

Fabiana Raquel Coutinho Moreira de Oliveira

As técnicas de Genética Molecular no desenvolvimento e prescrição de fármacos

Universidade Fernando Pessoa

Porto, 2017

Fabiana Raquel Coutinho Moreira de Oliveira

As técnicas de Genética Molecular no desenvolvimento e prescrição de fármacos

Universidade Fernando Pessoa

Porto, 2017

© 2017
Fabiana Raquel Coutinho Moreira de Oliveira
TODOS OS DIREITOS RESERVADOS

Fabiana Raquel Coutinho Moreira de Oliveira

As técnicas de Genética Molecular no desenvolvimento e prescrição de fármacos

Fabiana Raquel Coutinho Moreira de Oliveira

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

Sumário

Desde a descoberta da molécula de DNA, como código universal de toda a vida biológica na Terra, a ciência procurou maneiras de sequenciar ou ler a informação genética para desvendar os muitos mistérios relacionados com a nossa origem. Esta procura de respostas permitiu desenvolver novas formas tecnológicas que tornaram possível a sequenciação da informação genética. Tais métodos revolucionaram os vários campos científicos, e ainda mantêm muitas promessas.

A reação em cadeia da polimerase (PCR) e *Next-generation Sequencing* (NGS) são as tecnologias mais utilizadas hoje em dia, e mostraram ser os pilares essenciais da prática clínica. O seu contínuo desenvolvimento e inovação reduziram significativamente o custo da sequenciação do DNA e, como tal, estão a tornar-se cada vez mais proeminentes nos cuidados de saúde, ao estabelecer uma nova ordem de como praticamos medicina e de como as indústrias farmacêuticas podem utilizar o seu potencial no desenvolvimento de medicamentos de forma a maximizar os seus lucros, maximizando simultaneamente o potencial terapêutico dos fármacos e minimizando as reações adversas que, de acordo com a administração dos alimentos e das drogas (FDA), são a principal causa de morte relacionada com fármacos.

Os esforços têm sido crescentes para incluir a farmacogenómica nos ensaios clínicos do desenvolvimento de fármacos, especialmente nas fases 2 e 3, onde as taxas de sucesso são muito baixas, tendo um impacto económico nas empresas farmacêuticas. A FDA e a agência europeia do medicamento (EMA) incentivam a sua integração nos programas de desenvolvimento clínico, incluindo o uso de biomarcadores que podem ter peso nas decisões clínicas, abrindo assim portas para a integração da medicina personalizada.

É crucial compreender como cada técnica funciona e pretende-se esclarecer o leitor sobre as vantagens e limitações associadas a essas tecnologias, fornecendo um breve vislumbre de algumas barreiras atuais que limitam o pleno potencial da farmacogenómica.

Palavras-chave: PCR, NGS, farmacognética, farmacogenómica, SNPs, CYP450

Abstract

Since the discovery of the DNA molecule as the universal code of all biological life on Earth, science has searched for ways to sequence or read that genetic information to unveil the many mysteries related to its origin. This search for answers allowed the development of new techniques that made sequencing genetic information possible. Such methods revolutionized many scientific fields, and still hold many promises.

The polymerase chain reaction (PCR) and Next-generation sequencing (NGS) are the most used technologies today and have proven to be essential pillars in clinical practice. Their continuous development and innovation have reduced significantly the cost of DNA sequencing and as such, they are becoming more prominent in healthcare by establishing a new order of how we perform medicine, and how pharmaceutical industries may use their potential in drug development to maximize their profits while at the same time maximizing the usefulness of drugs and minimizing adverse reactions which, according to the food and drugs administration (FDA), are the major cause of drug related deaths.

To this aim, there have been growing efforts to include pharmacogenomics in clinical trials for drug development, especially in Phase 2 and 3 where the success rates are very low, impacting drug companies economically. FDA and european medicines agency (EMA) encourage its integration in clinical development programs, including the use of biomarkers that may weight in clinical decisions thus opening doors for the integration of personalized medicine.

However, it is crucial to understand how each technique works and enlighten the reader about the advantages and limitations associated with these technologies and give a brief glimpse of some current barriers that limit the full potential of pharmacogenomics.

Keywords: PCR, NGS, pharmacogenetics, pharmacogenomics, SNPs, CYP450

Dedicatória

Aos meus pais, António e Filomena
ao meu irmão, Jorge Miguel e aos
meus avós

Agradecimentos

Apraz-me agradecer ao Professor Doutor José Manuel Cabeda, que empenhadamente me orientou e ajudou na elaboração desta monografia, contribuindo com o seu vasto conhecimento para a execução do meu trabalho, de forma a fornecer-me total autonomia nas minhas investigações e ideias, sempre, evidentemente, dentro dos limites por ele estabelecidos. Em primeiro lugar, dedico o meu trabalho e a minha realização pessoal ao Professor Doutor José Manuel Cabeda que foi capaz de me proporcionar.

À minha mãe agradeço também a sua disponibilidade, ainda que à distância, noutro país, motivando-me e ajudando-me a dar o meu melhor, contribuindo para o crescente aumento do meu interesse nesta área da ciência ainda por concluir. A ela dedico este trabalho, que lhe proporcionou momentos, descritos por ela, “de muito interessantes” sob o ponto de vista pessoal e profissional, uma vez que também é professora de filosofia.

Ao meu pai, agradeço o seu esforço para a minha realização pessoal e o seu apoio incondicional no meu percurso académico que se realizou nesta faculdade.

Ao meu irmão, com quem eu vivo, embora mais novo, também vão os meus agradecimentos pela sua compreensão nos momentos mais tensos, ajudando-me a encontrar uma atmosfera familiar serena e consentânea, tomando a seu cargo as tarefas de rotina, libertando-me de algumas que me tomavam muito tempo.

À amiga de família, Cláudia Gomes, também docente, lhe presto o meu apreço e carinho por ter contribuído, presencialmente, para a minha motivação, empenho e dedicação na busca de um maior conhecimento.

Por último, agradeço à Universidade Fernando Pessoa, pela prestação do ensino de qualidade que os professores desta instituição me proporcionaram e a todos aqueles que nela trabalham e com quem eu me relacionei nestes anos de frequência inesquecíveis.

ÍNDICE GERAL

SUMÁRIO

ABSTRACT

DEDICATÓRIA

AGRADECIMENTOS

ÍNDICE GERAL

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

LISTA DE ABREVIATURAS

1. Introdução	1
1.1. Farmacogenómica e farmacogenética: uma medicina personalizada	1
2. Single Nucleotide Polimorphisms – A sua importância clínica	3
2.1. Tipos de SNPs	4
2.2. Efeitos clínicos dos SNPs	5
3. Citocromo P450	6
3.1. Citocromo P450: Uma perspectiva evolucionária.	7
3.2. Nomenclatura	9
3.3. Fatores que afetam a sua atividade e a sua função metabólica	9
4. A farmacogenómica no tratamento clínico	11
5. O contributo científico da PCR	19
6. Técnicas de sequenciação do DNA	23
6.1. Sequenciação de 1ª geração	23
6.1.1. Aplicação clínica	26
6.1.2. Limitações técnicas	27
6.2. Sequenciação de 2ª geração – Next-generation Sequencing	29
6.2.1. Aplicação clínica	40
6.2.2. Limitações técnicas	41
6.3. Sequenciação de 3ª e 4ª geração	44
6.3.1. Aplicação clínica	51
6.3.2. Limitações técnicas	52
7. Conclusão	54
8. Bibliografia	58

Índice de Figuras

Figura 1 – Sequenciação de Sanger.....	25
Figura 2 – Mutação G> GA de baixa frequência detetada pelo software <i>Mutation Surveyor</i>	29
Figura 3 – Visão geral do esquema de codificação e descodificação de cores SOLiD	34
Figura 4 – Design do sensor, poço e chip utilizados pela Ion Torrent	37
Figura 5 – Metodologia da Roche 454	39
Figura 6 – Sequenciação de uma única molécula pela HeliScope	45
Figura 7 – Sequenciação de uma única molécula de DNA em tempo real num nanowaveguide	47
Figura 8 – Detecção das bases de DNA modificadas durante a sequenciação SMRT	48
Figura 9 – Sequenciação por nanoporos.....	49
Figura 10 – Dispositivo MinION	50

Índice de Tabelas

Tabela 1 – Doses de manutenção esperadas da varfarina baseadas nos genótipos CYP2C9 e VKORC1	14
Tabela 2 – Vantagens e Desvantagens das plataformas NGS	42

Lista de Abreviaturas

A – Adenosina

ATP – do inglês: *Adenosine triphosphate*

Bp – *Base pair*

C – Citosina

CCD – do inglês: *Charge-coupled device*

CMOS – do inglês: *Complementary metal-oxide-semiconductor*

cSNPs – do inglês: *Coding SNPs*

CYP – Citocromo

CYP450 – Citocromo P450

dA – desoxinucleótido Adenosina

dC – desoxinucleótido Citosina

ddNTP – do inglês: *dideoxynucleotide triphosphate*

dG – desoxinucleótido Guanosina

DNA – do inglês: *Deoxyribonucleic acid*

dNTPs – do inglês: *Deoxynucleotides triphosphates*

dPCR - PCR digital

dT – Desoxinucleótido Timidina

ECA – Enzima Conversora de Angiotensina

ELISA – *Enzyme-linked immunosorbent assay*

EMA – do inglês: *European medicines agency*

emPCR – PCR em emulsão

FDA – Administração dos alimentos e das drogas

G – Guanosina

Gb – Gigabase

GINA – Ato de Não-discriminação por informação genética

gSNPs – SNPs associados ao gene

GWAS - Estudo de associação pangenômico

HCV – do inglês: *Hepatitis C Virus*

HIV – do inglês: *Human Immunodeficiency Virus*

INR – Razão normalizada internacional

Mb – Megabase

mRNA – RNA mensageiro

ONT – do inglês: *Oxford nanopore technology*

PAG – do inglês: *Polyacrylamide gel*

PCR – Reação em cadeia da polimerase

PGM – do inglês: *Personal Genome Machine*

PPi – do inglês: *Inorganic pyrophosphate*

pSNPs – SNPs relevantes para o fenótipo

qPCR – do inglês: *Quantitative PCR*

RNA – Ácido ribonucleico

rSNPs – *random SNPs*

SBS – *Sequencing by Synthesis*

SMRT – *Single Molecule Real Time*

SNPs – *Single nucleotide polymorphisms*

SOLiD – *Sequencing by Oligo Ligation Detection*

T – Timidina

tSMS – *True Single-molecule Sequencing*

ZMW – *Zero mode waveguide*

1. Introdução

1.1. Farmacogenómica e farmacogenética: uma medicina personalizada

O estudo do genoma humano, nomeadamente a sua completa sequenciação e caracterização de genes expressos nos humanos, criou novas oportunidades no desenvolvimento de novos fármacos pela indústria farmacêutica (Grenet, 2001). A intensa procura de novas técnicas de sequenciação e de novas tecnologias capazes de fornecer informações de um modo preciso e rápido (Moreno, 2013), tornaram-se numa das principais prioridades das indústrias farmacêuticas para o desenvolvimento de novos fármacos. O conhecimento de todos os genes humanos e as suas respetivas funções pode dar origem a um vasto leque de medidas preventivas eficazes contra várias doenças, nomeadamente de carácter genético e mudar o curso das estratégias utilizadas na investigação de novos fármacos, assim como os seus processos de desenvolvimento (Grahame-Smith, 1999).

Os avanços recentes na investigação genómica possibilitaram uma melhor compreensão da importância das técnicas da genética molecular durante as diferentes fases do desenvolvimento do fármaco, dando origem a um novo campo de investigação, a farmacogenómica (Gupta and Jhawar, 2016). Este termo foi recentemente introduzido na área científica e advém do termo farmacogenética. Alguns cientistas defendem que a farmacogenética teve origem nos anos 90, a partir do Projeto do Genoma Humano (*Human Genome Project*), enquanto que outros recordam que já Pitágoras, por volta do ano 510 A.C, reconhecia “os perigos de alguns, mas não outros, indivíduos que comem feijão fava” (Meletis and Konstantopoulos, 2004), associado a uma reação adversa a qual se sabe hoje que se deve à anemia hemolítica nos indivíduos que apresentam deficiências na enzima glucose-6-fosfato desidrogenase (Munir Pirmohamed, 2001).

Atualmente, estas duas áreas científicas (farmacogenética e farmacogenómica) estão inter-relacionadas e possuem objetivos idênticos. Ambas estudam a contribuição dos fatores genéticos na variabilidade interindividual da eficácia e segurança terapêutica. Enquanto que o termo farmacogenética é amplamente utilizado em relação aos genes, que

determinam o metabolismo do fármaco, a farmacogenómica engloba todos os genes do genoma que podem determinar a resposta farmacológica numa doença (Scott, 2011).

Sabe-se hoje em dia que cada ser humano é geneticamente distinto e, por esse motivo, responde de maneira diferente aos fatores causadores da doença assim como aos fármacos usados para o tratamento de doenças. Tais variações ou polimorfismos genéticos são estudados na farmacogenética, visando compreender qual o papel que a composição genética de um indivíduo desempenha na sua resposta medicamentosa, bem como os efeitos secundários suscetíveis de ocorrer, melhorando a nossa capacidade de identificar as causas genéticas das doenças e das respostas a medicamentos e procurar novos alvos farmacológicos (Ma and Lu, 2011).

Embora exista um número vasto de diferentes tipos de marcadores polimórficos, hoje em dia o maior interesse científico recai sobre os *single nucleotide polymorphisms* (SNPs) e no seu potencial para a determinação do perfil da resposta individual farmacológica.

Os SNPs (explicados com maior detalhe no capítulo 2) são pequenas variações genéticas com uma frequência de 1 em 1000 pares de base (bp) e ocorrem normalmente ao longo do DNA de um indivíduo. Cada SNP representa uma diferença única em apenas um nucleótido. Por exemplo, um SNP pode substituir o nucleótido citosina pelo nucleótido timina num local do DNA. Estes podem ser utilizados como marcadores biológicos, auxiliando assim na localização de genes associados a certas doenças, como o cancro, diabetes, entre outras (Wolf *et al.*, 2000). Quando estas variações genéticas ocorrem dentro de um gene ou numa região reguladora perto de um gene, os SNPs podem estar associados diretamente à doença ao afetar a função do gene. Nem todos os SNPs têm efeito na saúde ou na fase de desenvolvimento no humano, contudo estes SNPs poderão ser extremamente importantes no estudo da saúde humana, uma vez que podem auxiliar a prever uma resposta individual farmacológica, a suscetibilidade a fatores ambientais, como toxinas, e o risco de desenvolvimento de certas doenças (Komar, 2009).

A busca de um tratamento individualizado é considerada como o Santo Graal da prática médica. As reações adversas dos fármacos devido, em parte, à variação genética individual, são os principais problemas clínicos de maior relevo. Na prática, inúmeros pacientes recebem a mesma dosagem de um certo fármaco, sendo este eficaz na grande

maioria dos doentes, de pouca ou nenhuma consequência para outro conjunto de doentes, e tóxico, até mesmo fatal, num grupo mais restrito (Kalow, 2006).

A farmacogenómica não só traz benefícios para o modo como se executa a medicina, mas também para a indústria farmacológica. Atualmente, os fármacos que estão em desenvolvimento têm uma aprovação para o mercado muito baixa. De acordo com um estudo conduzido pela BIO e BioMedTracker, verificou-se que a taxa de sucesso dos fármacos na transição da fase I foi de 63,2% enquanto que a taxa de sucesso na transição da fase II foi de 30,7%, sendo esta a fase com a taxa de sucesso mais baixa quando comparada com as quatro fases. É nesta etapa que a indústria farmacêutica decide se deseja investir em estudos morosos e dispendiosos de fase III, ou optar por terminar o processo de investigação caso não haja viabilidade comercial, ou outros fatores (David W. Thomas, 2016). A farmacogenómica intervém na análise dos fármacos, no desenvolvimento clínico e no mercado e aplica uma abordagem sistemática da genómica para acelerar a descoberta de marcadores de resposta ao fármaco. Para além disso, a sinergia da farmacogenética e farmacogenómica nos ensaios clínicos permitirá que a doença possa ser tratada de acordo com o perfil genético de cada indivíduo em termos de segurança, eficácia, farmacocinética e farmacodinâmica (Burt and Dhillon, 2013).

2. Single nucleotide polymorphisms – A sua importância clínica

Os SNPs são a forma mais simples de variação de DNA entre os indivíduos. Estas alterações simples podem ser transições ou transversões, ocorrendo ao longo do genoma numa frequência de 1 em cada 1000 bp (Shastri, 2009).

Estes polimorfismos genéticos são transmitidos de geração em geração e particularmente responsáveis pela diversidade genética da população humana, afetando a evolução do genoma humano, as respostas farmacológicas de cada indivíduo a um determinado fármaco, e podem sobretudo ditar a suscetibilidade de um indivíduo para uma determinada doença, como por exemplo diabetes, obesidade, hipertensão e transtornos psiquiátricos (Shastri, 2009).

A maior parte destes SNPs ocorrem em zonas do genoma que não codificam proteínas, sendo considerados estáveis e sem efeitos deletérios para os organismos. Como já foi dito no capítulo 1, o grande interesse científico destes SNPs recai na possibilidade de usá-los como marcadores biológicos na identificação de doenças (Shastry, 2009) com a promessa de simplificar bastante o diagnóstico dos doentes e desencadear explicações acerca das características funcionais das mutações.

A distribuição genómica dos SNPs não é, contudo, homogénea, e ocorre com maior frequência nas regiões não codificantes do que nas regiões codificantes (Guo and Jamison, 2005; Varela and Amos, 2010). Quando os SNPs ocorrem em regiões codificantes, podem ser sinónimos ou não sinónimos (Hunt *et al.*, 2009). Os SNPs sinónimos não causam alteração no aminoácido, designados em geral como SNPs silenciosos sob o argumento de que a sua presença é inócua para o gene e para o seu produto correspondente. No entanto, alguns estudos demonstraram que estes polimorfismos sinónimos podem afetar o splicing do RNA, a estabilidade e estrutura das proteínas. De acordo com um estudo publicado (Kimchi-Sarfaty *et al.*, 2007), os indivíduos que expressam SNPs silenciosos no gene MDR1 codificante da glicoproteína P revelaram uma farmacocinética alterada da glicoproteína P, mostrando que o termo silencioso deve ser geralmente evitado para este tipo de SNPs.

Por sua vez, os SNPs não sinónimos produzem alterações nos aminoácidos que podem provocar o aparecimento de doenças, e conseqüentemente podem estar sujeitos à seleção natural. Estes polimorfismos podem provocar variações nas proteínas que contribuem para o polimorfismo genético, podendo influenciar a atividade do promotor (expressão genética), a conformação do RNA mensageiro (mRNA), a localização subcelular de mRNAs e/ou proteínas que podem produzir doenças (Shastry, 2009).

2.1. Tipos de SNPs

Os polimorfismos genéticos podem subdividir-se em 4 grupos: Random SNPs (rSNPs), SNPs associados aos genes (gSNPs), *coding SNPs* (cSNPs) e SNPs relevantes para o fenótipo (pSNPs). Os rSNPs constituem apenas 10% do nosso genoma. Estes SNPs são

bastante improváveis de terem um efeito no nosso fenótipo ou constituição, e muitos destes SNPs são usados como marcadores de genes no genoma (Singh, 2015).

Os gSNPs estão situados ao lado de genes, ou em regiões de um gene que não codificam proteínas nem fazem parte do template da proteína. O facto destes SNPs serem hereditários com estes genes, torna-os úteis na associação entre o gene e as suas variações nos fenótipos. O mapeamento dos gSNPs pode ser funcionalmente relevante se influenciarem elementos importantes de controlo de genes, aumentando ou diminuindo a transcrição do gene (Singh, 2015).

Relativamente aos cSNPs, situam-se nos exões (regiões codificadoras de proteínas) de um gene podendo ter uma grande influência na função da proteína, se existir incorporação de um aminoácido alternativo (Singh, 2015).

Por último, os pSNPs, tal como o próprio nome indica, estão associados ao fenótipo. Tanto os gSNPs como os cSNPs podem influenciar o fenótipo de um indivíduo – o gSNP na quantidade, e o cSNP na forma da proteína que codifica. Os pSNPs são o tipo de SNP mais importante do ponto de vista clínico, sendo considerados como as fundações do ramo da farmacogenética, no que diz respeito à sua influência da variação do gene na efetividade e tolerância dos fármacos (Singh, 2015).

2.2. Efeitos clínicos dos SNPs

Como foi explicado, os SNPs podem ter efeitos genéticos consoante a sua localização no genoma. Os efeitos destas variações genéticas também podem ser divididos em duas categorias que são relevantes na medicina, a influência da ação dos fármacos e o seu envolvimento no desenvolvimento e progressão das doenças (Komar, 2009).

As variações quanto à ação dos fármacos, podem explicar o porquê destes terem melhores efeitos em certos indivíduos e de certos efeitos secundários se manifestarem apenas em alguns indivíduos. O conhecimento da posição e dos efeitos dos SNPs, podem ser considerados no desenvolvimento de formas mais precisas no tratamento de doentes. Já as variações que têm influência no desenvolvimento e progressão de doenças, intervêm na suscetibilidade a certas patologias (Komar, 2009).

É importante salientar que os SNPs não são mutações e o termo mutação sugere que há uma mudança no nucleótido que previne a proteína de exercer a sua função normal. Estes tipos de mutações são raras, ocorrendo em menos de 1% da população (Karki *et al.*, 2015), e quando ocorrem são quase sempre severas, sendo por isso causadoras da doença. Em contraste com os SNPs, que são mais comuns, a sua presença apenas faz com que certos indivíduos tenham um maior risco de desenvolver doenças não sendo causadores da mesma. Estas podem ocorrer na presença de um certo ambiente, por combinações genéticas, fármacos e muitos outros fatores (Karki *et al.*, 2015).

