

Universidade de Trás-os-Montes e Alto Douro

**Genotipagem do *Human Leucocyte Antigen (HLA)*:
Técnicas atuais e aplicabilidade da abordagem de
*Next Generation Sequencing***

Dissertação de Mestrado em:
Biotecnologia para as Ciências da Saúde

Ana Clara Nunes Barbosa

Orientadora: Maria José Sá Carneiro de Paiva Peixoto

Co-Orientadora: Estela Maria Bastos Martins de Almeida



Vila Real, 2019

Genotipagem do *Human Leucocyte Antigen* (HLA):
Técnicas atuais e aplicabilidade da abordagem de *Next Generation Sequencing*

Universidade de Trás-os-Montes e Alto Douro

**Genotipagem do *Human Leucocyte Antigen* (HLA):
Técnicas atuais e aplicabilidade da abordagem de
*Next Generation Sequencing***

Dissertação de Mestrado em:
Biotecnologia para as Ciências da Saúde

Nome da candidata
Ana Clara Nunes Barbosa

Orientadores
Maria José Sá Carneiro de Paiva Peixoto
Estela Maria Bastos Martins de Almeida

Composição do Júri:

Doutora Maria do Rosário Lima Viseu de Carvalho Pinto Leite

Professora Doutora Estela Maria Bastos Martins de Almeida

Engenheira Maria José Sá Carneiro de Paiva Peixoto

Vila Real, 2019

Declara-se sob compromisso de honra que este trabalho foi expressamente elaborado pelo autor, como dissertação original, para o efeito da obtenção do grau de Mestre em Biotecnologia para as Ciências da Saúde, na Universidade de Trás-os-Montes e Alto Douro. Todas as contribuições não originais foram devidamente identificadas com indicação da fonte.

Genotipagem do *Human Leucocyte Antigen* (HLA):
Técnicas atuais e aplicabilidade da abordagem de *Next Generation Sequencing*

Genotipagem do *Human Leucocyte Antigen* (HLA):
Técnicas atuais e aplicabilidade da abordagem de *Next Generation Sequencing*

**Um dia quando olhares para trás,
verás que os dias mais belos
foram aqueles em que lutaste.”**

Sigmund Freud

Genotipagem do *Human Leucocyte Antigen* (HLA):
Técnicas atuais e aplicabilidade da abordagem de *Next Generation Sequencing*

Agradecimentos

Para a realização deste trabalho e concretização desta dissertação, foram várias as pessoas que ajudaram, direta ou indiretamente, com o seu apoio e disponibilidade, às quais quero demonstrar o meu honesto agradecimento.

À **Universidade de Trás os Montes e Alto Douro**, na pessoa do **Magnífico Reitor, professor Doutor António Fontainhas Fernandes**, pela disponibilidade e facilidade concebidas para a realização deste trabalho.

À **Direção de Mestrado**, pela aceitação deste projeto.

A todos os **Professores do Mestrado** por todo o apoio que nos deram durante estes dois anos e por todos os ensinamentos.

Agradeço do fundo do coração à **Professora Doutora Estela Bastos**, por mais uma vez me ter encontrado um estágio. Por tudo aquilo que me ensinou, pela compreensão, disponibilidade e apoio excecional na execução desta dissertação. Obrigada pelo apoio a nível pessoal e acima de tudo obrigada por estar sempre lá nestes anos todos. Continue a ser a excelente professora e pessoa que é.

À **Engenheira Maria Peixoto**, por ter aceite ser minha orientadora e me proporcionar enormes conhecimentos. Obrigada por toda a ajuda e por todo o tempo que despendeu comigo. Mesmo não tendo nada em troca e não me conhecendo, sempre deu o seu melhor e estarei eternamente grata por isso.

A todos as pessoas do **CENTRO DE SANGUE E TRANSPLANTAÇÃO DO PORTO-UNIDADE DE TRANSPLANTAÇÃO**. Todos eles, sem exceção, obrigado pela simpatia e boa disposição. Obrigada especialmente à equipa do Laboratório de Genética Molecular onde me integrei, por tudo o que me ensinaram, por todos os sorrisos e acima de tudo por me terem aceitado no vosso espaço. E às minhas companheiras de almoço pelas conversas partilhadas.

Obrigada aos meus **Clientes** que se tornaram amigos, por todos as gargalhadas, partilhas e preocupação. Sem dúvida os melhores que irei encontrar.

A todos os meus **Confidentes**, que sempre estiveram ao meu lado. Aos que Vila Real me deu: Anícia, Daniela, Fátima, Lígia, Patrícia e Rodolfo. Aos que quero que permaneçam sempre na minha vida: Andreia, Alexandre, Andreia, Alfredo, Tomás, Marina, Mónica, Nadine, Zé, Afonso, Inês, Mónica, Rosa e Águeda. Obrigada por todos os momentos, sorrisos e por acreditarem em mim!

Toda a **Família do meu Macaquinho**, que são uma segunda casa para mim. Um agradecimento especial aos meus **Sogros** por fazerem parte desta minha caminhada, e por me fazerem sempre sentir em casa.

Aos meus **Padrinhos de Batismo**, que mesmo morando na casa em frente sempre compreenderam as minhas ausências à sua casa que se prolongavam por semanas. Obrigada por todo o orgulho que têm em mim e por terem contribuído para a minha educação e para a pessoa que sou hoje! Obrigada por todos estes 23 anos de carinho!

Ao meu **Tio Néelson, Tia Fátima, Tia Conceição, Primos António Agostinho e Olga**, e à minha **Princesa Catarina** por estarem sempre lá quando eu precisei sem pedir nada em troca! E por me acompanharem sempre nestes 23 anos.

Ao meu **Cunhado Thomas** e à minha **Cunhada Cristina** por fazerem parte da minha família. E por me apoiarem neste projeto.

Aos meus **Patrões** e **Amigos** Sr. Vitorino e D.^a Manuela. Obrigada pela amizade e obrigada por me compreenderem e me auxiliarem nos meus horários. Obrigada pelos sorrisos, pelo que me ensinaram para ser uma excelente profissional em qualquer trabalho que vá realizar na minha vida e por me ajudarem a crescer. Obrigada do fundo do coração.

Aos meus **Sobrinhos Luana, Lara, Diego e Pierre** por serem o meu bem mais precioso e os meus diamantes. Ao meu **Irmão José Luís** e à minha **Irmã Isabel** por tudo o que fizeram por mim durante estes seis anos e durante toda a minha vida! Sei que estarão sempre lá para mim sem pedir nada em troca. Sem vocês estes seis anos não seriam possíveis! Obrigada pelo carinho e proteção, por tudo o que me deram! Agradecer o suficiente nunca será possível.

Ao meu **Pai** e à minha **Mãe**, que sem eles nada disto era possível. Por todo o esforço e enormes sacrifícios que fizeram para eu poder continuar a estudar, um grande OBRIGADO. Não chegam todas as palavras do mundo para vos agradecer.

E a Ti, meu **Macaquinho**. Por durante estes seis anos me dares a maior força do mundo. Por acreditares sempre em mim, principalmente quando eu não acreditava! Por toda a ajuda que me deste! Sabes o valor que tens para mim! Obrigada por não me deixares desistir!

A TODOS, UM GRANDE OBRIGADA!

