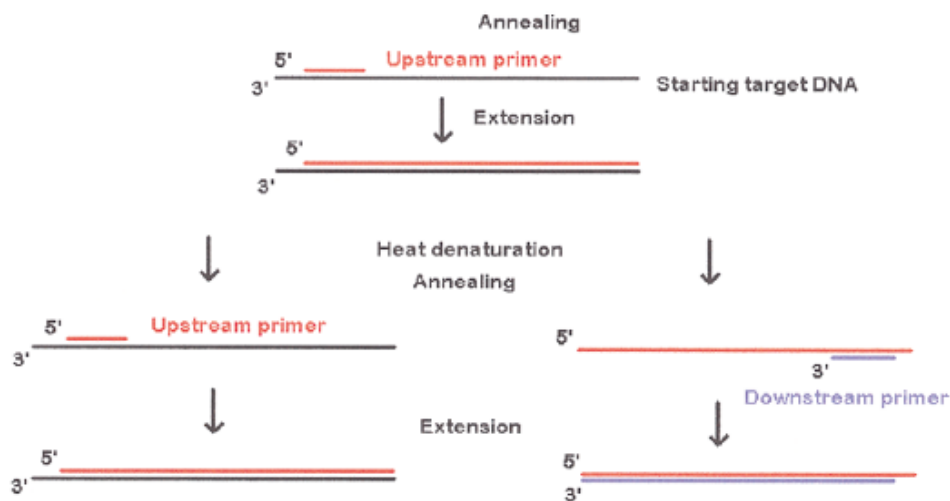


Figura 2 – esquema do PCR (Van Gelder, 2001)

Sequências pequenas são amplificadas mais facilmente, sendo normalmente a distância inferior a 500 pares de base (bp). Contudo, com protocolos especiais foi já possível amplificar fragmentos de >10 kbp de comprimento. Se a sequência a amplificar é conhecida, o desenvolvimento de primers específicos é relativamente fácil de conseguir graças ao extenso número de ferramentas informáticas de elevada qualidade disponíveis para o efeito. Se a sequência por outro lado não for conhecida será mais difícil de sintetizar os primers específicos, no entanto estes ainda podem ser obtidos por meio de primers degenerados (Bermingham, 2003). A *Polymerase Chain Reaction* tem vindo a ser uma grande ajuda na investigação e no diagnóstico de inúmeras doenças. Tornou-se também uma técnica padrão em todos os laboratórios que utilizam ácidos nucleicos nos seus protocolos. Assim, pode-se verificar que esta técnica é usada por um crescente número de cientistas de diversas áreas. (Joshi & Deshpande, 2010)

Consequentemente verificou-se um crescente número de aplicações para esta técnica tão versátil. O procedimento básico do PCR provou ser valioso para o diagnóstico de doenças já que, sequências específicas de DNA podem deste modo ser amplificadas em

grande número. De facto, uma das primeiras aplicações foi o diagnóstico de uma doença genética, a anemia das células falciformes, já que a técnica de PCR necessita apenas de uma pequena amostra, ao contrário das outras técnicas padrão utilizadas. Deve no entanto ter-se em atenção a presença de contaminações, sendo o PCR extremamente sensível, a presença de uma pequena quantidade de contaminantes será fatal para o resultado, verificando-se logo aqui uma grande limitação para esta técnica básica de PCR (Guyer, 2007).

O PCR pode ser usado para amplificar quantidades infinitesimais de material genético de agentes patogénicos que possam estar presentes em amostras de sangue, células, água, alimentos e outras amostras biológicas ou ambientais. Testes baseados em técnicas de PCR podem ser postos em prática para deteção de agentes infecciosos que são extremamente difíceis de cultivar em laboratório, como por exemplo a doença de Lyme (febre da carraça) ou o vírus da SIDA. No diagnóstico e investigação do cancro, o PCR pode detetar a mutação oncogénica ou indicar quais os genes que são expressos ou suprimidos, já que as moléculas de RNA mensageiro associadas a esses genes podem ser convertidos em DNA complementar, o qual pode ser amplificado. (Guyer, 2007)

A PCR é também utilizada em laboratórios forenses, sendo aqui particularmente útil, já que muitas vezes a amostra é muitíssimo pequena, uma gota de sangue ou um único cabelo, possui uma quantidade mínima de DNA original, no entanto até nestes casos é possível obter uma quantidade suficiente de DNA para identificação do suspeito (Joshi & Deshpande, 2010).

Outras aplicações incluem a identificação de sujeitos desaparecidos, testes de paternidade, encontrar dadores de órgãos compatíveis com o doente, restabelecimento

de relações evolucionárias entre organismos, pela extração de DNA de múmias ou ossos (Guyer, 2007).

Apesar de todas as vantagens subjacentes ao uso do PCR, denotam-se determinadas limitações na técnica. Uma das mais importantes limitações refere-se à potencial amplificação de DNA contaminante, originando falsos positivos. (Evans, 2009) A fonte predominante de contaminação deve-se a produtos “carry-over” provenientes das reações anteriores, estes produtos podem ser transmitidos de outras reações por meio de pipetas, tubos, reagentes ou das superfícies do laboratório. De modo a evitar esta contaminação é necessária a implementação de regras estritas dentro do laboratório. Estas contaminações tornam-se mais pronunciadas aquando do uso de primers universais (Yang & Rothman, 2004). Outra fonte de falsos positivos provém de contaminações internas de DNA latente no hospedeiro. Várias espécies virais, como por exemplo os retrovírus, tornam-se parte do genoma do hospedeiro, bastando a deteção de uma pequena quantidade deste para se obter um resultado falso positivo. (Van Gelder, 2001)

A necessidade da realização de uma eletroforese ou de uma hibridação com sondas após a reação de PCR constitui mais uma desvantagem. Devido a esta necessidade, a obtenção dos resultados da análise torna-se mais demorada. Também o facto dos resultados se basearem apenas na discriminação do tamanho das sequências torna a técnica menos precisa (Roche, 2003).

Resumindo, o PCR apresenta diversos problemas incluindo: probabilidade de falsos positivos, sensibilidade por vezes ainda insuficiente, baixa resolução, discriminação baseada apenas no tamanho das sequências, resultados apenas qualitativos, baixo nível de automatização, processamento moroso pós PCR (Roche, 2003).

De modo a superar tais limitações da técnica original foram desenvolvidas técnicas derivadas as quais se tornaram mais vantajosas e fidedignas do que a técnica original (Roche, 2003).

2.2.PCR Multiplex

Chamberlain et. all. descreveram pela primeira vez o PCR Multiplex, demonstrando a sua eficácia e utilidade para análises simultâneas de diferentes *loci* no gene da distrofia muscular (Chamberlain *et al.*, 1988). Desde então, o PCR multiplex tem sido otimizado e utilizado em diversas áreas, na deteção e análise de genes, análise de polimorfismos, mutações, análise quantitativa, no contexto do RT-PCR (*reverse transcriptase PCR*), no diagnóstico de doenças infecciosas para identificação de vírus, bactérias e parasitas (Elnifro *et al.*, 2000).

O PCR multiplex é uma extensão do PCR convencional. A vantagem do seu uso é a capacidade de amplificar mais do que uma sequência alvo simultaneamente, tornando-o mais económico e mais rápido (Ravikumar *et al.*, 2011). Para ser possível a amplificação simultânea de várias sequências basta adicionar mais do que um par de primers na reação (Elnifro *et al.*, 2000).

No entanto, esta técnica possui algumas dificuldades que necessitam ser ultrapassadas. A presença de mais do que um par de primers na reação aumenta a probabilidade de se obterem produtos inespecíficos devido à formação de *primer-dimers* (Elnifro *et al.*, 2000). Este fenómeno acontece devido à proporção de *primer – template*. Se a proporção for muito elevada há formação de *primer dimers*, assim como no caso de haver uma diluição muito grande de *template* e excesso de *primer*. Os *primers* devem estar 10^7 molar em excesso em relação ao *template*, no entanto a concentração do

primer não deve exceder os 0.5 μ M devido à formação de *primer-dimers*. Assim sendo, o que determina a razão *primer – template* é a complexidade do template usado na reação. Deste modo, se esta razão for baixa o produto não se vai acumular exponencialmente, já que as novas moléculas sintetizadas vão renaturar depois da desnaturação (baixa o rendimento) ou vão inibir a formação do produto. Nestas condições os *primer-dimers* serão amplificados com maior eficiência do que o alvo pretendido, consumindo componentes da reação. A otimização da reação do PCR multiplex visa minimizar estas interações não específicas, tomando especial atenção no desenho dos *primers* especialmente no comprimento, conteúdo de Guanina e Citosina e concentração (Markoulatos *et al.*, 2002). Geralmente, os primers devem ser desenhados de modo a que as suas temperaturas de fusão (T_m) sejam próximas dos outros primers usados na reação multiplex. Apesar de ser possível o cálculo das concentrações de primer e das temperaturas de annealing, as condições perfeitas para a reação devem ser definidas empiricamente. Para além dos primers, os outros componentes adicionados à reação (por exemplo: Mg^{2+} , dNTP, polimerase) e as condições cíclicas (por exemplo tempo de extensão, temperatura de annealing e hot start) devem ser também otimizadas. Deste modo, uma otimização completa das condições de reação torna-se muito demorada, o que torna os novos testes de diagnóstico mais caros (Shuber *et al.*, 1995). Apesar de tudo isto o uso do PCR multiplex possui vantagens, nomeadamente a possibilidade de reconhecer e discriminar simultaneamente infeções mistas, ou de caracterizar simultaneamente mais que um fator de resistência a fármacos (Messmer *et al.*, 2000).

O PCR multiplex é também frequentemente utilizado para implementar um controlo interno nos ensaios diagnósticos. O controlo interno, em contraste com um controlo

positivo externo, é uma sequência de DNA presente no mesmo tubo que contem a amostra o qual é amplificada simultaneamente com a sequência alvo. Numa reação onde não seja integrado um controlo interno, um resultado negativo (sem banda ou sinal) pode significar que não existe a sequência alvo na amostra, ou a presença de inibidores. No entanto, quando é usado um controlo interno a ausência de sinal em todos os produtos da reação significa que a reação foi inibida, devido à presença de substâncias inibitórias na amostra. Consequentemente, numa reação que possua um controlo interno, este deve sempre produzir um sinal, mesmo que não exista a sequência alvo do ensaio. Este método é bastante útil, já que pode revelar uma falha na reação de PCR que de outro modo só poderia ser revelada por um controlo de inibição, o qual é muitíssimo mais dispendioso, uma vez que exige a duplicação do número de ensaios (Hoorfar *et al.*, 2004).

2.3.Nested PCR

O Nested PCR é um método usado para aumentar a sensibilidade do PCR. É uma forma de re-amplificação de um produto de PCR. É executado um primeiro PCR o qual usa um par de primers que produzem uma grande quantidade de produto, o qual é depois usado como um molde para a segunda amplificação (Bermingham, 2003). É retirada uma pequena alíquota do produto inicial para uma segunda reação de PCR, onde os primers são desenhados de modo a amplificar uma região de DNA mais pequena, aqui os segundos primers são “*nested*” (Evans, 2009). Obtêm-se assim um produto mais pequeno, comparando com o original. Estima-se que o Nested PCR aumenta a sensibilidade da reação em 10^4 vezes, aquando da deteção do produto pretendido. Mesmo que na primeira amplificação a sequência pretendida esteja pouco visível entre

todo o produto amplificado não específico, é possível identifica-la recorrendo a primers específicos que possibilitam a sua amplificação por meio do Nested PCR. Em contrapartida, todos os produtos não específicos têm muito pouca afinidade para os primers *nested*, logo estes não irão ser amplificados e não irão contaminar o produto obtido na segunda amplificação (Mc Pherson, 2006).

Para evitar contaminações as duas reações de PCR podem ser executadas num único tubo. Os produtos de PCR podem depois ser analisados por eletroforese, esta irá revelar a primeira e segunda amplificação (Mc Pherson, 2006).

Em qualquer experiência que envolva o uso da reação de PCR é importante a execução dos controlos adequados para assegurar a especificidade da PCR. No caso do Nested PCR é ainda mais importante a execução de controlos devido à sua elevada sensibilidade (Mc Pherson, 2006).

2.4.RT-PCR

A *Reverse Transcription-polymerase Chain Reaction* (RT-PCR) permite estudar o RNA via PCR. Este método é altamente sensível e específico para a deteção de cópias de RNA, principalmente quando a quantidade de amostra é limitada (Roche, 2003).

As DNA polimerase termoestáveis usadas nas reações de PCR requerem DNA como substrato, assim o primeiro passo para a análise de RNA é a conversão total deste em cDNA (DNA complementar) por meio de enzimas de transcriptase reversa (Birmingham, 2003). As duas enzimas mais utilizadas são a *Avian myeloblastoma virus* (AMV) e a *Moloney murine leukaemia virus* (M-MuLV). Existem três diferentes mecanismos de ação usados para a síntese do cDNA: um primer oligo-dT para a amplificação específica de mRNA; nucleótidos aleatórios *hexamer* ou *nonamer*; ou um

primer específico para o gene pretendido. Depois de se obter o cDNA este submete-se a uma reação de PCR normal, com primers específicos para o gene de interesse (Evans, 2009).

O método original para esta reação utilizava uma transcriptase reversa mesofílica. No entanto, verificou-se que o uso de altas temperaturas apresentava mais vantagens. Estas vantagens incluem: o aumento da especificidade da ligação do primer. Destabilização da estrutura secundária do RNA alvo permitindo uma síntese completa melhorada, e a capacidade de incluir dUTP e UNG para prevenção do efeito *carryover* (Roche, 2003).

2.5. PCR em Tempo Real (qPCR)

O PCR em tempo real constitui para a comunidade científica uma das mais úteis ferramentas tecnológicas dos últimos tempos. Esta tecnologia deriva da PCR convencional, desenvolvida por Kary Mullis, mas procura ultrapassar algumas das sérias limitações que esta apresenta. Ao proceder à amplificação da sequência de DNA e só depois se analisar o produto, a quantificação tornava-se extremamente difícil. Esta limitação foi ultrapassada em 1992 com o desenvolvimento do PCR em tempo real por Higuchi et al. (Higuchi *et al.*, 1992).

No PCR em tempo real a quantidade de produto formado é analisado ao longo da reação, com a introdução de fluorocromos ou de sondas fluorescentes. A monitorização da fluorescência permite quantificar o produto formado (Kubista *et al.*, 2006).

Assumindo uma determinada eficiência de amplificação, (100% corresponde à obtenção do dobro das moléculas a cada ciclo) é possível determinar o número de moléculas de DNA da sequência amplificada que estavam originalmente presentes na amostra inicial (Kubista *et al.*, 2006).

Mais detalhadamente, na década de 90, investigadores demonstraram que a atividade exonucleolítica 5' da Taq DNA polimerase poderia ser explorada como um método indireto para aceder à quantidade de material genético amplificado com o uso de sondas fluorescentes específicas, eliminando assim a necessidade de execução de uma eletroforese (necessária na PCR convencional). Adicionalmente foi descoberto que o uso de fluorocromos capazes de se ligarem ao DNA poderia proporcionar uma monitorização da reação em tempo real, o PCR cinético. Com estas duas descobertas foi possível entrar na era do PCR em tempo real (Walker, 2007).

Existem métodos indiretos e diretos para monitorizar a fluorescência durante o ciclo de PCR (Figura 3).

Métodos diretos referem-se aos sistemas em que a fluorescência resulta da ligação de uma molécula fluorescente ao produto amplificado ou de uma incorporação direta de uma sonda fluorescente interferente no produto amplificado. Um exemplo de um método direto consiste no uso de um fluorocromo como o *SYBR Green*. Neste caso, a ligação do fluorocromo ao DNA de cadeia dupla origina a fluorescência. Enquanto que a marcação por *SYBR Green* tem a vantagem de ser aplicável a qualquer produto de PCR, eliminando assim a necessidade de desenhar um reagente específico (sonda). No entanto isto pode significar a falta de especificidade ou seja uma fluorescência de um produto não pretendido, originando um erro na quantificação (Walker, 2001).

Os métodos indiretos recorrem a reagentes específicos (sondas) para detetar os produtos amplificados. Estas sondas de hibridização específicas podem ser desenhadas de modo a só emitirem fluorescência quando se encontram ligadas ao gene de interesse (Walker, 2001). Exemplo de um método indireto é o sistema Taqman (Figura 3), um dos sistemas mais extensamente usados (Walker, 2001).