Nos tratamentos clínicos atuais, devido ao metabolismo, os fármacos não terão a eficácia esperada em muitos doentes, tendo muitas vezes reações adversas que podem ser fatais. O facto de não se conseguir prever a eficácia exata e o efeito adverso de um fármaco, dificultam o desenvolvimento da indústria farmacêutica (Chen and Wei, 2015). Numerosas pesquisas genómicas sobre o CYP450 (McGraw and Waller, 2012) apontam para a existência de SNPs no citocromo P450, que contribuem para a razão pela qual as pessoas têm diferentes reações medicamentosas e taxas de eficácia de fármacos superiores, ou inferiores (Preissner *et al.*, 2013).

Em suma, este conhecimento pode fornecer um ponto de partida para o desenvolvimento de novos e úteis, marcadores SNPs para testes médicos e uma medicação individualizada mais segura, assim como a compreensão da metabolização de fármacos para tratar os vários distúrbios que são mais comuns (Komar, 2009).

3. Citocromo P450

O citocromo P450 constituiu a superfamília de enzimas heme-tiolatos encontradas em quase todos os organismos do mundo procariota e eucariota e são responsáveis pelas várias reações químicas num grande número de substâncias (Lewis, 2005). Geralmente os organismos anaeróbios e alguns microaerófilos, como os parasitas *Plasmodium* ou *Giardia*, não apresentam genes CYP, mas nos organismos aeróbios o número de genes CYP varia desde 2 até 400 genes, encontrados por exemplo na *Schizosaccharomyces Pombe* e na batata (Nelson *et al.*, 2013).

A família do citocromo P450 apresenta níveis elevados de complexidade e diversidade. A sua complexidade verifica-se quando uma pequena mudança num único aminoácido pode ser o suficiente para alterar a sua regioselectividade na hidroxilação de duas enzimas intimamente relacionadas com o mesmo substrato (Schalk and Croteau, 2000).

As sequências do P450 são tão diversas que não apresentam um resíduo universalmente conservado em toda a família, levantando questões sobre a origem da família de genes P450. Esta procura de respostas acerca da sua origem ofereceu um leque vasto de campos de investigação sobre mecanismos de diversificação da família de genes (Schalk and Croteau, 2000).

3.1. Citocromo P450: uma perspetiva evolucionária

Pensa-se que o primeiro gene P450 emergiu na evolução das primeiras formas de vida. Foram propostas várias hipóteses que incluem a sua função como redutase, sobretudo da reação óxido nítrico redutase em ambientes anaeróbicos antes de adquirirem a capacidade de se ligar ao dioxigénio (O₂) (Deng *et al.*, 2007).

Durante o curso evolutivo, a família P450 sofreu algumas modificações para se poder adaptar às mudanças ambientais e modificações celulares que daí advieram (Jiang *et al.*, 2012).

O aparecimento das células eucarióticas, há cerca de 2 mil milhões de anos, intensificou a quantidade de oxigénio na atmosfera terrestre e conseqüentemente o aparecimento do metabolismo aeróbico. O nível crescente de oxigénio ajudou a produção de energia numa maneira mais eficiente nos metabolismos aeróbicos, produzindo mais adenosina trifosfato por hexose de açúcar do que nos metabolismos anaeróbicos. Este passo importante permitiu a formação de novos metabolitos, como os esteroides, alcaloides e isoflavonoides (Jiang *et al.*, 2012).

O novo ambiente aeróbico despoletou também o desenvolvimento de sistemas protetivos que garantissem a sobrevivência das primeiras formas de vida. Embora o oxigénio seja essencial para muitos organismos, alguns dos seus produtos metabólicos são tóxicos (Lewis, 2005).

O O_2 é uma molécula paramagnética, tendo na sua estrutura molecular dois elétrões com *spins* paralelos. Isso torna as reações com O_2 difíceis, uma vez que o dador orgânico tem que sofrer uma inversão lenta de *spins* para que este possa doar os seus elétrões. Tal processo não acontece com o dióxigénio singuleto, sendo este uma forma de alta energia do oxigénio. Para contornar esta necessidade, o O_2 é transformado em H_2O ao produzir um anião radical superóxido (O_2^-), o peróxido de hidrogénio (H_2O_2) e o radical hidroxilo (OH^\cdot) (Cederbaum, 2015). Estes intermediários (radicais livres) são sérias ameaças para as células uma vez que estes compostos podem interagir diretamente com compostos orgânicos e interferir nas ligações duplas de carbono, danificando as membranas celulares através da reação de radicais livres com ácidos gordos. Uma vez que estes produtos são tóxicos para as células, foi necessário desenvolver enzimas que possuíssem metais de transição para ajudarem a evitar a produção destes compostos intermediários (Thannickal, 2009).

Os CYPs desempenharam esta função protetora ao tornarem-se ativos nas reações de redução do oxigénio. Estes utilizavam assim o poder químico inerente da molécula O_2 e controlavam a sua ativação por via de duas reduções consecutivas do H_2O_2 (Lewis, 2005). Para além do seu papel ativo nas reações de redução do oxigénio, o CYP cataboliza vários tipos de reações, incluindo a hidroxilação, oxidação, sulfonação, entre outras, de um elevado número de compostos endógenos e exógenos (Tomaszewski *et al.*, 2008).

No P450 humano, devido aos polimorfismos genéticos, existem defeitos genéticos nos P450s individuais que originaram mudanças significativas na capacidade de metabolização de certos substratos, onde se incluem os fármacos de uso clínico. Certos SNPs mostraram ter um impacto significativo na atividade dos CYPs. Deste modo, os CYPs potenciam a resposta interindividual, e a sua variabilidade genética deve ser tida em conta na medicina personalizada. Existe um interesse considerável no rastreio de novos compostos farmacológicos para que as reações adversas sejam evitadas, tendo como base o conhecimento dos genótipos dos diferentes pacientes e a sua eficácia seja maximizada (Zhou *et al.*, 2009).

3.2. Nomenclatura

Devido à grande multiplicidade genética e à falta de sequências primárias entre os membros da superfamília do gene P450, um sistema de nomenclatura foi estabelecido para nomear e atribuir genes individuais em famílias e subfamílias com base na porcentagem de identidade de sequências de aminoácidos. As enzimas que partilham uma identidade maior ou igual a 40% são atribuídas a uma família particular designada por um número, enquanto que aquelas que partilham mais ou menos 50% de identidade, constituem uma subfamília particular designada por uma letra (Nebert *et al.*, 2013).

O sistema da nomenclatura utiliza o símbolo CYP como abreviatura para o citocromo P450 e uma designação alfanumérica que é empregue para nomear as famílias P450, as subfamílias e proteínas individuais. Por exemplo, o CYP2D6 pertence à família 2, à subfamília 2D e ao polipéptido 6 (Gopisankar, 2017).

O genoma humano contém 18 famílias CYP, divididas em 41 subfamílias codificadoras de proteínas, que por sua vez são codificadas por 57 genes (Sim and Ingelman-Sundberg, 2010).

As primeiras 3 famílias (CYP1-3) são responsáveis pelo metabolismo de substâncias exógenas, como fármacos, enquanto que as famílias CYP com maior número relacionam-se mais com substâncias endógenas (Sim and Ingelman-Sundberg, 2010).

Dentro destas 3 famílias, salientam-se seis isoenzimas P450 diferentes - CYP1A2, CYP2C19, CYP2C9, CYP2D6, CYP2E1 e CYP3A4 – que estão relacionadas com a metabolização de fármacos (Lynch and Price, 2007).

3.3. Fatores que afetam a sua atividade e a sua função metabólica

Os CYPs humanos são principalmente proteínas associadas às membranas, sendo estas expressas de uma forma ubíqua na maioria dos tecidos. Contudo, a distribuição da expressão tecidual de CYPs particulares varia, indicando que a eficiência do fármaco está dependente da expressão dos CYPs no tecido alvo e do próprio fármaco em si (Pelkonen and Raunio, 1997).

A maioria dos CYPs metaboliza mais de um fármaco e, de forma semelhante, um fármaco é frequentemente metabolizado por vários CYPs. Os fármacos também podem inibir ou induzir a atividade destes, ou por interação direta com a enzima ou na alteração da sua expressão. A caracterização destas interações é importante porque permite determinar combinações de fármacos utilizados na terapia que sejam compatíveis entre si e poderá permitir um melhor ajuste da concentração ótima do fármaco para cada paciente (Lin and Lu, 1998).

A indução enzimática é um processo pelo qual a exposição a certos substratos (sejam eles fármacos, poluentes ambientais, dieta, etc.) resulta numa biotransformação acelerada com uma redução correspondente do fármaco não metabolizado. Muitos dos fármacos podem exibir um decréscimo na sua eficácia devido ao metabolismo rápido, mas fármacos com metabolitos ativos podem exibir um aumento do efeito do fármaco e/ou toxicidade devido à indução enzimática (Sweeney and Bromilow, 2006).

A inibição enzimática também afeta a biotransformação de um fármaco. Este processo ocorre quando dois fármacos, que partilham o metabolismo através da mesma isoenzima, competem pelo mesmo recetor. O inibidor mais potente predominará, resultando num decréscimo no metabolismo do fármaco competidor. Para muitos fármacos, pode levar a um incremento de níveis da entidade não metabolizada e conseqüentemente a um maior potencial de toxicidade. Para fármacos em que a sua atividade farmacológica está dependente da biotransformação de um pró-fármaco, a inibição pode levar a um decréscimo da sua eficácia (Sweeney and Bromilow, 2006).

A inibição e a indução enzimática têm também ligações genéticas para além dos fatores exógenos que alteram a função dos CYPs. Muitos doentes têm uma variação num gene que afeta a função das enzimas metabolizadoras de fármacos. A função alterada da enzima pode mudar a taxa de metabolismo do fármaco (Pinto and Dolan, 2011).

Os doentes com variantes genéticas são mais suscetíveis de experimentar toxicidade de fármacos ou falta de eficácia. A maioria das variantes causam uma perda da função enzimática e aqueles que apresentem estas variantes são metabolizadores pobres ou intermediários. Por outro lado, alguns indivíduos apresentam variantes que aumentam a

função enzimática. Esses denominam-se de metabolizadores ultra-rápidos (Bertilsson *et al.*, 2002).

Uma vez que os genes autossômicos existem nas células aos pares, um originário de um parente diferente, um doente pode apresentar dois genes normais (sem variantes), uma variante, ou duas variantes. Os que não possuem variantes têm enzimas normais do citocromo P450, mas aqueles com uma ou duas variantes podem metabolizar mais ou menos do que as enzimas normais. Isso afeta o modo como o corpo metaboliza os fármacos e, por sua vez, afeta os níveis de um medicamento no sangue. Níveis muito altos, podem causar efeitos adversos e níveis muito baixos podem tornar o tratamento ineficaz (Bertilsson *et al.*, 2002).

Embora os polimorfismos genéticos possam explicar em grande parte a variabilidade para algumas enzimas (em particular CYP2D6 e CYP2C19), muitas destas enzimas são controladas multifatorialmente, incluindo polimorfismos adicionais de genes regulatórios e fatores não genéticos tais como: género, idade, doença, influências hormonais e muitos outros fatores (Coleman, 2010).

4. A Farmacogenómica no tratamento clínico

A compreensão do impacto dos fatores genéticos e não genéticos é importante na terapia farmacológica para determinar qual a concentração de fármaco a administrar nos pacientes de modo a evitar efeitos adversos. Existem fármacos que possuem uma variabilidade interindividual ampla e uma janela terapêutica estreita, necessitando de uma constante verificação dos níveis plasmáticos para que estes não sejam tóxicos.

- **Varfarina**

A varfarina é utilizada no tratamento de doenças trombóticas (Rettie and Tai, 2006) e uma das grandes desvantagens deste agente, também conhecido como Coumadin, é o facto de ser difícil de administrar a dose correta devido à sua janela terapêutica estreita e variabilidade interindividual, podendo causar hemorragias (Jiayi Li *et al.*, 2009).

A varfarina é um anticoagulante que atua reduzindo a atividade de fatores de coagulação dependentes da vitamina K. Este fármaco é geralmente usado na prevenção e tratamento de doenças trombóticas a longo prazo (Kuruvilla and Gurk-Turner, 2001).

A vitamina K desempenha um papel crítico na coagulação do sangue. Esta vitamina é anabolizada por um produto de genes da membrana dos hepatócitos do complexo vitamina K epóxido redutase subunidade 1 (VKORC1). Por outras palavras, o gene VKORC1 codifica a enzima epóxido redutase da vitamina K, sendo esta o alvo da varfarina. A varfarina exerce o seu efeito ao inibir a enzima codificada pelo gene VKORC1 que catalisa a conversão do epóxido da vitamina K para a forma ativa reduzida da vitamina K – vitamina K hidroquinona (Flockhart *et al.*, 2008).

A vitamina K hidroquinona é um cofator essencial na síntese de vários fatores de coagulação, promovendo a síntese de ácido γ -carboxiglutâmico. A diminuição da disponibilidade da vitamina K hidroquinona leva à diminuição da atividade dos fatores de coagulação II, VII, IX e X (Flockhart *et al.*, 2008).

No tratamento de doentes, a varfarina é administrada como uma mistura racémica dos estereoisómeros R e S. A (S)-varfarina é duas a cinco vezes mais potente do que a (R)-varfarina e é principalmente metabolizada pelo CYP2C9 (Munir Pirmohamed, 2006). Por sua vez, a (R)-varfarina é principalmente metabolizada pelo CYP3A4 com o envolvimento de outras enzimas do citocromo P450 (Jiayi Li *et al.*, 2009).

Dos alelos CYP2C9 conhecidos, existem três variantes alélicas para o CYP2C9 com maior relevância no tratamento de doentes com varfarina, sendo estas o CYP2C9*1, o CYP2C9*2 e o CYP2C9*3. O CYP2C9*1 é o alelo do tipo selvagem e está associado à atividade enzimática normal e ao fenótipo do metabolizador normal (Dean, 2012).

As duas variantes alélicas comuns associadas, com atividade enzimática reduzida são o CYP2C9*2 e CYP2C9*3. Em comparação com os metabolizadores normais, os doentes que herdaram uma ou duas cópias do *2 ou *3 são mais sensíveis, isto é, requerem doses mais baixas e correm maior risco de hemorragia durante o início da terapia (Dean, 2012).

As frequências alélicas do CYP2C9 variam em diferentes grupos étnicos. O alelo *2 é mais comum nas populações caucasianas do que nas populações asiáticas ou africanas. Já

o alelo *3 é menos comum e extremamente raro nas populações africanas (Scott *et al.*, 2010).

Em relação ao gene VKORC1, os doentes portadores do polimorfismo -1639G>A na região promotora do gene VKORC1 são mais sensíveis à varfarina e requerem uma dose terapêutica mais baixa. A sua variação alélica também varia nos vários grupos étnicos mas prevalece mais nas populações asiáticas, e pode ser um fator que contribui para menores requisitos de dose da mesma, frequentemente observados em pacientes de descendência asiática (Dean, 2012).

Assim sendo, a variabilidade interindividual é um obstáculo na terapia de muitos agentes terapêuticos, sobretudo no uso da varfarina. Neste fármaco, o cenário é particularmente dramático porque os doentes são todos encaminhados para o mesmo parâmetro farmacológico que consiste num aumento, de duas a três vezes, do tempo de coagulação. O tempo para a dosagem estável com varfarina pode demorar vários meses, pondo em risco a vida dos pacientes, uma vez que podem ser vítimas de sangramento excessivo se as doses de iniciação forem muito altas, ou formação de coágulos se as doses forem demasiado baixas (Kuruvilla and Gurk-Turner, 2001).

Para contornar este problema, a implementação da genotipagem nos doentes de modo a descobrir qual a variante alélica que esses doentes transportam, pode fornecer uma salvação já que este tipo de abordagem pode oferecer uma terapia individualizada e assim evitar os efeitos adversos provocados pela varfarina (Dean, 2012).

Existem estudos publicados (Flockhart *et al.*, 2008; Poopak *et al.*, 2015) que confirmam que o genótipo CYP2C9 e VKORC1 de um doente pode ser utilizado para ajudar a determinar a dose inicial ótima de varfarina. O gene CYP2C9 codifica uma das principais enzimas envolvidas no metabolismo da varfarina, e as suas várias variantes estão associadas a uma redução na atividade enzimática verificando-se taxas de depuração mais baixas de varfarina. Os doentes que transportam pelo menos uma cópia de um alelo variante (como CYP2C9*2 e CYP2C9*3) têm um metabolismo reduzido, levando a maiores concentrações de varfarina. Em média, estes pacientes requerem uma dose diária mais baixa de varfarina do que os doentes homozigóticos para o alelo CYP2C9*1 (Flockhart *et al.*, 2008). A eficácia da anticoagulação é monitorizada pela razão

normalizada internacional (INR), funcionando como um índice do seu efeito terapêutico. Assim, utiliza-se o teste INR para determinar a dose correta de varfarina, tendo como objetivo a avaliação da coagulação externa e o diagnóstico de fatores de deficiência hereditários e não genéticos, que se baseia no método de tentativa e erro (Poopak *et al.*, 2015).

Numa tentativa de controlar a sua dosagem, a FDA fornece uma tabela (tabela 1.) de dosagem baseada nos genótipos CYP2C9 e VKORC1. Se o genótipo CYP2C9 e/ou VKORC1 do doente é conhecido, então deve-se considerar os intervalos das doses iniciais indicados na tabela 1 (Dean, 2012). As informações fornecidas pela FDA também afirmam que os pacientes com as variantes CYP2C9 *1/*3, *2/*2, *2/*3 e *3/*3 podem necessitar de mais tempo (entre 2 a 4 semanas) para obter o efeito INR máximo para um determinado regime de dosagem do que os doentes sem estas variantes CYP (Dean, 2012).

Tabela 1. Doses de manutenção esperadas da varfarina baseadas nos genótipos CYP2C9 e VKORC1 (Dean, 2012)

VKORC1	CYP2C9					
	*1/*1	*1/*2	*1/*3	*2/*2	*2/*3	*3/*3
<i>GG</i>	5-7 mg	5-7 mg	3-4 mg	3-4 mg	3-4 mg	0,5-2 mg
<i>AG</i>	5-7 mg	3-4 mg	3-4 mg	3-4 mg	0,5-2 mg	0,5-2 mg
<i>AA</i>	3-4 mg	3-4 mg	0,5-2 mg	0,5-2 mg	0,5-2 mg	0,5-2 mg

Os intervalos são derivados de vários estudos clínicos e a variante VKORC1 -1639G>A (rs9923231) é usada nesta tabela. Outras variantes de VKORC1 co-herdadas podem também ser importantes determinantes na dose da varfarina (Dean, 2012).

Assim sendo, a dosagem inicial e de manutenção da varfarina deve ser individualizada para cada paciente. O objetivo da terapia com a mesma é conseguir um rácio normalizado num leque de objetivos para que esta condição seja tratada. Tal processo envolve a seleção da dose inicial, seguida do teste regular INR para que a dose de varfarina possa ser

ajustada até ser determinada a dose de manutenção diária apropriada a cada doente. Ao ter conhecimento dos genótipos de cada paciente, pode-se administrar a dose ótima para cada indivíduo, reduzindo assim o tempo necessário para atingir um INR estável e eliminando o risco de obter um INR elevado (risco de hemorragia) ou de um INR baixo (risco de trombose). Em geral, a duração da terapia anticoagulante varia de acordo com a indicação clínica e deve ser continuada até que não se apresente perigo de trombozes ou embolismos (Dean, 2012).

Os indivíduos que são mais propensos a beneficiar destes testes genéticos são aqueles que ainda não iniciaram a terapia com varfarina. A sua administração baseada no genótipo não é ainda o método padrão em vários sistemas de saúde e, apesar do esforço para conceber uma estratégia que reduzisse os efeitos adversos relacionados com a varfarina, a eficácia deste método ainda não foi verificada. Embora um ensaio randomizado (M. Pirmohamed *et al.*, 2013) tenha verificado que a dose de varfarina de acordo com o genótipo tenha melhorado o controlo do INR após iniciação desta, um outro estudo de maior escala (Furie, 2013), concluiu que não existia nenhum benefício neste método, sendo por isso ainda controverso, mas a existência de variantes genéticas raras podem ter sido responsáveis pela falha da previsão da dose ótima em certos doentes (Dean, 2012).

O que parece ser a maior barreira para a não genotipagem dos doentes é o custo associado a este método. Nos Estados Unidos, os testes normalmente custam algumas centenas de dólares e podem levar dias até que se conheça os seus resultados. Por exemplo, na Cleveland Clinic, os testes CYP2C9 e VKORC1 podem ser executados internamente a um custo que ronda os 700\$. Geralmente, estes custos não têm reembolso quando os testes não foram aprovados previamente. Mas, se estiverem prontamente disponíveis (sendo pagos por companhias de seguros) e, se os resultados forem obtidos entre as 24 e 48 horas, podem ser uma ferramenta valiosa na orientação e administração da dose de varfarina. Estes testes também podem ter um maior impacto na decisão dos doentes que são pró-ativos (doentes bem informados e que procuram informação fidedigna), reduzindo o risco dos efeitos adversos (Rouse *et al.*, 2013).

No caso dos doentes em que o seu genótipo não é conhecido, a norma geral é administrar uma dose inicial de varfarina, de 2 a 5 mg uma vez por dia. As necessidades de dosagem

de cada doente são depois determinadas através da monitorização rigorosa da resposta INR, procedendo a uma dose de manutenção diária de 2 a 10 mg (Dean, 2012).

- **Hipertensão**

Um outro exemplo em que a Farmacogenómica não apresenta benefícios, é no caso da terapia para a hipertensão. A hipertensão é um fator de risco modificável mais prevalente para a carga global de doenças (*global disease burden*) e muitos estudos farmacogenómicos tentaram identificar as variantes genéticas responsáveis pelas respostas altamente variadas, a uma ou outra medicação anti-hipertensiva. Apesar dos esforços, esses estudos ainda não obtiveram resultados práticos. Tais falhas provavelmente refletem a extrema heterogeneidade e complexidade da doença hipertensiva (Menni, 2015).