Resumo

O complexo MHC (*Major Histocompatibility Complex*) está presente em todos os mamíferos e, no homem é designado HLA (*Human Leucocyte Antigen*). Este *cluster* de genes está localizado no braço curto do cromossoma 6 (6p21) e possui mais de 220 genes, sendo uma das regiões mais intensivamente estudadas no genoma humano há mais de 20 anos, sendo que ainda nos dias de hoje é alvo de vários estudos. O complexo génico HLA é um dos mais polimórficos do genoma humano e divide-se em três regiões: HLA classe I, II e III. A principal função das moléculas de MHC é a apresentação de fragmentos peptídicos provenientes de agentes patogénicos e exibi-los à superfície das células de forma a serem reconhecidos e eliminados adequadamente pelos linfócitos T.

Desde a descoberta do HLA, vários estudos contribuíram para a melhor compreensão da resposta imunitária. As primeiras associações entre a região HLA e determinadas doenças reforçaram a ideia já existente, da complexidade deste *cluster* de genes e o seu extenso polimorfismo. Com o passar do tempo, foi possível verificar que determinados genótipos HLA estão associados com um grande número de doenças, bem como a reações adversas a certos fármacos e com a compatibilidade HLA na transplantação. O estudo desta região permitiu perceber muitas das suas características únicas, despertando um grande interesse tanto a nível clínico como científico, de tal forma que várias técnicas de genotipagem HLA foram desenvolvidas. Inicialmente a genotipagem era feita através de métodos serológicos, no entanto não eram completamente satisfatórios. Foi necessário desenvolver métodos de tipagem mais robustos e específicos: PCR-SSP (*Sequence Specific Primer*), PCR-SSO (*Sequence Specific Oligonucleotide*) e PCR-SBT (*Sequencing based typing*). No entanto, estes também apresentam resultados com ambiguidades, dependendo do tipo de metodologia e *kit* usado. Mais recentemente o desenvolvimento da tecnologia NGS (*Next Generation Sequencing*) permitiu um maior nível de resolução entre outras vantagens discutidas ao longo deste trabalho.

O presente trabalho teve como objetivos a aplicação das técnicas clássicas de genotipagem de HLA e a revisão sobre a aplicabilidade de NGS nesta área.

Neste estudo, após utilização das técnicas, foram calculadas as frequências alélicas para as amostras de doadores de medula e de sangue do cordão umbilical. Para os genes HLA-A, -B, -C e -DRB1, os alelos mais frequentes foram os mesmos em ambos os tipos de amostras, tendo sido o HLA-A*02 com frequências alélicas de 26,83% e 27,55%; HLA-B*44 com 14,34% e 13,96%; HLA-C*07 com 24,66% e 25,66%; e o gene HLA-DRB1*07 com

18,09% e 16,60%, para as amostras de dadores de medula e sangue do cordão, respetivamente. Estas frequências alélicas descritas relativamente às amostras de dadores de medula óssea voluntários e amostras de cordão a nível do Norte de Portugal mostram uma semelhança com as frequências alélicas obtidos por outros autores relativamente à população Portuguesa.

Com base na literatura, a introdução da tecnologia NGS nos laboratórios clínicos trará inúmeros benefícios, principalmente ao nível da compatibilidade dador-recetor. Esta tecnologia acarreta menos ambiguidades, menos custos associados à medida que a tecnologia evolui e para além disso, permitirá aumentar o nosso conhecimento relativamente à relevância biológica não só das regiões codificantes, mas principalmente das regiões não codificantes, já que permite a sequenciação mais alargada dos genes. Os estudos apresentados revelam dados bastantes satisfatórios quanto à aplicabilidade a nível laboratorial do NGS, com taxas de precisão acima dos 97% na genotipagem HLA.

Palavras-chave: HLA; genotipagem; PCR-SSO; PCR-SSP; PCR-SBT; NGS.

Abstract

The Major Histocompatibility Complex (MHC) is present in all mammals and is called HLA (Human Leucocyte Antigen) in humans. This gene cluster is located in the short arm of chromosome 6 (6p21) and has more than 220 genes, being one of the most intensively studied regions in the human genome for over 20 years, and is still the subject of many studies today. The HLA gene complex is one of the most polymorphic in the human genome and is divided into three regions: HLA class I, II and III. The primary function of MHC molecules is to present peptide fragments from pathogens and display them on the cell surface in order to be recognized and properly deleted by T lymphocytes.

Since the discovery of HLA, several studies have contributed to a better understanding of the immune response. The first associations between the HLA region and certain diseases reinforced the existing idea of the complexity of this gene cluster and its extensive polymorphism. Over time, it has been found that certain HLA genotypes are associated with a large number of diseases, as well as adverse reactions to certain drugs and HLA compatibility in transplantation. The study of this region has made it possible to see many of its unique characteristics, arousing great interest both clinically and scientifically, so that various HLA genotyping techniques have been developed. Initially the genotyping was done by serological methods, however they were not completely satisfactory. More robust and specific typing methods were needed: PCR-SSP (Sequence Specific Primer), PCR-SSO (Sequence Specific Oligonucleotide) and PCR-SBT (Sequencing Based Typing). However, these also yield ambiguous results, depending on the type of methodology and kit used. More recently, the development of Next Generation Sequencing (NGS) technology has enabled a higher level of resolution among other advantages discussed throughout this work.

The present work aimed to apply the classic HLA genotyping techniques and to review the applicability of NGS in this area.

In this study, after using the techniques, allele frequencies were calculated for bone marrow and cord blood donor samples. For the HLA-A, -B, -C and -DRB1 genes, the most frequent alleles were the same in both types of samples, being HLA-A*02 with allele frequencies of 26.83% and 27.55%; HLA-B*44 with 14.34% and 13.96%; HLA-C*07 with 24.66% and 25.66%; and the HLA-DRB1*07 gene with 18.09% and 16.60% for cord and cord blood donor samples, respectively. These allelic frequencies described for samples from volunteer bone marrow donors and cord samples in Northern Portugal show a similarity to the allelic frequencies obtained by other authors from the Portuguese population.

Based on the literature, the introduction of NGS technology in clinical laboratories will bring numerous benefits, particularly in terms of donor-recipient compatibility. This technology leads to less ambiguity, less cost as technology evolves, and furthermore, it will increase our knowledge regarding the biological relevance of polymorphisms not only of coding sequences, but especially non-coding regions, as it allows for broader gene sequencing. The studies presented reveal quite satisfactory data regarding the NGS laboratory applicability, with accuracy rates above 97% in HLA genotyping.

Keywords: HLA; genotyping; PCR-SSO; PCR-SSP; PCR-SBT; NGS.