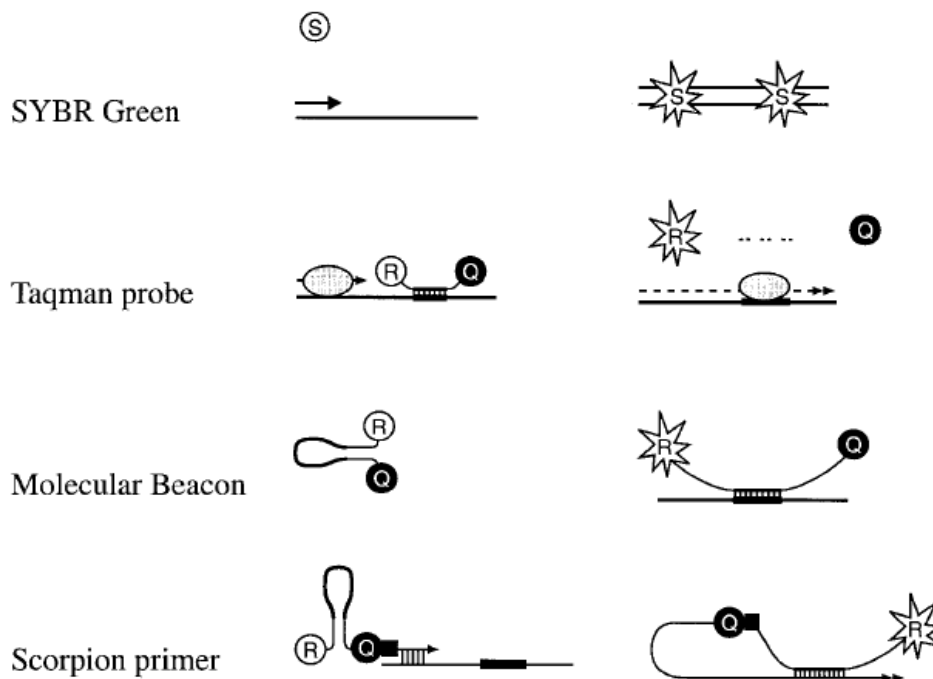


Figura 3: Sistemas de fluorescência. Sistemas que usam um fluorocromo como o SYBR Green (S) ou uma sonda que possui fluorocromos: o repórter (R) e o quencher (Q). A separação do *quencher* da proximidade do *reporter* permite a medição da fluorescência do repórter (Walker, 2001).

A reação de Taqman requer uma sonda de hibridização marcada com dois fluorocromos fluorescentes. Um é o fluorocromo *reporter* na extremidade 5' (por exemplo FAM – 6-carboxy-fluorescein ou VIC; Watzinger *et al.*, 2006) e a outra é o *quencher* na extremidade 3' (por exemplo TAMRA – 6-carboxy-tetramethyl-rhodamine; (Heid *et al.*, 1996, Gibson *et al.*, 1996). Quando a sonda está intacta, a emissão proveniente do fluorocromo *reporter* é suprimida devido à ligação física que existe entre este e o *quencher*. No entanto, durante a fase de extensão no ciclo de PCR, a atividade exonucleolítica da DNA polimerase hidrolisa a sonda de hibridização separando assim o fluorocromo do *quencher*. O aumento relativo da fluorescência é monitorizado em

tempo real durante a fase de amplificação, recorrendo a um detetor, presente no termociclador (Gibson *et al.*, 1996).

As sondas *molecular beacon* (Figura 3) possuem complementaridade entre as suas extremidades, o que lhes confere uma estrutura característica em forma de gancho de cabelo (de onde é derivado o seu nome) a qual mantem os fluorocromos próximos, isto permite a extinção da fluorescência do *reporter* por proximidade física com o *quencher* (Walker, 2001). A linearização da sonda subsequente à ligação é usada para atenuar o efeito do *quencher*. Este tipo de sonda foi utilizada para o desenvolvimento de sistemas de sondas unimoleculares como os primers *Sunset* e *Scorpion*, estes são incorporados na sequencia amplificada durante a reação (Walker, 2001).

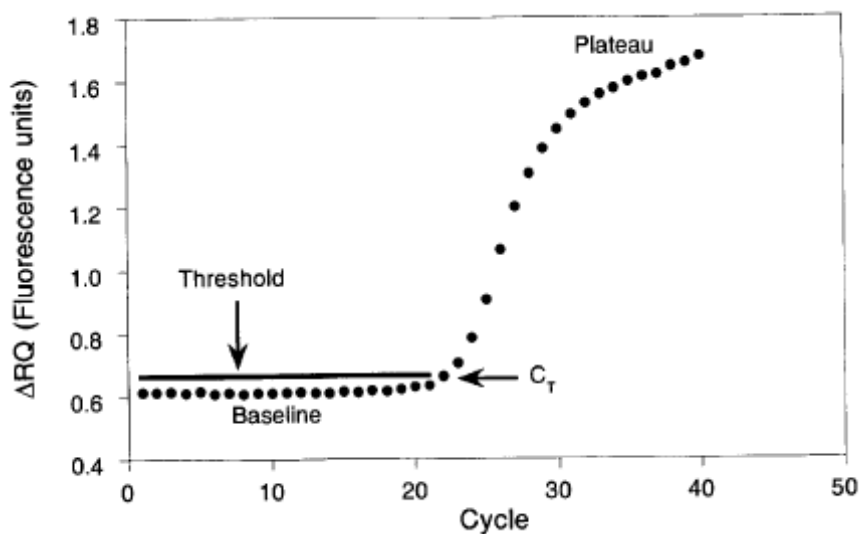


Figura 4: Detecção do produto do PCR em tempo real: o software constrói gráficos com os dados obtidos da fluorescência, na fase de amplificação. (Heid *et al.*, 1996)

O termociclador de qPCR não é mais que uma combinação de um termociclador convencional com uma fonte de luz, detetor de fluorescência e software o qual deteta e quantifica automaticamente os produtos de PCR (figura 4; Gibson *et al.*, 1996).

Um algoritmo computadorizado compara a quantidade de fluorocromo *reporter* (R) com a emissão do *quencher* (Q) a cada ciclo, durante a fase de amplificação do ciclo de PCR, gerando um valor de ΔR_n (R/Q). O valor de ΔR_n reflete a quantidade de sonda de hibridização que foi degradada. O software constrói então uma função exponencial em função dos últimos três valores de ΔR_n , de cada ciclo de extensão. Um limiar de emissão de fluorescência relativa é definido na linha de base do ΔR_n durante os primeiros 10 a 15 ciclos. O algoritmo calcula o ciclo no qual cada amplificação atinge um limiar significativo (C_t). O valor de C_t é proporcional ao número de cópias alvo presentes na amostra, ou seja é uma medida quantitativa do número de cópias alvo que se encontram em cada amostra (Gibson *et al.*, 1996).

Os ensaios feitos por PCR em tempo real são caracterizados por uma grande sensibilidade e precisão. Outra vantagem deste método é o não haver processamento pós PCR, evitando assim possíveis contaminações dos produtos. Também por possuir tempos de resposta rápidos e baixo custo, o PCR em tempo real revolucionou o diagnóstico molecular, por exemplo foram desenvolvidos novos sistemas capazes de detectar microrganismos em menos de 30 minutos. Sondas *Molecular beacon* têm também sido usadas para quantificação de agentes patogênicos por possuírem a vantagem de usarem um quencher que não é fluorescente e por possuírem quatro fluorocromos diferentes os quais podem ser quantificados simultaneamente. No entanto as sondas *molecular beacon* são altamente sensíveis a desajustes na região de ligação, devendo assim ser usados onde existem regiões muito conservadas (Klein, 2002).

Uma das limitações do uso da técnica é a necessidade de já existirem dados referentes ao gene de interesse. Consequentemente, o qPCR só pode ser usado para análise de genes conhecidos. Até recentemente informação sobre a sequência de interesse tem sido

derivada de fragmentos de genes de organismos cultivados *in vitro* ou de clones documentados oficialmente, usando primers baseados em sequências conhecidas. Por isso, estes métodos estão inevitavelmente limitados a análises de sequências relacionadas com outras que já foram anteriormente caracterizadas (Smith & Osborn, 2009).

Resumindo, a introdução do PCR em tempo real no campo do diagnóstico molecular tem vindo a simplificar a quantificação de ácidos nucleicos, gerando enormes quantidades de dados rapidamente. No entanto, é necessário proceder à análise crítica dos dados obtidos. Apensar dos ensaios por PCR em tempo real serem relativamente precisos e reprodutíveis, a exatidão dos dados obtidos depende de vários fatores como a preparação da amostra, qualidade dos ensaios padrão e escolha do gene. Deste modo, para serem obtidos dados fiáveis estes devem ser comparados com outros ensaios consagrados. Este é o maior desafio no diagnóstico molecular, garantir dados fidedignos no futuro (Klein, 2002).

2.6. PCR digital

O PCR digital é um exemplo do poder do PCR, fornecendo a oportunidade de um diagnóstico molecular sem precedentes. (Pohl & Shih Ie, 2004)

Esta técnica foi desenvolvida por Vogelstein e Kinzler na tentativa de encontrar mais aplicações para o PCR convencional. Esta tecnologia baseia-se na aplicação de condições ótimas de PCR para a amplificação de uma única molécula de DNA alvo, depois da amplificação procede-se à deteção de sequências específicas obtidas pelo PCR, com o objetivo de contar os alelos obtidos. (Pohl & Shih Ie, 2004)

Este procedimento experimental traduz-se da seguinte forma (Pohl & Shih Ie, 2004) :

-
- A amostra de DNA é diluída em placas contendo vários poços
 - O PCR é realizado em condições ótimas desenhadas para amplificar uma única cópia de DNA alvo
 - As sequências amplificadas sofrem hibridação com sondas fluorescentes (sinalizadores moleculares), as quais permitem a detecção de produtos específicos, pela utilização de diferentes fluorocromos
 - O PCR digital conta diretamente, uma por uma, quantos alelos (paterno versus materno ou sequência mutante versus sequência selvagem) estão presentes na amostra.

Após este processo, várias análises estatísticas podem ser executadas para medir a distribuição dos alelos mutantes. Esta abordagem fornece uma base estatística rigorosa tornando possível uma análise sem precedentes do estado dos alelos em amostras de tecidos e fluidos corporais. Assim o PCR digital transforma o sinal analógico convencional num sinal linear e digital capaz de ser partilhado mundialmente. (Pohl & Shih Ie, 2004)

Para executar o PCR digital, amostras de DNA genómico, do tecido ou do fluido corporal, são diluídas por 384 poços em placas de PCR, para haver em média, aproximadamente 0,5 moléculas de DNA alvo (equivalente genómico) por poço. A diluição ótima das amostras de DNA pode ser calculada com kits de quantificação de DNA, para determinar a quantidade de equivalentes genómicos nas amostras iniciais. (Pohl & Shih Ie, 2004)

Assim que os produtos da amplificação de uma única molécula de DNA alvo possuem uma sequência homogénea, uma variedade de técnicas convencionais podem ser usadas para verificar a sua presença. Reagentes com sondas fluorescentes, os quais podem ser

usados diretamente nos mesmos poços onde se procedeu à amplificação, são particularmente adequados para este propósito. Atualmente, sondas e outros análogos são muito usados para a detecção dos produtos de PCR em ensaios de PCR digital. (Pohl & Shih Ie, 2004)

Para análise de mutações, é desenvolvido um par de sondas que hibridizam especificamente uma com a sequência normal e outra com a sequência mutada . Assim, o estado mutacional de um alelo específico num determinado poço de análise é determinado pela intensidade da fluorescência emitida por cada sonda naquele poço específico. Ao ser efetuada a contagem dos outros poços, o PCR digital pode ser usado para detetar mutações presentes em quantidades muito pequenas nas amostras em análise. A sensibilidade da detecção da mutação depende do número de poços incluídos para análise e da taxa de mutação intrínseca da polimerase usada para amplificação. (Pohl & Shih Ie, 2004)

Para avaliar o desequilíbrio alélico, são usados SNPs (single nucleotide polymorphisms) para representar os alelos maternal e paternal. Um par de primers de PCR e um par de sondas são usadas para cada SNP (Figura 5). O PCR digital é executado usando um marcador SNP para cada paciente homocigótico. Os produtos resultantes são então analisados usando as sondas para determinar a representação alélica. (Pohl & Shih Ie, 2004)

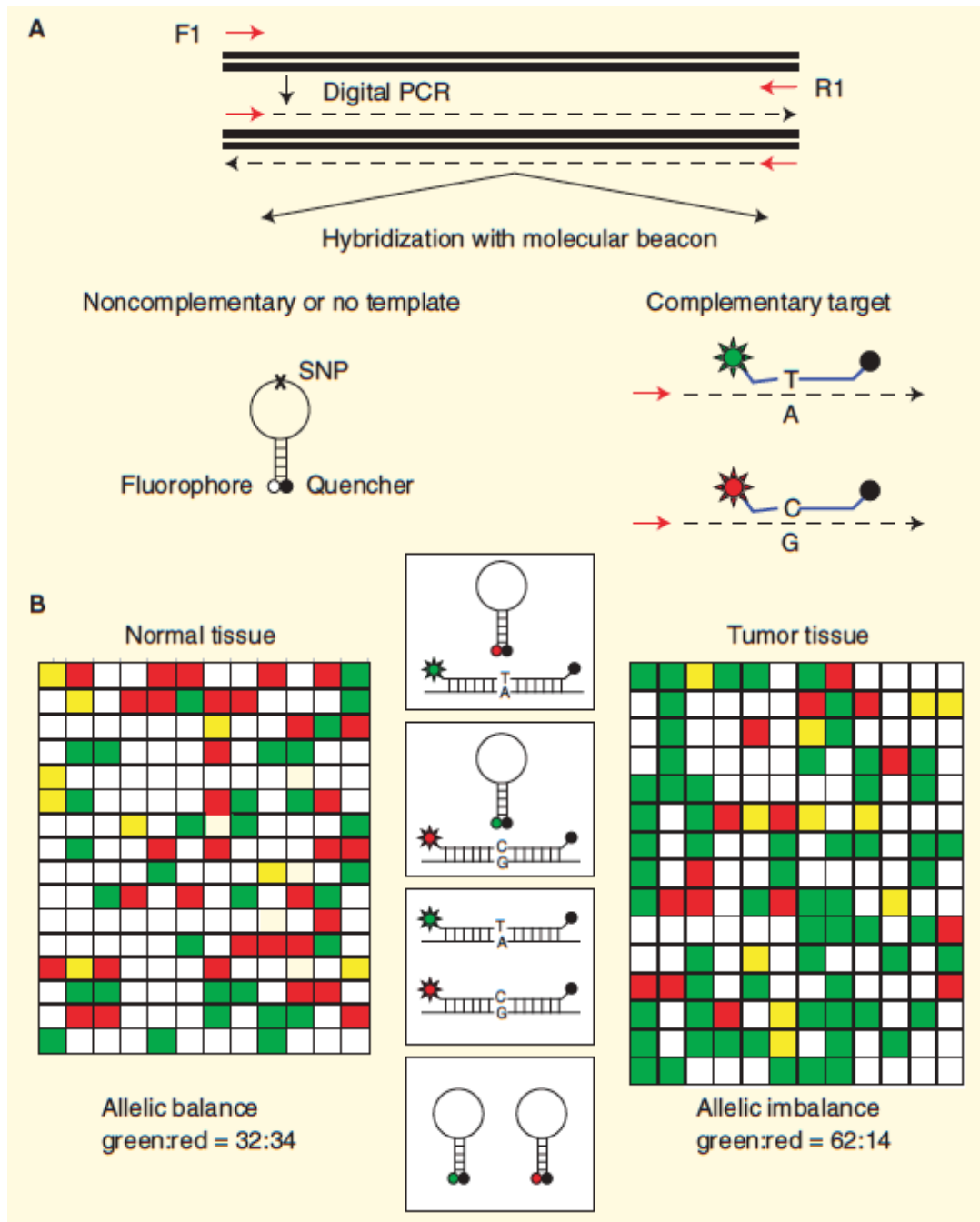


Figura 5 – Análise por PCR digital para verificar desequilíbrio alélico (Pohl & Shih Ie, 2004)

3. Diagnóstico baseado na sequenciação

3.1. Sequenciação de 1ª geração

A sequenciação de ácidos nucleicos é uma das bases mais importantes para a investigação biológica. Existem várias gerações de tecnologias de sequenciação de DNA, as quais podem ser caracterizadas segundo o método ou segundo os resultados que elas fornecem.