A elevada pressão arterial tem múltiplas etiologias e envolve mudanças fisiológicas variadas ao longo da vida. As variantes do estudo de associação pangenómico (GWAS) identificadas e associadas à pressão arterial são muito menos do que aquelas associadas à concentração de glicose ou de lípidos no sangue, e variantes de pressão arterial identificadas também explicam a menor variação fenotípica (<3%) (Munroe *et al.*, 2013), que é, provavelmente, mais uma evidência consistente com um alto grau de complexidade para distúrbios hipertensivos.

Além disso, as múltiplas classes de fármacos anti-hipertensivos (diuréticos, inibidores da ECA, bloqueadores dos recetores angiotensina II, bloqueadores dos canais de cálcio, betabloqueadores), quer isoladamente quer combinados, complicam ainda mais os estudos farmacogenómicos, assim como outras variações devido a flutuações da pressão sanguínea intrínseca à própria pessoa, como por exemplo, mudanças na dieta, exercício físico, hora do dia, combinações de fármacos prescritos, o uso de álcool, entre outras. Vários GWAS farmacogenómicos reportaram variantes em genes plausíveis associados a diuréticos tiazídicos (Turner *et al.*, 2008; Turner *et al.*, 2013; Chittani *et al.*, 2015), bloqueadores dos recetores de angiotensina II (Turner *et al.*, 2012; Frau *et al.*, 2013) e beta bloqueadores (Gong *et al.*, 2015; Gong *et al.*, 2016). Contudo, a maioria destes

estudos envolvem amostras pouco representativas, que demonstram pouca eficácia no que diz respeito à detecção de associações que têm significado genómico. Grandes consórcios internacionais foram estabelecidos, com o objetivo de aumentar as oportunidades de descoberta e replicação através da montagem de amostras de maior dimensão (Cooper-DeHoff and Johnson, 2016). No entanto, mesmo que os grandes estudos em consórcio sejam capazes de identificar associações de variantes genéticas no futuro, é difícil imaginar que, mesmo em combinação, qualquer uma dessas variantes de pequeno efeito, recentemente identificadas, tenha um poder clínico preditivo importante.

- **Antidepressivos**

De forma semelhante, o GWAS da resposta ao tratamento com antidepressivos (Biernacka *et al.*, 2015) não produziu resultados promissores. Uma limitação particular, nos estudos farmacogenómicos de medicamentos psicotrópicos, é a avaliação de condições de doença e medidas quantitativas de respostas ao tratamento, as quais, ao contrário da pressão sanguínea ou níveis séricos de colesterol, são todas baseadas em critérios de diagnósticos psiquiátricos multidimensionais (Zandi and Judy, 2010).

Esse método tem sido criticado devido à falta de confiabilidade, como por exemplo, grau de consenso sobre um diagnóstico ou resposta ao tratamento visto pelos diferentes clínicos, a redação e o modo como os questionários são respondidos pelos pacientes (Aboraya *et al.*, 2006). Essa complexidade, comparando o fenótipo equívoco versus o fenótipo inequívoco, já foi realçada num estudo (Nebert *et al.*, 2008).

O mesmo é especialmente verdadeiro para o distúrbio depressivo maior (Robert Freedman *et al.*, 2013). Os critérios psiquiátricos com base clínica não refletem diretamente as condições biologicamente homogéneas, o que torna a definição fenotípica das condições psiquiátricas altamente heterogéneas e a medição das respostas ao tratamento muito imprecisas (Zandi and Judy, 2010).

Devido, em parte, a estas dificuldades, o GWAS do distúrbio depressivo maior, com uma amostra razoavelmente grande (> 9.000 casos), não conseguiu identificar qualquer associação que fosse replicável (Ripke *et al.*, 2013). Em contraste, um estudo

farmacogenômico avaliou um fármaco original (citalopram e escitalopram) e as concentrações plasmáticas de metabolitos em 435 doentes com distúrbio depressivo maior. Este estudo (Ji *et al.*, 2014) identificou associações altamente significativas com as variantes CYP2C19 e CYP2D6. Esta importante distinção entre fenótipos clínicos psiquiátricos difusos e medidas quantitativas de fenótipos de sangue ou de fármacos urinários/metabólicos, também já foi enfatizada (Nebert *et al.*, 2008). Por conseguinte, torna-se problemático discernir como estes estudos de associação farmacológica para o metabolismo, poderão auxiliar a predição do risco genético de muitas respostas clínicas complexas (Ji *et al.*, 2014).

- **Interferão α**

Os exemplos referidos anteriormente mostram que estes métodos implementados para o tratamento de doentes são ainda controversos. Apesar disso, a farmacogenômica já foi particularmente útil em alguns agentes terapêuticos. Um exemplo bem-sucedido inicial de GWAS da eficácia de fármaco veio de estudos sobre a resposta variável do tratamento com o interferão α na infecção provocada pelo vírus da hepatite C (HCV). Em 2009, três estudos independentes (Ge *et al.*, 2009; Suppiah *et al.*, 2009; Tanaka *et al.*, 2009) reportaram variantes genéticas, próximas do gene IL28B, associadas a respostas de tratamento com o interferão α . O efeito é medido por resposta virológica sustentada, isto é, ausência de vírus detetável no final do seguimento. A mesma variante também foi reportada como estando correlacionada com a espontânea ausência do HCV (David L. Thomas *et al.*, 2009), sugerindo que a variante pode influenciar a resposta imune do hospedeiro, com ou sem interferência do fármaco.

Outro aspeto a salientar é a importância do alelo C, que é mais frequente nas populações caucasianas e asiáticas do que nas populações africanas. Esta diferença provavelmente explica o relatório anterior (Hoofnagle *et al.*, 2009), cujos resultados revelaram a baixa eficácia entre a população afro-americana infetada com HCV nos ensaios de tratamento com o interferão α .

Os novos agentes antivirais diretos, em associação com a combinação do interferão peguilado/ribavirina, é o novo padrão de tratamento do HCV, sendo necessário reavaliar a utilidade clínica do genótipo IL28B (Alexander J. Thompson and McHutchison, 2012). Embora o teste do genótipo IL28B continue a fornecer informações clínicas importantes sobre a resposta esperada ao tratamento, a genotipagem do IL28B poderá perder a sua utilidade, uma vez que a terapia com HCV afasta-se dos regimes baseados no interferão (Jensen and Pol, 2012).

5. O contributo científico da PCR

A PCR transformou radicalmente as áreas científicas ao tornar possível a identificação específica e produção de grandes quantidades de DNA. Não só revolucionou como impulsionou enormes esforços científicos, nomeadamente o Projeto do Genoma Humano. Esta técnica é amplamente utilizada no diagnóstico de doenças, assim como nas várias técnicas de sequenciação (capítulo 6.), e realiza ainda estudos quantitativos do foro genómico de uma maneira rápida e precisa (Garibyan and Avashia, 2013). Além disso, a PCR permitiu revelar mutações responsáveis pelas doenças hereditárias ou variantes oncogénicas, auxiliando o profissional médico a tomar decisões informadas acerca do tratamento. Já no desenvolvimento de novos fármacos, as informações provenientes desta técnica permitiram uma melhor compreensão da definição molecular da doença e dos processos associados à eficácia e toxicidade destes novos fármacos (Goodsaid, 2004).

A metodologia da PCR, combinada com os sistemas da expressão de genes, baseia-se na incrementação da disponibilidade de proteínas recombinantes em quantidades suficientes para a caracterização bioquímica e farmacológica. A PCR é um processo enzimático que permite a amplificação de um fragmento específico de DNA de um conjunto complexo (Garibyan and Avashia, 2013). A técnica da PCR necessita da presença de um molde de DNA, de *primers*, nucleótidos e uma DNA polimerase. A DNA polimerase é a enzima chave neste processo uma vez que liga os vários nucleótidos (adenosina, timidina, citosina e guanosina) a um conjunto para formar produtos de PCR. Os primers aqui referenciados, são curtos fragmentos de DNA, que servem como um ponto de iniciação para a DNA

polimerase e têm uma sequência complementar às extremidades do DNA alvo. Os componentes previamente mencionados, são misturados em tubos de ensaios e colocados num termociclador, que provoca a desnaturação da cadeia dupla de DNA e possibilita, posteriormente, a adição de nucleótidos pela DNA polimerase após a ligação do primer à cadeia de DNA (Garibyan and Avashia, 2013).

Neste trabalho destacam-se alguns ensaios PCR que são bastante úteis no diagnóstico e sequenciação e permitem a amplificação de quantidades infinitesimais de material genético.

- **PCR Multiplex**

Os sistemas PCR multiplex são muito usados nos estudos biológicos e médicos porque permitem a amplificação simultânea de vários fragmentos de DNA numa única reação. Esta capacidade de reduzir o número de reações necessárias numa amostra para alvos diferentes ajuda a poupar no tempo e nos custos, o que torna os sistemas multiplex muito úteis quando é necessário rastrear grandes quantidades de amostra. Apesar disso, esta técnica apresenta algumas limitações que necessitam de ser ultrapassadas. Hoje em dia, os fabricantes dos kits PCR multiplex anunciam os seus kits como *ready-to-use*, não sendo necessária nenhuma otimização. Embora isso seja normalmente verdadeiro para os kits diagnóstico, nos kits genéricos ou kits para investigação, contendo reagentes como tampões, dNTPs ou DNA polimerases, torna-se quase imprescindível de otimizar a concentração de primers e das condições da termociclagem para se obter reações equilibradas e estáveis (Dieffenbach and Dveksler, 2003). Isto porque a presença de vários primers nesta técnica leva à formação de produtos inespecíficos devido à formação de dímeros de primers (Elnifro *et al.*, 2000). Tal fenómeno acontece quando a proporção primer-template é muito elevada, ou quando há uma diluição excessiva de template face ao excesso de primers (Moreno, 2013). Nestas condições, estes dímeros são amplificados com maior eficiência do que o alvo de interesse, com a agravante do consumo dos componentes da reação. Para contornar este problema, foram feitas várias otimizações que se centram no design dos primers, nomeadamente o seu comprimento e o conteúdo

de guanina e citosina. Mas, a otimização completa das condições da reação é morosa, refletindo-se no custo acrescido por cada teste (Elnifro *et al.*, 2000). Mesmo assim, as vantagens da PCR multiplex suplantam as desvantagens, nomeadamente a sua utilidade na deteção de resistências a fármacos e no diagnóstico clínico (Messmer *et al.*).

- **Nested PCR**

Esta técnica destina-se a aumentar ainda mais a sensibilidade da PCR, reduzindo em paralelo a amplificação de produtos inespecíficos (Haff, 1994). Dois conjuntos de primers são utilizados nas duas execuções sucessivas de PCR. O primeiro conjunto de primers produz uma grande quantidade de produto para servir como molde na execução seguinte, e o segundo conjunto apenas se destina a reamplificar o primeiro produto, sendo designados como *nested*. Obtém-se, por sua vez, produtos mais pequenos que aumentam a sensibilidade da reação aquando da deteção do produto pretendido. Uma vez que os produtos não específicos têm pouca afinidade para estes primers, estes não são amplificados e, por sua vez, não contaminam o produto do segundo ciclo (Haff, 1994). Esta técnica envolve a execução de controlos adequados que asseguram a sua especificidade (Mc Pherson and Moller, 2006).

- **PCR em Tempo Real**

A PCR em tempo real (qPCR) é a técnica de eleição dos cientistas interessados em detetar a dinâmica da expressão génica nos vários organismos. A sua elevada sensibilidade e especificidade tornam esta técnica imprescindível, por exemplo, na deteção de SNPs. Para além disso, a qPCR permitiu a genotipagem quantitativa a partir de pequenas amostras e contornou limitações da PCR convencional, dado que o seu poder quantitativo era limitado e de complexa execução.

Na qPCR, a quantidade de produto formado é analisada ao longo da reação ao usar fluorocromos que permitem monitorizar a fluorescência e assim quantificar o produto formado (Kubista *et al.*, 2006). Esta técnica utiliza-se agora em combinação com a PCR multiplex que usa sondas combinadas e primers para as sequências específicas tornando-

-se muito útil na deteção de SNPs e genes associados ao cancro, visto que possibilita a formação de painéis de genes para a maioria das doenças mais comuns (Deepak *et al.*, 2007).

Embora a qPCR seja uma técnica bastante poderosa em comparação com as outras técnicas, esta possui limitações que dependem do uso correto da calibração e do controlo. A aplicação bem-sucedida e rotineira dos diagnósticos de PCR é muitas vezes limitada pela falta de qualidade do template devido a metodologias ineficientes de extração de RNA e/ou presença de compostos que podem ser inibitórios na amostra (Deepak *et al.*, 2007). Uma outra limitação que se aponta, é o facto da qPCR ser só usada para análise de genes conhecidos (Moreno, 2013).

- **PCR Digital**

A PCR digital (dPCR) não é a técnica mais popular para medir a presença e concentração de uma sequência de DNA, sendo este título atribuído a qPCR. Mas, a PCR digital detém uma maior sensibilidade e precisão quando comparada com a qPCR. Esta utiliza os mesmos primers e sondas que a qPCR, e é capaz de identificar alelos que ocorrem numa frequência de 1 em milhares. Outra característica a salientar é o facto da dPCR não necessitar de calibração e controlos como a qPCR (Kuypers and Jerome, 2017). No entanto, as suas desvantagens fazem com que esta técnica não seja a predileta para a quantificação de DNA. Estas incluem a limitação de volume da amostra por reação, que por sua vez, restringe o limite inferior de deteção e a quantificação baixa devido ao abandono molecular, já que nem todos os templates são amplificados. Esta limitação é particularmente problemática quando se pretende quantificar RNA porque o passo da transcrição reversa é incompleto para alguns alvos de RNA. Enquanto que a qPCR utiliza 4 a 6 componentes fluorescentes nas sondas que se podem ligar a diferentes alvos na amostra, a dPCR apenas utiliza 2 fluorocromos, limitando a sua capacidade de deteção dos diferentes alvos na amostra. O custo da dPCR é superior ao da qPCR, e a instrumentalização assim como os seus reagentes são mais caros o que faz com que esta técnica não seja tão empregue como a qPCR (Kuypers and Jerome, 2017).

- **PCR em emulsão**

A PCR em emulsão (emPCR) é um método comumente usado na amplificação de templates nas múltiplas plataformas de sequenciação NGS. Este método baseia-se na compartimentação de templates nas gotículas de água numa emulsão água/óleo. Idealmente, cada gotícula de água contém apenas uma única molécula template e funciona como um micro reator de PCR. Porém a formação de subprodutos é um grande obstáculo para a amplificação por PCR convencional eficaz. A utilização da PCR em emulsão permite circundar estes obstáculos e ser bastante utilizada nas tecnologias NGS (Shao *et al.*, 2011).

6. Técnicas de sequenciação do DNA

O termo sequenciação no contexto da biologia é, na maioria das vezes, sinónimo da sequenciação de DNA, o que significa identificar os nucleotídeos de uma molécula de DNA de modo a obter-se uma sequência de letras, isto é, uma leitura. O conhecimento das sequências do DNA tornou-se indispensável na investigação científica uma vez que proporcionam grande conhecimento de informação genética que é utilizada nos diversos campos da medicina, como no diagnóstico médico, tratamento clínico, biotecnologia, virologia, entre outros (Kircher and Kelso, 2010).

Com a técnica da PCR (Mullis, 1990), desencadeou-se o desenvolvimento de novas técnicas revolucionárias, como a sequenciação, que auxiliaram a comunidade científica a descobrir os genes que estão associados a doenças, ao fenótipo e potenciais alvos terapêuticos no desenvolvimento de fármacos.

6.1. Sequenciação de 1ª geração

Ao longo dos tempos, as técnicas de sequenciação do DNA sofreram várias evoluções, podendo estas ser caracterizadas segundo o método, ou segundo os resultados provenientes dessas tecnologias (Moreno, 2013).

- **Técnica de Sanger**

A técnica de Sanger iria alterar para sempre o progresso da tecnologia de sequenciação do DNA. A sequenciação clássica de Sanger publicada em 1977 (Sanger *et al.*, 1977), baseia-se em terminações de cadeia específicas de base em quatro reações separadas (“A”, “G”, “C” e “T”) correspondentes aos quatro nucleotídeos diferentes na composição do DNA (Ari and Arikan, 2016). Na presença de todos os quatro desoxinucleotídeos trifosfatos (dNTPs), um dideoxinucleotídeo trifosfatado (ddNTP) específico é adicionado a cada reação; por exemplo, ddATP para a reação “A” e assim por diante (Heather and Chain, 2016). O uso de ddNTPs nas reações de sequenciação foi uma abordagem que deu resultados muito superiores em comparação com a técnica protótipo de 1975, *plus and minus* (Ari and Arikan, 2016), desenvolvida pela mesma equipa. A extensão de uma cadeia de DNA recentemente sintetizada termina cada vez que um ddNTP correspondente é incorporado. Como o ddNTP está presente em quantidades mínimas, a terminação acontece raramente e estocasticamente (Heather and Chain, 2016), resultando num aglomerado de produtos de extensão em que cada posição de uma base N resulta num produto correspondente terminado pela incorporação do ddNTP na extremidade 3’.

O segundo aspeto do método foi o uso de moléculas de fósforo radioativas incorporadas na cadeia de DNA recentemente sintetizada através de um precursor marcado (dNTP ou primer de sequenciação), tornando cada produto detetável na radiografia (Ari and Arikan, 2016). Como cada reação resultava numa mistura complexa de produtos DNA radioativos, esta técnica sofreu melhorias, como a utilização do gel poliacrilamida (PAG), que permitiu a separação individual destas moléculas por eletroforese, como se constata na figura 1. (A) (França *et al.*, 2002).

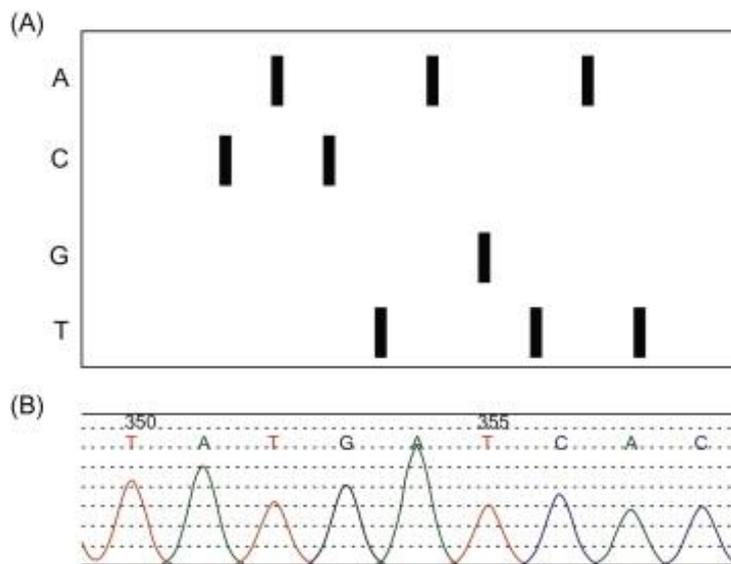


Figura 1. - Sequenciação de Sanger. (A) Modelo dos resultados da eletroforese em gel para sequenciação de Sanger da molécula DNA 50-TATGATCAC-30. A seqüência pode ser lida do fim do gel para o princípio (dos fragmentos mais pequenos para os maiores). (B) Eletroferograma para a sequenciação de Sanger da mesma molécula. Os resultados são lidos da esquerda para a direita (Hagemann, 2015).

Por mais inovador que este método fosse, há cerca de 30 anos, os PAGs eram muito lentos e laboriosos não sendo suficiente para realizar os desafios estabelecidos pelo Human Genome Project (França *et al.*, 2002). Assim, foi desenvolvida a sequenciação moderna de Sanger (sequenciação fluorescente automatizada, sequenciação *dye-terminator*) que usa ddNTPs marcados fluorescentemente que permitem que o passo de amplificação seja realizado numa única reação (Ari and Arikian, 2016). O produto da reação é uma mistura de fragmentos de DNA de cadeia simples de vários comprimentos, cada um marcado numa extremidade com um fluoróforo acoplado ao ddNTP incorporado e que indica a identidade do ddNTP incorporado. A reação é separada por eletroforese capilar, sendo os produtos separados por tamanho. O registo contínuo da intensidade de fluorescência (Karger and Guttman, 2009) de quatro cores na extremidade do capilar, resulta num

eletroferograma (Figura 1. (B)) que pode ser interpretado por um software especializado, como o Mutation Surveyor (Minton *et al.*, 2011).

A sequenciação de Sanger hoje em dia usa o método fluorescente *dye-terminator* e é realizada com kits comercialmente disponíveis. O BigDye v3.1 (*BigDye Terminator v3.1 Cycle Sequencing Kit*, 2002) é recomendado pelo fornecedor para a maioria das aplicações que incluem longos comprimentos de leitura.

6.1.1. Aplicação clínica

Apesar das vantagens das técnicas de sequenciação da *next-generation* (abordadas na secção 6.2.) em que a taxa de transferência é de maior magnitude (Ari and Arikan, 2016), a sequenciação de Sanger mantém um lugar essencial na genética clínica por, pelo menos, duas razões específicas. Em primeiro lugar, a sequenciação de Sanger serve como um método ortogonal de confirmação para as variantes de sequências identificadas pela NGS (Mu *et al.*, 2016). Ao validar os testes clínicos da NGS, os materiais de referência sequenciados pelas abordagens de Sanger fornecem a verdade fundamental contra a qual o ensaio NGS pode ser comparado (Mu *et al.*, 2016). Estes materiais podem incluir reagentes bem caracterizados publicamente disponíveis, tais como linhas celulares estudadas no projeto HapMap (Spencer *et al.*, 2014) ou amostras clínicas previamente testadas pela técnica de Sanger. Sendo um método ortogonal, a sequenciação de Sanger providencia uma forma de confirmar variantes identificadas pela NGS (Mu *et al.*, 2016). Seria impraticável confirmar cada variante pelo método de Sanger dado ao grande número de primers, reações e interpretações que seriam necessárias para este método. No entanto, pode haver casos em que a veracidade de uma variante específica é duvidosa, ou seja, variantes biologicamente implausíveis ou suspeitas de serem espúrias e deste modo, a sequenciação de Sanger é o método mais fácil de resolver essas incertezas sendo, apesar de tudo, um protocolo inestimável em qualquer laboratório de genética clínica (Bartlett *et al.*, 2014).