Índice Geral

Agradecimentos	vii
Resumo	ix
Abstract	xi
Índice geral	xiii
Índice de figuras	xv
Índice de tabelas	xvii
Lista de abreviaturas	xviii
Capítulo 1- Introdução	1
1.1. Considerações gerais sobre o MHC	1
1.2. Estrutura e conteúdo génico do HLA	1
1.3. Contexto Imunológico	3
1.4. Estrutura das proteínas MHC, processamento e apresentação antigénica	3
1.5. Nomenclatura do HLA	7
1.6. HLA e doenças humanas	8
1.6.1. Doença celíaca	8
1.6.2. Diabetes Mellitus tipo 1	9
1.6.3. Espondilite anquilosante	10
1.6.4. Hipersensibilidade ao abacavir	10
1.7. HLA e transplantação	11
1.8. Métodos para genotipagem do HLA	12
1.8.1. Genotipagem Serológica	12
1.8.2. Genotipagem Genética	13
1.8.2.1. PCR-SSP (<i>Sequence Specific Primer</i>)	13
1.8.2.2. PCR-SSO (<i>Sequence Specific Oligonucleotide</i>)	14
1.8.2.3. Reverse PCR-SSO (<i>Reverse Sequence Specific Oligonucleotide</i>)	15
1.8.2.4. PCR-SBT (<i>Sequencing based typing</i>)	16
1.9. Tecnologia NGS e o HLA	17
1.9.1. <i>Roche454 System</i>	18
1.9.2. <i>SOLiD (Sequencing by oligonucleotide ligation and detection)</i>	19
1.9.3. <i>Ion Torrent PGM (Personal Genome Machine)</i>	20
1.9.4. <i>Illumina</i>	21
1.9.5. <i>Pacific Biosciences</i>	23
1.9.6. <i>Oxford Nanopore Technologie</i>	23
1.9.7. Bioinformática e considerações finais	24
Capítulo 2- Material e Métodos	25
2.1. Colheita das amostras biológicas e triagem	25
2.2. Extração de DNA	25
2.3. Quantificação e Análise da pureza do DNA	27
2.4. Genotipagem HLA Classe I e II pelo <i>kit OLERUP SSP</i>	27
2.4.1. Procedimento	27
2.4.1.1. Amplificação por PCR	27
2.4.1.2. Pós-Amplificação	28
2.4.2. Cálculo e análise dos resultados	28
2.4.3. Validação dos resultados	29
2.4.4. Interpretação do PCR-SSP pelo <i>software SCORE</i>	29
2.5. Genotipagem HLA pela metodologia rSSO usando o <i>kit Labtype rSSO</i> através da tecnologia <i>Luminex</i>	29

2.5.1. Procedimento	29
2.5.1.1. Amplificação por PCR	29
2.5.1.2. Desnaturação/Neutralização	30
2.5.1.3. Hibridização	30
2.5.1.4. Marcação	30
2.5.2. Validação	31
2.5.3. Leitura e interpretação dos dados com o <i>software</i> HLA <i>Fusion</i>	31
2.6. Genotipagem HLA pela metodologia PCR-SBT HLA usando os <i>kits</i> <i>HLAssure Texas BioGene</i>	31
2.6.1. Procedimento	32
2.6.1.1. Amplificação por PCR	33
2.6.1.2. Purificação das reações de PCR	34
2.6.1.3. Reação de sequenciação	34
2.6.1.4. Purificação das reações por precipitação alcoólica	34
2.6.1.5. Eletroforese Capilar	34
2.6.2. Validação	34
2.7. Genotipagem HLA por Sequenciação <i>SeCore</i>	35
2.7.1. Procedimento	35
2.7.1.1. Amplificação por PCR	35
2.7.1.2. Purificação das reações de PCR	35
2.7.1.3. Reação de sequenciação	35
2.7.1.4. Purificação das reações por precipitação alcoólica	36
2.7.1.5. Eletroforese Capilar	36
2.7.2. Validação	36
2.7.3. Interpretação de sequências pelo <i>software</i> <i>uTYPE</i>	37
2.8. Cálculo das frequências alélicas	37
Capítulo 3- Resultados e Discussão	39
Capítulo 4- Conclusão	61
Referências Bibliográficas	63
Anexos	

Índice de Figuras

Figura 1- Estrutura génica do MHC humano.	1
Figura 2- Número de alelos para alguns genes HLA. Dados recolhidos em julho de 2019.	2
Figura 3. a) Estrutura das moléculas de MHC classe I. b) Estrutura da molécula de MHC classe II.	4
Figura 4. Apresentação antigénica por parte das moléculas de MHC classe I.	6
Figura 5. Apresentação antigénica por parte das moléculas de MHC classe II.	7
Figura 6. Nomenclatura de um alelo HLA.	8
Figura 7. Interpretação pelo software <i>SCORE</i> no programa <i>HELMBERG-SCORE</i> , para a genotipagem HLA Classe I e II pelo kit <i>OLERUP SSP</i> .	45
Figura 8. Histogramas da interpretação da tecnologia <i>Luminex</i> através do software HLA <i>Fusion</i> da <i>One Lambda</i> .	46
Figura 9. Genotipagem em alta resolução obtida pelos kits <i>HLAssure Texas BioGene</i> e interpretação pelo software <i>ACCUTYPE</i> .	47
Figura 10. Genotipagem HLA pelos kits <i>SeCore</i> e interpretação no software <i>uTYPE</i> .	48

Genotipagem do *Human Leucocyte Antigen* (HLA):
Técnicas atuais e aplicabilidade da abordagem de *Next Generation Sequencing*

Índice de Tabelas

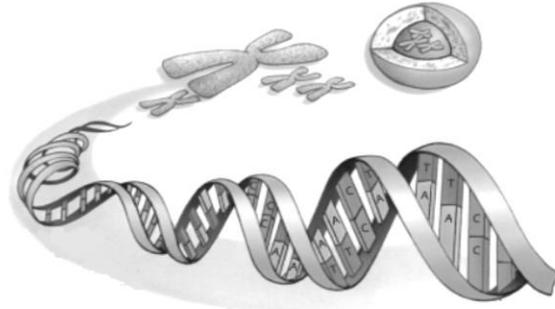
Tabela 1. Frequências alélicas, em baixa resolução, das amostras de doadores de medula óssea do Norte de Portugal, genotipados desde 2017 até julho de 2018.	40
Tabela 2. Frequências alélicas em baixa resolução para as amostras doadas de sangue do cordão genotipados desde 2013 até julho de 2018, na região Norte de Portugal.	41
Tabela 3. Frequências alélicas para os loci HLA-A, -B, -C e -DRB1 em estudos de genotipagem de doadores de medula óssea em diferentes populações.	44
Tabela 4. Vantagens e desvantagens que levaram à evolução da genotipagem serológica para a genotipagem genética.	50
Tabela 5. Resumo das vantagens e desvantagens para os métodos convencionais e para a tecnologia NGS.	50
Tabela 6. Visão geral sobre as plataformas NGS usadas em genotipagem HLA e comparação de suas características.	52

Lista de Abreviaturas

- aa**- aminoácido
ACD- Ácido citrato dextrose
APC- Células apresentadoras de antígenos (*Antigen Presenting Cell*)
Ala- Alanina
Asp- Ácido Aspártico
bp- pares de bases
DNA- Ácido desoxirribonucleico
CEDACE- Centro Nacional de Dadores de Células de Medula Óssea, Estaminais ou de Sangue do Cordão
cDNA- DNA complementar
CLIP- Péptido da cadeia invariante associado com a Classe II (*Class II – associated Invariant chain Peptide*)
CTLs- Linfócitos T citotóxicos
DII- Doença inflamatória intestinal
ddNTPs- Didesoxinucleotídeos
dNTPs- Desoxinucleotídeos
dNMPs- Desoxinucleosídeos monofosfato
EA- Espondilite Anquilosante
EDTA- Ácido etilenodiamina tetra-acético
ERAP 1- Aminopeptidase 1 do retículo endoplasmático (*Endoplasmic Reticulum Aminopeptidase 1*)
ERp57- Proteína dissulfeto isomerase 57 do retículo endoplasmático
FA- Frequência alélica
GSA- *Primers* de amplificação para PCR de um grupo específico
Indels- Inserções e deleções
IPST- Instituto Português do Sangue e da Transplantação
KIR- *Killer-cell immunoglobulin-like receptors*
LSA- *Primers* de amplificação para PCR de um locus específico
HLA- *Human Leukocyte Antigen*
MHC- *Major Histocompatibility Complex*
MIIC- Compartimento do MHC classe II (*MHC class II compartment*)
miRNA- Micro RNAs
mRNA- RNA mensageiro
NGS- *Next Generation Sequencing*
NKT- Linfócitos T *Natural Killer*
PCR- Reação em cadeia da polimerase (*Polymerase Chain Reaction*)
PCR-SBT- *Sequencing based typing*
PCR-SSO- *Sequence Specific Oligonucleotide*
PCR-SSP- *Sequence Specific Primer*
PE- R-ficoeritrina
PGM- *Personal Genome Machine*
PLC- Complexo de carga peptídico (*Peptide Loading Complex*)
RE- Retículo Endoplasmático
RCLB- *Red cell lysis buffer*
RNA- Ácido ribonucleico
SAPE- R-ficoeritrina conjugada à estreptavidina
Ser- Serina
SI- Sistema Imunitário
SMRT- Sequenciação em tempo real de uma única molécula (*Single Molecule Real Time Sequencing*)
SNPs- Polimorfismo de nucleótido único (*Single Nucleotide Polymorphism*)
SNVs- Variantes de um único nucleótido (*Single Nucleotide Variants*)
SOLiD- *Sequencing by Oligonucleotide Ligation and Detection*
TAP- Transportador associado ao processamento antigénico (*Transporter associated with Antigen Processing*)
TBE- Tris/Borato/EDTA
TCR- Recetor das células T
T1DM- Diabetes mellitus tipo 1 (*Type 1 Diabetes Mellitus*)
Th- Linfócito T *Helper*
Treg- Linfócitos T reguladores
UTR- Regiões não traduzidas (*Untranslated region*)
UV- Radiação ultravioleta
VIH- Vírus da imunodeficiência humana
α- Alfa
β- Beta
φ- Phi