Historicamente em 1975, Sanger introduziu o conceito de um método de sequenciação de DNA em “*The Croonian Lecture*” (Sanger, 1975), mais tarde publicou um método rápido para a determinação de sequências no DNA por meio de síntese primerizada com DNA polimerase (Sanger & Coulson, 1975). No ano de 1977, dois artigos excepcionais sobre sequenciação de DNA foram publicados: a sequenciação dideoxy enzimática de Frederick Sanger, baseada nas terminações dideoxinucleotídicas análogas da sequência (Sanger *et al.*, 1977); e outro por Allan Maxam e Walter Gilbert sobre uma técnica de sequenciação por degradação química, onde fragmentos de DNA são clivados quimicamente em bases específicas e separadas por eletroforese (Maxam & Gilbert, 1977).

O método de Sanger viria a tornar-se a base das técnicas de sequenciação nos 30 anos subsequentes à sua primeira publicação. Neste método, o DNA a ser sequenciado é fragmentado, separado e clonado em plasmídeos depois usados para transformar a *Echerichia coli* (Figura 6). Para cada reação, uma única colônia de bactérias é escolhida e o plasmídeo é isolado (Shendure & Ji, 2008).

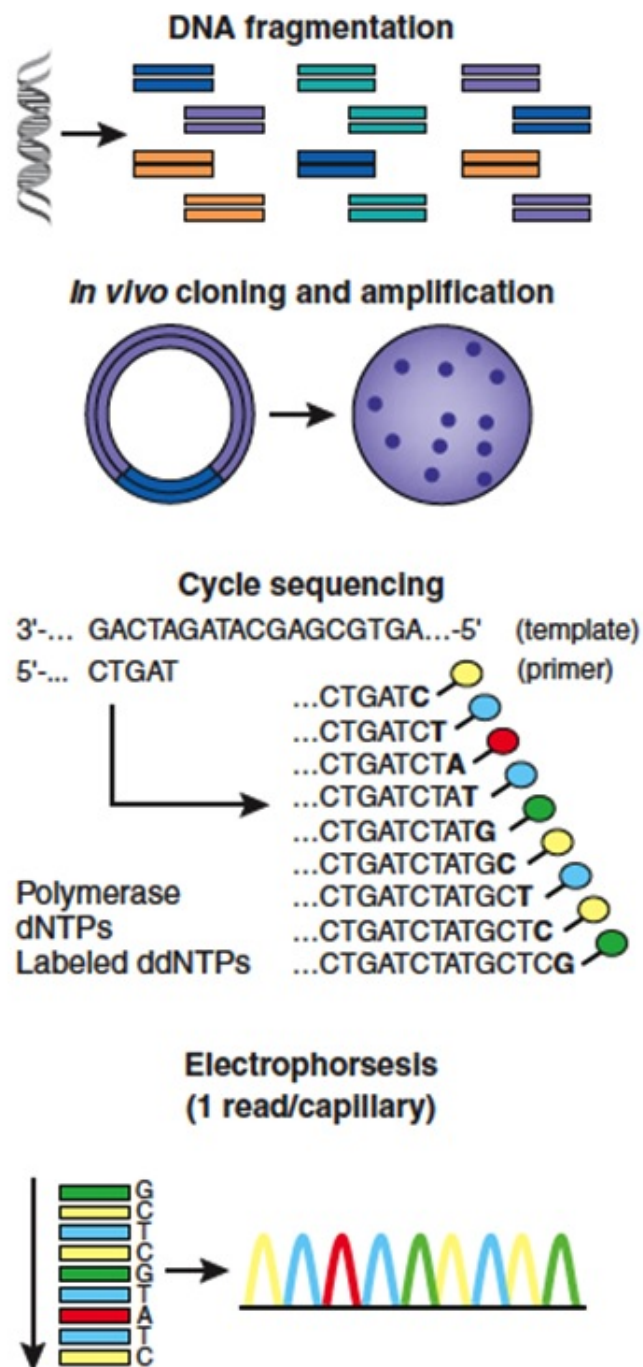


Figura 6 – Esquema do processo de sequenciação *shotgun de novo* (Shendure & Ji, 2008)

Este processo de síntese foi sofrendo melhorias sucessivas e hoje em dia envolve a incorporação de nucleótidos normais (os 2' – desoxinucleotídeos naturais – dNTPs) e

terminadores de síntese marcados com fluorocromos fluorescentes (2', 3' – dideoxynucleotídeos – ddNTPs; Sanger *et al.*, 1977). A razão dNTP/ddNTP na reação de sequenciação determina a frequência de terminação da cadeia, e portanto, a distribuição dos comprimentos das cadeias terminadas. A sequência é determinada após ser submetida a uma eletroforese de alta resolução em tubos capilares, onde os produtos são separados segundo o tamanho (Shendure & Ji, 2008).

Assim que estes fragmentos marcados saem do capilar passam pelo detetor, onde um laser excita os fluorocromos, produzindo emissões de fluorescência de quatro cores diferentes. A leitura da cor fluorescente emitida pelos fragmentos é processada por software especializado, que revela a sequência de DNA. (Smith *et al.*, 1987).

As melhorias mais significativas foram desenvolvidas durante o *Human Genome Program* (HGP), o qual foi iniciado em 1990 e prolongou-se durante 13 anos, numa busca internacional pelo conhecimento do genoma humano (McGinn & Gut, 2012). Mais de três bilhões de dólares foram gastos para serem finalmente conhecidas as diferenças no genoma de cada indivíduo, já que são estas diferenças individuais que ditam a suscetibilidade genética a doenças, a farmacogenômica e a resposta a fármacos, por exemplo (Chan, 2005).

O progresso na metodologia de sequenciação foi impulsionada pela ambição de sequenciar alvos de complexidade crescente. O grupo de investigação de Sanger conseguiu sequenciar o genoma humano mitocondrial de 16.5kb (Anderson *et al.*, 1981), seguido pelos 48.5kb do genoma completo do fago lambda (Sanger *et al.*, 1982). Depois de Sanger se aposentar, o seu pupilo Bart Barrell conduziu a sequenciação do genoma do citomegalovírus de 237kb (Baer *et al.*, 1984). Durante os 15 anos seguintes foi possível fazer leituras de bandas até 400kb, com o método dideoxy. Esta evolução

foi possível devido ao uso de géis de sequenciação extremamente finos e ao uso de marcadores ^{35}S , os quais fornecem bandas mais nítidas do que os ^{32}P (devido ao baixo nível de energia emitido pelas partículas β (Hutchison, 2007).

Depois de três décadas de avanços tecnológicos, a técnica de Sanger pode ser aplicada para conseguir leituras de fragmentos com comprimentos até $\cong 1,000\text{bp}$, e isto com uma precisão de 99.9%. No contexto de um elevado rendimento na sequenciação genómica por shotgun, a sequenciação de Sanger tem-se tornado uma técnica cada vez mais económica (Shendure & Ji, 2008).

3.2. Sequenciação de 2ª geração

Na sequenciação de 2ª geração moldes de DNA clonal individual são gerados e sequenciados em paralelo, pelo uso de micelas. As reações de sequenciação são realizadas com ciclos de adição de bases que podem ser feitas recorrendo ao uso de polimerização ou ligação de DNA, e a visualização é feita por quimioluminescência (pirosequenciação), fluorescência (*Illumina Genome Analyser*), ou sequenciação por oligo-ligações e fluorescência (SOLiD). A chave da sequenciação de 2ª geração é o uso de vários templates clonais em paralelo e um processo de determinação da sequência com uso de replicação enzimática (Mcginn & Gut, 2012).

- **Roche/454 FLX Pyrosequencer**

A plataforma de sequenciação 454 foi a primeira plataforma de alto rendimento no mercado, sendo introduzida em 2004. Esta baseia-se numa tecnologia alternativa, a pirosequenciação, desenvolvida por Pal Nyrén e Mostafa Ronaghi, do Royal Institute of Technology de Estocolmo, em 1996 (Ronaghi *et al.*, 1996).

Na pirosequenciação, cada nucleótido introduzido pela DNA polimerase resulta na libertação de um pirofosfato, o qual inicia uma série de reações em cadeia que produzem luz (pela quebra da oxiluciferina pela luciferase). A quantidade de luz produzida é proporcional ao número de nucleótidos incorporados (até ao ponto de saturação do detetor). A abordagem da Roche/454 sugere que em vez de serem sequenciados em tubos distintos, as moléculas de DNA da biblioteca sejam sequenciadas em paralelo e em massa após amplificação por PCR, as reações ocorrem na superfície das milhares de esferas de agarose (Mardis, 2008a). A superfície destas esferas possui milhões de oligonucleotideos, complementares às sequências específicas do adaptador 454 na biblioteca de fragmentos, assim, cada esfera fica associada a um único fragmento (Mardis, 2008b).

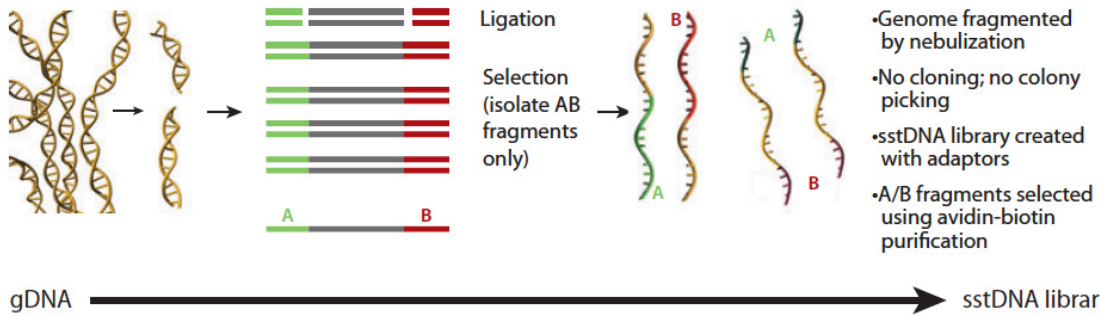
Cada um dos complexos esfera/fragmento é isolado em micelas óleo/água individuais, as quais também possuem reagentes de PCR na fase aquosa, depois do ciclo térmico cada micela produz em apenas algumas horas aproximadamente um milhão de moléculas prontas para sequenciação (Figura 7). Assim, cada esfera de agarose contém na sua superfície até 1,000,000 de cópias do fragmento de DNA original para se produzir um sinal detetável. Estas moléculas amplificadas são depois sequenciadas em paralelo (Mardis, 2008b).

Várias centenas de milhares de esferas (cada uma com um fragmento único amplificado) são dispostas na superfície da 454 PTP (*picotiter plate*), a qual é uma estrutura capilar de sílica onde existem poços individuais que fornecem um local fixo para as esferas, onde cada reação de sequenciação pode ser monitorizada. Subsequentemente, esferas magnéticas de látex muito mais pequenas (com 1µm de diâmetro), ligadas às enzimas necessárias para a pirosequenciação, são adicionadas para

rodear as esferas de agarose (que possuem o DNA) no PTP. Já que o PTP também é o local por onde os reagentes de pirosequenciação fluem, é incorporado no sequenciador. Assim, os nucleótidos e as soluções dos reagentes são lá adicionados passo a passo (Mardis, 2008a).

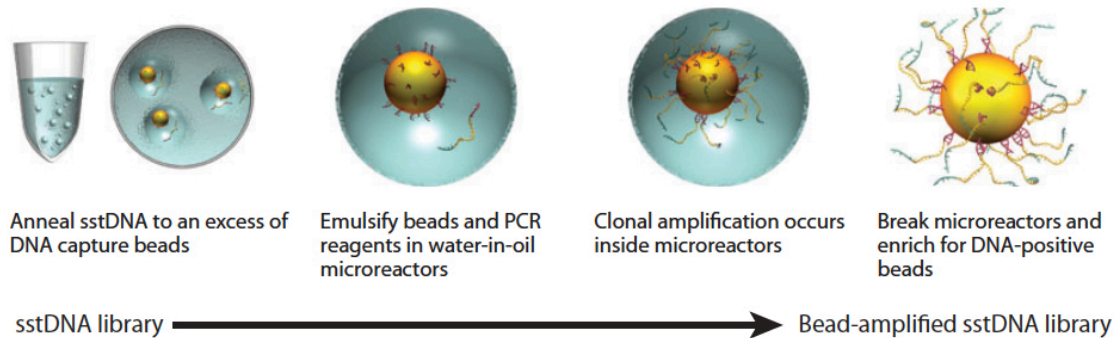
DNA library preparation

4.5 hours



Emulsion PCR

8 hours



Sequencing

7.5 hours

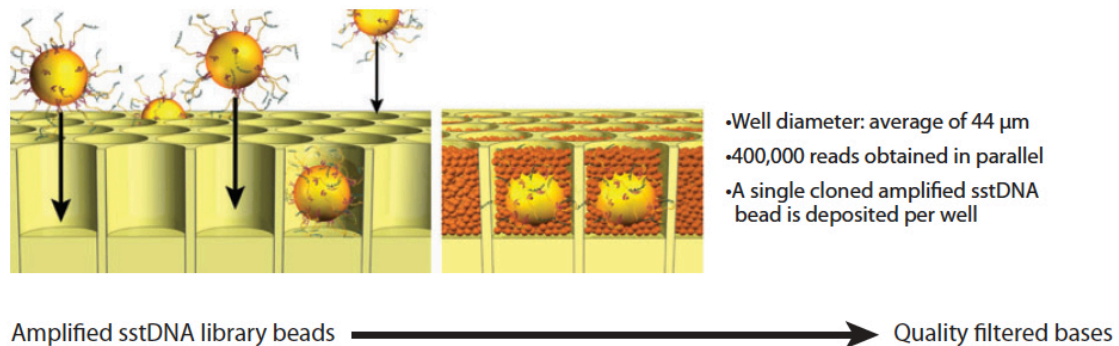


Figura 7 – Esquema do processo de sequenciação do Roche 454. (Mardis, 2008b)

Do lado oposto ao PTP está uma câmara que regista a luz emitida por cada esfera.

A visualização dos flashes de luz da atividade da luciferase regista quais os poços que estão a adicionar cada nucleótido específico, e a luz emitida é diretamente proporcional à quantidade de um determinado nucleótido incorporado.

Os primeiros quatro nucleótidos (T, C, G e A) no fragmento adjacente ao primer de sequenciação adicionados na construção da biblioteca correspondem ao fluxo sequencial de nucleótidos na célula de fluxo. Esta estratégia permite que o software do 454 calibre a luz emitida por cada incorporação. No entanto, esta calibração não consegue interpretar extensões longas de nucleótidos múltiplos (homopolímeros), podendo provocar erros de inserção ou deleção de bases. Em contrapartida, já que cada passo de incorporação é específico de um nucleótido, erros de substituição são raramente encontrados na leitura do sequenciador Roche/454 (Mardis, 2008b).

O sistema atual (GS FLX+) permite leituras médias de 750-800bp (85%>500bp; 45%>700bp e algumas com até 1000bp). Estas leituras são depois processadas pelo software de análise do 454 e depois rastreados por vários filtros de modo a remover sequencias com baixa qualidade, sequencias misturadas (mais do que um fragmento de DNA por esfera) e sequencias sem a sequencia inicial TCGA. Deste processamento resulta um volume de dados de cerca de 700 Mb. Em comparação, o ABI 3730 (sequenciador automático de primeira geração baseado no método de Sanger) quando programado para sequenciar 24 x 96 placas de poços por dia produz 440 kb de dados da sequencia em 7 horas, com um comprimento médio de leitura de 650 bp por amostra (Mardis, 2008a).

- **Illumina genome analyzer**

Introduzido em 2006, o *Illumina Genome Analyzer* baseia-se no conceito de “sequenciação por síntese” (SBS) o qual é semelhante à sequenciação de Sanger. Como acontece no 454 da Roche, o protocolo de sequenciação do Illumina requer que as sequências a serem determinadas sejam convertidas numa biblioteca de sequenciação especial, a qual permite a amplificação e imobilização das sequências para serem submetidas à sequenciação. Para este propósito, dois adaptadores diferentes são adicionados às terminações 5’ e 3’ de todas as moléculas (Figura 8; Kircher & Kelso, 2010).

Esta biblioteca de cadeia dupla é desnaturada para obter DNAs de cadeia única. Estas cadeias simples são dispostas em concentrações muito baixas pelos canais de uma célula de fluxo. Esta “*flow cell*” possui na sua superfície dois tipos de oligonucleotídeos imobilizados complementares aos dois adaptadores, utilizados para produzir a biblioteca de sequenciação. Estes oligonucleotídeos hibridizam com as moléculas das cadeias das bibliotecas. Por síntese reversa, começando pela zona hibridizada, a nova molécula que está a ser criada encontra-se covalentemente ligada à *flow cell* (Kircher & Kelso, 2010).