Em segundo lugar, a sequenciação de Sanger fornece um meio para corrigir a cobertura de regiões que estão mal cobertas pela NGS. Podem existir regiões que são resistentes à

sequenciação pela NGS devido à má captação, amplificação ou outras idiossincrasias (Rehm *et al.*, 2013). Uma maneira de restaurar a cobertura destas áreas é aumentar a quantidade de DNA no seu input, mas a quantidade disponível pode ser limitada. Quando a área a ser preenchida é pequena, uma abordagem muito prática consiste na utilização da sequenciação de Sanger para abranger as regiões mal cobertas pela NGS (Rehm *et al.*, 2013).

Quando a sequenciação de Sanger é usada para preencher dados provenientes da NGS, os dados da NGS e de Sanger devem ser integrados em conjunto para fins de análise e relatórios, o que representa um desafio já que esses dados são obtidos por métodos diferentes e não têm correspondência mútua (Pandey *et al.*, 2016). Por exemplo, as medidas de qualidade de sequência que são significativas para a NGS, não são aplicáveis a Sanger. O conceito de profundidade de cobertura só pode ser aplicado indiretamente aos dados de Sanger. As frequências de alelos são imprecisamente verificadas no método de Sanger, porque são observadas a partir das alturas de pico em vez das contagens de leitura (Roman K Thomas *et al.*, 2006). Embora a NGS possa ser potencialmente validada para permitir a interpretação de uma variante significativa a partir de uma única leitura sem referência, a sensibilidade analítica da sequenciação de Sanger é baixa. Cada vez mais os requisitos para a detecção de variantes são de uma sensibilidade analítica para alelos <20%, em particular quando se avalia o estado mutacional de amostras de tumores heterogêneos (Davidson *et al.*, 2012) ou de quasi espécies infetantes em indivíduos com HIV (Fisher *et al.*, 2015). As variantes com uma frequência alélica mais baixa podem ser indistinguíveis do ruído ou de erros provenientes da sequenciação. Deste modo, o desempenho e um ensaio da NGS pode ser alterado nas áreas emendadas por Sanger, e estes desvios no desempenho devem ser documentados e/ou excluídos (Arsenic *et al.*, 2015).

6.1.2. Limitações técnicas

Apesar do método de Sanger ser considerado pela comunidade científica um método padrão para a sequenciação, esta técnica tem várias limitações. Uma das limitações é a

sua parcialidade biológica. O método é baseado na clonagem de DNA estranho em vetores que têm de ser compatíveis com a maquinaria de replicação utilizada nas células *E. coli* (França et al., 2002). No entanto, esta limitação pode ser superada ao sequenciar diretamente produtos de PCR, mas esta é uma abordagem complexa para sequenciação de novo, apesar de ser rotineiro e de diagnóstico.

Uma outra limitação da técnica de Sanger é a sua habilidade restrita para sequenciar e analisar frequências alélicas (Arsenic *et al.*, 2015). Ao contrário das tecnologias NGS, nas quais cada sequência lida é originada a partir de uma única molécula de DNA, os resultados da sequenciação de Sanger representam as características agrupadas de todas as moléculas molde. Isto não apresenta dificuldade se a amostra for de uma população homogênea. O DNA genômico representa um conjunto de dois haplótipos do doente, pelo que as posições nas quais o doente é heterozigótico vão dar origem a resultados ambíguos se a análise heterozigótica não estiver especificamente ativada (Davidson *et al.*, 2012).

Por exemplo, a heteroplasmia mitocondrial é um cenário clínico em que se verifica variantes de baixa frequência alélica. Esta patologia é clinicamente significativa, uma vez que a gravidade de muitas doenças mitocondriais são dependentes da razão entre as sequências do tipo selvagem e sequências mutantes, com sintomas progressivamente mais graves que podem emergir se a razão dos genomas mutados for superior à dos genomas normais (Sondheimer *et al.*, 2011).

As bases variantes de baixa frequência alélica aparecem em eletroferogramas como picos baixos que podem ser indistinguíveis do ruído da linha de base (Men *et al.*, 2008). A sensibilidade da sequenciação de Sanger deve ser validada em cada laboratório, mas geralmente esta ronda os 20% (Dong and Yu, 2011).

Os alelos variantes abaixo desta frequência podem estar presentes na amostra e são fielmente identificados pela NGS, mas não podem ser confirmados de forma confiável pela sequenciação de Sanger (Sandmann *et al.*, 2017). Porém, peças de software como o Mutation Surveyor têm a sensibilidade necessária para detetar essas baixas frequências alélicas (Figura 2.).

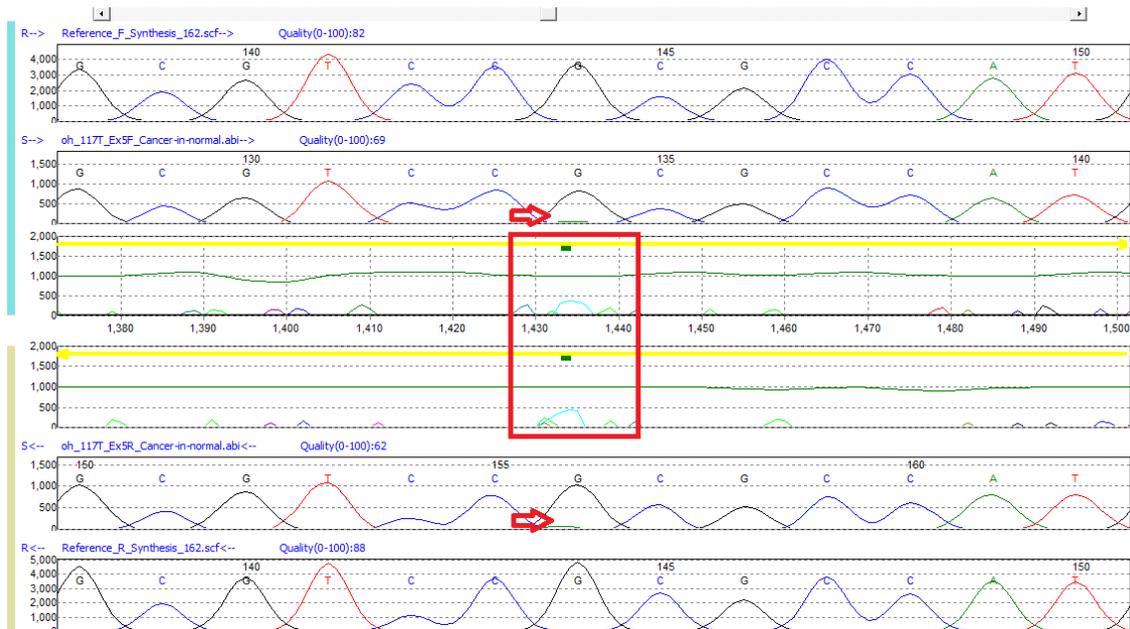


Figura 2. - Mutação G> GA de baixa frequência detetada pelo software *Mutation Surveyor*. A seta na figura indica a presença de uma possível variante de baixa frequência detetada no exão 5 do gene tp53 (Softgenetics, 2016).

6.2. Sequenciação de 2ª geração – Next-generation sequencing

O projeto do genoma humano ajudou a promover o desenvolvimento de técnicas de sequenciação de DNA mais rápidas e baratas. Como já foi descrito, a sequenciação de Sanger (Sanger *et al.*, 1977) dominava a indústria levando a uma série de realizações monumentais, incluindo a conclusão da sequenciação do genoma humano. Apesar das conquistas tecnológicas durante esta era, as limitações da técnica de Sanger, nomeadamente a quantidade de dados de produção, velocidade, qualidade de sequenciação (Janitz, 2011) e os seus custos elevados, impulsionaram o desenvolvimento de novos métodos. Sobretudo, a necessidade de conduzir projetos de sequenciação em grande escala de forma mais económica promoveu o desenvolvimento da tecnologia NGS (Woollard *et al.*, 2011).

Em comparação com o método de Sanger, as plataformas NGS são conhecidas como métodos de sequenciação paralela devido à sua capacidade de sequenciar centenas de milhões de fragmentos de DNA simultaneamente em paralelo, capturando uma quantidade vasta de informação genómica a um baixo custo (Gagan and Van Allen, 2015).

O seu comprimento de leitura, isto é, o número real de bases contínuas sequenciadas, para a NGS é muito mais curto do que o alcançado pelo método Sanger. Atualmente, a tecnologia NGS permite a sequenciação rotineira de 50-500 bp (mas com picos bem superiores) em contraste com o método de Sanger que, após uma série de inovações tecnológicas, atingiu a capacidade de rotineiramente ler 1000-1200 bp (Zhang *et al.*, 2011).

Contudo, esta menor capacidade de leitura sequencial de cada fragmento é ultrapassada pela massiva quantidade de fragmentos lidos em paralelo, permitindo com isso a sequenciação total de todo um genoma ou, inclusivamente, a obtenção de informações completas em misturas de sequências como são os casos da do microbioma (Jovel *et al.*, 2016), ou em misturas heterogêneas de células como acontece com infeções pelo vírus da imunodeficiência humana (HIV) (Capobianchi *et al.*), ou em casos de cancro (Gagan and Van Allen, 2015). Por este motivo, nos últimos anos a NGS tem tido um maior impacto no estudo do cancro porque esta tecnologia permite o rastreio do genoma do cancro, melhorando significativamente a nossa compreensão dos mecanismos que impulsionam a iniciação, progressão, manutenção, resistência e gestão clínica do cancro (Serrati *et al.*, 2016). A NGS fornece assim um catálogo abrangente de sequências genómicas dentro das células cancerosas, permitindo a deteção da heterogeneidade interpaciente e intratumoral que facilita as decisões de tratamento para a terapia personalizada do cancro, e consequentemente, com um impacto positivo nos resultados clínicos (Serrati *et al.*, 2016).

Para além do progresso na área da oncologia com a tecnologia NGS, a informação genómica proveniente desta técnica também pode ser utilizada para a determinação da resistência a fármacos (antibióticos e quimioterapêuticos), prever a resposta a intervenções médicas e dirigir e direcionar o desenvolvimento de novos agentes

farmacêuticos que afetam alvos biológicos associados à progressão da doença (Yadav *et al.*, 2014).

Com todo este progresso tecnológico, a NGS tem sido especialmente vantajosa para a indústria farmacêutica, permitindo uma maior gestão de recursos e providenciando resultados de sequenciação confiáveis (Nielsen *et al.*, 2011).

As indústrias farmacêuticas já adotam as tecnologias NGS e já se pode ver o resultado do impacto desta tecnologia no desenvolvimento de novos fármacos e vacinas (Woollard *et al.*, 2011). A sua alta sensibilidade na detecção de sequências presentes em níveis baixos, significa que a NGS pode ser usada como ferramenta de controlo de qualidade nas vacinas detetando vírus adventícios ao utilizar uma abordagem metagenómica (Victoria *et al.*, 2010). As vacinas contra a poliomielite, a rubéola, o sarampo, a papeira e o rotavírus já foram analisadas desta forma, e as empresas já oferecem este tipo de abordagem como um serviço de despistagem para as indústrias farmacêuticas e biotecnológicas. Neste caso, a NGS é utilizada no início do desenvolvimento das vacinas garantindo que quaisquer contaminantes nos processos, se presentes, são detetados precocemente, e para além disso, permite identificar as diferentes vias pelas quais os patógenos ativam respostas imunológicas protetoras que podem aumentar potencialmente essas respostas ao vacinar pessoas com base nas suas assinaturas genéticas, identificadas até à data, que preveem a sua imunogenicidade e segurança (Furman and Davis, 2015).

A detecção da baixa abundância das variantes HIV resistentes a fármacos é um outro exemplo da utilização da NGS no desenvolvimento de fármacos. A NGS permite detetar variantes em vários pacientes que são responsáveis pela resistência aos fármacos antirretrovirais usados, e conseqüentemente, permitiu aos clínicos elaborar planos de regimes antirretrovirais, submetendo esses mesmos pacientes a uma terapia diferente (Woollard *et al.*, 2011).

Um outro campo em que a NGS começa a ter aplicação é no desenvolvimento de produtos biofarmacêuticos, mais concretamente, no desenvolvimento de bibliotecas de anticorpos. Vários grupos investigaram como é que esta tecnologia podia complementar a biblioteca de anticorpos. De acordo com um estudo publicado em 2009 (Dias-Neto *et al.*, 2009), a NGS foi usada juntamente com a PCR em tempo real numa tentativa de remodelar o

modelo tradicional da ação dos fagos. Em comparação com os métodos tradicionais, envolvendo a técnica da ELISA e a sequenciação de Sanger, este novo método mostrou ser mais rápido e produzir mais informação. Embora tenha sido inicialmente aplicado ao estudo das interações proteína-proteína, o método também pode ser aplicável ao estudo da interação anticorpo-antígeno (Di Niro *et al.*, 2010).

Para além da NGS possibilitar todos estes novos progressos na indústria farmacêutica, esta também permitiu que a sequenciação do genoma humano baixasse para um preço que ronda os 1.000\$ (Christensen *et al.*, 2015).

- **Illumina sequencing**

Os sequenciadores e reagentes desenvolvidos pela Solexa, e mais tarde comercializados pela Illumina, Inc., tornaram-se uma das plataformas mais utilizadas na genómica clínica devido à sua versatilidade e às suas vantagens de custo/velocidade/rendimento (Buermans and den Dunnen, 2014). A característica mais distintiva da sequenciação Illumina está no modo de como as bibliotecas são preparadas. Estas bibliotecas são híbridas com uma superfície bidimensional de uma célula de fluxo (*flow cell*) que é posteriormente submetida a uma amplificação em ponte que resulta na criação de um agrupamento localizado (uma colónia de PCR) com cerca de 2000 fragmentos de biblioteca idênticos com um diâmetro de $\sim 1\mu\text{m}$ (Buermans and den Dunnen, 2014). Estes fragmentos são sequenciados no local por incorporação sucessiva de nucleótidos com terminação reversível marcados com fluorescência (Hodkinson and Grice, 2015). Depois de cada etapa de incorporação, a superfície da célula de fluxo é interpretada por um dispositivo de carga acoplada (CCD) que consulta cada posição para a identidade do nucleótido incorporado mais recentemente (Hodkinson and Grice, 2015). Após a formação de imagens, o grupo fluorescente é clivado, o terminador reversível é desativado e as cadeias do template estão prontas para o próximo ciclo de incorporação. Estes ciclos sucessivos de incorporação e de imagem resultam numa série de ficheiros de imagem grandes, que são subsequentemente analisados para determinar a sequência em cada polónia (Hodkinson and Grice, 2015).

Os sequenciadores Illumina de 2ª geração mais utilizados que utilizam a sequenciação por síntese (SBS) são os sistemas HiSeq e MiSeq. O HiSeq é um instrumento de maior escala enquanto que o MiSeq é conceitualizado como um sequenciador pessoal. Dependendo da plataforma e dos reagentes, atualmente pode-se obter até 300 e 500 ciclos de dados de sequências nas plataformas HiSeq e MiSeq respectivamente (Kozich *et al.*, 2013).

O sistema HiSeq (modelo 2500) aceita uma ou duas células de fluxo, cada uma com oito vias. Neste sistema, uma única célula de fluxo pode gerar até 300Gb de dados de 2x100 bp. Demora aproximadamente 11 dias para prosseguir automaticamente da geração do cluster para a sequência final. No modo de execução rápida, uma célula de fluxo pode gerar 60 Gb de leituras de 2x100 bp ou 90 Gb de leituras de 2x100 bp, necessitando de 27 horas e 40 horas respectivamente (Wang, 2016).

O sistema MiSeq aceita uma célula de fluxo de uma única via. Com os kits de reagentes atuais e referências padrão, o MiSeq consegue gerar 5,1 Gb de leituras de 2x150 bp em cerca de 24 horas. Utiliza-se o MiSeq quando a quantidade total de sequência desejada é insuficiente para justificar o uso de uma célula de fluxo inteira do sistema HiSeq, que pode ser devido ao baixo número de amostras a serem agrupadas em conjunto, ao tamanho pequeno da região alvo a ser sequenciada, ou devido à baixa cobertura (Wang, 2016).

- **SOLiD Sequencing**

A plataforma *Sequencing by Oligo Ligation Detection* (SOLiD), em vez de realizar a sequenciação por síntese de um nucleótido de cada vez, envolve a hibridação e ligação de sondas e sequências âncora a uma cadeia de DNA (Tomkinson *et al.*, 2006). As sondas codificam uma ou duas bases conhecidas que fazem ligação complementar entre a sonda e o template. As sequências âncora codificam uma sequência conhecida que é complementar a uma sequência adaptadora, proporcionando um local para iniciar a ligação (Landegren *et al.*, 1988). Após a ligação, a base ou bases conhecidas na sonda são identificadas e um novo ciclo começa, após a remoção completa do complexo âncora-

sonda ou por clivagem para remover o fluoróforo e promover regeneração do local de ligação (Moorthie *et al.*, 2011).

A plataforma SOLiD utiliza sondas codificadas com duas bases, em que cada sinal fluorométrico representa um dinucleótido (Valouev *et al.*, 2008). Conseqüentemente, o seu output não está diretamente associado à incorporação de um nucleótido conhecido. Uma vez que as 16 possíveis combinações de dinucleótidos (Figura 3.) não podem ser individualmente associadas com fluoróforos espectralmente resolúveis, utilizam-se quatro sinais fluorescentes. Cada subconjunto de quatro combinações de dinucleótidos é representado por um tipo de fluoróforo (*color space*). Assim, cada sinal de ligação representa um dos vários possíveis dinucleótidos que devem ser processados por imagem durante a análise de dados (Massingham and Goldman, 2012; Lee *et al.*, 2015).

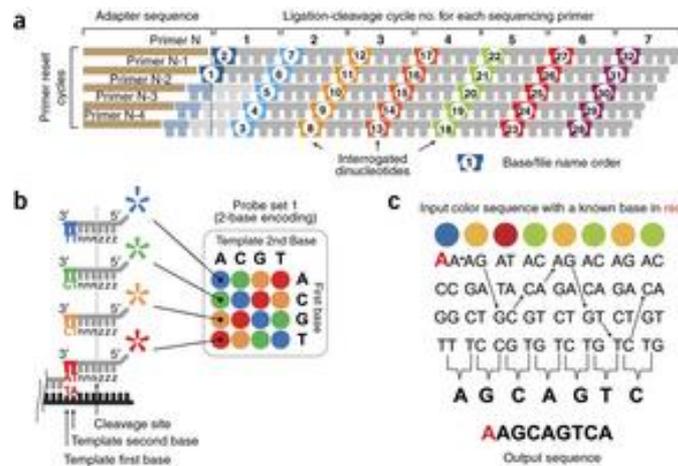


Figura 3. Visão geral do esquema de codificação e decodificação de cores SOLiD. (a) A posição de base dentro da sequência de templates é delimitada por círculos brancos e dever ser usada para nomear os arquivos de imagem, e os números de ciclos de seqüência reais são observados em ambos os lados. Cada extensão de ligação é mostrada a cores diferentes. (b) Esquema de codificação de color space SOLiD. (c) Esquema de decodificação SOLiD. Desde que qualquer uma das identidades de base seja conhecida (a vermelho), a seqüência color space pode ser convertida numa seqüência de nucleótidos (Lee *et al.*, 2015).

A sequenciação SOLiD tem várias vantagens. Uma delas é utilização dos fluoróforos aos subconjuntos que trás uma maior especificidade e maior redução de erros quando comparada com o emparelhamento de bases de um único nucleótido (Massingham and Goldman, 2012). A metodologia de alongação do primer utilizada neste processo, é também altamente tolerante a sequências repetitivas. Isto porque, na alongação do primer realizam-se cinco rondas de alongação e indicadores, cada uma das quais é seguida por uma reposição do primer. Em cada ciclo, um primer universal é hibridizado com as cadeias template, posicionando uma extremidade no início da região a ser sequenciada. Esta posição é deslocada posteriormente por um nucleótido em cada ciclo subsequente da alongação do primer. São depois realizados vários ciclos de ligação, detecção e clivagem (Massingham and Goldman, 2012). Em cada passo de ligação é introduzido um conjunto de oligonucleótidos já marcados, e a cor do fluoróforo ligado a cada molécula corresponde à sequência dos dois primeiros nucleótidos. Os restantes seis nucleótidos são desnaturados. Realiza-se uma reação de ligação de modo que o fluoróforo ligado a cada bead indica a sequência dinucleotídica em *downstream* da do primer de sequenciação. O substrato é lavado, e uma imagem digital é adquirida para documentar a cor ligada a cada bead (Massingham and Goldman, 2012).

Então, a codificação *color space* dos nucleotídeos permite um certo grau de proteção contra erros. Uma vez que cada nucleótido é sondado duas vezes, as mutações pontuais manifestam-se sempre com duas mudanças de cor em relação ao padrão nas duas rodadas sucessivas da alongação do primer. Uma única mudança de cor normalmente reflete um erro de sequenciação (Massingham and Goldman, 2012).

A sequenciação SOLiD pode ser realizada na plataforma *Applied Biosystems series 5500 Genetic Analyzer* que utiliza um ou dois FlowChips. Cada FlowChip tem seis colunas que permite gerar cerca de 120 Gb de dados no modo de sequenciação de fragmentos 1x75 bp. No modo 2x50 bp, cada chip gera rotineiramente 160 Gb. Existem outras plataformas mais antigas disponíveis, mas têm uma menor capacidade de sequenciação (Sara El-Metwally *et al.*, 2014).

- **Sequenciação Ion Proton**

Todas as plataformas acima descritas envolvem a tradução da sequência de DNA para um evento físico detetável. Enquanto que a sequenciação de Sanger, a sequenciação Illumina e SOLiD são baseadas na detecção com base na fluorescência, a plataforma *Ion Torrent* da *Life Technologies* baseia-se na conversão da sequência em pequenas alterações de pH que ocorrem como resultado da libertação do H^+ quando os nucleótidos são incorporados numa sequência de elongação (Merriman and Rothberg, 2012; Yuan *et al.*, 2016).