Capítulo 1

Introdução



1.1. Considerações gerais sobre o MHC

O complexo MHC (*Major Histocompatibility Complex*) está presente em todos os mamíferos. O MHC humano, designado HLA (*Human Leucocyte Antigen*) está localizado no braço curto do cromossoma 6 (6p21) e engloba mais de 220 genes, destacando-se genes que codificam proteínas envolvidas na mediação de respostas imunitárias a doenças infecciosas, influenciando igualmente o resultado na transplantação de órgãos e células (Robinson, 2016). Esta é uma das regiões mais estudadas do genoma, o que permitiu evidenciar características únicas e um grande interesse científico e clínico, de tal forma que várias técnicas foram desenvolvidas para genotipagem desta região (Allcock, 2012).

A principal função das moléculas HLA é a apresentação de péptidos provenientes de agentes patogénicos ou mesmo do próprio organismo e exibi-los à superfície das células de forma a serem reconhecidos e eliminados por linfócitos T (Robinson, 2016).

1.2. Estrutura e conteúdo génico do HLA

O complexo HLA apresenta uma elevada densidade génica. É um dos locais mais polimórficos do genoma humano e divide-se em três regiões: classe I, II e III (Choo *et al.*, 2007), tal como descrito na figura 1.

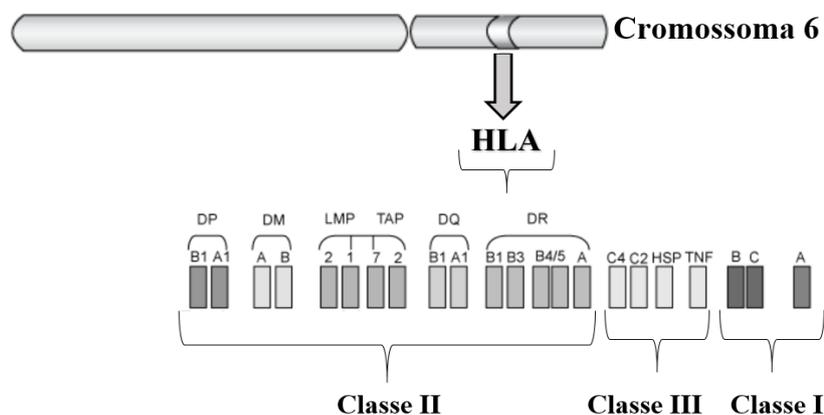


Figura 1. Estrutura génica do MHC humano. O complexo HLA localiza-se no cromossoma 6 e divide-se em três classes: **classe I** (HLA-A, -B, -C, entre outros) numa posição mais telomérica, **classe II** (HLA-DP, -DM, -DO, -DQ, -DR, entre outros) numa posição mais centromérica e **classe III**, cujos genes codificam componentes do sistema do complemento (Adaptado de Silva *et al.*, 2008).

Dada a importância do HLA em doenças e na transplantação, os genes pertencentes a este complexo têm sido alvo dos mais variados estudos, tendo a genotipagem em alta resolução dos principais *loci* do HLA evidenciado informação essencial, permitindo criar um banco de dados com todas as sequências oficiais do HLA, o IMGT/HLA (www.ebi.ac.uk/imgt/hla), cujo conteúdo é constantemente atualizado sempre que um novo alelo ou uma nova função de determinado alelo são descobertos (Allcock, 2012). O anexo A apresenta o aumento exponencial na identificação de novos alelos HLA desde 1987 e julho de 2019. A figura 2 apresenta a distribuição destes alelos pelos genes mais importantes do HLA (<https://www.ebi.ac.uk/ipd/imgt/hla/stats.html>).

NÚMERO DE ALELOS HLA			
Alelos HLA		23907	
HLA classe I		17191	
HLA classe II		6716	

HLA CLASSE I			
Genes	Alelos	Proteínas	Nulos
A	5266	3552	286
B	6537	4494	3359
C	5140	3359	236
E	43	11	1
F	44	6	0
G	69	19	3

HLA CLASSE II			
Genes	Alelos	Proteínas	Nulos
DRA	7	2	0
DRB	3171	2226	124
DRB1	2581	1834	78
DRB3	304	233	15
DRB4	233	152	16
DRB5	96	70	15
DQA1	183	76	6
DQB1	1718	1151	72
DPA1	132	52	1
DPB1	1449	960	75

Figura 2. Número de alelos para alguns genes HLA. Dados recolhidos em julho de 2019. No total, existem **23.907** alelos HLA, dentro dos quais **17.191** de classe I e **6.716** de classe II. Na classe I, o gene mais polimórfico é o **B** e o menos polimórfico é o E. Na classe II, o **DRB** é o mais polimórfico (Adaptado de <https://www.ebi.ac.uk/ipd/imgt/hla/stats.html>, acessado em julho de 2019).

A classe I codifica as proteínas clássicas (A, B, C) e não clássicas (E, F, G) e inclui genes com funções intrínsecas ao SI (Sistema Imunitário) e genes com funções fora do SI. A classe II inclui sobretudo genes que codificam proteínas HLA-II (subunidades α e β) e um pequeno número de genes envolvidos no processamento de antígenos para as proteínas de classe I. A classe III possui maior densidade génica, com genes com funções variáveis dentro e fora do SI, mas apresenta menor variabilidade (Allcock, 2012).