Esta nova molécula dobra-se e liga-se a outro oligonucleotídeo complementar ao segundo adaptador que não está ligado à placa, podendo ser usado para sintetizar uma segunda cadeia ligada também covalentemente à placa. Este processo de dobra da molécula e de síntese reversa, chamada de amplificação em ponte, é repetida várias vezes e cria aglomerados de vários milhares de cópias da sequência original, muito próximos na célula de fluxo (Kircher & Kelso, 2010).

Estes aglomerados distribuídos aleatoriamente contêm cópias idênticas da mesma sequência. Deste modo as bibliotecas estão prontas para serem sequenciadas.

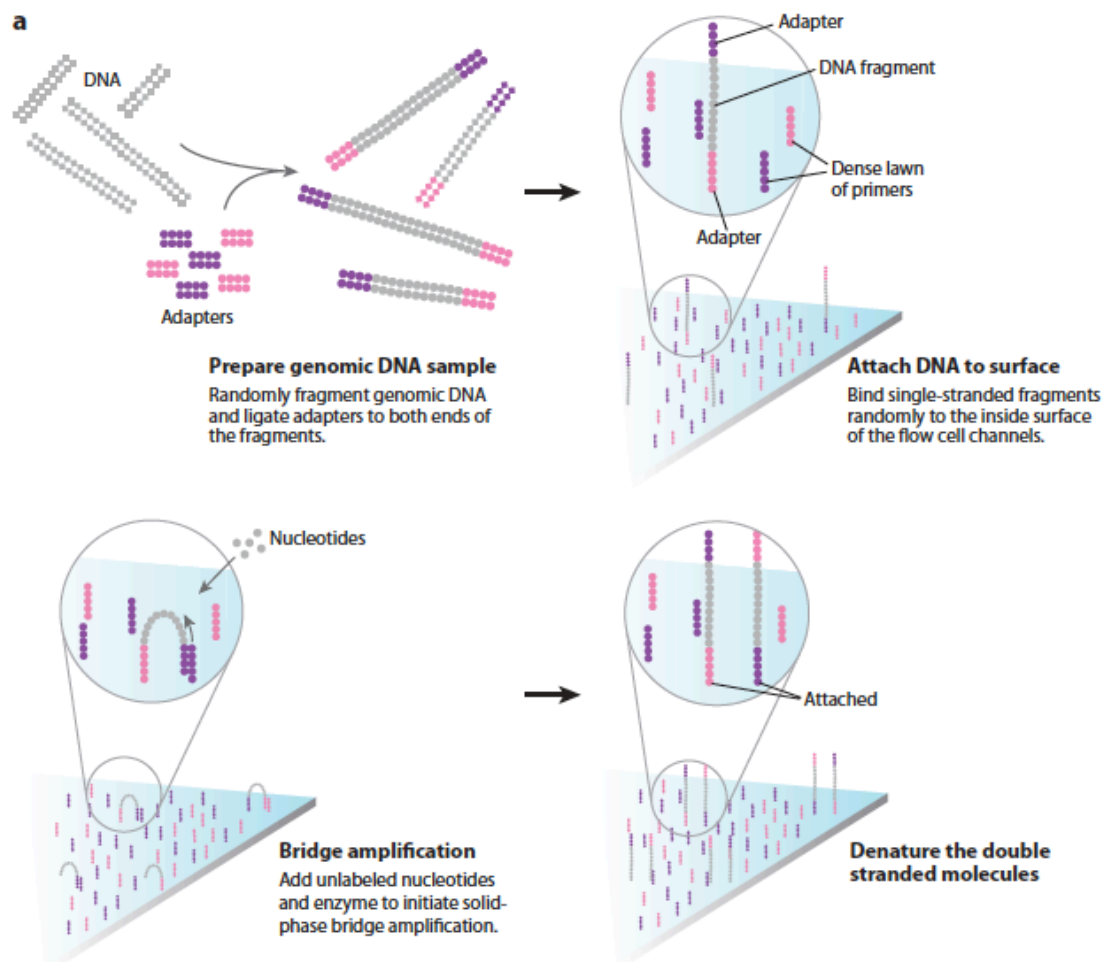


Figura 8 – O processo de sequenciação do Illumina (Mardis, 2008b)

No analisador de genoma, milhões de aglomerados são sequenciados simultaneamente, as moléculas de DNA são sequenciadas base a base em paralelo usando quatro nucleótidos marcados com fluorescência. As quatro bases completam umas com as outras para se ligarem ao alvo, esta competição natural garante a alta precisão. Depois de cada síntese os fluorocromos são excitados por um laser, a cor obtida identifica a base que foi adicionada. Este fluorocromo é depois retirado para que a próxima base se possa ligar ao template, é lida então cada base adicionada em cada ciclo (Kircher & Kelso, 2010).

A “Solexa Sequencing” como foi introduzida a 2007, inicialmente permitia a sequenciação simultânea de vários milhões de seqüências bastante pequenas numa só reação. Anos mais tarde foram feitas várias atualizações dando origem ao atual *Illumina Genome Analyzer* (Kircher & Kelso, 2010).

- **Applied Biosystems SOLiD**

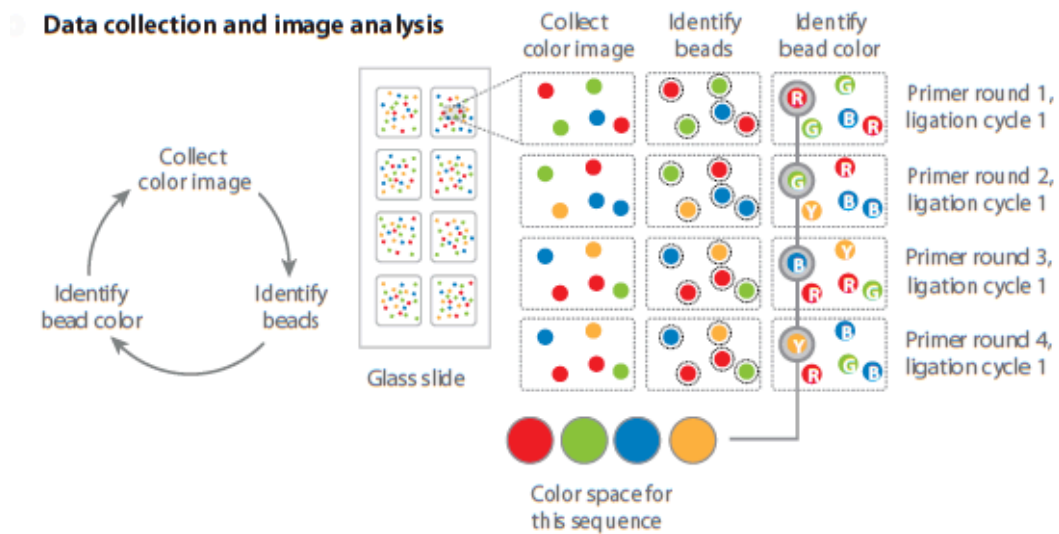
A plataforma SOLiD (Sequencing by Oligo Ligation and Detection) usa uma biblioteca adaptador-ligando semelhante aos sistemas anteriores e usa uma emulsão de PCR com pequenas esferas magnéticas para amplificar fragmentos para sequenciação. Ao contrário das outras plataformas de sequenciação de segunda geração, a SOLiD usa a DNA ligase em vez da DNA polimerase (Mardis, 2008b). Uma mistura de sondas fluorescentes com 4 cores distintas competem para a ligação ao primer, cada cor representa 4 das 16 sequencias possíveis (cada uma com 4 nucleótidos). Neste processo, um primer de sequenciação é hibridizado com as cópias de cadeia simples das moléculas da biblioteca a serem sequenciadas e é ligado pela DNA ligase. Depois de ser medida a fluorescência parte da sonda fluorescente é clivada da extremidade 5' deixando 2 nucleótidos da sonda e um grupo fosfato livre, disponível para se dar a próxima reação de ligação (Kircher & Kelso, 2010).

Os comprimentos de leitura para o SOLiD estão definidos entre 25 a 35 bp, e cada processo tem um rendimento de 2 a 4 Gb de dados sobre a sequencia de DNA (Mardis, 2008b).

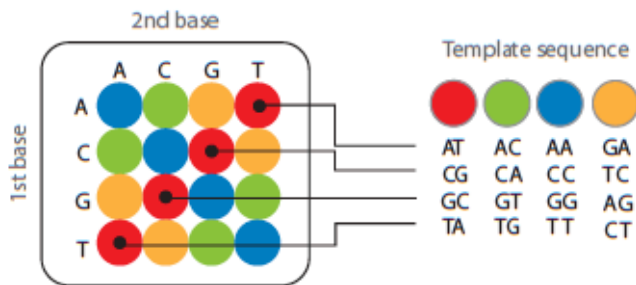
Depois de múltiplas ligações (normalmente durante 10 ciclos) os fragmentos sintetizados são removidos por desnaturação. Uma segunda ronda de sequenciação é iniciada com a hibridização de um novo primer universal, seguindo-se novas rondas de

sequenciação. Este processo denominado “*2 base encoding*”, realiza uma leitura dupla de cada base ao longo de vários ciclos (Figura 9), o que aumenta a precisão da leitura (Mardis, 2008a).

O sistema SOLiD atualmente permite a sequenciação de mais de 300 milhões de esferas em paralelo, com um comprimento de leitura entre 25 e 75 bp. Assim ambos os sistemas, ABI SOLiD e Illumina GA, são comparáveis em termos de rendimento e preço ($\approx 5,000$ Mb/dia; $\approx 0,50$ \$/Mb). A taxa de erros depende, no entanto, da disponibilidade de um genoma de referência, para correção de erros (Kircher & Kelso, 2010). As duas diferenças mais significativas, que refletem a utilidade das técnicas de segunda geração em detrimento das de primeira, referem-se ao comprimento da leitura das sequências ser muito mais pequena e a capacidade de correção de erros ser muito mais alta do que nos sequenciadores capilares. Ambas afetam a forma como as leituras são utilizadas em análises bioinformáticas, dependendo da aplicação (Mardis, 2008b).



Possible dinucleotides encoded by each color



Double interrogation

With 2 base encoding each base is defined twice



Decoding

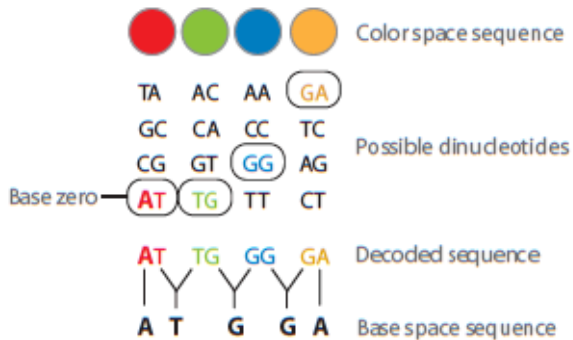


Figura 9 – Princípios do “2 base encoding”. (Mardis, 2008b)

- **Life Technologies's Ion Proton**

O primeiro instrumento do género, comercializado pela *Life Technologies*, o sequenciador *IonTorrent Personal Genome Machine* (PGM™), usa a incorporação enzimática de bases semelhante à tecnologia da *Roche*. A PGM™ usufrui do poder da tecnologia semicondutora, detetando os prótons libertados aquando da incorporação dos nucleótidos durante o processo de síntese. Fragmentos de DNA com sequências-adaptador específicas são ligadas e posteriormente amplificadas por clonagem em emulsão de PCR, na superfície de esferas de 3 micrómetros de diâmetro. Estas esferas, conhecidas por *Ion Sphere Particles*, são transferidas para os poços sensíveis a prótons e a sequenciação é iniciada a partir de um local específico na sequência. Assim que se inicia a sequenciação, cada uma das quatro bases (C, G, A e T) é introduzida sequencialmente (Quail *et al.*, 2012). Cada incorporação bem sucedida é determinada pela variação de pH devido à libertação de H⁺ durante a incorporação de bases. Um sistema com numerosos sensores de pH é usado em paralelo para detetar de que base se trata (Mcginn & Gut, 2012). O PGM™ tornou-se em pouco tempo o aparelho de sequenciação mais vendido em todo o mundo.

Depois do sucesso do PGM™ a *Life Technologies* lança o *IonProton™ Sequencer* que funciona de um modo ainda mais simplificado, por meio de *chips*. Neste caso, transístores de efeito de campo (FET's) são responsáveis pela deteção das variações de pH numa estrutura de micro poços. Para aumentar o rendimento, o *chip* de sequenciação do *IonProton™* usa uma matriz muito densa de micropoços, onde cada poço atua como uma câmara individual de polimerização de DNA contendo uma DNA polimerase e um fragmento para sequenciação. Imediatamente sob esta camada de micropoços está um estrato sensível a iões, seguida por uma subcamada, composta por uma matriz de

detetores FETs, alinhada com as camadas superiores (Figura 10). Os FET detetam as variações de pH durante a incorporação de bases e convertem este sinal em voltagem, gravando a leitura no software (Niedringhaus *et al.*, 2011).

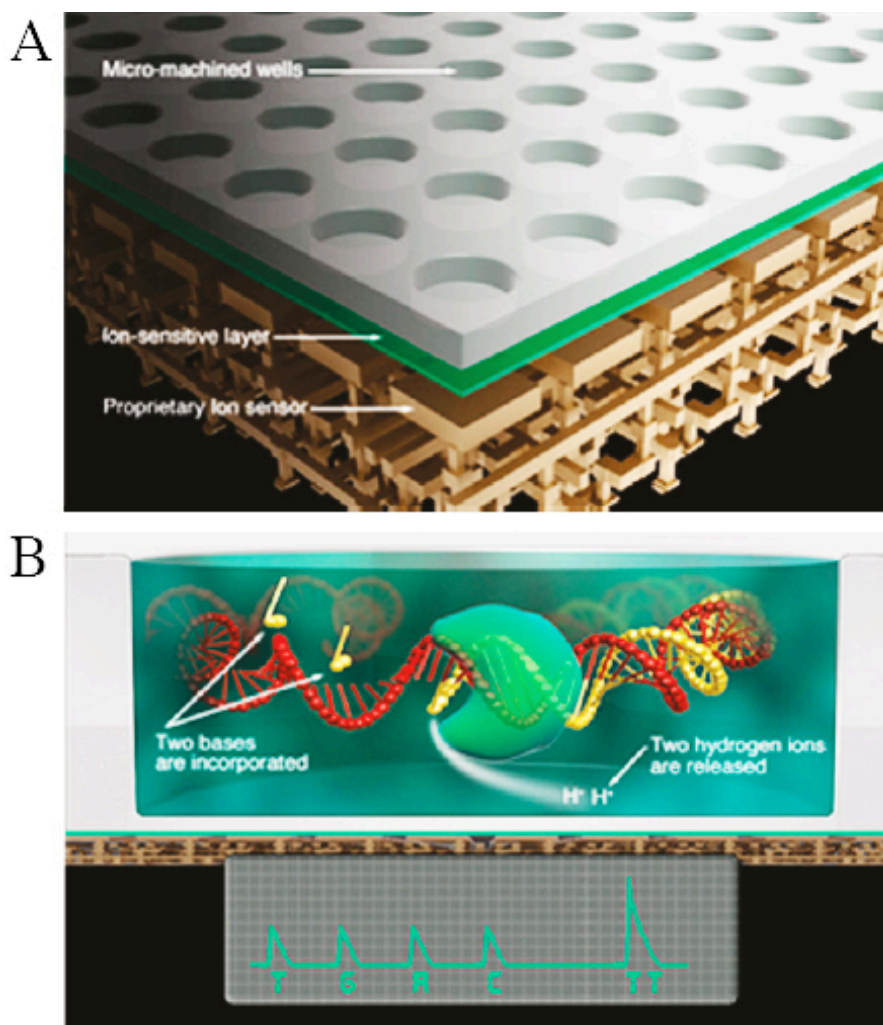


Figura 10: Esboço do *chip* do *IonProton™*. (A) uma vista cama a camada do chip, revelando a estrutura. A camada superior contem poços individuais de reação de polimerização e as duas camadas inferiores são compostas pelos sensores FET. Cada poço possui um FET correspondente que deteta a mudança no pH; (B) reação a ocorrer em cada poço: incorporação, pela DNA polimerase, de uma repetição de dois nucleótidos com a liberação de 2 íons de hidrogênio; (C) Gráfico de leitura que permite a identificação das bases (Niedringhaus *et al.*, 2011).

O primeiro *chip* comercializado é descartável e intitula-se *Ion 314 sequencing chip*. Este possui 1.2 milhões de poços e produz 10 Mb de informação, com leituras na ordem das 100 bases. De modo a aumentar ainda mais o rendimento foram lançados em 2011 dois novos chips, o *Ion 316* e o *318*, com 165 e 660 milhões de poços respetivamente. O *Ion 318* pode produzir até 1 Gb de informação e ler para além das 200 bases, sendo o *chip* de sequenciação mais evoluído e barato presente no mercado (Niedringhaus *et al.*, 2011).