A preparação da biblioteca para esta plataforma envolve a fragmentação, a ligação do adaptador e a seleção do tamanho. As moléculas da biblioteca são ligadas à superfície dos beads e amplificadas por emPCR de modo que cada bead é revestido com uma população homogénea de moléculas. Cada bead é então disperso num poço à superfície de um chip de matriz de sensores semicondutores (Figura 4.) construídos pela tecnologia complementar de semicondutores de óxido metálico (CMOS) (Merriman and Rothberg, 2012).

O substrato tem a propriedade de que cada poço, que contém um transístor sensível a iões, funciona como um medidor de pH que é extremamente sensível às flutuações de pH. A sequenciação da *Ion Torrent* é realizada inundando a placa com uma espécie de desoxinucleótido (dA, dC, dG, dT) de cada vez. Em cada poço, a incorporação de um nucleótido provoca a libertação de um pirofosfato e um H^+ , que produz uma alteração no pH. A mudança na concentração de H^+ fornece uma mudança na voltagem na porta do transístor, permitindo que a corrente passe através do transístor, produzindo um sinal. Os homopolímeros, isto é, nucleótidos múltiplos, causam a incorporação de uma maior número de nucleótidos, com uma mudança de pH correspondente maior e, consequentemente, um sinal maior (Merriman and Rothberg, 2012).

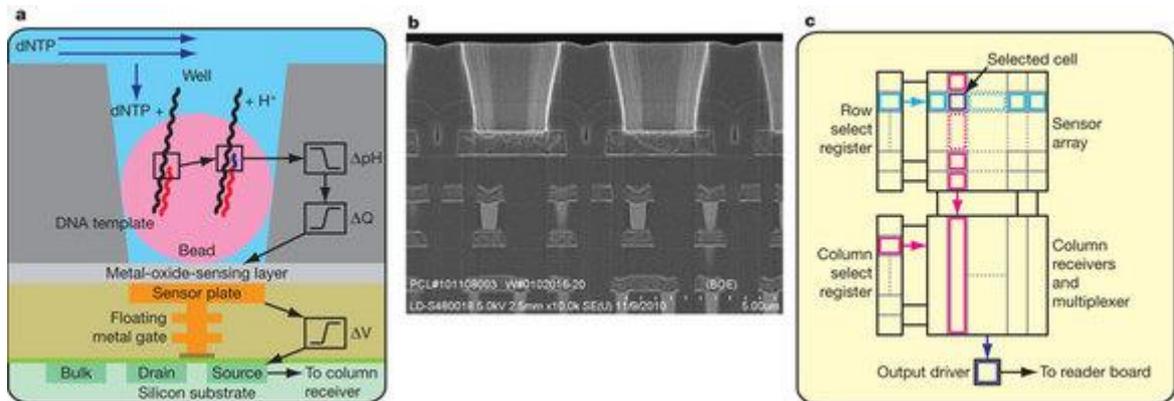


Figura 4. Design do sensor, poço e chip utilizados pela *Ion Torrent*. (a) Desenho simplificado de um poço e um bead que contém um template de DNA, com o sensor e eletrônica subjacente. Os prótons (H^+) são libertados quando os desoxinucleótidos (dNTP) são incorporados nas cadeias de DNA em crescimento, alterando o pH do poço (ΔpH). Isto induz a uma alteração no potencial de superfície da camada de detecção do óxido de metal, e uma alteração no potencial (ΔV) do terminal da fonte do transistor. (b) Micrografia dos eletrões mostrando o alinhamento dos poços sobre a placa de sensor de metal ISFET e as camadas eletrônicas subjacentes. (c) Os sensores são dispostos numa matriz bidimensional. Um registo da seleção de linha permite que cada sensor acione a sua tensão da fonte para uma coluna e um registo da seleção da coluna seleciona uma das colunas para o output que é registado eletronicamente (Rothberg *et al.*, 2011).

A reação da sequenciação é altamente análoga à pirosequenciação, com uma diferença de que o pH é detetado em vez da luz, e a taxa de transferência do método depende do número de poços por chip, que por sua vez, está relacionado com a restrição de fabricação dos semicondutores. Uma série de chips Proton transporta cerca de 154 milhões de poços, cada um com $1,25 \mu m$ de diâmetro com um espaçamento de $1,68 \mu m$ (Merriman and Rothberg, 2012), no entanto já foram lançados novos chips, Ion 316 e Ion 318 que contêm 165 e 660 milhões de poços respetivamente (Moreno, 2013) aumentando o seu rendimento.

Na gama alta de desempenho da *Ion Torrent Personal Genome Machine* (PGM™), as referências atuais são de 45,5 milhões com leituras de 400 bp (1,2 Gb de sequência) e com um tempo de execução de 7,3h (Ion PGM System Specifications 2013).

A utilização de desoxinucleótidos naturais, acima referidos, em vez dos seus derivados, torna este método único ao reduzir a parcialidade na sequenciação relacionada com a incorporação de nucleótidos não naturais. Em comparação com as outras plataformas, as leituras são relativamente longas, e os tempos de reação são muito curtos (cerca de 3h para 300 bases) (Merriman and Rothberg, 2012). O software associado à sequenciação do *Ion Torrent* é distribuído como um produto de código aberto que contribuiu para a sua adesão relativamente ampla.

- **Sequenciadores genómicos Roche 454**

As plataformas Roche 454 são baseadas na pirosequenciação. Na configuração convencional, isto é, *non-array-based*, a pirosequenciação é um método altamente sensível para a sequenciação por síntese. Uma cadeia template de DNA é imobilizada e exposta a desoxinucleótidos individuais, na presença da DNA polimerase, na luciferase, e na ATP sulurilase, que está ilustrado na figura 5 (Mardis, 2008).

A luciferina está presente como um substrato da luciferase, e o fosfosulfato de adenosina 50 apresenta-se aqui como o substrato para a ATP sulfurilase. Cada nucleótido incorporado provoca a libertação de uma porção de pirofosfato inorgânico (PPi), e este PPi é combinado com o fosfosulfato de adenosina 50 para formar adenosina trifosfato (ATP) numa reação catalisada pela ATP sulfurilase. Na presença de ATP, a luciferase converte a luciferina em oxiluciferina, emitindo luz, que resulta num sinal que pode ser captado. Para evitar o consumo de dATP pela luciferase, resultado na emissão de sinais aberrantes, o dATP α S é utilizado como fonte de nucleótidos de desoxiadenosina. A apirase é adicionada mais tarde para degradar os trifosfatos para nucleótidos monofosfatados com um fosfato inorgânico Pi, iniciando-se um novo ciclo (Mardis, 2008).

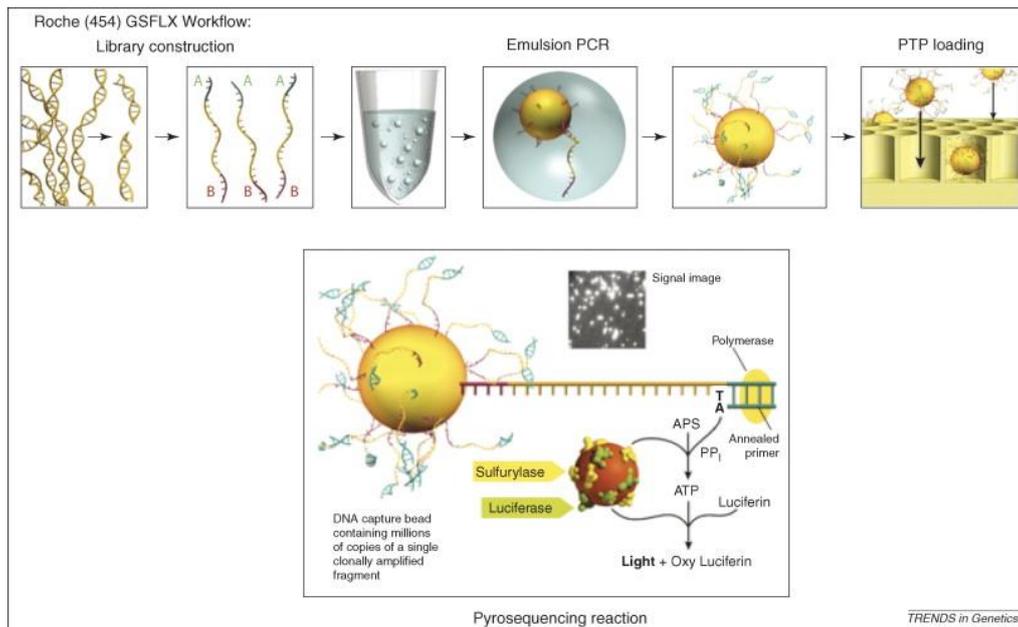


Figura 5. Metodologia da Roche 454: Preparação da biblioteca, PCR em emulsão, e carregamento de chips seguindo-se a sequenciação (Mardis, 2008).

A pirosequenciação foi mais tarde adaptada pela corporação 454, para ter um rendimento mais alto, desenvolvendo uma plataforma paralela altamente multiplexada. Os fragmentos de DNA da biblioteca são capturados na superfície de pequenos beads, com uma diluição limitante que assegura que cada bead receba apenas uma única molécula (Figura 5.). Estas moléculas de DNA são amplificadas numa emulsão de óleo em água (emPCR) na superfície dos beads, e em seguida são immobilizadas na superfície da placa PicoTiter dentro de um poço. A fibra ótica utilizada neste processo, permite que a emissão da luz em cada poço seja detetada (Mardis, 2008).

O 454 GS20 *Genome Sequencer*, lançado em 2005, foi a primeira plataforma NGS a tornar-se comercialmente disponível. As plataformas atuais incluem o sistema GS FLX1 e o GS Junior, que é uma versão mais baixa (Sara El-Metwally *et al.*, 2014). Estes instrumentos incorporam a ótica e reagentes necessários para executar a reação de sequenciação e capturar os dados resultantes deste método. O GS FLX1, em combinação química do GS FLX Titanium (Gilles *et al.*, 2011), é frequentemente citado como tendo

comprimentos de leitura até 1000 bp, e uma produção de 700 Mb por execução de 23 horas. Tem uma precisão reportada de 99,9% com uma cobertura de 15x. O GS Junior, com a química do GS Junior Titanium, tem comprimentos de leitura de 400 bp, e uma taxa de transferência de 35 Mb em 10 horas de execução. A sua precisão situa-se nos 99% (Sara El-Metwally *et al.*, 2014). A grande vantagem que esta plataforma 454 oferece incluiu a produção de um comprimento de leitura longo e um tempo de resposta curto (Sara El-Metwally *et al.*, 2014).

6.2.1. Aplicação clínica

A tecnologia NGS é comumente usada para o teste de mutações humanas herdadas nas doenças genéticas e também, para mutações adquiridas nas aplicações de oncologia, incluindo o diagnóstico de cancro, prognóstico e previsão de resposta à quimioterapia (Shen *et al.*, 2015). A sequenciação de genes isolados ou porções de genes que transportam mutações específicas, é eficiente em termos de custo e tempo e tornou a sequenciação acessível nos laboratórios clínicos dos grandes centros médicos (Shen *et al.*, 2015). Os painéis de genes pré-concebidos estão atualmente comercialmente disponíveis em kits dedicados para a sequenciação amplicon (pedaço de DNA ou RNA que é o produto de um processo de amplificação ou replicação), incluindo os painéis de cancro *Ion AmpliSeq Ready-to-Use (Life Technologies)* (Ion AmpliSeq™ Library Preparation for Human Identification Applications 2015), o painel *TruSeqsAmplicon Cancer* (Illumina) (TruSeq® Amplicon - Cancer Panel 2012), e o painel *NimbleGenSeqCap EZ Comprehensive Cancer* (Roche) (van der Werf *et al.*, 2015). Para além disso, estes painéis de genes personalizados podem ser usados noutras plataformas de sequenciação. Os painéis AmpliSeq e TruSeq dependem do PCR multiplex tanto para a seleção do alvo como para a amplificação, enquanto que o sistema NimbleGenSeqCap combina métodos de captura de hibridização da solução com a amplificação de PCR secundária para a formação das bibliotecas de amplicon. A implementação da tecnologia NGS num laboratório clínico é, no entanto, complexa e requer conhecimentos significativos em aspetos clínicos, técnicos e bioinformáticos (Gullapalli *et al.*, 2012;

Loman *et al.*, 2012). Os laboratórios que têm esta tecnologia implementada, aplicam rigorosamente todas as medidas de qualidade para o desenvolvimento, validação e garantia da qualidade da NGS sob a orientação rigorosa de vários regulamentos (Endrullat *et al.*, 2016).

6.2.2. Limitações técnicas

Apesar das conquistas científicas que as plataformas de 2ª geração proporcionaram, existem limitações que, de alguma forma, não permitem que se explore todo o potencial que estas oferecem, abrindo caminho para o desenvolvimento de técnicas para a sequenciação de DNA da 3ª e 4ª geração, descritas com mais detalhe a seguir (Park and Kim, 2016).

Começando pela plataforma SOLiD, a sua desvantagem incluiu a sua complexidade conceitual (Shendure and Ji, 2008). Os comprimentos de leitura curtos tornam a montagem da sequência difícil, particularmente para a *de novo assembly* (Liu *et al.*, 2012). Esta tem sido muito útil nos estudos de transcriptomas e genomas (Cao *et al.*, 2015) e esta limitação faz com que esta plataforma não seja ideal para esses estudos.

Procedendo para a plataforma *Ion Torrent*, uma desvantagem frequentemente citada é a sua taxa de erro relativamente elevada em comparação com as plataformas baseadas na Illumina, particularmente nas regiões homopoliméricas. Algumas frações e leituras são rejeitadas devido a ensaios de qualidade insuficientes, incluindo problemas na razão sinal/ruído, a aglomeração de beads, os beads não clonais, ou falha na incorporação de um bead num dado sensor. Uma vez excluídas estas leituras, uma precisão de 99,5% foi relatada para as leituras de 250 bp (Merriman and Rothberg, 2012).

A fonte de erro mais problemática está relacionada com os homopolímeros, cujo comprimento deve ser deduzido a partir da altura do pico. Mesmo uma precisão de 99,3% (Merriman and Rothberg, 2012) dentro dos homopolímeros significa que haverá leituras corretas em 96,5% do tempo, o que é significativo dado que leituras com este comprimento não são incomuns no genoma. As dificuldades de sequenciação através dos homopolímeros causam incertezas na sequência da própria leitura e também causam um

desfasamento dos nucleotídeos subsequentes devido à extensão incompleta, embora a ordem em que os nucleotídeos são fluidos através do chip possa ser alterada para atenuar estes problemas (Massingham and Goldman, 2012).

Na plataforma Roche 454, o custo por base sequenciada para esta plataforma é extremamente alto, o que faz com que este facto seja a maior desvantagem desta plataforma. O custo ronda sensivelmente os 10 dólares por milhão de bases (Liu *et al.*, 2012). Para além disso, o passo da sequenciação, que envolve os homopolímeros, é uma outra limitação significativa. Tal como a sequenciação pelo *Ion Torrent*, os homopolímeros são detetados como eventos de incorporação de grande magnitude à medida que os nucleótidos são incorporados (Shendure and Ji, 2008). A relação entre o comprimento do percurso e a intensidade do sinal, isto é, a área sob a curva, não é estritamente estequiométrica, no que resulta em dificuldades acrescidas para a sequenciação nos homopolímeros com mais de 5 bases de comprimento (Sara El-Metwally *et al.*, 2014).

Para a plataforma Illumina, o que se verifica é que os comprimentos de leitura são limitados devido a uma multiplicidade de fatores que causam a degradação do sinal. A clivagem incompleta de marcadores fluorescentes contribuiu para o desfasamento do sinal. Uma outra desvantagem que se aponta é o facto dos instrumentos Illumina serem caros comparativamente com os das outras plataformas (Shendure and Ji, 2008).

Na tabela que se segue, apresenta-se de uma forma resumida as vantagens e desvantagens de plataformas comercialmente disponíveis.

Tabela 2. – Vantagens e desvantagens das plataformas NGS (Adaptado de Sara El-Metwally *et al.*, 2014)

Instrumento	Principais Vantagens	Principais Desvantagens
454 FLX Titanium	Longo comprimento de leitura.	Custos elevados por MB.

Continuação da tabela 2.

454 FLX +	Dobro máximo do comprimento de leitura do Titanium.	Alto custo de capital e por MB; Problemas na atualização.
Illumina HiSeq 2000	Possui duas células de fluxo e menor custo por MB em todas as plataformas.	Elevado custo de capital e necessidade elevada de computação.
Illumina MiSeq	Custo do instrumento moderado; custos por MB moderados; maior comprimento de leitura da Illumina.	Baixas leituras e maior custo por MB quando comparado com o HiSeq.
Ion Torrent – PGM™	Instrumento fácil de usar; instrumento de baixo custo com chips descartáveis.	Maior taxa de erro que a Illumina; maior custo por MB em comparação com o MiSeq.
Continuação da tabela 2. Ion Torrent – Proton	Instrumento de custos moderados; custo similar ao MiSeq mas com mais comprimentos de leitura.	Maior taxa de erro que a Illumina; leituras mais curtas que o MiSeq e HiSeq 2500 Rapid Run.
SOLiD 5500xl	Cada flow-chip pode ser executado independentemente; grande taxa de precisão.	Longevidade da plataforma; curtas leituras; maior dificuldade de processamento de dados que a Illumina; grande custo de capital.

6.3. Sequenciação de 3ª e 4ª geração

Como foi descrito anteriormente, os métodos de segunda geração requerem a preparação de uma biblioteca seguindo-se a sua amplificação em PCR. Estes passos são laboriosos, consomem tempo e recursos, e certas regiões que são amplificadas resultam em erros de PCR que se refletem nos dados analisados (Rhoads and Au, 2015).

Os métodos de 3ª geração contornam estes problemas, uma vez que adotam uma abordagem de sequenciação de uma única molécula e detêm a capacidade de detetar sinais muito pequenos sem comprometer a sua precisão. Ou seja, não têm a necessidade de amplificar o template, embora o próprio passo de sequenciação ainda envolva a sequenciação por síntese (Roberts *et al.*, 2013). As plataformas atuais da terceira geração incluem as seguintes plataformas – PacBioRS (Ferrarini *et al.*, 2013) fabricada pela *Pacific Biosciences*, e *Heliscope Sequencer* (John F. Thompson and Steinmann, 2010) produzida pela Helicos BioSciences.

Nesta parte, as abordagens de 4ª geração serão definidas como aquelas que envolvem a análise da sequência de moléculas de DNA únicas, não envolvendo etapas de amplificação. A sequenciação é realizada sem a síntese de DNA, e por essa razão, não há marcação dos nucleótidos nos passos de deteção. As tecnologias baseadas em nanoporos são os melhores exemplos das plataformas de 4ª geração, sendo a *Oxford Nanopore Technology* (ONT) (Feng *et al.*, 2015) a única comercialmente disponível.

Existem outras tecnologias em desenvolvimento, como a *Transmission Electron Microscopy* (Bell *et al.*, 2012) e a *Electronic Sequencing* (Fuller *et al.*, 2016), mas estas ainda não são usadas no uso clínico e não estão disponíveis comercialmente. No entanto, as suas características e designs inovadores faz com que valha a pena serem mencionadas.

- **Helicos HeliScope**

HeliScope baseia-se na tecnologia de sequenciação de uma única molécula e foi a primeira plataforma capaz de sequenciar moléculas individuais. Esta foi desenvolvida pela, agora defunta, Helicos Biosciences e tal como nas outras plataformas a HeliScope necessita de uma biblioteca de DNA. Esta é construída pela fragmentação, antes da adição

da cadeia poli-A (Figura 6.) em cada molécula, aleatória da amostra do genoma e estes fragmentos são posteriormente desnaturados e hibridizados na célula de fluxo com oligómeros poli-T ligados à superfície. As polimerases passam na célula de fluxo com um tipo de nucleótido de cada vez (A, T, C e G) e as sequências reversas pelas polimerases que começam pelos oligonucleótidos poli-T (Kircher and Kelso, 2010). Os ciclos de sequenciação por síntese prosseguem com recurso a nucleóticos marcados fluorescentemente que são depois visualizados com uma câmara CCD. É lida primeiro a fluorescência das adenosinas que se ligam às sequências, depois as bases citosina, marcadas também fluorescentemente que complementam a sequência. A leitura é novamente feita e o processo repete-se para todas as outras bases até se atingir o comprimento de leitura desejável (Kircher and Kelso, 2010). Um terminador de cadeia previne a adição de nucleótidos extra em cada ciclo o que eliminava os erros que pudessem advir desse passo.

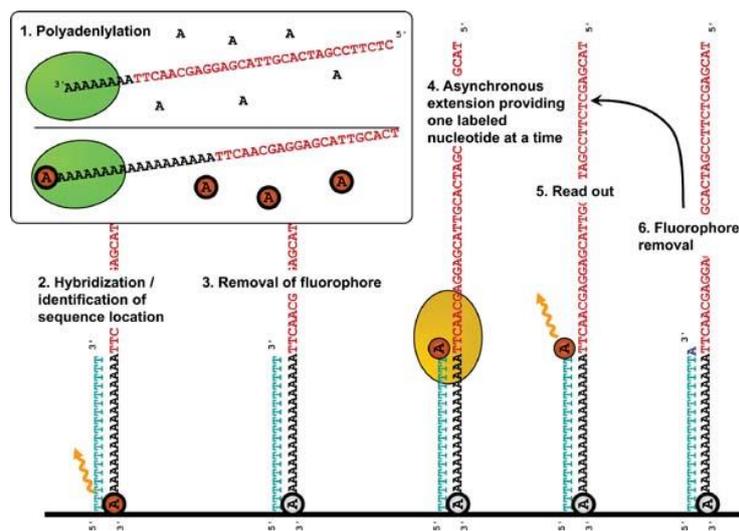


Figura 6. – Sequenciação de uma única molécula pela HeliScope (Kircher and Kelso, 2010).

Uma vez que esta plataforma não requer um passo de amplificação antes da sequenciação, devido à abordagem *true single-molecule sequencing* (tSMS) (Pareek *et al.*, 2011) que esta emprega, evita-se assim mutações artificiais de DNA introduzidas pelos ciclos de PCR. Para além disso, a abordagem tSMS pode ser usada para as aplicações RNA-seq diretas, o que representa uma grande vantagem sobre os outros métodos de sequenciação paralelos.