1.3. Contexto imunológico

As moléculas HLA-I expressam-se na superfície da maioria das células nucleadas enquanto HLA-II expressa-se unicamente nas células apresentadoras de antígenos (APC- *Antigen Presenting Cell*), como os linfócitos T ativados, os linfócitos B, os monócitos/macrófagos e células dendríticas (Choo *et al.*, 2007). Os linfócitos subdividem-se em T e B, e são células imunologicamente competentes que participam, juntamente com os neutrófilos e monócitos, na defesa do organismo contra infecções e invasão por corpos estranhos. Os linfócitos B reconhecem o antígeno mesmo na sua forma nativa, ao contrário das células T que apenas reconhecem antígenos apresentados sob a forma de péptidos pelo MHC (Kallon, 2011). Os linfócitos T reconhecem uma grande variedade de antígenos provenientes de vírus, bactérias e outros microrganismos. O reconhecimento pelas células B é feito através das imunoglobulinas, que podem ser segregadas pela célula sob forma solúvel, os anticorpos, ou podem ficar ancoradas à membrana. Os linfócitos T dividem-se em auxiliares (*helper*- Th), citotóxicos (CTLs), linfócitos T de memória, linfócitos T reguladores (Treg) e *Natural Killer* (NKT). Os linfócitos Th e CTLs expressam à sua superfície os coreceptores CD4 e CD8, respetivamente. Os T-CD4⁺ reconhecem péptidos ligados a HLA-II e os T-CD8⁺ reconhecem péptidos ligados a HLA-I, sendo este reconhecimento fundamental para a resposta efetiva das células T (Janeway *et al.*, 2001).

1.4. Estrutura das proteínas MHC, processamento e apresentação antigénica

As moléculas HLA-I são glicoproteínas que consistem numa cadeia α codificada por genes da classe I do HLA em associação por ligações não covalentes com um péptido, a β 2 microglobulina, que é codificada por um gene com *locus* no cromossoma 15. A cadeia α divide-se em três domínios extracelulares (α 1, α 2 e α 3) uma região transmembranar e uma cauda citoplasmática. Tal como apresentado na figura 3a, os domínios α 1 e α 2 formam um sulco que permite a ligação do péptido, habitualmente com 8 a 10 aminoácidos. A região transmembranar possui características hidrofóbicas permitindo a ancoragem da molécula à membrana da célula alvo através da cauda citoplasmática. A β 2 microglobulina é altamente conservada e liga-se ao coreceptor CD8. A constituição aminoacídica do sulco de ligação ao antígeno é responsável pelo elevado polimorfismo de HLA-I, condicionando o tipo de péptido ao qual a molécula se vai ligar (Anaya *et al.*).

As moléculas HLA-II são glicoproteínas que se encontram ligadas à membrana por duas cadeias polimórficas associadas de forma não covalente, cadeia α e β codificadas por

genes HLA-II. Tal como descrito na figura 3b, os domínios extracelulares $\alpha 1$ e $\beta 1$ formam o sulco de ligação ao péptido; a região transmembranar com aminoácidos hidrofóbicos permite a ancoragem à membrana através da cauda citoplasmática e os domínios $\alpha 2$ e $\beta 2$ são o local de ligação de CD4 das células T. Os péptidos que se ligam à molécula de HLA-II são maiores (12 a 24 aminoácidos) e o elevado polimorfismo deve-se aos diferentes aminoácidos que revestem o sulco onde se liga o péptido (**Janeway *et al.*, 2001**).

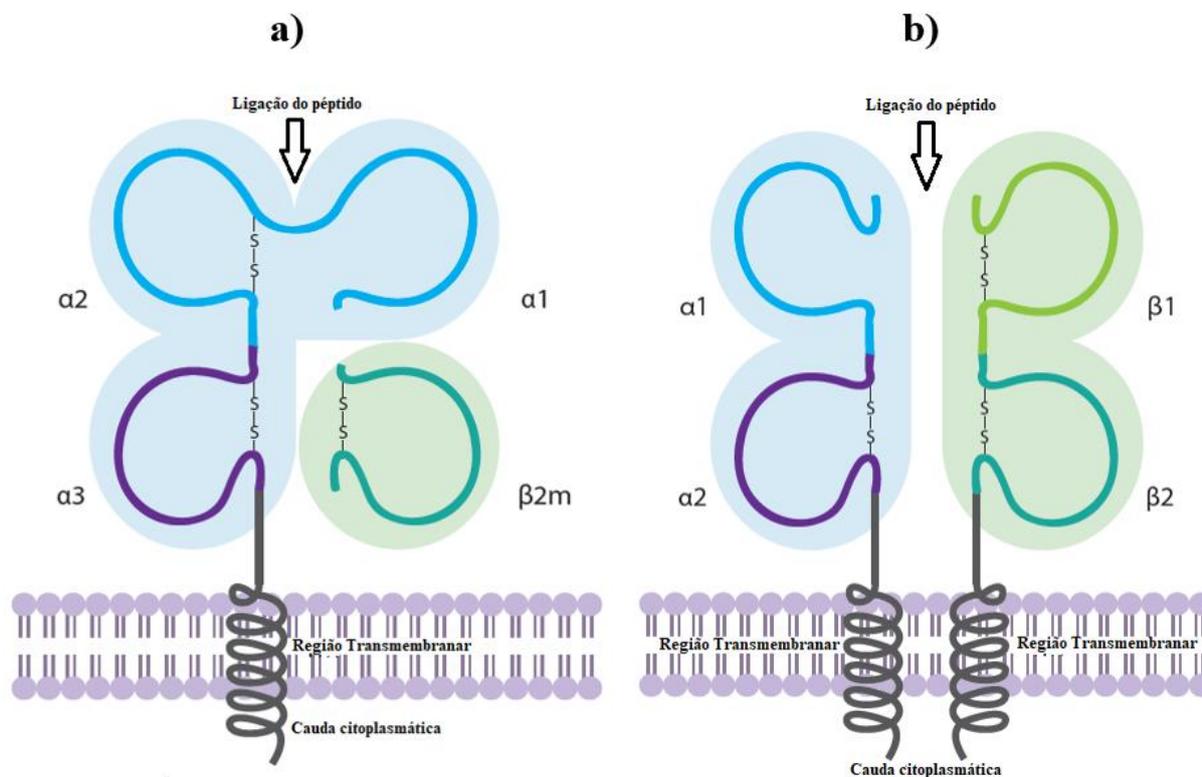


Figura 3: a) Estrutura das moléculas de HLA-I. b) Estrutura da molécula de HLA-II (**Anaya *et al.*, 2013**).

As vias de apresentação de antígenos são essenciais para monitorização dos compartimentos celulares com agentes patogénicos alojados, reportando a sua presença às células T. Os antígenos, originados endogenamente no citosol, são apresentados às células T-CD8⁺ como péptidos ligados a moléculas HLA-I. Esses linfócitos reconhecem-nos e eliminam as células infetadas. Os péptidos obtidos de antígenos exógenos, que são digeridos nos compartimentos endocíticos dos macrófagos, células dendríticas e células B, são apresentados às células T-CD4⁺ via HLA-II (**Rock, 2016**).