Entre o sequenciador *Ion PGM™* e o *IonProton™* (Figura 11), a tecnologia *Ion Torrent* pode cobrir qualquer aplicação. O primeiro é ideal para sequenciar genes, genomas pequenos ou para traçar perfis de expressão de genes, por apenas 99 dólares por *chip*.



Figura 11: Sequenciadores comercializados pela Life Technologies, o primeiro *Ion Personal Genome Sequencer* e o segundo *IonProton™ Sequencer* (Life Technologies, 2013).

A sua rapidez e sensibilidade tornam-no ideal para auxiliar o avanço na investigação clínica. A *Life Technologies* procura agora aprovação pela FDA.

Quanto ao sequenciador *IonProton™*, o seu *chip* Ion 316, ou *IonProton™* I é ideal para a sequenciação de exomas – regiões de DNA que codificam proteínas – e o Ion 318, ou *IonProton™* II, para sequenciar genomas humanos inteiros (Figura 12).

A sequenciação de um genoma humano, um processo que anteriormente poderia demorar meses ou semanas, torna-se deste modo potencialmente acessível à prática clínica corrente. Uma vez aprovado o seu uso para além dos limites da investigação científica, será possível desenhar terapêuticas dirigidas ao perfil genético de cada indivíduo (Life Technologies, 2013).

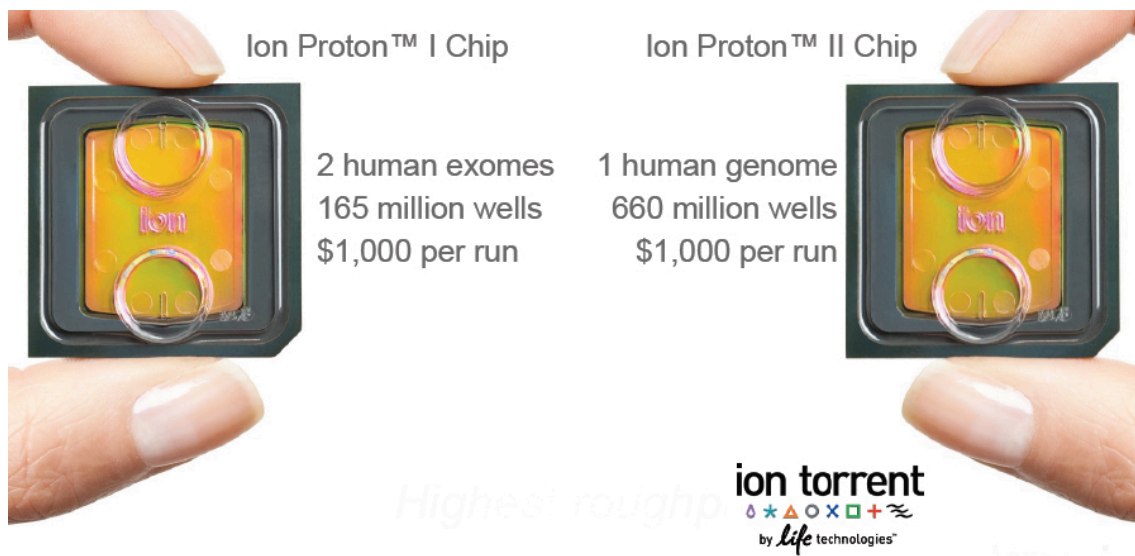


Figura 12: Visualização dos dois chips de sequenciação usados pelo *IonProton™* Sequencer (Life Technologies, 2013).

3.3. Sequenciação de 3ª geração

Ao contrário das tecnologias de segunda geração, que dependem do PCR para obter as sequências, da sua ligação a placas sólidas e da subsequente imagiologia da sequenciação, a terceira geração apresenta tecnologias que não necessitam de qualquer amplificação, sendo a sequenciação feita em moléculas únicas (Mccarthy, 2010). Deste modo a terceira geração conseguiu superar as duas gerações anteriores em termos de comprimento de leitura da sequência (de dezenas de bases para dezenas de milhares de bases por cada leitura), rapidez do processo (de dias para horas ou minutos), alto rendimento, tempo de resposta mais rápido (10,000 a 20,000 vezes mais rápido), quantidades mínimas de amostra são necessárias e o seu baixo custo (Pareek *et al.*, 2011). Comparações entre os vários sistemas estão representados na Tabela 2.

Uma das maiores vantagens da terceira geração é a sua capacidade de, em tempo real, detetar eventos biológicos a nível molecular, dando a melhor resolução possível para perceber o que está a acontecer no organismo.

A sequenciação hoje em dia tem um papel importantíssimo no diagnóstico, por exemplo no mapeamento do genoma de cancros de modo a se compreender quais as alterações que se verificaram nas células tumorais e porque é que elas se comportam de determinada maneira, ou no estudo de genomas para se descobrir variações genéticas ligadas a doenças. Com este método de sequenciação de moléculas únicas, estende-se a investigação de maneira a ser possível a deteção de mutações raras entre indivíduos em termos de saúde ao nível molecular (Mccarthy, 2010).

Tabela 2: Comparação de plataformas de segunda e terceira geração. (Adaptado de (Pareek *et al.*, 2011))

Empresa	Roche GS GLX	Illumina-Sollexa	Life Technologies	Helicos Biosciences	Pacific Biosciences
Site	www.454.com/index.asp	www.solexa.com	www3.appliedbiosystems.com/AB_Home	www.helicosbio.com	www.pacificbiosciences.com
Plataforma	GS FLX Titanium	HiSeq 2000, Genome Analyzer IIE	ABI SOLiD, SOLiD 4	HeliScope	SMRT
Preparação	ePCR clonal na superfície de esferas	Amplificação clonal enzimática em vidro	ePCR clonal na superfície de esferas	Deteção de moléculas únicas	Deteção de moléculas únicas
Característica	Luz emitida de reações secundárias pela libertação de pirofosfato	Emissão fluorescente de nucleotídeos marcados com fluorocromos	Emissão fluorescente de nucleotídeos marcados com fluorocromos	Deteção em tempo real da fluorescência	Deteção em tempo real da fluorescência
Método	Emulsão de PCR/esferas	Amplificação isotermal na "flow cell"	Emulsão de PCR/esferas	Sequenciação de moléculas únicas	Sequenciação de moléculas únicas
Química	Pirosequenciação	Terminadores fluorescentes reversíveis	Ligação de sondas a oligonucleotídeos	Terminadores fluorescentes reversíveis	Nucleotídeos fluorescentes fosfo-ligados
Método de sequenciação	Pirosequenciação	Terminadores fluorescentes reversíveis	Por ligases	Uma base de cada vez	Por síntese
Comprimentos de leitura (bases)	400	36	35	>1000	>1000
Tempo por leitura	10 h	2 – 3 dias	6 dias	12	<1
Rendimento bases/leitura (Gb)	0,035	3 – 6	10 – 20	28	100 /hora

- **Helicos HeliScope**

A empresa Helicos foi a primeira a comercializar um aparelho capaz de sequenciar moléculas individuais, não havendo assim a necessidade de um processo de amplificação. Este tipo de sequenciação tem a vantagem de não ser afetada por erros na preparação de biblioteca ou na amplificação e facilita a sequenciação de quantidades mínimas de material genético. O HeliScope, foi introduzido no mercado em Março de

2008, no entanto passado um ano e meio apenas quatro aparelhos tinham sido instalados mundialmente. Isto pode dever-se ao elevado custo de cada aparelho (um milhão de dólares aproximadamente) e o baixo número de investidores. Estes problemas foram entretanto parcialmente mitigados, e tanto o custo por equipamento, como o custo por leitura diminuíram significativamente. Esta tecnologia baseia-se na tecnologia “*true single molecule sequencing*” – tSMS (Pareek *et al.*, 2011). O processo pode ser iniciado por um único técnico, num tubo ou numa micro lâmina. A amostra de DNA é fragmentada antes da adição de *poly-A-tail* (cauda de várias bases Adenina) em cada molécula. No último passo da poliadenilação é adicionada uma adenina marcada com fluorescência (Kircher & Kelso, 2010).

De seguida a biblioteca é transferida para a “*flow cell*” do HeliScope, a qual possui na sua superfície milhões de nucleotídeos *poly-T* universais. Cada molécula transferida hibridiza com as terminações *poly-T*, desencadeando as reações de sequenciação. Assim que as amostras estão ligadas à “*flow cell*”, são colocadas no HeliScope para leitura. Um laser ilumina a superfície da “*flow cell*” revelando a localização de cada template fluorescente. Com estas coordenadas identificadas, a adenina marcada com fluorescência, na terminação 3’, é retirada e a sequenciação começa. Polimerases passam pela “*flow cell*” com um tipo de nucleótido de cada vez, A, C, G e T, com fluorescência. As sequências reversas vão ser então estendidas pelas polimerases que começam pelos oligonucleotídeos poli-T. É lida a fluorescência das adenosinas que se ligaram às sequências antes de ser removido o fluoróforo (Figura 13), de seguida são as bases citosina marcadas com fluorescência que se vão ligar complementando a sequência, a leitura é novamente feita e o processo repete-se para as outras bases até se atingir o comprimento de leitura desejado (Kircher & Kelso, 2010).

Assim, cada sequência é única e sequenciada independentemente proporcionando um elevado rendimento.

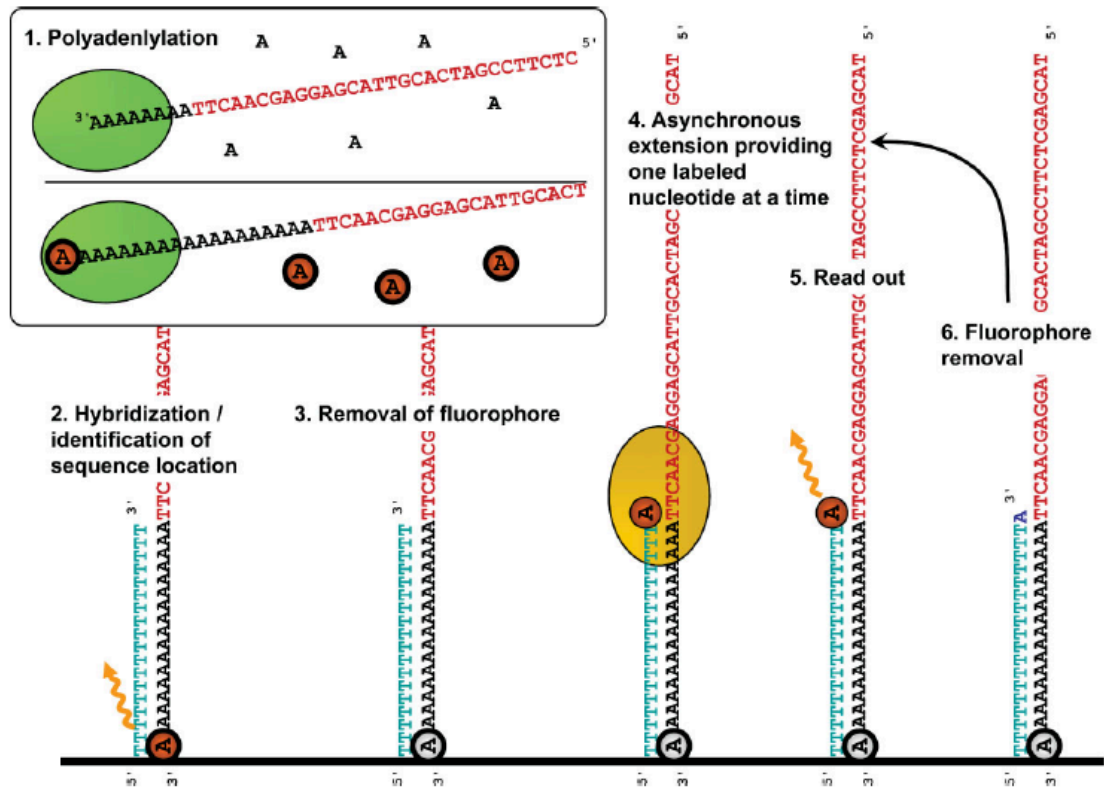


Figura 13: Sequenciação executada pelo HeliScope (Kircher & Kelso, 2010)

- **Pacific Biosciences SMRT**

A tecnologia da Pacific Biosciences fornece fundamentalmente um novo tipo de leitura, com o potencial de ultrapassar algumas limitações das tecnologias de última geração disponíveis atualmente, distinguindo-se pela maior quantidade de informação disponível em menor tempo, menos custos e menor quantidade de erros. A tecnologia de sequenciação SMRT (*Single Molecule Real Time*) possibilita, pela primeira vez, a observação de uma síntese natural de DNA pela DNA polimerase, prometendo tornar-se

uma tecnologia transformadora que permitirá um novo paradigma na análise genómica (Carneiro *et al.*, 2012).

A reação de sequenciação de um fragmento de DNA é feita por uma única DNA polimerase ligada à base de cada câmara de visualização ZMW (*Zero Mode Waveguide*; Figura 14).

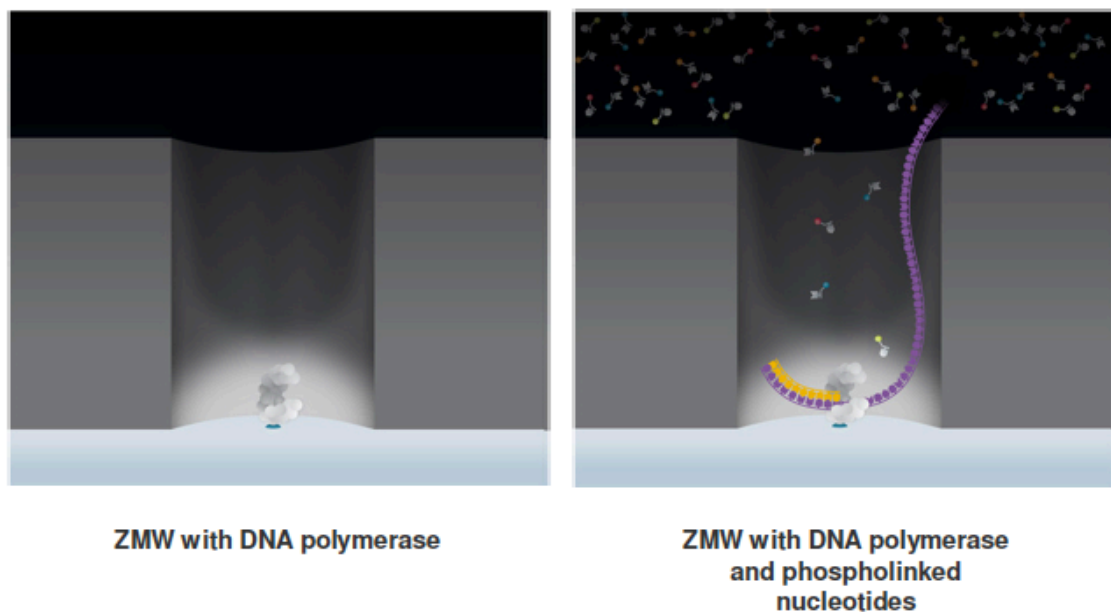


Figura 14: Sequenciação SMRT pela DNA polimerase (Pacific Biosciences, 2013)

A ZMW é uma câmara cilíndrica de metal, com aproximadamente 9 nm de diâmetro que possui uma base de vidro por onde passa a luz, para que seja possível a leitura. Durante a sequenciação, o DNA da amostra é alongado pela polimerase usando nucleótidos marcados na terminação fosfato com 4 tipos de fluorescência (uma cor para cada base) presentes na solução. A marcação no grupo fosfato, ao contrário da marcação feita na base (Figura 15), permite que a enzima clive a fluorescência como parte do processo de incorporação do nucleótido na cadeia nova de DNA, deixando assim uma cadeia sintetizada natural. Esta inovação demonstra a utilidade da técnica pela

exploração das propriedades da DNA polimerase. À medida que a polimerase alonga a sequência vai adicionar um nucleótido de cada vez, durante este período de incorporação é possível ler a fluorescência emitida pelo nucleótido individualmente. Subsequentemente, a leitura simultânea e contínua da fluorescência permite a determinação da sequência de DNA em cada uma das milhares de câmaras ZMW (Pareek *et al.*, 2011).