- **Single-Molecule Real-Time (SMRT) da Pacific Biosciences**

A tecnologia SMRT baseia-se em três processos-chave que incluem, a célula SMRT que permite observar a incorporação dos ácidos nucleicos individuais em tempo real; a utilização de ácidos nucleicos fosforilados, que permite a obtenção de longos comprimentos de leitura; e uma nova plataforma que permite a deteção de uma única molécula (Yanhu *et al.*, 2015). De um modo resumido, para sequenciar o DNA, utiliza-se as células SMRT que contêm um chip de sequenciação, que por sua vez, integra milhares de guias de ondas óticas (ZMWs – Zero-mode waveguides). Estes guias, também conhecidos por nanowaveguide, são produzidos por técnicas de fabricação de semicondutores que utilizam estruturas físicas com um revestimento de alumínio num substrato de sílica transparente com capacidade de criar ondas eletromagnéticas presentes no espectro visível (Foquet *et al.*, 2008). Portanto, a sequenciação de uma única molécula de DNA é realizada apenas por uma DNA polimerase que se encontra ligada ao fundo de cada um dos ZMW (Figura 7.) (Pile, 2009).

À medida que a luz se desloca através da pequena abertura, e também devido às suas características físicas, torna-se possível medir a atividade de uma única molécula de DNA polimerase. Esta molécula está ligada ao fundo do guia de ondas, e como o volume do tampão iluminado num guia ZMW individual contém apenas 20 zeptolitros (20×10^{-21} litros), permite a observação da emissão a partir das bases individuais que são incorporadas na cadeia de DNA numa matriz do dispositivo CCD. Mesmo assim, a abordagem SMRT ainda envolve a sequenciação por síntese que utiliza dNTPs marcados com fluorescência (Korlach *et al.*, 2008).

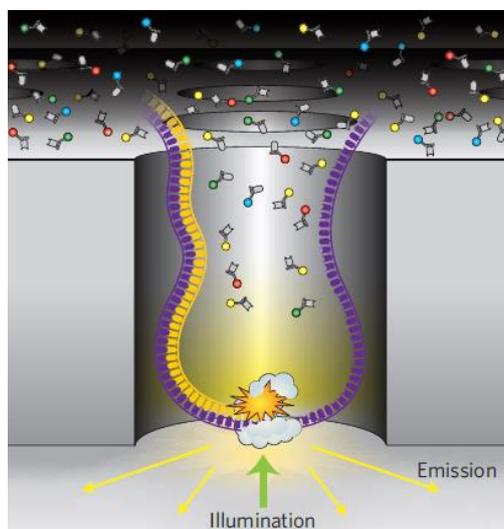


Figura 7. – Sequencição de um única molécula de DNA em tempo real num nanowaveguide (Pile, 2009).

Esta técnica tem bastantes vantagens, e uma delas distingue-se pelo facto de produzir leituras com comprimentos rotineiramente de 3000 bp, podendo também atingir comprimentos que rondem os 20.000 bp (Roberts *et al.*, 2013). Esses longos comprimentos de leitura tornam a *de novo assembly* de *contigs* (conjunto de segmentos de DNA sobrepostos que, em conjunto, representam uma região consensual de DNA) e genomas muito mais simples. Estas leituras com longos comprimentos permitem resolver variantes estruturais como inserções e zonas repetitivas de guanina e timina que facilita a localização de uma leitura no genoma de referência (Tattini *et al.*, 2015). Embora essas leituras individuais contenham uma percentagem significativamente maior de erros, quando comparado com as leituras provenientes da Illumina e outras plataformas (Goodwin *et al.*, 2016). Esses comprimentos de leitura fornecem sequências concisas que são estatisticamente relevantes com um alto grau de precisão.

Uma outra vantagem que se realça é o facto desta técnica possibilitar a identificação direta de modificações epigenéticas. Como a reação de sequenciação rastreia a taxa de polimerização do DNA, pequenas mudanças que aconteçam na cinética da sua polimerização causadas por modificações epigenéticas, como metilações, resultam na mudança do timing de quando os sinais fluorescentes são observados (Figura 8.).

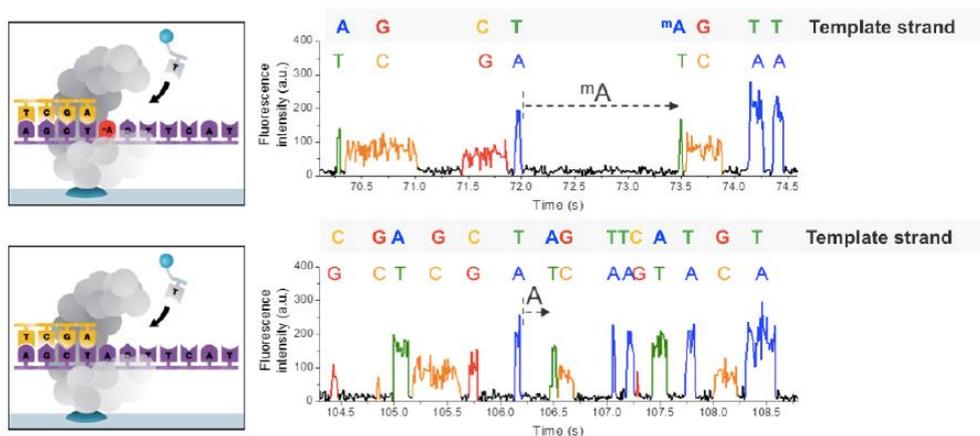


Figura 8. – Detecção das bases de DNA modificadas durante a sequenciação SMRT. A presença da base modificada (a vermelho) na cadeia de DNA (no topo) resulta numa incorporação tardia do nucleótido T quando comparada com o template de DNA controlo (em baixo) (Flusberg *et al.*, 2010).

Foi recentemente lançada uma versão melhorada da plataforma de sequenciação SMRT denominada PacBioRS2. Esta versão tem um maior número de células ZMW, duplicando assim o seu rendimento na sequenciação (Rhoads and Au, 2015).

- **Sequenciação Nanopore**

Em contraste com todos os outros métodos já referenciados, a sequenciação de DNA baseada na tecnologia de nanoporos não envolve SBS, sendo por isso designada como sequenciação de 4ª geração (Feng *et al.*, 2015). A sequenciação baseada em nanoporos,

utilizada pela plataforma ONT, depende de variações nas correntes elétricas que resultam da translocação das moléculas de DNA através de nanoporos que perfuram a membrana. Embora existam várias variações técnicas na sua abordagem, todas elas dependem de uma pequena voltagem (100mV) (Branton *et al.*, 2008) que é aplicada através da membrana que separa duas câmaras cheias de eletrólitos aquosos. Estes poros podem ser fabricados a partir de materiais à base de carbono, como a grafite, ou por proteínas biológicas como a α -hemossilina (Feng *et al.*, 2015). A utilização dos poros biológicos possibilita a fácil modificação e a produção em massa destes, com tamanhos bem definidos, enquanto que os nanoporos à base de carbono possibilitam a estabilização da membrana e o posicionamento das proteínas (Moreno, 2013).

Cada nanoporo possui localmente a cadeia de DNA enrolada que permite que os nucleótidos sejam transferidos sequencialmente através do poro (Figura 9. a)). Uma vez que a molécula de translocação bloqueia parcialmente a corrente de iões através do nanoporo, as alterações da corrente são utilizadas para deduzir a sequência dos nucleótidos da cadeia de ácido nucleico (Branton *et al.*, 2008).

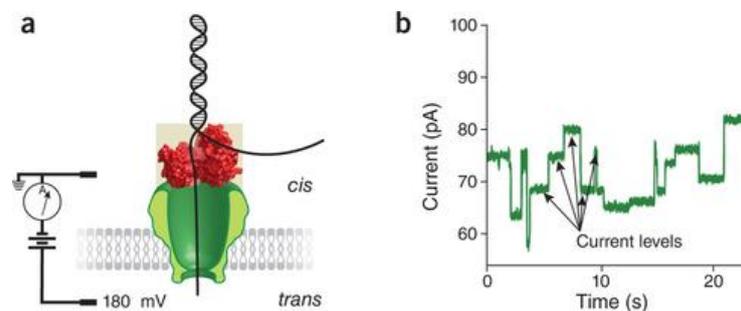


Figura 9. – Sequenciação por nanoporos. a) Translocação do polinucleótido (a preto) através do nanoporo controlado por uma enzima (a vermelho). b) Corrente iônica registada através do nanoporo (Deamer *et al.*, 2016).

Foi necessário a incorporação de enzimas e, posteriormente, algumas manipulações a nível enzimático para que a sequenciação fosse possível. No início, a identificação das

bases era difícil porque estas atravessavam a região de detecção do poro em menos de 10 μ s. A identificação dos nucleótidos usando as pequenas modificações de corrente necessitava de uma maior razão sinal/ruído, exigindo que cada base residisse na zona de detecção do nanoporo por, pelo menos, 100 μ s. A utilização da phi29 DNA polimerase por exemplo, permitiu o controlo da velocidade da deposição das bases no poro, tornando possível a leitura individualizada dos nucleótidos presentes na sequência (Figura 9. b)) (Deamer *et al.*, 2016).

A ONT desenvolveu recentemente as plataformas GridION e MinION (Figura 10.) que competem com outros métodos de sequenciação tanto a nível de precisão como de custo. A MinION é um dispositivo portátil e do tamanho de uma *pen-drive*, sendo esta a única plataforma descartável em miniatura disponível no mercado (Moreno, 2013).

O interesse pela sequenciação por nanoporos reside sobretudo pelo facto deste método fornecer uma sequência de moléculas de DNA individuais com comprimentos de leitura muito longos (podem atingir os 50.000 bp) (Branton *et al.*, 2008) e a portabilidade que a plataforma MinION oferece.



Figura 10. – Dispositivo MinION (Oxford Nanopore Technologies, 2013).

- **Transmission Electron Microscopy**

Uma vez que os nucleótidos numa cadeia de DNA não podem ser diretamente detetados por microscopia eletrónica, a sequenciação baseada na microscopia eletrónica requer a modificação individual das bases por referências que têm um maior número atómico. Numa abordagem que é análoga à utilização de fluoróforos para marcar o DNA no

método de Sanger, utilizam-se passos de amplificação para marcar as diferentes bases da cadeia de DNA que têm um contraste diferente na microscopia eletrónica. Tem uma capacidade potencial de produzir comprimentos de leitura que ultrapassem os 100.000 bp (Bell *et al.*, 2012), mas o passo de sequenciação não é inteiramente direto porque é necessário um espaçamento de base a base uniforme para obter leituras confiáveis. Este tipo de sequenciação é ainda muito teórico e não existem ainda plataformas comercialmente disponíveis.

- **Electronic Sequencing**

Este método inclui o uso de campos magnéticos e eletrónicos para imobilizar beads que transportam os ácidos nucleicos para serem sequenciados. Algumas etapas de sequenciação eletrónica foram demonstradas em sistemas modelos, mas esta tecnologia ainda está em desenvolvimento (Fuller *et al.*, 2016).

6.3.1. Aplicação clínica

As sequenciações de 3^a e 4^a geração têm tido um grande impacto nos estudos epigenéticos que, conseqüentemente, têm vindo a auxiliar o desenvolvimento de novos fármacos. Como os perfis epigenéticos são dependentes do tipo de célula e são criados por fatores genéticos e ambientais, a exploração destes mecanismos epigenéticos levou à descoberta de novos biomarcadores que são úteis no desenvolvimento de fármacos (Heerboth *et al.*, 2014). Já existem biomarcadores disponíveis baseados na epigenética, como por exemplo, a associação da metilação do septo 9 ao cancro colorretal (Y. Li *et al.*, 2014), e fármacos, como o Vorinostat (Bubna, 2015) que é um inibidor da histona desacetilase para o tratamento do linfoma cutâneo das células T, que foram, entretanto, desenvolvidos recorrendo ao auxílio das técnicas de 3^a geração (Ozsolak, 2012).

O recurso à sequenciação da 3^a geração também não só permitiu a observação do estado de metilação e deteção direta das modificações dos nucleótidos, como também possibilitou que outros componentes epigenéticos fossem estudados, nomeadamente a

organização da cromatina e modificações de histonas (Heerboth *et al.*, 2014). A integração dos estudos epigenéticos e os mecanismos transcriptómicos possibilitaram a descoberta de novos biomarcadores que fossem aplicados nas vacinas. As diferenças nas respostas induzidas pelas vacinas só podiam ser detetadas nos estudos de células únicas e não com as abordagens tradicionais que envolviam a medição de um sinal médio num grande número de células. As técnicas da 2ª geração eram aplicadas nas medições de células únicas, mas os passos necessários para a manipulação de células diminuíram significativamente a qualidade dos dados. Daí que as técnicas de 3ª geração providenciaram uma solução mais direta para os estudos que envolviam populações de células raras, tais como as células tumorais circulantes (Yu *et al.*, 2011) e a determinação dos mecanismos de expressão de genes nas células T CD8 no desenvolvimento de vacinas (Flatz *et al.*, 2011).

Para que o potencial máximo da sequenciação da 3ª e 4ª geração seja atingido para o desenvolvimento de fármacos, as empresas detentoras das plataformas necessárias para tal, necessitam de alcançar viabilidade económica (Ozsolak, 2012). Os sequenciadores de alto rendimento da 2ª geração de baixo custo são as preferíveis para ambientes de pesquisa e diagnóstico o que dificulta a introdução das plataformas de 3ª geração no mercado. Isto quer dizer que a conquista e viabilidade comercial destas empresas dependerá, em grande parte, da sua capacidade de diferenciação e oferta, assim como as suas vantagens económicas (Branton *et al.*, 2008).

6.3.2. Limitações técnicas

Embora as técnicas de 3ª geração apresentem um grande número de vantagens face às técnicas de 2ª geração, estas ainda continuam a ser otimizadas de modo que as suas limitações sejam superadas e o seu custo justifique a sua adesão nas várias áreas científicas (Heather and Chain, 2016).

A *nanopore sequencing*, apesar da sua elevada sensibilidade e baixo custo, a maior limitação deste método consiste na destruição da amostra no processo da sequenciação que pode ser considerado como um obstáculo para as aplicações com amostras preciosas,

não permitindo um segundo ciclo de leitura para redução de erros (Kircher and Kelso, 2010).

Em relação à plataforma Heliscope que utilizava a abordagem tSMS, a sua desvantagem mais significativa residia no comprimento de leitura, que se situava entre os 25 e 150 bp (King, 2015). O tempo da reação da sequenciação, que demorava sensivelmente cerca de 8 dias (Donaldson *et al.*, 2015) a ser completada, também era uma grande desvantagem principalmente para os grandes centros clínicos que pretendessem utilizar esta tecnologia. O sistema PacBio SMRT parece incrível, mas não o é devido a algumas limitações. Um dos principais problemas que apresenta é a sua célula de fluxo não ter um rendimento tão alto quando comparado com a plataforma Illumina (Goodwin *et al.*, 2016). Nem todas as ZMWs realizarão reações de sequenciação bem-sucedidas. Alguns dos poços simplesmente não apresentam o template DNA o que influencia o rendimento da plataforma (Ghorbanpour *et al.*, 2017). A PacBio tem leituras longas, mas com menos rendimento, do que a Illumina, que tem leituras mais curtas mas com um melhor rendimento já que utiliza mais do que uma célula de fluxo (Roberts *et al.*, 2013). Um outro problema que apresenta é a sua taxa de erro ser mais alta do que a da Illumina. São precisos muitos passos na metodologia para adquirir uma precisão de 99%, o que depois se reflete no custo da plataforma (Ghorbanpour *et al.*, 2017).

7. Conclusão

O desenvolvimento de novas técnicas de sequenciação aliadas à inovação das PCRs, contribuiu significativamente para o conhecimento do genoma humano. Todo este esforço científico abriu portas para a terapia personalizada ser empregue na prática clínica. A premissa de um medicamento para todos está cada vez mais recuada sempre que se desvenda mais um mistério genético através da tecnologia acessível e simples que nos é apresentada hoje em dia. A sequenciação de 3ª e 4ª geração mostram-se cada vez mais prometedoras na área da investigação clínica ao revelar projetos ambiciosos que visam a compreensão da informação codificada nos genomas inteiros, assim como a completa caracterização dos transcriptomas e uma melhor compreensão da relação entre as proteínas e DNA. Tal esclarecimento nestas áreas auxiliam a indústria farmacêutica no desenvolvimento de novos fármacos que podem ser aplicados na terapia individual dos doentes, de forma a solucionar os presentes desafios relacionados com a metabolização de fármacos e os efeitos adversos intrínsecos a estes. Porém, este sonho da terapia individualizada está ainda muito longe de ser concretizado visto que emergem questões de carácter económico e, sobretudo, de carácter ético.

O impacto financeiro dos fármacos sempre foi um aspeto relevante na indústria farmacêutica. Atualmente, cerca de 30 grandes empresas farmacêuticas estão a investir na farmacogenómica, como por exemplo a Roche e a Pfizer, e o principal interesse das empresas na farmacogenómica é melhorar a sua eficácia na descoberta e desenvolvimento de novos medicamentos. Para alguns, o conceito da farmacogenómica no mercado adaptado a uma subpopulação com um genótipo específico tem sido visto muito desfavoravelmente porque pode excluir doentes com genótipos particulares e levar ao aumento de reações adversas a medicamentos, já que o fármaco é apenas comercializado para um grupo restrito de indivíduos. Mas é provável que as empresas estejam mais interessadas em identificar possíveis reações adversas, ao ter em conta as enormes implicações financeiras da retirada do produto do mercado. As empresas farmacêuticas parecem estar menos interessadas nos fármacos com licença, exceto quando podem lucrar com uma eventual extensão da licença, como já aconteceu com o fármaco abacavir usado

no tratamento do HIV. Foi reportado que alguns doentes apresentam sérias reações de hipersensibilidade ao abacavir, pelo que é necessária uma interrupção no tratamento com esse fármaco. No entanto, a recente disponibilidade dos testes genéticos para a hipersensibilidade pode reduzir substancialmente esse risco, o que equivale à continuação da comercialização desse medicamento. Dado que os custos para o desenvolvimento de um novo fármaco são exorbitantes (em média rondam os mil milhões), os custos dos testes genéticos são bastante acessíveis e, por isso, mais atraentes para as empresas.

Embora exista o risco de segmentação no mercado, a maioria das empresas estão interessadas em desenvolver medicamentos eficazes e seguros para todos os subgrupos.

A farmacogenómica pode ajudar a recuperar medicamentos que são ineficazes na população em geral, mas que são eficazes num subgrupo. Porém, é improvável que os medicamentos retirados do mercado por questões de segurança sejam resgatados. A maioria desses fármacos estará fora da patente e, por consequência, deixam de ser financeiramente atrativos. Os dados genéticos necessários para o desenvolvimento de testes adequados possivelmente deixarão de estar disponíveis. Contudo, mesmo que a indústria não tenha interesse em desenvolver fármacos para um determinado subgrupo, o seu investimento é salvaguardado uma vez que podem vender os seus direitos a empresas mais pequenas e inovadoras que estão dispostas a apostar na continuação do desenvolvimento do medicamento para esse subgrupo.

Em termos de diagnóstico, como já foi previamente discutido, embora existam biomarcadores genéticos que os clínicos podem usar para prever a eficácia e/ou toxicidade, o verdadeiro desafio estará na resolução das mais variadas questões que surgirão, a saber: se os biomarcadores devem ser utilizados na avaliação do doente e saber quando utilizar os testes de diagnóstico. Muitos cientistas defendem que a utilidade clínica dos biomarcadores é apropriado nos doentes e que estes testes são um investimento na saúde pública, prometendo melhorar o resultado clínico e a partilha do conhecimento científico nos vários centros. Outros, defendem que é difícil prever até que ponto a informação nas várias variantes genéticas é relevante no tratamento clínico. As barreiras comerciais, económicas e sociais, que são alimentadas pelos acionistas com diferentes agendas, poderão impedir a sua integração na indústria farmacêutica.

Além disso, algumas preocupações foram levantadas pela comunidade científica sobre o direito individual à privacidade, bem como a potencial discriminação e ilegitimidade para o emprego e seguros. O ato de não-discriminação por informação genética (GINA), lançado em 2008, proíbe especificamente o uso indevido de informações genéticas nas decisões de emprego, coberturas de seguro, bem como na decisão dos países sobre quem pode residir no seu país. Uma outra intenção do GINA é encorajar os indivíduos a participarem na pesquisa genética, mas as questões de partilha de informações e confidencialidade não satisfizeram as partes interessadas, nomeadamente os doentes. Discute-se quais as instâncias não governamentais que devem ter propriedade sobre os materiais genéticos e o direito de aceder e de saber qual delas deve ter direito a essas informações. Estas questões são primordiais para assegurar a confidencialidade e identidade da pessoa humana. Esta abordagem ainda não foi satisfatoriamente concluída. Ora, estas questões poderão ser suficientes para barrar o potencial do desenvolvimento da farmacogenómica.

É importante, no entanto, ter em conta que a análise clínica de um certo indivíduo se reflete no momento enquanto que a análise genómica providencia dados definitivos. A análise genómica é baseada no conceito de probabilidades e, embora haja genes que possam ter um impacto patológico no indivíduo, muitas das vezes essas patologias nem sequer se chegam a manifestar e, neste aspeto, a farmacogenómica estaria a influenciar o livre arbítrio individual, assim como os direitos básicos de cada pessoa.

A área da medicina preventiva também pode beneficiar da farmacogenómica uma vez que ao executar testes genéticos nas várias populações pode reduzir o custo na saúde. No entanto, o que distingue a farmacogenómica de outros instrumentos que são utilizados na medicina preventiva é o facto dos clínicos não saberem o que fazer quanto aos achados genéticos acidentais. O que fazer com esta informação? Trata-se de informação que é propriedade do indivíduo e a comunidade e a legislação ainda não estão preparadas para lidar com tal panorama.