As moléculas HLA-I apresentam péptidos endógenos provenientes de proteínas sintetizadas por células que expressam sequências mutantes, proteínas microbianas (provenientes de infeções, por exemplo virais) ou proteínas polimórficas estranhas (provenientes de transplantes). Estes péptidos são apresentados às células T CD8⁺ que

destroem estas células “anormais”. Estas proteínas são ubiquitinadas, ficam marcadas para degradação pelo proteossoma, sendo digeridas pelas peptidases. No entanto, algumas conseguem escapar e são conduzidas até ao retículo endoplasmático (RE) através da proteína TAP (*Transporter associated with antigen processing*) que é uma glicoproteína inserida na membrana do RE. Em conjunto com as chaperonas tapasina, calreticulina, a proteína dissulfeto isomerase (ERp57) e moléculas vazias de HLA-I, formam um complexo de carga peptídico (PLC-*Peptide Loading Complex*). A TAP é constituída pela TAP1 e TAP2, e no conjunto é responsável pela translocação do péptido pelo lúmen do RE até às moléculas de MHC-I. Para além desta, também a tapasina desempenha importantes funções. A tapasina permite carregar o lúmen do RE com as moléculas de MHC-I sintetizadas pelos ribossomas e mantê-las num estado recetivo para ligação aos péptidos, auxiliando a ligação péptido-MHC I. Na maioria dos casos, os péptidos são demasiados longos e necessitam ser cortados pela aminopeptidase ERAP 1 na extremidade N-terminal de forma a se ligarem corretamente ao HLA-I. O primeiro passo deste processamento ocorre quando a tapasina se liga ao heterodímero de HLA-I e esta chaperona liga-se ao transportador TAP, estabilizando-o, criando assim a ponte de passagem dos péptidos no lúmen do RE até ao HLA-I. Por último, a tapasina liga-se à ERp57 dando estabilidade estrutural ao complexo de carga de péptidos e criando condições ótimas de ligação do péptido. A estabilização do complexo de carga leva à saída da molécula de HLA-I carregada com o péptido via RE e Complexo de Golgi até à superfície da célula, onde após exocitose ocorrerá a apresentação antigénica pelas células T CD8⁺. Os péptidos que não se ligam a HLA-I são transportados para o citosol e degradados. Alguns podem ser protegidos deste processo de degradação até ao momento em que a molécula de HLA-I está livre e possa ser carregada novamente com outro péptido (figura 4) (Anaya *et al.*, 2013, Hermann *et al.*, 2015 e Rock, 2016).

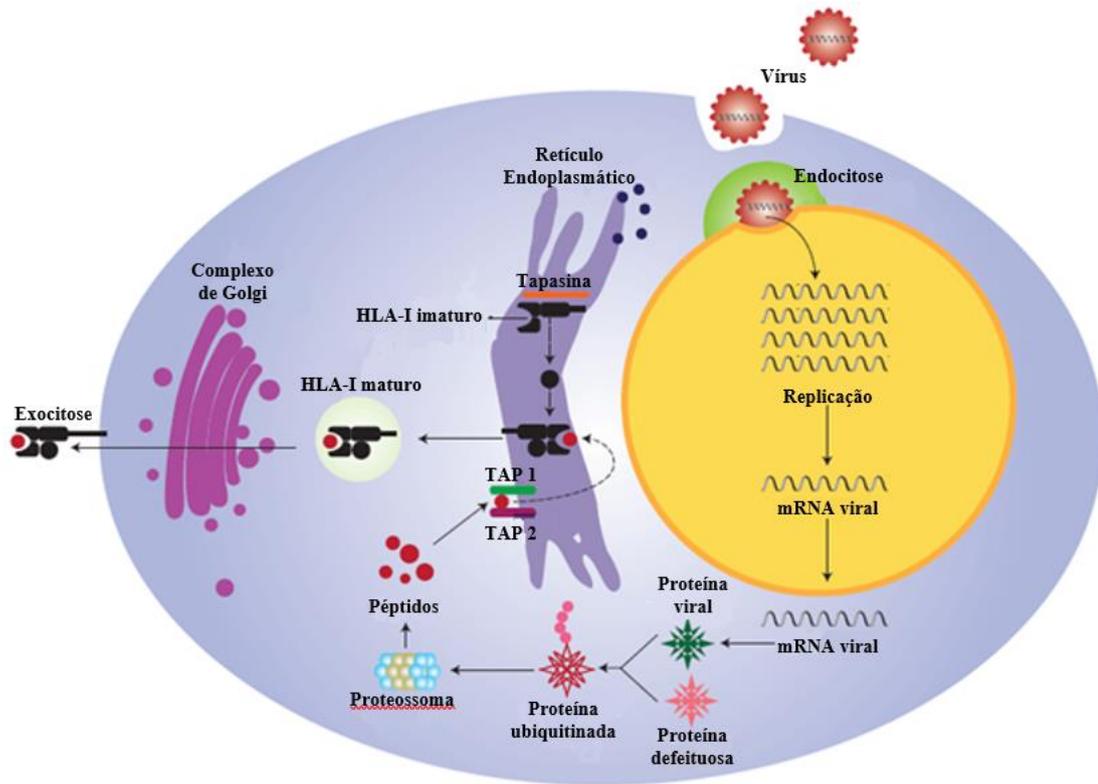


Figura 4. Apresentação antigénica por parte das moléculas de MHC-I (adaptado de Anaya *et al.*, 2013).

As moléculas HLA-II apresentam péptidos derivados de proteínas extracelulares e auto proteínas degradadas via endossomal. Durante a montagem da molécula de HLA-II no RE, esta associa-se com a cadeia invariante Ii, que funciona como pseudopéptido e preenche o sulco de ligação do péptido. Adicionalmente, a cadeia invariante permite uma saída eficiente do HLA-II do RE e a sua translocação via complexo de Golgi até ao compartimento endossomal designado MIIC (*MHC class II compartment*). Neste compartimento, a cadeia invariante é degradada por proteases endossomais, nomeadamente a catapsina L e catapsina S. No entanto um fragmento da cadeia invariante, CLIP (*Class II – associated Invariant chain Peptide*), é inacessível às proteases, logo permanece no sulco até ser trocado por um péptido adequado, através do auxílio da chaperona HLA-DM. Esta abre o sulco de ligação de forma a libertar péptidos de baixa afinidade como o CLIP, permitindo a ligação de péptidos de elevada afinidade. As moléculas de HLA-II carregadas com o péptido deslocam-se para a membrana plasmática, via transporte vesicular ou sob a forma de túbulos e dá-se a apresentação às células T CD4⁺ (figura 5) (Anaya *et al.*, 2013 e Rock, 2016).

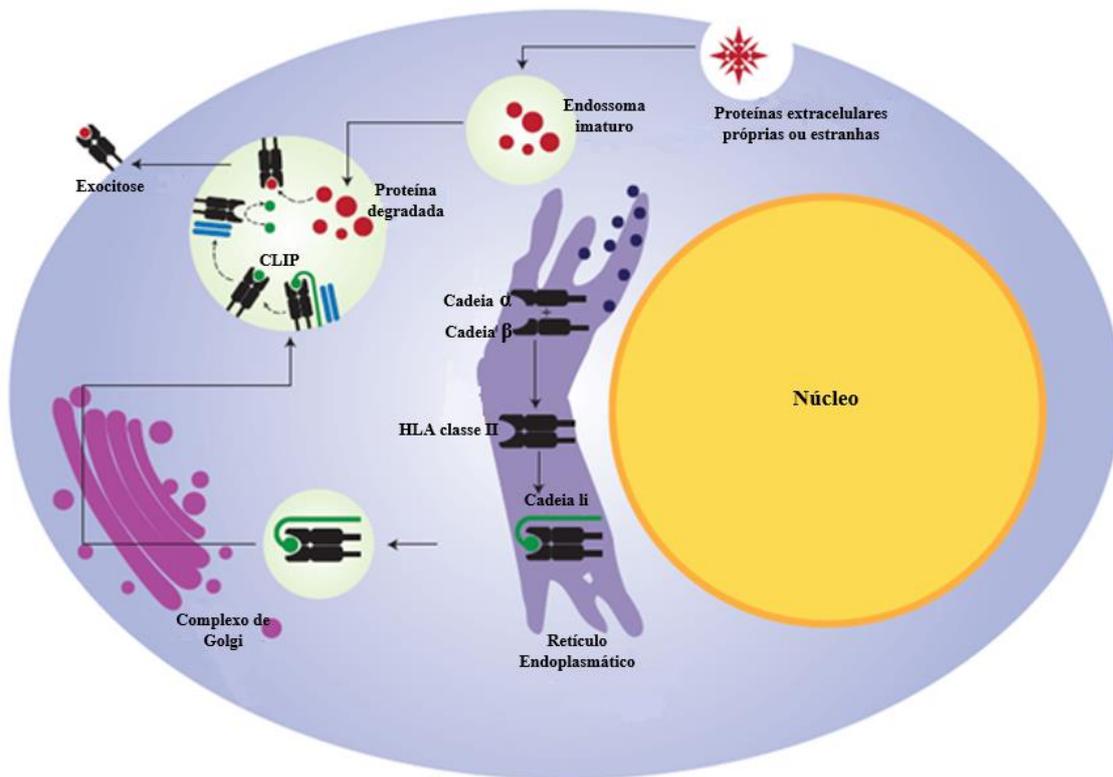


Figura 5. Apresentação antigénica por parte das moléculas de MHC-II (adaptado de Anaya *et al.*, 2013).