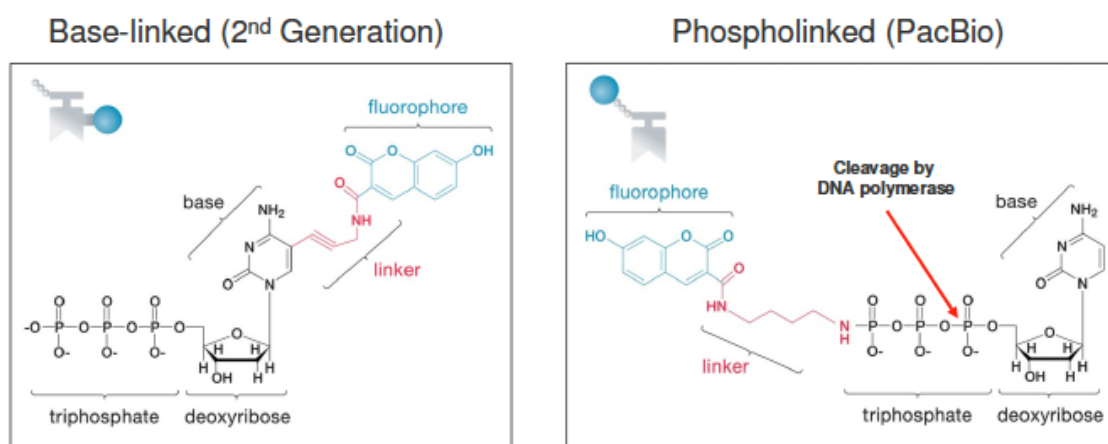


Figura 15: Diferença da marcação fluorescente nas duas gerações de sequenciação (Pacific Biosciences, 2013).

- **Oxford Nanopore Technology**

O princípio básico por detrás da sequenciação por nanoporos é extremamente simples e foi demonstrada pela primeira vez em 1996. Esta abordagem, ainda em desenvolvimento, está a tornar-se uma potencial tecnologia de 4^a geração. Esta classificação torna-se pertinente já que a identificação da sequência passa a ser feita sem recorrer à deteção ótica, juntamente com a eliminação das etapas de lavagem (Niedringhaus *et al.*, 2011). A tecnologia *Oxford Nanopores* oferece a possibilidade de identificar modificações individuais entre os nucleótidos – consegue, por exemplo, fazer a distinção entre a 5-metil-citosina e a citosina (Kircher & Kelso, 2010).

A sequenciação por nanoporos envolve o uso de uma fina membrana, com elevada resistência elétrica, que contém poros com aproximadamente 1.5 – 2 nm de diâmetro, imersa numa solução salina. É aplicada uma corrente elétrica ao sistema que conduz os iões pelos nanoporos, provocando uma mudança de potencial. Uma biomolécula carregada negativamente, tal como o DNA, atravessa ou bloqueia o canal em direção à carga positiva, criando uma resistência no fluxo de iões. Esta resistência leva a alterações na corrente elétrica, na ordem dos picoamperes (pA), que podem ser medidas num circuito elétrico (figura 16). Moléculas de diferentes tamanhos vão bloquear o poro de forma diferente, ou durante espaços de tempo diferentes, possibilitando a discriminação entre as quatro bases de DNA (T, G, C e A) e assim a identificação da sequência (Mcginn & Gut, 2012, Kircher & Kelso, 2010, Gupta, 2008). Adicionalmente, esta tecnologia consegue distinguir polinucleótidos que possuem composição e comprimento similar, como na razão GC:AT, mas possuem posições diferentes nem que seja apenas numa única posição (Gupta, 2008).

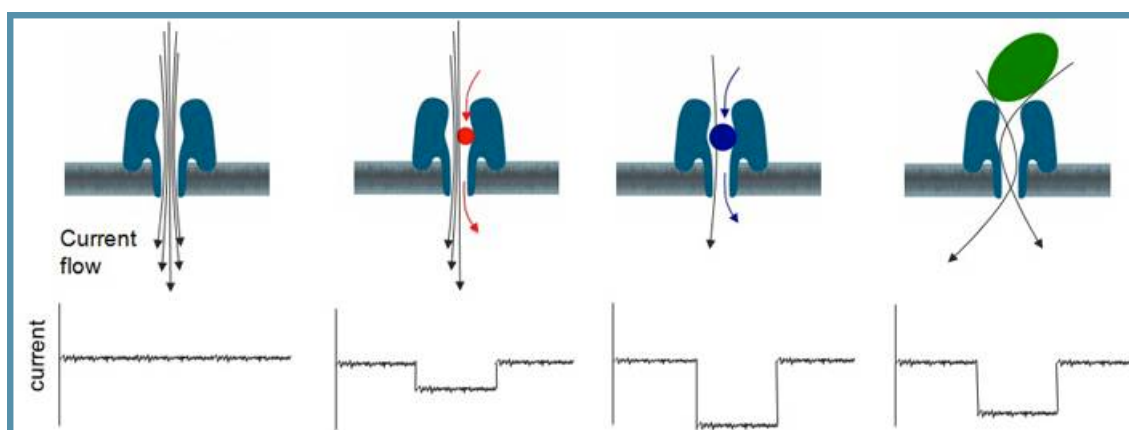


Figura 16: Moléculas diferentes provocam alterações na corrente elétrica, proporcionando a sua identificação (Oxford Nanopore Technologies, 2013)

No entanto, para esta tecnologia ser aplicada à sequenciação, uma enzima, a exonuclease deve ser ligada covalentemente à entrada do nanoporo. Esta tem a função de degradar a dupla hélice de DNA depositando no poro os nucleótidos, proporcionando a leitura individualizada de cada nucleótido presente na sequência (Kircher & Kelso, 2010).

Atualmente existem dois tipos de nanoporos, os biológicos e os *solid-state*. Os nanoporos biológicos usam a proteína α -hemolisina, originalmente presente em membranas celulares durante o processo de lise. As vantagens do uso de nanoporos biológicos incluem a possibilidade de serem facilmente modificados, o seu tamanho estar bem definido e poderem ser produzidos em massa, sem alterar o tamanho e estrutura. No caso da α -hemolisina, deve ser salientado que para alcançar um sistema totalmente funcional, capaz de distinguir as quatro bases de DNA, o nanoporo teve de ser modificado geneticamente e um adaptador teve de ser introduzido no canal (Mcginn & Gut, 2012). Este adaptador tem um papel fundamental na desaceleração das bases assim que estas entram no poro, já que caso um nucleótido passe a uma velocidade muito elevada pelo poro, demasiado rápida para efetuar uma leitura individualizada, a sensibilidade do aparelho diminui (Gupta, 2008).

Os nanoporos *solid-state* são estruturas com um orifício à escala nanométrica formulados numa membrana sintética de sílica ou derivados (SiN_x ou SiO_2). O uso destes poros sintéticos alivia a dificuldade de estabilizar a membrana e de posicionar as proteínas, desvantagens dos biológicos. São habitualmente fabricados por feixes de iões ou elétrões, podendo assim o tamanho do poro ser alterado. No entanto, são ainda necessários elevados progressos na síntese para atingir a precisão atómica, que é intrínseca aos poros biológicos (Mcginn & Gut, 2012).

Apesar da sua extrema sensibilidade, este método tem a desvantagem de destruir a amostra em estudo, o que impossibilita uma segunda leitura para redução de erros (Kircher & Kelso, 2010).

Apesar de tudo, a *Oxford Nanopore Technologies* oferece plataformas extremamente evoluídas e vantajosas. Comercializada recentemente, a plataforma de sequenciação “GridION” possui uma irmã em tamanho mais reduzido, a “MinION”. Esta última, é a primeira e única plataforma de sequenciação descartável em miniatura disponível no mercado. Do tamanho de uma pen USB, custa aproximadamente 900 dólares, e é fornecida a clientes selecionados (Oxford Nanopore Technologies, 2013)

A companhia acredita que a sua tecnologia inovadora irá dar início a uma nova era em medicina, farmacologia, agricultura, energia, biodiversidade, biologia evolucionária, entre outras áreas, discriminadas na figura 17, presente no site da ONT.

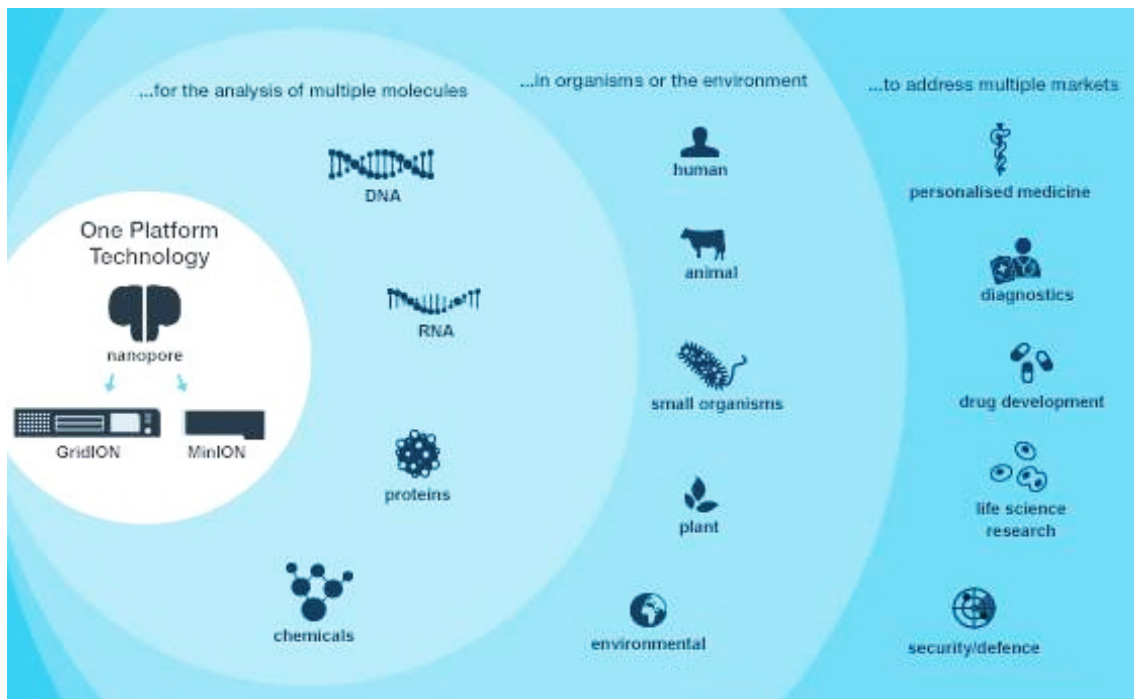


Figura 17 – Leque de aplicações das plataformas GridION e MinION. (Oxford Nanopore Technologies, 2013)

4. O Diagnóstico Molecular

O diagnóstico molecular está atualmente a ser revolucionado pelo desenvolvimento de técnicas inovadoras de análise genómica. Estas técnicas, aliadas à bioinformática, potenciam o diagnóstico rápido e eficaz de doenças genéticas, infecciosas ou oncológicas, por métodos cada vez mais sofisticados e não invasivos, mesmo antes de estas se manifestarem. Uma das vantagens dos métodos de diagnóstico baseados na análise de ácidos nucleicos é a sua sensibilidade, revelando-se especialmente útil quando se possui uma amostra muito reduzida para análise (Irving, 1999).

Nos laboratórios clínicos, a análise de amostras provenientes de indivíduos suspeitos de possuírem uma doença passível de ser diagnosticada por métodos moleculares, é feita diariamente. Quando a sequência do gene em questão é conhecida, é preferível a análise sequencial por PCR, caso contrário é feita uma sequenciação direta de modo a caracterizar a nova região de DNA. Deste modo, o aparecimento de novas técnicas cada vez mais sensíveis e rápidas e com aplicações bastante alargadas, facilitam o trabalho tanto de geneticistas como de investigadores, oncologistas ou microbiólogos. (Bernard, 2001).

4.1. Diagnóstico Genético

É facilmente perceptível a importância das técnicas de PCR e sequenciação aplicadas ao diagnóstico genético, tendo estas, desde cedo, merecido um lugar de destaque quer no diagnóstico, quer no prognóstico de doenças genéticas (por exemplo a fibrose cística)

ou oncológicas. Para esse efeito, uma vasta gama de amostras de tecidos (sangue, sémen, cabelo ou raspagens citológicas) pode ser processada e analisada por PCR (Bermingham, 2003).

Em termos de testes genéticos, a informação pode ser adquirida de duas formas: por ensaios diretos, onde a amostra é testada para verificar se o individuo possuiu ou não um genótipo específico; ou por rastreamento de genes, onde marcadores são usados para demonstrar sequências conhecidas em estudos familiares. Os ensaios diretos são os mais usados em laboratórios de diagnóstico e o seu papel é ainda mais importante quando é necessária a identificação de desordens genéticas muitíssimo raras. Para estes ensaios o gene e a doença devem ser conhecidos, assim como a sequência normal ou selvagem. (Bermingham, 2003)

- **Hereditariedade**

No caso da hereditariedade, diz-se que uma desordem é familiar quando esta se verifica com maior frequência numa determinada família, comparando com a população em geral. No entanto, uma situação de natureza genética não significa que esta seja hereditária, ou seja, a alteração genética pode ocorrer após o nascimento, como no caso de mutações induzidas, que se verificam na maioria dos tumores.

As desordens de natureza hereditária podem ser classificadas como monogénicas e poligénicas. A hereditariedade autossómica recessiva (AR) é monogénica e é responsável pela transmissão de doenças ou caracteres associados a genes localizados num dos cromossomas autossómicos. A fibrose cística é um exemplo bastante representativo de uma doença AR, onde mais de 1300 mutações alélicas têm sido identificadas. A doença é provocada por deficiência dos transportadores de cloro da membrana celular, devido a mutações do gene codificador CFTR localizado em 7q31.2.

A triagem padrão de mutações pode ser feita por PCR multiplex, seguido de hibridização reversa. Outras desordens AR como a hemocromatose hereditária, a doença de Tay-Sachs ou a atrofia muscular espinhal podem também ser detetadas por métodos moleculares (Evans, 2009).

Quanto à hereditariedade autossômica dominante (AD), os genes relacionados com as doenças localizam-se num dos 22 pares de cromossomas autossômicos. É também do tipo monogénica e neste caso transmite-se de um modo “vertical”, não saltando habitualmente nenhuma geração. Nas desordens AD, como na acondroplasia e nas doenças associadas ao gene FGFR pode-se confirmar a presença da mutação por PCR. Há também uma série de doenças AD caracterizadas por unidades de repetição anormal de nucleotídeos associada a doenças neuromusculares como a doença de Huntington. Outro exemplo é o síndrome de Marfan podendo ser feito o diagnóstico pela pesquisa da mutação do gene FBN1 localizado em 15q21.1 que codifica uma proteína do tecido conjuntivo, a fibrina. Esta tem uma frequência de 1 para 10.000 habitantes e cerca de 25% dos casos são devidos a mutações *de novo*.

Quanto às desordens ligadas ao cromossoma X, estas dizem respeito a doenças determinadas por genes localizados neste cromossoma. Neste caso, os homens são quase sempre os únicos afetados, sendo o alelo mutado transmitido aos filhos através de mulheres portadoras não afetadas. Duplicações de genes podem ser encontradas em distrofias musculares como a distrofia muscular de Becker ou de Duchenne. Neste caso, é necessário verificar se a mãe é portadora da mutação patogénica no gene DMD localizado em Xp21, já que os filhos do sexo masculino possuem um risco de 50% de ocorrência. Esta distrofia é das mais frequentes e habitualmente 1 em 3000 recém-nascidos do sexo masculino são afetados e 1 em 1500 mulheres são portadoras. A morte

ocorre antes dos 20 anos de idade, para a maioria dos doentes. São também exemplos de doenças relacionadas com o cromossoma X, a agamaglobulinémia, síndrome de Hunter e hemofilia A e B (Evans, 2009).

4.2. Diagnóstico oncológico

O cancro, como doença multifatorial, desenvolve-se pela interação de fatores ambientais e de fatores genéticos. Uma determinada mutação pode conferir suscetibilidade que apenas se exprime face à exposição a fatores desencadeantes ambientais (como o fumo do tabaco) devido a defeitos na reparação do DNA. Por outro lado, certas mutações presentes numa família (portanto hereditárias) criam um aumento de suscetibilidade, podendo a percentagem de indivíduos doentes aproximar-se de 50%. Deste modo torna-se importante a identificação de genes associados à suscetibilidade oncológica de cada indivíduo.

Há hoje em dia uma longa lista de cancros associados a mutações hereditárias específicas, sendo as mutações *APC*, *BRCA1* e *2*, *TP53*, *Rb1* e *KRAS* associadas a polipose adenomatosa familiar (FAP), cancros dos ovários e da mama, síndrome Li-Fraumeni, retinoblastoma e cancro do cólon, apenas alguns exemplos (Evans, 2009).