Uma outra temática que deve ser abordada refere-se ao modo de procedimento acerca do agrupamento dos indivíduos para tratamento, tendo como base a sua informação genética. Um grupo poderá conter indivíduos que respondam bem ao medicamento e, outro grupo

poderá conter indivíduos que possuem genes que não lhes permitam adquirir uma reação positiva ao fármaco. Este cenário levanta várias questões que são pertinentes na utilização da farmacogenómica na indústria farmacêutica. Em primeiro lugar, trata-se de determinar quais os grupos que devem ser alvo desta indústria e saber o que acontece aos indivíduos que não pertencem ao grupo dos seus interesses. Assim, torna-se necessário questionar quais os grupos a ser tratados. Exemplificando: imagine-se que existe um grupo pequeno de pessoas que melhor reage ao medicamento e que esse grupo pequeno corresponde a indivíduos mais ricos e com interesse nesta indústria.

Temos legitimidade para pensar, segundo este exemplo, que a farmacogenómica pode revolucionar o sistema de saúde, mas ao mesmo tempo também pode excluir indivíduos de potenciais tratamentos se for gerida com outros interesses em conta.

Para finalizar, a farmacogenómica ainda não é universalmente empregue na prática clínica porque não existem muitas seguradoras ou outras instâncias, que estejam interessadas em pagar estes custos, uma vez que o argumento apresentado se baseia na probabilidade das variantes detetadas terem relevância clínica. Na prática, continua-se a utilizar as técnicas da PCR, que são mais económicas e auxiliam bastante no diagnóstico clínico. A medicina individualizada encontra-se ainda muito distante de ser concretizada, mas quando a farmacogenómica se tornar uma ciência indispensável na prática clínica, esta será definitivamente um fator decisivo na prática médica com o potencial de salvar muitas vidas.

8. Bibliografia

- Aboraya, A., Rankin, E., France, C., El-Missiry, A. and John, C. (2006). The Reliability of Psychiatric Diagnosis Revisited: The Clinician's Guide to Improve the Reliability of Psychiatric Diagnosis, *Psychiatry (Edgmont)*, 3(1), pp. 41-50.
- Ari, S. and Arikan, M. (2016). Next-Generation Sequencing: Advantages, Disadvantages, and Future, *In Plant Omics: Trends and Applications*, pp. 109-135.
- Arsenic, R., Treue, D., Lehmann, A., Hummel, M., Dietel, M., Denkert, C. and Budczies, J. (2015). Comparison of targeted next-generation sequencing and Sanger sequencing for the detection of PIK3CA mutations in breast cancer, *BMC Clinical Pathology*, 15, p. 20.
- Bartlett, J. M. S., Shaaban, A., Schmitt, F. and Press, C. U. (2014). *Molecular Pathology: A Practical Guide for the Surgical Pathologist and Cytopathologist*. Cambridge University Press.
- Bell, D. C., Thomas, W. K., Murtagh, K. M., Dionne, C. A., Graham, A. C., Anderson, J. E. and Glover, W. R. (2012). DNA base identification by electron microscopy, *Microsc Microanal*, 18(5), pp. 1049-1053.
- Bertilsson, L., Dahl, M.-L., Dalén, P. and Al-Shurbaji, A. (2002). Molecular genetics of CYP2D6: Clinical relevance with focus on psychotropic drugs, *British Journal of Clinical Pharmacology*, 53(2), pp. 111-122.
- Biernacka, J. M., Sangkuhl, K., Jenkins, G., Whaley, R. M., Barman, P., Batzler, A., Altman, R. B., Arolt, V., Brockmüller, J., Chen, C. H., Domschke, K., Hall-Flavin, D. K., Hong, C. J., Illi, A., Ji, Y., Kampman, O., Kinoshita, T., Leinonen, E., Liou, Y. J., Mushiroda, T., Nonen, S., Skime, M. K., Wang, L., Baune, B. T., Kato, M., Liu, Y. L., Praphanphoj, V., Stingl, J. C., Tsai, S. J., Kubo, M., Klein, T. E. and Weinshilboum, R. (2015). The International SSRI Pharmacogenomics Consortium (ISPC): a genome-wide association study of antidepressant treatment response, *Translational Psychiatry*, 5(4), p. e553.
- BigDye Terminator v3.1 Cycle Sequencing Kit* (2002). USA: Biosystems, A.

-
- Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Ling, X. S., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. V., Ramsey, J. M., Riehn, R., Soni, G. V., Tabard-Cossa, V., Wanunu, M., Wiggin, M. and Schloss, J. A. (2008). The potential and challenges of nanopore sequencing, *Nat Biotech*, 26(10), pp. 1146-1153.
- Bubna, A. K. (2015). Vorinostat—An Overview, *Indian Journal of Dermatology*, 60(4), pp. 419-419.
- Buermans, H. P. J. and den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications, *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842(10), pp. 1932-1941.
- Burt, T. and Dhillon, S. (2013). Pharmacogenomics in early-phase clinical development, *Pharmacogenomics*, 14(9), pp. 1085-1097.
- Cao, H., Wu, H., Luo, R., Huang, S., Sun, Y., Tong, X., Xie, Y., Liu, B., Yang, H., Zheng, H., Li, J., Li, B., Wang, Y., Yang, F., Sun, P., Liu, S., Gao, P., Huang, H., Sun, J., Chen, D., He, G., Huang, W., Huang, Z., Li, Y., Tellier, L. C., Liu, X., Feng, Q., Xu, X., Zhang, X., Bolund, L. and Krogh, A. (2015). De novo assembly of a haplotype-resolved human genome, 33(6), pp. 617-622.
- Capobianchi, M. R., Giombini, E. and Rozera, G. Next-generation sequencing technology in clinical virology, *Clinical Microbiology and Infection*, 19(1), pp. 15-22.
- Cederbaum, A. I. (2015). Molecular mechanisms of the microsomal mixed function oxidases and biological and pathological implications, *Redox Biology*, 4, pp. 60-73.
- Chen, Q. and Wei, D. (2015). Human cytochrome P450 and personalized medicine, *Adv Exp Med Biol*, 827, pp. 341-351.
- Chittani, M., Zaninello, R., Lanzani, C., Frau, F., Ortu, M. F., Salvi, E., Fresu, G., Citterio, L., Braga, D., Piras, D. A., Carpini, S. D., Velayutham, D., Simonini, M., Argiolas, G., Pozzoli, S., Troffa, C., Glorioso, V., Kontula, K. K., Hiltunen, T. P., Donner, K. M., Turner, S. T., Boerwinkle, E., Chapman, A. B., Padmanabhan, S., Dominiczak, A. F., Melander, O., Johnson, J. A., Cooper-Dehoff, R. M., Gong,

-
- Y., Rivera, N. V., Condorelli, G., Trimarco, B., Manunta, P., Cusi, D., Glorioso, N. and Barlassina, C. (2015). TET2 and CSMD1 genes affect SBP response to hydrochlorothiazide in never-treated essential hypertensives, *Journal of hypertension*, 33(6), pp. 1301-1309.
- Christensen, K. D., Dukhovny, D., Siebert, U. and Green, R. C. (2015). Assessing the Costs and Cost-Effectiveness of Genomic Sequencing, *Journal of Personalized Medicine*, 5(4), pp. 470-486.
- Coleman, M. D. (2010). Factors Affecting Drug Metabolism, in *Human Drug Metabolism*. John Wiley & Sons, Ltd, pp. 159-212.
- Cooper-DeHoff, R. M. and Johnson, J. A. (2016). Hypertension pharmacogenomics: in search of personalized treatment approaches, *Nature reviews. Nephrology*, 12(2), pp. 110-122.
- David W. Thomas, J. B., John Audette, Adam Carroll, Corey Dow-Hygelund, Michael Hay (2016). Clinical Development Success Rates 2006-2015 - BIO, Biomedtracker, Amplion 2016, *BIO Industry Analysis*, p. 26.
- Davidson, C. J., Zeringer, E., Champion, K. J., Gauthier, M.-P., Wang, F., Boonyaratanakornkit, J., Jones, J. R. and Schreiber, E. (2012). Improving the limit of detection for Sanger sequencing: A comparison of methodologies for KRAS variant detection, *Biotechniques*, 53(3), pp. 182-188.
- Deamer, D., Akeson, M. and Branton, D. (2016). Three decades of nanopore sequencing, *Nat Biotech*, 34(5), pp. 518-524.
- Dean, L. (2012). Warfarin Therapy and Genotypes CYP2C9 and VKORC1, *Medical Genetics Summaries*
- Deepak, S. A., Kottapalli, K. R., Rakwal, R., Oros, G., Rangappa, K. S., Iwahashi, H., Masuo, Y. and Agrawal, G. K. (2007). Real-Time PCR: Revolutionizing Detection and Expression Analysis of Genes, *Current Genomics*, 8(4), pp. 234-251.
- Deng, J., Carbone, I. and Dean, R. A. (2007). The evolutionary history of Cytochrome P450 genes in four filamentous Ascomycetes, *BMC Evolutionary Biology*, 7, pp. 30-30.

-
- Di Niro, R., Sulic, A. M., Mignone, F., D'Angelo, S., Bordoni, R., Iacono, M., Marzari, R., Gaiotto, T., Lavric, M., Bradbury, A. R., Biancone, L., Zevin-Sonkin, D., De Bellis, G., Santoro, C. and Sblattero, D. (2010). Rapid interactome profiling by massive sequencing, *Nucleic Acids Res*, 38(9), p. e110.
- Dias-Neto, E., Nunes, D. N., Giordano, R. J., Sun, J., Botz, G. H., Yang, K., Setubal, J. C., Pasqualini, R. and Arap, W. (2009). Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis, *PLoS One*, 4(12), p. e8338.
- Dieffenbach, C. W. and Dveksler, G. S. (2003). *PCR Primer: A Laboratory Manual, Second Edition (2003) Paperback*. Cold Spring Harbor Laboratory Press.
- Donaldson, P., Daly, A., Ermini, L. and Bevitt, D. (2015). *Genetics of Complex Disease*.
- Dong, C. and Yu, B. (2011). Mutation Surveyor: An In Silico Tool for Sequencing Analysis, in Yu, B. and Hinchcliffe, M. (eds.) *In Silico Tools for Gene Discovery*. Totowa, NJ: Humana Press, pp. 223-237.
- Elnifro, E. M., Ashshi, A. M., Cooper, R. J. and Klapper, P. E. (2000). Multiplex PCR: optimization and application in diagnostic virology, *Clin Microbiol Rev*, 13(4), pp. 559-570.
- Endrullat, C., Glökler, J., Franke, P. and Frohme, M. (2016). Standardization and quality management in next-generation sequencing, *Applied & Translational Genomics*, 10, pp. 2-9.
- Feng, Y., Zhang, Y., Ying, C., Wang, D. and Du, C. (2015). Nanopore-based Fourth-generation DNA Sequencing Technology, *Genomics, Proteomics & Bioinformatics*, 13(1), pp. 4-16.
- Ferrarini, M., Moretto, M., Ward, J. A., Surbanovski, N., Stevanovic, V., Giongo, L., Viola, R., Cavalieri, D., Velasco, R., Cestaro, A. and Sargent, D. J. (2013). An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome, *BMC Genomics*, 14, p. 670.
- Fisher, R. G., Smith, D. M., Murrell, B., Slabbert, R., Kirby, B. M., Edson, C., Cotton, M. F., Haubrich, R. H., Kosakovsky Pond, S. L. and Van Zyl, G. U. (2015). Next generation sequencing improves detection of drug resistance mutations in infants

-
- after PMTCT failure, *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, 62, pp. 48-53.
- Flatz, L., Roychoudhuri, R., Honda, M., Filali-Mouhim, A., Goulet, J.-P., Kettaf, N., Lin, M., Roederer, M., Haddad, E. K., Sékaly, R. P. and Nabel, G. J. (2011). Single-cell gene-expression profiling reveals qualitatively distinct CD8 T cells elicited by different gene-based vaccines, *Proceedings of the National Academy of Sciences of the United States of America*, 108(14), pp. 5724-5729.
- Flockhart, D. A., Orkane, D., Williams, M. S., Watson, M. S., Flockhart, D. A., Gage, B., Gandolfi, R., King, R., Lyon, E., Nussbaum, R., Orkane, D., Schulman, K., Veenstra, D., Williams, M. S. and Watson, M. S. (2008). Pharmacogenetic testing of CYP2C9 and VKORC1 alleles for warfarin, *Genet Med*, 10(2), pp. 139-150.
- Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J. and Turner, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing, *Nature Methods*, 7, pp. 461-465.
- Foquet, M., Samiee, K. T., Kong, X., Chaudhuri, B. P., Lundquist, P. M., Turner, S. W., Freudenthal, J. and Roitman, D. B. (2008). Improved fabrication of zero-mode waveguides for single-molecule detection, *Journal of Applied Physics*, 103(3), p. 034301.
- França, L. T. C., Carrilho, E. and Kist, T. B. L. (2002). A review of DNA sequencing techniques, *Quarterly Reviews of Biophysics*, pp. 169-200.
- Frau, F., Citterio, L., Ortu, M. F. and Zaninello, R. (2013). Genomic Association Analysis Identifies Multiple Loci Influencing Antihypertensive Response to an Angiotensin II Receptor Blocker, *Hypertension*, 61(1), pp. e5-e5.
- Fuller, C. W., Kumar, S., Porel, M., Chien, M., Bibillo, A., Stranges, P. B., Dorwart, M., Tao, C., Li, Z., Guo, W., Shi, S., Korenblum, D., Trans, A., Aguirre, A., Liu, E., Harada, E. T., Pollard, J., Bhat, A., Cech, C., Yang, A., Arnold, C., Palla, M., Hovis, J., Chen, R., Morozova, I., Kalachikov, S., Russo, J. J., Kasianowicz, J. J., Davis, R., Roeber, S., Church, G. M. and Ju, J. (2016). Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array, *Proc Natl Acad Sci U S A*, 113(19), pp. 5233-5238.

-
- Furie, B. (2013). Do pharmacogenetics have a role in the dosing of vitamin K antagonists?, *N Engl J Med*, 369(24), pp. 2345-2346.
- Furman, D. and Davis, M. M. (2015). New approaches to understanding the immune response to vaccination and infection, *Vaccine*, 33(40), pp. 5271-5281.
- Gagan, J. and Van Allen, E. M. (2015). Next-generation sequencing to guide cancer therapy, *Genome Medicine*, 7(1), p. 80.
- Garibyan, L. and Avashia, N. (2013). Research Techniques Made Simple: Polymerase Chain Reaction (PCR), *The Journal of investigative dermatology*, 133(3), pp. e6-e6.
- Ge, D., Fellay, J., Thompson, A. J., Simon, J. S., Shianna, K. V., Urban, T. J., Heinzen, E. L., Qiu, P., Bertelsen, A. H., Muir, A. J., Sulkowski, M., McHutchison, J. G. and Goldstein, D. B. (2009). Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance, *Nature*, 461(7262), pp. 399-401.
- Ghorbanpour, M., Manika, K. and Varma, A. (2017). *Nanoscience and Plant–Soil Systems*. Springer International Publishing.
- Gilles, A., Megléc, E., Pech, N., Ferreira, S., Malausa, T. and Martin, J.-F. (2011). Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing, *BMC Genomics*, 12, pp. 245-245.
- Gong, Y., McDonough, C. W., Beitelshes, A. L., Roubey, N. E., Hiltunen, T. P., O’Connell, J. R., Padmanabhan, S., Langaee, T. Y., Hall, K., Schmidt, S. O. F., Curry, R. W., Gums, J. G., Donner, K. M., Kontula, K. K., Bailey, K. R., Boerwinkle, E., Takahashi, A., Tanaka, T., Kubo, M., Chapman, A. B., Turner, S. T., Pepine, C. J., Cooper-DeHoff, R. M. and Johnson, J. A. (2015). PTPRD gene associated with blood pressure response to atenolol and resistant hypertension, *Journal of hypertension*, 33(11), pp. 2278-2285.
- Gong, Y., Wang, Z., Beitelshes, A. L., McDonough, C. W., Langaee, T. Y., Hall, K., Schmidt, S. O. F., Curry, R. W., Gums, J. G., Bailey, K. R., Boerwinkle, E., Chapman, A. B., Turner, S. T., Cooper-DeHoff, R. M. and Johnson, J. A. (2016). Pharmacogenomic Genome-Wide Analysis of Blood Pressure Response to Beta-blocker in Hypertensive African Americans, *Hypertension*, 67(3), pp. 556-563.

-
- Goodsaid, F. (2004). Quantitative real time polymerase chain reaction in drug development, *Drug Development Research*, 62(2), pp. 151-158.
- Goodwin, S., McPherson, J. D. and McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies, *Nat Rev Genet*, 17(6), pp. 333-351.
- Gopisankar, M. G. (2017). CYP2D6 pharmacogenomics, *Egyptian Journal of Medical Human Genetics*.
- Grahame-Smith, D. G. (1999). How will knowledge of the human genome affect drug therapy?, *British Journal of Clinical Pharmacology*, 47(1), pp. 7-10.
- Gullapalli, R. R., Lyons-Weiler, M., Petrosko, P., Dhir, R., Becich, M. J. and LaFramboise, W. A. (2012). Clinical Integration of Next Generation Sequencing Technology, *Clinics in laboratory medicine*, 32(4), pp. 585-599.
- Guo, Y. and Jamison, D. C. (2005). The distribution of SNPs in human gene regulatory regions, *BMC Genomics*, 6, pp. 140-140.
- Gupta, S. and Jhawar, V. (2016). Quality by design (QbD) approach of pharmacogenomics in drug designing and formulation development for optimization of drug delivery systems, *J Control Release*, 245, pp. 15-26.
- Haff, L. A. (1994). Improved quantitative PCR using nested primers, *PCR Methods Appl*, 3(6), pp. 332-337.
- Hagemann, I. S. (2015). Chapter 1 - Overview of Technical Aspects and Chemistries of Next-Generation Sequencing A2 - Kulkarni, Shashikant, in Pfeifer, J. (ed.) *Clinical Genomics*. Boston: Academic Press, pp. 3-19.
- Heather, J. M. and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA, *Genomics*, 107(1), pp. 1-8.
- Heerboth, S., Lapinska, K., Snyder, N., Leary, M., Rollinson, S. and Sarkar, S. (2014). Use of Epigenetic Drugs in Disease: An Overview, *Genetics & Epigenetics*, 6, pp. 9-19.
- Hodkinson, B. P. and Grice, E. A. (2015). Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research, *Advances in Wound Care*, 4(1), pp. 50-58.

-
- Hoofnagle, J. H., Wahed, A. S., Brown, J. R. S., Howell, C. D. and Bellw, S. H. (2009). Early Changes in Hepatitis C Virus (HCV) Levels in Response to Peginterferon and Ribavirin Treatment in Patients with Chronic HCV Genotype 1 Infection, *The Journal of Infectious Diseases*, 199(8), pp. 1112-1120.
- Hunt, R., Sauna, Z. E., Ambudkar, S. V., Gottesman, M. M. and Kimchi-Sarfaty, C. (2009). Silent (synonymous) SNPs: should we care about them?, *Methods Mol Biol*, 578, pp. 23-39.
- Ion AmpliSeq™ Library Preparation for Human Identification Applications (2015).
- Ion PGM System Specifications (2013). Life Technologies.
- Janitz, M. (2011). *Next-Generation Genome Sequencing: Towards Personalized Medicine*. Wiley.
- Jensen, D. M. and Pol, S. (2012). IL28B genetic polymorphism testing in the era of direct acting antivirals therapy for chronic hepatitis C: ten years too late?, *Liver International*, 32, pp. 74-78.
- Ji, Y., Schaid, D. J., Desta, Z., Kubo, M., Batzler, A. J., Snyder, K., Mushiroda, T., Kamatani, N., Ogburn, E., Hall-Flavin, D., Flockhart, D., Nakamura, Y., Mrazek, D. A. and Weinshilboum, R. M. (2014). Citalopram and escitalopram plasma drug and metabolite concentrations: genome-wide associations, *British Journal of Clinical Pharmacology*, 78(2), pp. 373-383.
- Jiang, Y.-Y., Kong, D.-X., Qin, T., Li, X., Caetano-Anollés, G. and Zhang, H.-Y. (2012). The Impact of Oxygen on Metabolic Evolution: A Chemoinformatic Investigation, *PLOS Computational Biology*, 8(3), p. e1002426.
- Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A. L., Madsen, K. L. and Wong, G. K. S. (2016). Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics, *Frontiers in Microbiology*, 7, p. 459.
- Kalow, W. (2006). Pharmacogenetics and pharmacogenomics: origin, status, and the hope for personalized medicine, *Pharmacogenomics J*, 6(3), pp. 162-165.
- Karger, B. L. and Guttman, A. (2009). DNA Sequencing by Capillary Electrophoresis, *Electrophoresis*, 30(Suppl 1), pp. S196-S202.

-
- Karki, R., Pandya, D., Elston, R. C. and Ferlini, C. (2015). Defining “mutation” and “polymorphism” in the era of personal genomics, *BMC Medical Genomics*, 8, p. 37.
- Kimchi-Sarfaty, C., Oh, J. M., Kim, I. W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V. and Gottesman, M. M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity, *Science*, 315(5811), pp. 525-528.
- King, G. (2015). *Venoms to Drugs: Venom as a Source for the Development of Human Therapeutics*. Royal Society of Chemistry.
- Kircher, M. and Kelso, J. (2010). High-throughput DNA sequencing concepts and limitations, *Methods, Models & Techniques*. doi: 10.1002/bies.200900181.
- Komar, A. A. (2009). *Single Nucleotide Polymorphisms - Methods and Protocols*. Edited by Komar, A. A. Cleveland: Humana Press.
- Korlach, J., Marks, P. J., Cicero, R. L., Gray, J. J., Murphy, D. L., Roitman, D. B., Pham, T. T., Otto, G. A., Foquet, M. and Turner, S. W. (2008). Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures, *Proc Natl Acad Sci U S A*, 105(4), pp. 1176-1181.
- Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. and Schloss, P. D. (2013). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform, *Applied and Environmental Microbiology*, 79(17), pp. 5112-5120.
- Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., Sindelka, R., Sjoberg, R., Sjogreen, B., Strombom, L., Stahlberg, A. and Zoric, N. (2006). The real-time polymerase chain reaction, *Mol Aspects Med*, 27(2-3), pp. 95-125.
- Kuruvilla, M. and Gurk-Turner, C. (2001). A review of warfarin dosing and monitoring, *Proceedings (Baylor University. Medical Center)*, 14(3), pp. 305-306.
- Kuypers, J. and Jerome, K. R. (2017). Applications of Digital PCR for Clinical Microbiology, *J Clin Microbiol*.
- Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988). A ligase-mediated gene detection technique, *Science*, 241(4869), pp. 1077-1080.