1.5. Nomenclatura do HLA

A elevada variação alélica nos *loci* HLA e a constante modernização das técnicas de sequenciação exigem um sistema de nomenclatura especializado (Torres e Moraes, 2011), regulado pelo Comité de Nomenclatura do sistema HLA da organização Mundial de Saúde (*World Health Organization Nomenclature Committee for Factors of the HLA System*). A base de dados IPD-IMGT/HLA tem como missão compilar os dados de novas sequências, que são alvo de uma criteriosa avaliação para a sua inclusão. Os alelos HLA possuem um número constituído por quatro conjuntos de dígitos tal como descrito na figura 6. Cada alelo começa com o nome do *locus* (ex: HLA-A, em que HLA corresponde ao prefixo e “A” ao gene, o hífen separa o prefixo do gene), seguido de dois conjuntos de números. A separar o nome do *locus* dos dígitos é usado um asterisco “*”, e cada conjunto de números é separado por dois pontos “:”. O primeiro conjunto corresponde ao grupo alélico, sendo o que define um alelo em baixa resolução. O segundo conjunto corresponde a um subtipo desse alelo, traduzindo-se numa proteína específica, sendo o que define um alelo em alta resolução. O terceiro conjunto corresponde a polimorfismos na região codificante que resultam em substituições sinónimas e o quarto conjunto é usado para revelar diferenças nas regiões não codificantes (intrões ou UTRs- regiões não traduzidas).

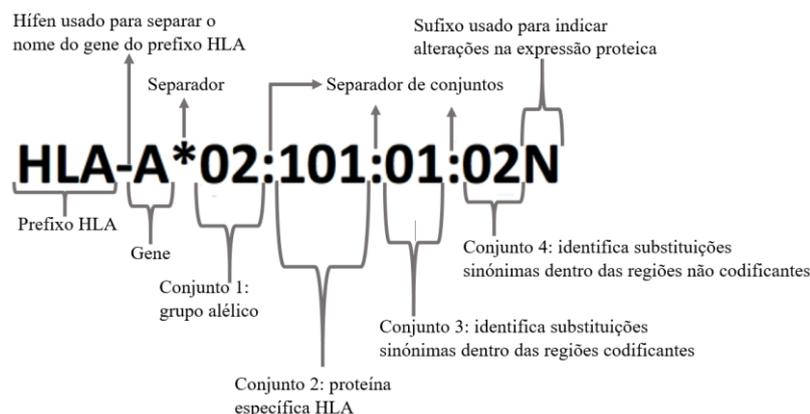


Figura 6. Nomenclatura de um alelo HLA (Adaptado de <http://hla.alleles.org/nomenclature/naming.html>).

Adicionalmente é possível incluir sufixos com informação quanto à expressão proteica, como: “N”- alelos que não são expressos ou alelos nulos; “L”- baixa (*low*) expressão à superfície da célula quando comparados com níveis normais; “S”- alelos cuja proteína é expressa sob forma secretada (*S-secreted*) e solúvel e não está na superfície da célula; “A”- se há dúvidas sobre a forma como a proteína é expressa, sendo considerada aberrante (*A-aberrant*); “Q”- expressão do alelo questionável (*Q-questionable*). Estas regras de nomenclatura estão ilustradas na figura 6 (Paunic *et al.*, 2012 e <http://hla.alleles.org/nomenclature/naming.html>).

1.6. HLA e doenças humanas

O sistema HLA foi descoberto há mais de 60 anos, tendo contribuído para a compreensão do sistema imunitário e patogénese de algumas doenças (Bodis *et al.*, 2018). As primeiras associações entre o HLA e determinadas doenças reforçaram a ideia já existente, da complexidade desta região e extenso polimorfismo. Alguma destas doenças estão associadas ao sistema imunológico, outras relacionam-se com doenças infecciosas e até reações adversas a fármacos (Howell, 2014). Ao longo desta revisão serão dados exemplos de doenças associadas com o HLA.

1.6.1. Doença Celíaca

A doença celíaca é uma doença autoimune que, aquando a ingestão de glúten, leva à produção de autoanticorpos e atrofia das vilosidades do intestino delgado. Esta reação inflamatória provoca danos no intestino e consequente má absorção de nutrientes, diarreias crónicas e perda de peso (Howell, 2014 e Mannon, 2018).

Os alelos que manifestam a principal associação do HLA com esta doença são HLA-DQA1*05:01/DQB1*02:01 e HLA-DQA1*02:01/DQB1*02:02, conhecidos como DQ2 e DQ8, respetivamente. Estima-se que aproximadamente 90% dos pacientes com esta doença autoimune expressam as moléculas HLA-DQ2 e os 10% restantes expressam principalmente a molécula HLA-DQ8 (**Monos e Winchester, 2018**).

A genotipagem HLA é extremamente importante quando há suspeita de doença celíaca, principalmente em crianças em que no caso de resultados positivos para HLA-DQ2 e DQ8 evita-se a realização de uma biópsia (**Howell, 2014**). Em adultos, a genotipagem HLA-DQ2 e DQ8 também pode ser útil, no entanto de forma mais limitada já que a presença destes alelos não são regra para se desenvolver a doença, já que há muitos indivíduos que os possuem e não desenvolvem a doença (**Mannon, 2018**).

Apesar de um teste negativo ser preditivo para excluir a doença, um resultado positivo apenas demonstra suscetibilidade genética. A regra dos quatro em cinco pode ser aplicada a todas as faixas etárias para o diagnóstico da doença celíaca. Nesta regra quatro dos cinco critérios têm de ser observados: sintomas típicos, aumento significativo dos anticorpos da doença celíaca, genotipagem HLA-DQ2 ou HLA-DQ8, biópsia com resultado característico e resposta a uma dieta livre de glúten (**Bodis et al., 2018**).

1.6.2. Diabetes Mellitus tipo 1

A diabetes mellitus tipo 1 (T1DM), conhecida como diabetes mellitus dependente de insulina, é uma doença autoimune que afeta a produção de insulina pelas células β , em consequência da destruição destas dentro das ilhotas de Langerhans, no pâncreas, após um processo inflamatório que leva à deficiência de insulina (**Monos e Winchester, 2018**).