No caso do cancro do cólon não polipomatoso hereditário (HNPCC) ou no síndrome de Lynch a instabilidade de microssatélite ocorre em cerca de 95% dos casos e pode ser detetada por PCR (Evans, 2009). As mutações dos genes *MSH2*, *MLH1*, *PMS1*, *PMS2* e *MSH6* envolvidos na reparação de erros de emparelhamento, têm sido associados a esta patologia. Também se encontram associados a este síndrome tumores do estômago, ovário ou do endométrio. Os indivíduos são sujeitos a este tipo de diagnóstico genético quando possuem pelo menos três familiares com carcinoma colorrectal, dois familiares (ou um diagnosticado antes dos 50 anos) em primeiro grau com neoplasia relacionada

com o HNPCC, ou pelo menos duas gerações consecutivas afetadas. Caso o doente seja portador de mutações patogénicas é importante proceder a intervenções preventivas como a colectomia total profilática ou histerectomia total no caso de tumores do endométrio (Regateiro, 2003).

Os sarcomas e leucemias são classificadas em termos de translocações definitivas, o que a nível molecular significa a fusão de genes que levam à expressão de cópias anormais do mesmo. Nestes casos é usado o RT-PCR para a deteção destas construções quiméricas de mRNA, auxiliando assim o diagnóstico de tumores. Também no caso dos rhabdomiosarcomas alveolares o diagnóstico pode ser feito por esta técnica pela deteção do transcriptoma quimérico. Este tipo de sarcomas são caracterizados por rearranjos t(2:13) (q35;q14) que justapõem parte do gene *PAX3* no cromossoma 2 com o *FKHR* no cromossoma 13. Outros sarcomas como o sinovial e o sarcoma *Ewing* podem ser também diagnosticados por RT-PCR. As leucemias e linfomas são também possíveis de detetar por RT-PCR, a leucemia mieloide crónica (CML) por exemplo, é caracterizada pelo rearranjo t(9;22) (q34;q11), do qual resulta a fusão do transcriptoma quimérico dos genes *BRC* e *ABL1* (Evans, 2009).

No caso de membros das famílias com história de FAP, a deteção molecular da mutação no gene APC, em linfócitos de sangue periférico, deve ser feita logo nos primeiros anos de vida, já que cerca de 100% dos doentes desenvolvem cancro colorrectal. Aqui, a doença começa a manifestar-se pela hiperproliferação do epitélio do cólon com subsequente mutação do protooncogene KRAS. Após o desenvolvimento do adenoma classe II dá-se a inativação do antioncogene DCC facilitando a mutação de outro antioncogene, o TP53, resultando em carcinoma e posteriormente em metástases. Neste caso em particular, o diagnóstico molecular é imprescindível para a deteção de

portadores, possibilitando uma intervenção precoce em termos de profilaxia e tratamento. (Regateiro, 2003).

No caso do cancro da mama existem duas formas: (1) – esporádica, responsável pela grande maioria dos tumores; e (2) – hereditária, responsável por 5 – 10% de todos os cancros da mama e a maioria antes dos 50 anos. Existem diversos genes relacionados com o cancro da mama hereditário, sendo os mais frequentes as mutações em *BRCA1* (mama 85%, ovário 45%) e *BRCA2* (mama 85%, ovário 15%). Após a avaliação do risco individual dá-se início ao rastreio e ao aconselhamento médico que pode passar pela profilaxia cirúrgica, com mastectomia total bilateral (opção a considerar a partir dos 35 anos), salpingo-ooforectomia bilateral (no caso de tumores nos ovários) ou quimioprevenção (mama – tamoxifeno; ovário – contraceção oral).

Depois da obtenção dos resultados analíticos que coloquem indivíduos perante um elevado risco de cancro, é necessária a interpretação integrada de cada caso no âmbito de uma consulta de Genética Tumoral.

Deve-se assegurar o acompanhamento e o apoio psicológico aos pacientes para que estes tomem medidas de prevenção mesmo sendo indivíduos saudáveis (Regateiro, 2003).

4.3. Farmacogenómica

Nos tempos que correm, cada vez mais se reconhece a importância das variações genéticas que alteraram a efetividade de certos fármacos. Diferenças hereditárias que afetam o seu efeito terapêutico foram documentadas pela primeira vez nos anos 50, dando sentido à expressão “farmacogenómica”. Esta área foi evoluindo ao longo das décadas seguintes e abrange agora todos os aspetos farmacocinéticos dos fármacos

(absorção, distribuição, metabolização e excreção) assim como os seus locais alvo (Day *et al.*, 2013).

A evolução de estratégias de sequenciação e a diminuição dos custos associados à sua prática tem proporcionado um aumento da sua utilização em oncologia. Nesta área, o tratamento era feito convencionalmente com base em estudos de efetividade na população em geral, e não tinha em conta a variação interindividual. A análise do DNA de tumores como auxílio na escolha do tratamento é uma estratégia que foi iniciada há mais de 20 anos. É cada vez mais comum os pacientes serem sujeitos a terapias específicas e individualizadas, com base na presença ou ausência de uma mutação específica. Isto tem proporcionado o controlo de vários cancros, incluindo mama, pulmão e cólon, tem também aumentado o tempo de vida e proporcionado a cura de cancros testiculares e leucemias linfoblásticas (McLeod, 2013).

Variações somáticas, variações encontradas no tumor, mutações germinativas e variações hereditárias encontradas em certos indivíduos, podem influenciar o prognóstico da doença e a resposta à terapia. Estas mutações ou biomarcadores tumorais podem ser classificados como marcadores de prognóstico ou marcadores de previsão, e são usados para identificar subpopulações de pacientes que são mais propensos a responder a uma dada terapia (Nelson *et al.*, 2012).

Desde o desenvolvimento bem sucedido do Imatinib para o tratamento de leucemia mieloide crónica, outros inibidores tirosina quinase têm sido desenvolvidos. Também no cancro da mama, um paciente com amplificação do gene *HER2* pode ser tratada com trastuzumabe, anticorpos monoclonais anti-*HER2*, ou com lapatinib, um inibidor da tirosina quinase do *HER2*. Tumores sólidos são agora rotineiramente analisados para a

pesquisa de mutações nos oncogenes como por exemplo nos *EGFR*, *PIK3CA* e *KRAS* já que eles preveem também a resposta a regimes de quimioterapia (Day *et al.*, 2013).

O acesso às mutações somáticas do DNA pode ser feito por PCR digital e tem tido um impacto muito positivo no tratamento dos doentes oncológicos. A identificação de mutações do *KRAS* nos codões 12 ou 13 em 30% dos pacientes com cancro do cólon sugere que não há qualquer benefício na terapia com anticorpos específicos para o recetor *EGFR*, um fator de crescimento epidermal. Por outro lado, o cancro do pulmão, os melanomas e doenças mieloproliferativas tendem a ser sensíveis a inibidores tirosina quinase quando estas possuem mutações nos genes *EGFR*, *BRAF* e *JAK2*, respetivamente (Mcleod, 2013). Também a discriminação alélica por qPCR com sondas *Taq-man* tem sido usada para analisar polimorfismos em enzimas que metabolizam fármacos e a suscetibilidade individual, como as glutatona S-transferases (*GSTM1*, *GSTT1*, *GSTP1*) e citocromos P450 (*CYP2C*, *CYP2D*), aldeído desidrogenase, N-acetiltransferase e triopurina metiltransferase (Walker, 2001).

Já que muitos dos mecanismos genéticos de resistência a fármacos são desconhecidos, a farmacogenómica clínica assume um papel fundamental no seu estudo e identificação, através da sequenciação de exomas específicos. A identificação da resistência a inibidores da tirosina quinase EGFR no cancro do pulmão, levou à implementação da sequenciação rotineira do gene EGFR para orientar as escolhas terapêuticas. Por outro lado, a identificação da mutação TSC1 que confere sensibilidade ao everolimus, permitiu a aplicação deste fármaco a casos de cancro da bexiga em que a mesma se encontra presente (Mcleod, 2013).

Apesar de tudo, apenas um número limitado de ensaios gene-específicos estão aprovados para uso clínico de rotina pela FDA – Food and Drug Administration

(McLeod, 2013). Adicionalmente, as células cancerígenas podem mutar e desenvolver resistência específica para os novos fármacos, resultando na proliferação de tumores multirresistentes. Esta heterogeneidade tumoral apresenta um novo desafio à noção de farmacoterapia individualizada (Day *et al.*, 2013).

4.4. Diagnóstico de doenças infecciosas

O diagnóstico molecular está a revolucionar a prática clínica em infeciologia. Os seus efeitos benéficos são significativos em ambientes de cuidados intensivos, onde as ferramentas de diagnóstico possuem a rapidez e precisão necessária para auxiliar a escolha da terapia e o prognóstico. A sequenciação e o PCR tem provado a sua aplicabilidade na clínica, proporcionando a deteção específica de agentes patogénicos, a avaliação de novas infeções emergentes, vigilância, deteção precoce de ameaças biológicas e perfis de resistência a antibióticos. Adicionalmente, a evolução das técnicas automatizadas permite uma diminuição do custo de cada análise, um aumento na precisão e especificidade, rapidez e capacidade de deteção de múltiplos alvos simultaneamente (Yang & Rothman, 2004).

O poder do PCR em tempo real cumpre-se na deteção e identificação de micro-organismos difíceis de semear *in vitro* ou que possuem um tempo de incubação demasiado extenso para um diagnóstico urgente. Por exemplo no caso da *Mycobacterium tuberculosis*, *Legionella pneumophila*, *Listeria monocytogenes* e *Neisseria gonorrhoeae*. Anteriormente, no caso de suspeita de infeções provocadas por este tipo de micro-organismos, o diagnóstico era feito retrospectivamente, ou seja, pela resposta à antibioterapia.

Também na análise de mutações que conferem resistência a antibióticos, em bactérias como a *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Helicobacter pylori*,

Enterococcus faecalis e *Enterococcus faecium* as técnicas moleculares têm sido indispensáveis. Não esquecendo também a capacidade de identificação de esporos como no caso da bactéria *Bacillus anthracis*, agente etiológico do antraz, uma potencial arma biológica (Valasek & Repa, 2005). Apesar de todas estas vantagens apenas um número limitado de métodos microbiológicos foram aprovados pela FDA (Yang & Rothman, 2004).

Uma detecção rápida e precoce permite a prescrição de uma antibioterapia direcionada e pode a longo prazo ajudar a reduzir o uso de antibióticos de largo espectro, diminuindo assim o problema do aparecimento de estirpes multirresistentes (Gilbert, 2002). Contudo, o aparecimento de genes que conferem determinada resistência a antibióticos continua a aumentar tornando-se cada vez mais um problema de saúde pública. Deste modo, foram iniciados programas de mapeamento de genes de resistência de modo a proporcionar à população uma terapêutica eficaz e personalizada, revolucionando o sistema de vigilância de saúde pública.

Adicionalmente, é necessária uma correta interpretação dos resultados. A presença de um determinado gene de resistência não implica a sua expressão e consequente resistência, significa sim, que quando este gene não está presente o micro-organismos é suscetível a um determinado antibiótico. Por exemplo, a resistência à meticilina é mediada pelo gene *mecA*. Uma característica distinta da resistência à meticilina é a sua expressão heterogênea, deste modo, quando se elaboram testes para avaliar a resistência de uma dada bactéria, esta pode parecer falsamente suscetível a certos antibióticos β -lactâmicos *in vitro*. Por esta razão é preferível a detecção direta do gene *mecA* por PCR, eliminando a necessidade de testes *in vitro* que na maioria dos casos fornecem resultados falsos que dificultam o tratamento do doente. Deste modo, o PCR tornou-se o

método de eleição para a detecção de MSRA (*Staphylococcus aureus* resistentes à meticilina) e para a identificação de resistências da *M tuberculosis* à rifampicina (Yang & Rothman, 2004).

Em virologia, o PCR permite determinar a carga viral, oferece a possibilidade de determinar a dinâmica da proliferação vírica, monitorizar a resposta ao tratamento e a distinção entre infecção ativa ou latente. Comercialmente já existe uma grande variedade de kits de PCR para análise qualitativa e quantitativa de vírus. (Watzinger *et al.*, 2006).

A tecnologia de PCR em tempo real permite a detecção e quantificação de vírus presentes em diferentes tipos de amostras: plasma, sangue periférico, líquido cefalorraquidiano, células mononucleares, saliva, expetoração, urina, fezes, entre outros tecidos sólidos. O isolamento do material genético, DNA ou do RNA viral, é feito inicialmente por meio de kits de extração de ácidos nucleicos. Não só o qPCR mas também o RT-PCR permitem que a análise seja também feita em tecidos embebidos em parafina, água, plantas e animais.

Devido à sua diversidade de aplicações, o RT-PCR deteta diversos vírus de RNA, como o vírus da imunodeficiência humana (HIV), o vírus da hepatite C (HCV), o vírus T-linfotrófico humano (HTLV-1 e HTLV-2), enterovírus (EV), vírus sincicial respiratório (RSV), o vírus da influenza (Inf-A e Inf-B), rotavírus e outros vírus que causam gastroenterites. O RT-PCR pode também ser empregue na detecção de RNA mensageiro, útil no diagnóstico de vírus de DNA que possuem uma fase latente durante o seu ciclo (Watzinger *et al.*, 2006).

Em análises de amostras reduzidas, o qPCR alia-se ao PCR multiplex para a detecção simultânea de mais de um alvo, principalmente em amostras de líquido cefalorraquidiano onde se suspeita da existência de múltiplos vírus neuro-invasivos.

Pelo mesmo procedimento é possível detetar o vírus da hepatite B e C (HBV, HCV), o HIV-1, vírus da parainfluenza (PIV1-4), o herpes simplex vírus (HSV), o citomegalovírus (CMV), o vírus da varicela zóster (VZV), o vírus de herpes humano 6 (HHV-6), adenovírus, o vírus de Epstein-Barr (EBV) ou o vírus da raiva (Elnifro *et al.*, 2000).

A deteção da carga viral é essencial para determinar a resposta à terapia, assim como para a identificação de estirpes resistentes, crucial em tratamentos de infeções provocadas pelo vírus da hepatite C e B ou na infeção por HIV. Assim, a execução rotineira de qPCRs seria vantajosa (Watzinger *et al.*, 2006).

Atualmente um dos maiores problemas das terapias antiretrovirais, é a elevada ocorrência de mutações no vírus do HIV que conferem resistência mesmo a fármacos de última linha. É crucial nestes casos amplificar e sequenciar o RNA viral para identificar estas mutações. É feito de seguida um plano terapêutico individualizado de modo a serem obtidos os melhores resultados. Todo este processo pode ser feito hoje em dia por RT-PCR seguido de sequenciação ou por PCR digital são detetadas e quantificadas as mutações raras mas mais significativas para a terapêutica (Yang & Rothman, 2004). Também no caso do vírus do HPV (*human papillomavirus*), é crucial aplicação de métodos moleculares para fenotipagem. Com identificação da estirpe presente na amostra citológica é possível avaliar o risco de cancro do colo uterino já que as estirpes 16, 18, 31 ou 45 são consideradas de risco elevado (Tsongalis & Silverman, 2006).

Um sistema sólido e universal de diagnóstico molecular de doenças infecciosas vai tornar-se uma ferramenta indispensável para a clínica do século XXI. Tanto em cenários clínicos ou cenários de catástrofe, ensaios rápidos e fiáveis serão indispensáveis para determinar a avaliação do risco dependendo do agente causador e da sua virulência. A

gestão de recursos será facilitada e permitirá uma triagem rápida de modo a ser administrada a terapia mais eficaz (Yang & Rothman, 2004).

4.5. Metagenômica

A flora humana, o vasto número de micro-organismos que vivem em estreito equilíbrio com o nosso organismo, é um campo importante mas ainda pouco explorado, passível de uma análise detalhada com vista a futuras aplicabilidades farmacológicas. A metagenômica ou genômica ambiental consiste na análise genômica de comunidades complexas de micro-organismos em contextos ecológicos variados (Dupre & O'malley, 2007).

Estudos recentes sugerem que novas estratégias terapêuticas e diagnósticas podem surgir pela compreensão da habilidade da flora comensal em afetar o metabolismo e predispor ao desenvolvimento de doenças. Aliando o conhecimento do genoma humano e dos fatores de risco ambientais, é possível compreender a sua variação interindividual. Nem sempre os biomarcadores clínicos e genéticos conseguem explicar a resposta terapêutica do doente. Um aspeto negligenciado mas crucial para o metabolismo de xenobióticos é a influência dos biliões de micro-organismos que habitam o nosso trato gastrointestinal, não só aumentando e diversificando as enzimas metabólicas como também alterando a expressão genética do hospedeiro. Esta interação hospedeiro – flora é recíproca, já que a exposição a xenobióticos, especialmente antibióticos, podem afetar a estrutura das nossas comunidades microbianas (Haiser, 2012).