-
- Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Ferrante, T. C., Terry, R., Turczyk, B. M., Yang, J. L., Lee, H. S., Aach, J., Zhang, K. and Church, G. M. (2015). Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues, *Nat. Protocols*, 10(3), pp. 442-458.
- Lewis, D. F. V. (2005). *Guide to Cytochromes P450 structure and Function*. 2 edn. London: Taylor & Francis Inc.
- Li, J., Wang, S., Barone, J. and Malone, B. (2009). Warfarin Pharmacogenomics, *Pharmacy and Therapeutics*, 34(8), pp. 422-427.
- Li, Y., Song, L., Gong, Y. and He, B. (2014). Detection of colorectal cancer by DNA methylation biomarker SEPT9: past, present and future, *Biomark Med*, 8(5), pp. 755-769.
- Lin, J. H. and Lu, A. Y. H. (1998). Inhibition and Induction of Cytochrome P450 and the Clinical Implications, *Clinical Pharmacokinetics*, 35(5), pp. 361-390.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M. (2012). Comparison of Next-Generation Sequencing Systems, *Journal of Biomedicine and Biotechnology*, 2012, p. 11.
- Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J. and Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms, *Nat Biotechnol*, 30(5), pp. 434-439.
- Lynch, T. and Price, A. (2007). The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects., *Am Fam Physician*, 76(3), pp. 391-396.
- Ma, Q. and Lu, A. Y. H. (2011). Pharmacogenetics, Pharmacogenomics, and Individualized Medicine, *Pharmacological Reviews*, 63(2), pp. 437-459.
- Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics, *Trends Genet*, 24(3), pp. 133-141.
- Massingham, T. and Goldman, N. (2012). Error-correcting properties of the SOLiD Exact Call Chemistry, *BMC Bioinformatics*, 13, pp. 145-145.
- Mc Pherson, J. and Moller, S. (2006). *PCR second edition*. Taylor & Francis Group.

-
- McGraw, J. and Waller, D. (2012). Cytochrome P450 variations in different ethnic populations, *Expert Opin Drug Metab Toxicol*, 8(3), pp. 371-382.
- Meletis, J. and Konstantopoulos, K. (2004). Favism - from the "avoid fava beans" of Pythagoras to the present, 7(1), pp. 17-21.
- Men, A. E., Wilson, P., Siemering, K. and Forrest, S. (2008). *Sanger DNA Sequencing*. WILEY-VCH Verlag GmbH & Co.
- Menni, C. (2015). Blood pressure pharmacogenomics: gazing into a misty crystal ball, *Journal of Hypertension*, 33(6), pp. 1142-1143.
- Merriman, B. and Rothberg, J. M. (2012). Progress in ion torrent semiconductor chip based sequencing, *Electrophoresis*, 33(23), pp. 3397-3417.
- Messmer, T., Tully, T. N., Jr., Ritchie, B. W. and Moroney, J. F. A tale of discrimination: Differentiation of chlamydiaceae by polymerase chain reaction, *Seminars in Avian and Exotic Pet Medicine*, 9(1), pp. 36-42.
- Minton, J. A., Flanagan, S. E. and Ellard, S. (2011). Mutation surveyor: software for DNA sequence analysis, *Methods Mol Biol*, 688, pp. 143-153.
- Moorthie, S., Mattocks, C. J. and Wright, C. F. (2011). Review of massively parallel DNA sequencing technologies, *The HUGO Journal*, 5(1-4), pp. 1-12.
- Moreno, A. C. A. (2013). *Diagnóstico Molecular na era da sequenciação de 3ª geração e da PCR digital*. Tese de Mestrado em Ciências Farmacêuticas - Universidade Fernando Pessoa, Porto, pp. 1-4.
- Mu, W., Lu, H. M., Chen, J., Li, S. and Elliott, A. M. (2016). Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing, *J Mol Diagn*, 18(6), pp. 923-932.
- Mullis, K. B. (1990). The unusual origin of the polymerase chain reaction, *Sci Am*, 262(4), pp. 56-61, 64-55.
- Munroe, P. B., Barnes, M. R. and Caulfield, M. J. (2013). Advances in Blood Pressure Genomics, *Circulation Research*, 112(10), pp. 1365-1379.
- Nebert, D. W., Wikvall, K. and Miller, W. L. (2013). Human cytochromes P450 in health and disease, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1612), p. 20120431.

-
- Nebert, D. W., Zhang, G. and Vesell, E. S. (2008). From Human Genetics and Genomics to Pharmacogenetics and Pharmacogenomics: Past Lessons, Future Directions, *Drug metabolism reviews*, 40(2), pp. 187-224.
- Nelson, D. R., Goldstone, J. V. and Stegeman, J. J. (2013). The cytochrome P450 genes locus: the origin and evolution of animal cytochrome P450s, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1612), p. 20120474.
- Nielsen, R., Paul, J. S., Albrechtsen, A. and Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data, *Nature reviews. Genetics*, 12(6), pp. 443-451.
- Oxford Nanopore Technologies [Em Linha]. Disponível em <www.nanoporetech.com>. [Consultado em 26/9/2017]
- Ozsolak, F. (2012). Third Generation Sequencing Techniques and Applications to Drug Discovery, *Expert Opinion on Drug Discovery*, 7(3), pp. 231-243.
- Pandey, R. V., Pabinger, S., Kriegner, A. and Weinhäusel, A. (2016). MutAid: Sanger and NGS Based Integrated Pipeline for Mutation Identification, Validation and Annotation in Human Molecular Genetics, *PLoS ONE*, 11(2), p. e0147697.
- Pareek, C. S., Smoczynski, R. and Tretyn, A. (2011). Sequencing technologies and genome sequencing, *Journal of Applied Genetics*, 52(4), pp. 413-435.
- Park, S. T. and Kim, J. (2016). Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing, *International Neuropsychology Journal*, 20(Suppl 2), pp. S76-83.
- Pelkonen, O. and Raunio, H. (1997). Metabolic activation of toxins: tissue-specific expression and metabolism in target organs, *Environmental Health Perspectives*, 105(Suppl 4), pp. 767-774.
- Pile, D. (2009). Biophotonics: Eavesdropping on DNA replication, *Nat Photon*, 3(2), pp. 79-80.
- Pinto, N. and Dolan, M. E. (2011). Clinically Relevant Genetic Variations in Drug Metabolizing Enzymes, *Current drug metabolism*, 12(5), pp. 487-497.

-
- Pirmohamed, M. (2001). Pharmacogenetics and pharmacogenomics, *British Journal of Clinical Pharmacology*, 52(4), pp. 345-347.
- Pirmohamed, M. (2006). Warfarin: almost 60 years old and still causing problems, *British Journal of Clinical Pharmacology*, 62(5), pp. 509-511.
- Pirmohamed, M., Burnside, G., Eriksson, N., Jorgensen, A. L., Toh, C. H., Nicholson, T., Kesteven, P., Christersson, C., Wahlstrom, B., Stafberg, C., Zhang, J. E., Leathart, J. B., Kohnke, H., Maitland-van der Zee, A. H., Williamson, P. R., Daly, A. K., Avery, P., Kamali, F. and Wadelius, M. (2013). A randomized trial of genotype-guided dosing of warfarin, *N Engl J Med*, 369(24), pp. 2294-2303.
- Poopak, B., Rabieipoor, S., Safari, N., Naraghi, E., Sheikhsofla, F. and Khosravipoor, G. (2015). Identification of CYP2C9 and VKORC1 polymorphisms in Iranian patients who are under warfarin therapy, *International Journal of Hematology-Oncology and Stem Cell Research*, 9(4), pp. 185-192.
- Preissner, S. C., Hoffmann, M. F., Preissner, R., Dunkel, M., Gewiess, A. and Preissner, S. (2013). Polymorphic Cytochrome P450 Enzymes (CYPs) and Their Role in Personalized Therapy, *PLoS ONE*, 8(12), p. e82562.
- Rehm, H. L., Bale, S. J., Bayrak-Toydemir, P., Berg, J. S., Brown, K. K., Deignan, J. L., Friez, M. J., Funke, B. H., Hegde, M. R., Lyon, E., the Working Group of the American College of Medical, G. and Genomics Laboratory Quality Assurance, C. (2013). ACMG clinical laboratory standards for next-generation sequencing, *Genetics in medicine : official journal of the American College of Medical Genetics*, 15(9), pp. 733-747.
- Rettie, A. E. and Tai, G. (2006). The pharmacogenomics of Warfarin closing in on personalized medicine, 6(5).
- Rhoads, A. and Au, K. F. (2015). PacBio Sequencing and Its Applications, *Genomics, Proteomics & Bioinformatics*, 13(5), pp. 278-289.
- Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P. and Weissman, M. M. (2013). A mega-analysis of genome-wide association studies for major depressive disorder, *Mol Psychiatry*, 18(4), pp. 497-511.

-
- Robert Freedman, David A. Lewis, Robert Michels, Daniel S. Pine, Susan K. Schultz, Carol A. Tamminga, Glen O. Gabbard, Susan Shur-Fen Gau, Daniel C. Javitt, Maria A. Oquendo, Patrick E. Shrout, Eduard Vieta and Joel Yager (2013). The Initial Field Trials of DSM-5: New Blooms and Old Thorns, *American Journal of Psychiatry*, 170(1), pp. 1-5.
- Roberts, R. J., Carneiro, M. O. and Schatz, M. C. (2013). The advantages of SMRT sequencing, *Genome Biology*, 14(7), pp. 405-405.
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., McKernan, K. J., Williams, A., Roth, G. T. and Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing, *Nature*, 475(7356), pp. 348-352.
- Rouse, M., Cristiani, C. and Teng, K. A. (2013). Should we use pharmacogenetic testing when prescribing warfarin?, *Cleveland Clinic Journal of Medicine*, 80(8), pp. 483-486.
- Sandmann, S., Graaf, A. O. d., Karimi, M., Reijden, B. A. v. d., Hellström-Lindberg, E., Jansen, J. H. and Dugas, M. (2017). Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data, *Sci. Rep.*, 7(43169).
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors, *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp. 5463-5467.
- Sara El-Metwally, M. S., Ouda, O. M. and Helmy, M. (2014). *Next Generation Sequencing Technologies and Challenges in Sequence Assembly*. Springer Science & Business.
- Schalk, M. and Croteau, R. (2000). A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (-)-limonene hydroxylase from a C6- to a C3-

-
- hydroxylase, *Proceedings of the National Academy of Sciences of the United States of America*, 97(22), pp. 11948-11953.
- Scott, S. A. (2011). Personalizing medicine with clinical pharmacogenetics, *Genetics in Medicine*, 13(12), pp. 987-995.
- Scott, S. A., Khasawneh, R., Peter, I., Kornreich, R. and Desnick, R. J. (2010). Combined CYP2C9, VKORC1 and CYP4F2 frequencies among racial and ethnic groups, *Pharmacogenomics*, 11(6), pp. 781-791.
- Serrati, S., De Summa, S., Pilato, B., Petriella, D., Lacalamita, R., Tommasi, S. and Pinto, R. (2016). Next-generation sequencing: advances and applications in cancer diagnosis, *Oncotargets and therapy*, 9, pp. 7355-7365.
- Shao, K., Ding, W., Wang, F., Li, H., Ma, D. and Wang, H. (2011). Emulsion PCR: A High Efficient Way of PCR Amplification of Random DNA Libraries in Aptamer Selection, *PLoS ONE*, 6(9), p. e24910.
- Shastry, B. S. (2009). SNPs: Impact on Gene Function and Phenotype, in Komar, A. A. (ed.) *Single Nucleotide Polymorphisms: Methods and Protocols*. Totowa, NJ: Humana Press, pp. 3-22.
- Shen, T., Pajaro-Van de Stadt, S. H., Yeat, N. C. and Lin, J. C. H. (2015). Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes, *Frontiers in Genetics*, 6, p. 215.
- Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing, *Nat Biotech*, 26(10), pp. 1135-1145.
- Sim, S. C. and Ingelman-Sundberg, M. (2010). The Human Cytochrome P450 (CYP) Allele Nomenclature website: a peer-reviewed database of CYP variants and their associated effects, *Human Genomics*, 4(4), pp. 278-281.
- Singh, R. (2015). *Bioinformatics: Genomics and Proteomics*. India: Vikas Publishing House.
- Softgenetics [Em linha]. Disponível em < www.softgenetics.com >. [Consultado em 25/04/2017]

-
- Sondheimer, N., Glatz, C. E., Tirone, J. E., Deardorff, M. A., Krieger, A. M. and Hakonarson, H. (2011). Neutral mitochondrial heteroplasmy and the influence of aging, *Human Molecular Genetics*, 20(8), pp. 1653-1659.
- Spencer, D. H., Tyagi, M., Vallania, F., Bredemeyer, A. J., Pfeifer, J. D., Mitra, R. D. and Duncavage, E. J. (2014). Performance of Common Analysis Methods for Detecting Low-Frequency Single Nucleotide Variants in Targeted Next-Generation Sequence Data, *The Journal of Molecular Diagnostics : JMD*, 16(1), pp. 75-88.
- Suppiah, V., Moldovan, M., Ahlenstiel, G., Berg, T., Weltman, M., Abate, M. L., Bassendine, M., Spengler, U., Dore, G. J., Powell, E., Riordan, S., Sheridan, D., Smedile, A., Fragomeli, V., Müller, T., Bahlo, M., Stewart, G. J., Booth, D. R. and George, J. (2009). IL28B is associated with response to chronic hepatitis C interferon-[alpha] and ribavirin therapy, *Nat Genet*, 41(10), pp. 1100-1104.
- Sweeney, B. P. and Bromilow, J. (2006). Liver enzyme induction and inhibition: implications for anaesthesia, *Anaesthesia*, 61(2), pp. 159-177.
- Tanaka, Y., Nishida, N., Sugiyama, M., Kurosaki, M., Matsuura, K., Sakamoto, N., Nakagawa, M., Korenaga, M., Hino, K., Hige, S., Ito, Y., Mita, E., Tanaka, E., Mochida, S., Murawaki, Y., Honda, M., Sakai, A., Hiasa, Y., Nishiguchi, S., Koike, A., Sakaida, I., Imamura, M., Ito, K., Yano, K., Masaki, N., Sugauchi, F., Izumi, N., Tokunaga, K. and Mizokami, M. (2009). Genome-wide association of IL28B with response to pegylated interferon-[alpha] and ribavirin therapy for chronic hepatitis C, *Nat Genet*, 41(10), pp. 1105-1109.
- Tattini, L., D'Aurizio, R. and Magi, A. (2015). Detection of Genomic Structural Variants from Next-Generation Sequencing Data, *Frontiers in Bioengineering and Biotechnology*, 3, p. 92.
- Thannickal, V. J. (2009). Oxygen in the Evolution of Complex Life and the Price We Pay, *American Journal of Respiratory Cell and Molecular Biology*, 40(5), pp. 507-510.
- Thomas, D. L., Thio, C. L., Martin, M. P., Qi, Y., Ge, D., O'huigin, C., Kidd, J., Kidd, K., Khakoo, S. I., Alexander, G., Goedert, J. J., Kirk, G. D., Donfield, S. M.,

-
- Rosen, H. R., Tobler, L. H., Busch, M. P., McHutchison, J. G., Goldstein, D. B. and Carrington, M. (2009). Genetic variation in IL28B and spontaneous clearance of hepatitis C virus, *Nature*, 461(7265), pp. 798-801.
- Thomas, R. K., Nickerson, E., Simons, J. F., Jänne, P. A., Tengs, T., Yuza, Y., Garraway, L. A., LaFramboise, T., Lee, J. C., Shah, K., O'Neill, K., Sasaki, H., Lindeman, N., Wong, K.-K., Borrás, A. M., Gutmann, E. J., Dragnev, K., DeBiasi, R., Chen, T.-H., Glatt, K. A., Greulich, H., Desany, B., Lubeski, C. K., Brockman, W., PabloAlvarez, Hutchison, S. K., Leamon, J. H., Ronan, M. T., Turenchalk, G. S., Egholm, M., Sellers, W. R., Rothberg, J. M. and Meyerson, M. (2006). Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing, *Nat. Med.* , 12, pp. 852–855.
- Thompson, A. J. and McHutchison, J. G. (2012). Will IL28B polymorphism remain relevant in the era of direct-acting antiviral agents for hepatitis C virus?, *Hepatology*, 56(1), pp. 373-381.
- Thompson, J. F. and Steinmann, K. E. (2010). Single Molecule Sequencing with a HeliScope Genetic Analysis System, *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, CHAPTER, pp. Unit7.10-Unit17.10.
- Tomaszewski, P., Kubiak-Tomaszewska, G. and Pachecka, J. (2008). Participation of CYP isoenzymes in the metabolism of endogenous substances and drugs, *Polish Pharmaceutical Society*, 65(3), pp. 307-318.
- Tomkinson, A. E., Vijayakumar, S., Pascal, J. M. and Ellenberger, T. (2006). DNA ligases: structure, reaction mechanism, and function, *Chem Rev*, 106(2), pp. 687-699.
- TruSeq® Amplicon - Cancer Panel (2012). Illumina, Inc.
- Turner, S. T., Bailey, K. R., Fridley, B. L., Chapman, A. B., Schwartz, G. L., Chai, H. S., Sicotte, H., Kocher, J.-P., Rodin, A. S. and Boerwinkle, E. (2008). Genomic Association Analysis Suggests Chromosome 12 Locus Influencing Antihypertensive Response to Thiazide Diuretic, *Hypertension*, 52(2), pp. 359-365.

-
- Turner, S. T., Bailey, K. R., Schwartz, G. L., Chapman, A. B., Chai, H. S. and Boerwinkle, E. (2012). Genomic Association Analysis Identifies Multiple Loci Influencing Antihypertensive Response To an Angiotensin II Receptor Blocker, *Hypertension*, 59(6), pp. 1204-1211.
- Turner, S. T., Boerwinkle, E., O'Connell, J. R., Bailey, K. R., Gong, Y., Chapman, A. B., McDonough, C. W., Beitelshes, A. L., Schwartz, G. L., Gums, J. G., Padmanabhan, S., Hiltunen, T. P., Citterio, L., Donner, K. M., Hedner, T., Lanzani, C., Melander, O., Saarela, J., Ripatti, S., Wahlstrand, B., Manunta, P., Kontula, K., Dominiczak, A. F., Cooper-DeHoff, R. M. and Johnson, J. A. (2013). Genomic Association Analysis of Common Variants Influencing Antihypertensive Response to Hydrochlorothiazide, *Hypertension*, 62(2), pp. 391-397.
- Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J. A., Costa, G., McKernan, K., Sidow, A., Fire, A. and Johnson, S. M. (2008). A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning, *Genome Res*, 18(7), pp. 1051-1063.
- van der Werf, I. M., Kooy, R. F. and Vandeweyer, G. (2015). A Robust Protocol to Increase NimbleGen SeqCap EZ Multiplexing Capacity to 96 Samples, *PLoS ONE*, 10(4), p. e0123872.
- Varela, M. A. and Amos, W. (2010). Heterogeneous distribution of SNPs in the human genome: Microsatellites as predictors of nucleotide diversity and divergence, *Genomics*, 95(3), pp. 151-159.
- Victoria, J. G., Wang, C., Jones, M. S., Jaing, C., McLoughlin, K., Gardner, S. and Delwart, E. L. (2010). Viral Nucleic Acids in Live-Attenuated Vaccines: Detection of Minority Variants and an Adventitious Virus, *Journal of Virology*, 84(12), pp. 6033-6040.
- Wang, X. (2016). *Next-Generation Sequencing Data Analysis*. CRC Press, Inc.
- Wolf, C. R., Smith, G. and Smith, R. L. (2000). Pharmacogenetics, *BMJ : British Medical Journal*, 320(7240), pp. 987-990.

-
- Woollard, P. M., Mehta, N. A. L., Vamathevan, J. J., Van Horn, S., Bonde, B. K. and Dow, D. J. (2011). The application of next-generation sequencing technologies to drug discovery and development, *Drug Discovery Today*, 16(11–12), pp. 512-519.
- Yadav, N. K., Shukla, P., Omer, A., Pareek, S. and Singh, R. K. (2014). Next Generation Sequencing: Potential and Application in Drug Discovery, *The Scientific World Journal*, 2014, p. 7.
- Yanhu, L., Lu, W. and Li, Y. (2015). [The principle and application of the single-molecule real-time sequencing technology], *Yi Chuan*, 37(3), pp. 259-268.
- Yu, M., Stott, S., Toner, M., Maheswaran, S. and Haber, D. A. (2011). Circulating tumor cells: approaches to isolation and characterization, *The Journal of Cell Biology*, 192(3), pp. 373-382.
- Yuan, Y., Xu, H. and Leung, R. K.-K. (2016). An optimized protocol for generation and analysis of Ion Proton sequencing reads for RNA-Seq, *BMC Genomics*, 17, p. 403.
- Zandi, P. P. and Judy, J. T. (2010). The Promise and Reality of Pharmacogenetics in Psychiatry, *The Psychiatric clinics of North America*, 33(1), pp. 181-224.
- Zhang, J., Chiodini, R., Badr, A. and Zhang, G. (2011). The impact of next-generation sequencing on genomics, *Journal of genetics and genomics = Yi chuan xue bao*, 38(3), pp. 95-109.
- Zhou, S.-F., Liu, J.-P. and Chowbay, B. (2009). Polymorphism of human cytochrome P450 enzymes and its clinical impact, *Drug Metabolism Reviews*, 41(2), pp. 89-295.