Distúrbios na flora geralmente removem ou matam uma fração da comunidade, fornecendo uma oportunidade para novos colonizadores aumentarem os seus números. O efeito dos antibióticos na flora intestinal serve de paradigma para estes distúrbios. Durante as últimas décadas a resistência a antibióticos tem crescido, e a preocupação

em encontrar novos e mais eficazes antibióticos tem também aumentado. Compreende-se agora o efeito que a antibioterapia pode ter sobre a flora intestinal, perturbando a interação entre o hospedeiro e os micro-organismos, e afetando a saúde e o bem estar do primeiro. No entanto, a resistência a antibióticos pode ser benéfica caso seja desenvolvida pela própria flora comensal. Mais concretamente, uma vez sujeita a terapia antimicrobiana, a flora habitual é dizimada, o que perturba a saúde do doente mesmo depois de o organismo patogénico alvo ser eliminado. Caso essa mesma flora desenvolvesse resistência à terapêutica antimicrobiana, esta poderia ser aplicada sem afetar o equilíbrio normal do doente (Costello *et al.*, 2012).

Contudo, a diversidade microbiana não está limitada só às bactérias. Outros micro-organismos, como os fungos, têm também papéis importantes na estabilidade das comunidades microbianas e na saúde humana. As metodologias de identificação de fungos têm sido limitadas, em comparação com as técnicas disponíveis para bactérias. Desde a descoberta da candidíase oral em 400 a.C. por Hipócrates, a comunidade científica começou a explorar o papel que a flora comensal tem na saúde e na doença. Estudos feitos em meios de cultura demonstraram que a *Malassezia*, *Rhodotorula*, *Debaromyces*, *Cryptococcus* e a *Candida* fazem parte da flora comensal fúngica. Infecções fúngicas cutâneas afetam um elevado número de indivíduos em todo o mundo, mas o papel destes dermatófilos em infeções comuns das unhas dos pés pode tornar difícil a identificação da flora recorrendo apenas a meios de cultura. Pela dificuldade em estabelecer condições ótimas de crescimento, os meios de cultura têm vindo a ser substituídos por métodos de sequenciação (Findley *et al.*, 2013).

Estudos como o de Findley, *et al* (2013) demonstram a complexidade dos ecossistemas da pele. Diferentes fatores determinam que tipo de micro-organismos vão habitar a

nossa pele, dependendo muitas vezes das suas propriedades fisiológicas. Por exemplo, espécies de *Malassezia* predominam em áreas do tronco e braços. Em contraste, os pés apresentam uma diversidade fúngica tremenda e possuem menor estabilidade, proporcionando uma oportunidade para fungos patogénicos criarem doença. Os pés, seja calcanhares, espaços interdigitais ou unhas, são locais comuns de infeções recorrentes de fungos. Também é relatado no estudo de Findley, *et al* a diversidade fúngica em locais anatómicos propensos ao desenvolvimento de outras infeções cutâneas, incluindo dermatites seborreicas, dermatites atópicas e *tinea cruris*. Com os avanços nas técnicas de sequenciação é possível compreender as interações entre os micro-organismos (bactéria – fungo; bactéria – bactéria; fungo – fungo) que residem nestes meios de elevada diversidade. Mais ensaios em pele sã devem ser feitos assim como estudos sobre doenças dermatológicas devem ser realizados de modo a explorar as interações hospedeiro – micro-organismo. Adicionalmente, novos tipos de tratamento devem ser desenvolvidos de modo a se poderem substituir os azoles, as equinocandinas e a amfotericina B, terapêuticas antifúngicas com efeitos adversos significativos (danos hepáticos e urológicos).

5. Conclusão

A evolução tecnológica, a par dos numerosos avanços científicos dos últimos 60 anos, têm transformado o modo como abordamos o mundo e os seus sistemas biológicos. O que inicialmente requeria um ambicioso investimento material e humano está hoje disponível de um modo surpreendentemente simples e acessível, passível de ser posto ao alcance da prática clínica e da saúde das comunidades. As imensas possibilidades dos métodos moleculares não se esgotam nos pontos abordados neste trabalho, e novas aplicações estão ainda por ser reveladas.

O PCR digital e a sequenciação de 3ª geração, são neste momento a promessa de um futuro próspero na área da investigação clínica. É importante que os profissionais de saúde estejam conscientes das vantagens associadas à utilização destas tecnologias e das possibilidades diagnósticas e terapêuticas a elas inerentes.

Tanto o desafio lançado pela National Human Genome Research Institute, como o empenho da comunidade científica em colocar as suas descobertas ao serviço da humanidade, impulsionaram o desenvolvimento de métodos de PCR e de sequenciação mais poderosos, baratos e acessíveis. A redução dos custos na sequenciação de DNA, em várias ordens de grandeza, está a democratizar a sua utilização. Assim, investigadores em todo o mundo têm a oportunidade de por em prática os seus projetos numa escala anteriormente possível apenas em grandes centros genómicos

Uma vez compreendida a influencia da diversidade humana no comportamento das doenças, foi possível dar inicio a uma nova era na investigação clínica, onde o PCR e a sequenciação são a chave para o desenvolvimento de terapêuticas direcionadas e individualizadas às exigências genéticas de cada individuo.

A redução dos custos e o aumento da mobilidade dos equipamentos abrem a porta ao diagnóstico de doenças e ao arranque de novos e mais detalhados estudos epidemiológicos em áreas remotas. Esta facilidade de acesso será um ponto de partida para a transferência de tecnologias científicas para países em desenvolvimento, respondendo às duas necessidades prementes.

Espera-se também que estas tecnologias mais avançadas e baratas entrem em meios educacionais, permitindo aos educadores científicos impulsionar e encorajar os seus alunos.

Apesar de ser difícil de prever o que o futuro reserva no campo da genética, é possível antever que o desenvolvimento tecnológico não estancará por aqui. Novos métodos e novas aplicações surgirão, ajudando-nos a desvendar definitivamente os segredos da célula.

6. Bibliografia

- Anderson, S., Bankier, A. T., Barrell, B. G., De Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R. e Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, 290, 457-65.
- Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., Gibson, T. J., Hatfull, G., Hudson, G. S., Satchwell, S. C., Seguin, C. e Et Al. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature*, 310, 207-11.
- Bermingham, N., Luettich, K. (2003). Polymerase chain reaction and its applications. *Current Diagnostic Pathology*, 159 - 164.
- Bernard, P. 2001. Mutation Detection by Fluorescent Hybridization Probe Melting Curves. *Rapid Cycle Real-Time PCR*. Salt Lake City, Utah, USA: Springer-Verlag Berlin Heidelberg New York.
- Carneiro, M. O., Russ, C., Ross, M. G., Gabriel, S. B., Nusbaum, C. e Depristo, M. A. (2012). Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics*, 13, 375.
- Chamberlain, J. S., Gibbs, R. A., Ranier, J. E., Nguyen, P. N. e Caskey, C. T. (1988). Deletion Screening of the Duchenne Muscular-Dystrophy Locus Via Multiplex DNA Amplification. *Nucleic Acids Research*, 16, 11141-11156.
- Chan, E. Y. (2005). Advances in sequencing technology. *Mutat Res*, 573, 13-40.
- Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. e Relman, D. A. (2012). The application of ecological theory toward an understanding of the human microbiome. *Science*, 336, 1255-62.
- Day, E., Dear, P. H. e Mccaughan, F. (2013). Digital PCR strategies in the development and analysis of molecular biomarkers for personalized medicine. *Methods*, 59, 101-7.
- Demidov, V. V. (2003). DNA diagnostics in the fifty-year retrospect. *Expert Rev Mol Diagn*, 3, 121-4.
- Dupre, J. e O'malley, M. A. (2007). Metagenomics and biological ontology. *Stud Hist Philos Biol Biomed Sci*, 38, 834-46.
- Elnifro, E. M., Ashshi, A. M., Cooper, R. J. e Klapper, P. E. (2000). Multiplex PCR: Optimization and application in diagnostic virology. *Clinical Microbiology Reviews*, 13, 559-+.
- Evans, M. (2009). The polymerase chain reaction and pathology practice. *Diagnostic Histopathology*, 15, 334 - 356.
- Findley, K., Oh, J., Yang, J., Conlan, S., Deming, C., Meyer, J. A., Schoenfeld, D., Nomicos, E., Park, M., Kong, H. H. e Segre, J. A. (2013). Topographic diversity of fungal and bacterial communities in human skin. *Nature*, 498, 367-70.
- Gibson, U. E., Heid, C. A. e Williams, P. M. (1996). A novel method for real time quantitative RT-PCR. *Genome Res*, 6, 995-1001.
- Gilbert, G. L. (2002). Molecular diagnostics in infectious diseases and public health microbiology: cottage industry to postgenomics. *Trends in Molecular Medicine*, 8, 280-7.
- Gupta, P. K. (2008). Single-molecule DNA sequencing technologies for future genomics research. *Trends Biotechnol*, 26, 602-11.

-
- Guyer, R. L. E. K. J., D. E. 2007. The Molecule of the Year. *In: Science*, T. a. a. F. T. a. O. (ed.) *PCR Colelection*. Washington D.C.: AAAS / Science Business Office.
- Haiser, H. T., P. 2012. Is It Time for a Metagenomic Basis of Therapeutics? *Science*.
- Heid, C. A., Stevens, J., Livak, K. J. e Williams, P. M. (1996). Real time quantitative PCR. *Genome Res*, 6, 986-94.
- Higuchi, R., Dollinger, G., Walsh, P. S. e Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. *Biotechnology (N Y)*, 10, 413-7.
- Hoorfar, J., Malorny, B., Abdulmawjood, A., Cook, N., Wagner, M. e Fach, P. (2004). Practical considerations in design of internal amplification controls for diagnostic PCR assays. *Journal of Clinical Microbiology*, 42, 1863-8.
- Hutchison, C. A., 3rd (2007). DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res*, 35, 6227-37.
- Irving, R. a. H., P. J. (1999). Molecular diagnostics: state of the art. *Molecular Medicine Today*, 5.
- Joshi, M. e Deshpande, J. D. (2010). Polymerase Chain Reaction: Methods, Principles and Application. *International Journal of Biomedical Research*, 81-97.
- Kircher, M. e Kelso, J. (2010). High-throughput DNA sequencing--concepts and limitations. *Bioessays*, 32, 524-36.
- Klein, D. (2002). Quantification using real-time PCR technology: applications and limitations. *Trends in Molecular Medicine*, 8, 257-260.
- Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., Sindelka, R., Sjoberg, R., Sjogreen, B., Strombom, L., Stahlberg, A. e Zoric, N. (2006). The real-time polymerase chain reaction. *Mol Aspects Med*, 27, 95-125.
- Life Technologies [Em Linha]. Disponível em <www.lifetechnologies.com>. [Consultado em 17/7/2013]
- Mardis, E. R. (2008a). The impact of next-generation sequencing technology on genetics. *Trends Genet*, 24, 133-41.
- Mardis, E. R. (2008b). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*, 9, 387-402.
- Markoulatos, P., Siafakas, N. e Moncany, M. (2002). Multiplex polymerase chain reaction: A practical approach. *Journal of Clinical Laboratory Analysis*, 16, 47-51.
- Maxam, A. M. e Gilbert, W. (1977). A new method for sequencing DNA. *Proc Natl Acad Sci U S A*, 74, 560-4.
- Mc Pherson, J., Moller, S. 2006. *PCR second edition*, Taylor & Francis Group.
- Mccarthy, A. (2010). Third generation DNA sequencing: pacific biosciences' single molecule real time technology. *Chem Biol*, 17, 675-6.
- Mcginn, S. e Gut, I. G. (2012). DNA sequencing - spanning the generations. *N Biotechnol*.
- Mcleod, H. L. (2013). Cancer pharmacogenomics: early promise, but concerted effort needed. *Science*, 339, 1563-6.
- Messmer, T., Tully, T. N., Ritchie, B. W. e Moroney, J. F. (2000). A tale of discrimination: Differentiation of Chlamydiaceae by polymerase chain reaction. *Seminars in Avian and Exotic Pet Medicine*, 9, 36-42.
- Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Review Genetics*, 11, 31-46.

-
- Moorthie, S., Mattocks, C. J. e Wright, C. F. (2011). Review of massively parallel DNA sequencing technologies. *Hugo J*, 5, 1-12.
- Nelson, M. R., Wegmann, D., Ehm, M. G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S. A., Fraser, D., Warren, L., Aponte, J., Zawistowski, M., Liu, X., Zhang, H., Zhang, Y., Li, J., Li, Y., Li, L., Woollard, P., Topp, S., Hall, M. D., Nangle, K., Wang, J., Abecasis, G., Cardon, L. R., Zollner, S., Whittaker, J. C., Chissoe, S. L., Novembre, J. e Mooser, V. (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science*, 337, 100-4.
- Niedringhaus, T. P., Milanova, D., Kerby, M. B., Snyder, M. P. e Barron, A. E. (2011). Landscape of next-generation sequencing technologies. *Anal Chem*, 83, 4327-41.
- Oxford Nanopore Technologies [Em Linha]. Disponível em <www.nanoporetech.com>. [Consultado em 17/7/2013]
- Pacific Biosciences [Em Linha]. Disponível em <www.pacificbiosciences.com>. [Consultado em 20/6/2013]
- Pareek, C. S., Smoczynski, R. e Tretyn, A. (2011). Sequencing technologies and genome sequencing. *J Appl Genet*, 52, 413-35.
- Pohl, G. e Shih Ie, M. (2004). Principle and applications of digital PCR. *Expert Review Molecular Diagnostics*, 4, 41-7.
- Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P. e Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13, 341.
- Ravikumar, G., Urs, S. R., Prakash, N. B. V., Rao, C. G. P. e Vardhana, K. V. (2011). Development of a multiplex polymerase chain reaction for the simultaneous detection of microsporidians, nucleopolyhedrovirus, and densovirus infecting silkworms. *Journal of Invertebrate Pathology*, 107, 193-197.
- Regateiro, J. 2003. *Manual de Genética Médica*, Coimbra, Imprensa da Universidade de Coimbra.
- Roche 2003. The Evolution of PCR. California: Roche Molecular Systems.
- Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. e Nyren, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem*, 242, 84-9.
- Sanders, S. 2007. Setting off a Chain Reaction. *PCR collection*. Washington D. C.: American Association for the Advancement of Science.
- Sanger, F. (1975). The Croonian Lecture, 1975. Nucleotide sequences in DNA. *Proc R Soc Lond B Biol Sci*, 191, 317-33.
- Sanger, F. e Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94, 441-8.
- Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. e Petersen, G. B. (1982). Nucleotide sequence of bacteriophage lambda DNA. *Journal of Molecular Biology*, 162, 729-73.
- Sanger, F., Nicklen, S. e Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceeding of the National Academy of Sciences U S A*, 74, 5463-7.

-
- Shendure, J. e Ji, H. (2008). Next-generation DNA sequencing. *Natural Biotechnology*, 26, 1135-45.
- Shuber, A. P., Grondin, V. J. e Klinger, K. W. (1995). A simplified procedure for developing multiplex PCRs. *Genome Res*, 5, 488-93.
- Smith, C. J. e Osborn, A. M. (2009). Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol Ecol*, 67, 6-20.
- Smith, L. M., Kaiser, R. J., Sanders, J. Z. e Hood, L. E. (1987). The synthesis and use of fluorescent oligonucleotides in DNA sequence analysis. *Methods Enzymol*, 155, 260-301.
- Tsongalis, G. J. e Silverman, L. M. (2006). Molecular diagnostics: A historical perspective.
- Valasek, M. A. e Repa, J. J. (2005). The power of real-time PCR. *Adv Physiol Educ*, 29, 151-9.
- Van Gelder, R. N. (2001). Applications of the polymerase chain reaction to diagnosis of ophthalmic disease. *Surv Ophthalmol*, 46, 248-58.
- Walker, N. J. (2001). Real-time and quantitative PCR: applications to mechanism-based toxicology. *J Biochem Mol Toxicol*, 15, 121-7.
- Walker, N. J. 2007. A Technique Whose Time Has Come. *PCR Collection*. Washington D. C.: The American Association for the Advancement of Science.
- Watzinger, F., Ebner, K. e Lion, T. (2006). Detection and monitoring of virus infections by real-time PCR. *Mol Aspects Med*, 27, 254-98.
- Yang, S. e Rothman, R. E. (2004). PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *Lancet Infect Dis*, 4, 337-48.