Mais de 90% dos indivíduos caucasianos com esta doença, apresentam alelos HLA de alto risco, são estes o HLA-DR3 (DRB1*03:01/DQA1*05:01/DQB1*02:01) e -DR4 (DRB1*04:01 ou *04:04 com DQA1*03:01/DQB1*03:02). Além destes existem alelos que conferem proteção à doença, são estes o HLA-DQ6 que codifica os alelos DQA1*01:02/DQB1*06:02 e o HLA-DR15 que codifica o alelo DRB1*15:01 (**Harrison, 2018**). Os alelos de proteção *versus* predisposição diferem no aminoácido na posição 57 da cadeia β da molécula DQB1. Se o aminoácido codificado por ambos os alelos for o ácido aspártico (Asp) é pouco provável que haja desenvolvimento da doença. Foi demonstrado que a presença de alanina (Ala) ou serina (Ser) na posição 52, confere suscetibilidade para a doença (**Anaya et al., 2013 e Howell, 2014**).

Apesar destas associações, a genotipagem HLA-DQ tem pouca utilidade no diagnóstico. No entanto, foi proposto que a genotipagem de vários *loci* adicionais em crianças com genótipos HLA de alto risco e com história familiar de T1DM, pode ser considerada uma estratégia de intervenção precoce (**Howell, 2014**). No anexo B observam-se os haplótipos HLA-DRB1, -DQA1 e -DQB1 associados à suscetibilidade e proteção contra a doença.

1.6.3. Espondilite Anquilosante

A espondilite anquilosante (EA) é uma doença que afeta a coluna e articulações sacroilíacas, sendo considerada uma artrite inflamatória crônica. Causa incapacidade significativa e está associada a outras patologias como a artrite periférica, uveíte anterior, psoríase e doença inflamatória do intestino (DII) (**Robinson e Brown, 2013**).

Em 1973 a frequência de especificidade para o HLA-B27 foi de 95% em indivíduos com espondilite anquilosante. Diferentes alelos B27 têm diferentes forças de associação com a doença, sendo que os alelos HLA-B*27:02 e -B*27:05 demonstram o maior grau de associação. Embora a associação do EA ao HLA-B*27 esteja entre as associações genéticas mais fortes com uma doença comum, o mecanismo de ação permanece incerto. Para além disso suspeita-se que outros genes dentro e fora do HLA estejam associados com a doença (**Monos e Winchester, 2018**). Em quase todas as populações em todo o mundo, o HLA-B27 está fortemente associado à EA, estimando-se que cerca de 90% dos pacientes com ancestralidade europeia possuam o alelo HLA-B*27 (**Dashti et al., 2018**).

A genotipagem HLA-B*27 é um teste requisitado a nível laboratorial no diagnóstico da EA, em que a presença deste alelo é um marcador confirmatório útil para a iniciação do tratamento (**Monos e Winchester, 2018**).

1.6.4. Hipersensibilidade ao Abacavir

O abacavir é um medicamento antiretroviral com seletividade para o VIH (vírus da imunodeficiência humana), inibindo a atividade do retrovírus. Este é muito usado no tratamento de pacientes com VIH. No entanto, muitos pacientes manifestam reações adversas, tendo sido demonstrado que a reação de hipersensibilidade destes pacientes deve-se à presença do alelo HLA-B*57:01, em consequência da ativação de células T CD8⁺ específicas capazes de reconhecer o abacavir quando apresentado pela molécula HLA-B*57:01 (**Monos e Winchester, 2018**). Esta reação de hipersensibilidade é severa e potencialmente fatal com

sintomas típicos de febre, erupções cutâneas, vômitos e falta de ar que se manifestam logo nas primeiras 42h de tratamento (**Dean, 2015**).

Tal como outros alelos HLA, a frequência alélica varia entre diferentes populações. No entanto, é de ressaltar que a associação alélica e a hipersensibilidade ao abacavir é bastante forte e de grande utilidade clínica. Apesar de nem todos os pacientes com este alelo positivo desenvolverem hipersensibilidade, é recomendada a genotipagem do alelo HLAB*57:01 em pacientes VIH-1 positivos, antes de iniciarem terapia com abacavir. Pacientes com este alelo (homozigotia ou heterozigotia) não podem ser tratados com abacavir (**Dean, 2015**).

1.7. HLA e transplantação

A compatibilidade HLA é de extrema importância na transplantação de órgãos, especialmente renal e medula óssea, definindo qual o dador com maior compatibilidade. Nos casos de fígado e coração, a compatibilidade do *locus* DR é a mais importante, no entanto encontram-se outras barreiras externas como tempos de isquemia e a disponibilidade de dadores (**Mahdi, 2013**).

A incompatibilidade HLA pode dar-se a nível antigénico ou alélico. O primeiro caso deve-se a substituições de aminoácidos tanto na ligação peptídica como nas regiões de reconhecimento das células T. Ao nível alélico caracteriza-se por substituição de aminoácidos apenas nas regiões de ligação peptídica. Num transplante, as moléculas HLA do dador são reconhecidas pelo sistema imunológico do recetor por mecanismos diretos e indiretos de alo-reconhecimento desencadeando uma resposta aloimune (**Mahdi, 2013**). O primeiro envolve o reconhecimento direto do complexo péptido-HLA do tecido do dador pelas células T do recetor, através das semelhanças estruturais entre as moléculas HLA do dador que permitem que o TCR (recetor das células T) interaja com o complexo péptido-HLA. O segundo envolve a apresentação “indireta” dos antígenos HLA do dador processados pelas células APC do recetor, criando péptidos apresentados pelas moléculas HLA do recetor para as células T do recetor (**Monos e Winchester, 2018**).

Utilizando agentes imunossupressores e terapias apropriadas, a ativação de células T pelas moléculas de HLA do dador após o transplante clínico pode ser controlada. No entanto, o principal problema a longo prazo é a presença de anticorpos específicos de dadores que se desenvolvem contra antígenos HLA incompatíveis, constituindo um grande desafio e necessidade de monitoramento contínuo (**Monos e Winchester, 2018**).

A compatibilidade HLA entre dador e recetor mostrou um efeito significativo na aceitação do enxerto. Os efeitos da compatibilidade HLA-B e –DR são os que têm maior impacto na rejeição do enxerto, sendo que efeitos das incompatibilidades HLA-DR são o mais importante nos primeiros 6 meses após transplante (Mahdi, 2013).

1.8. Métodos para genotipagem do HLA

O elevado polimorfismo genético do HLA permite que o sistema imunológico lute contra vários agentes patogénicos (Paunic *et al.*, 2012) e exige avanços na tecnologia e padronização de técnicas moleculares de genotipagem (Bontadini *et al.*, 2012). A genotipagem clássica do HLA era realizada através de testes serológicos com anticorpos específicos dos antígenos, permitindo identificar variantes de proteínas HLA na superfície das células. A serologia foi sendo substituída por métodos genéticos, pois não era capaz de identificar alelos específicos (Paunic *et al.*, 2012). Os testes genéticos são mais específicos pois os reagentes, como *primers* e sondas, têm por base uma sequência de nucleótidos específica; são mais flexíveis, pois são desenvolvidos novos reagentes considerando os alelos adicionados à base de dados e os testes podem ser realizados em vários níveis de resolução; são mais robustos, já que não requerem um tipo específico de célula ou viabilidade e é pouco afetado pela saúde do paciente (Bontadini *et al.*, 2012).

1.8.1. Genotipagem Serológica

Nesta técnica são usadas placas comerciais contendo anticorpos para uma multiplicidade de alelos HLA conhecidos. Os linfócitos recetores são ensaiados nos poços da placa conjuntamente com o complemento e o corante. Os anticorpos anti-HLA ligam-se aos antígenos do recetor, o complemento é ativado desencadeando-se assim a morte celular. A genotipagem serológica do recetor é determinada pela adição do corante que em consequência da morte celular consegue entrar nas células, sendo visível uma coloração avermelhada através de microscopia de contraste. Estes poços são denotados como positivos para determinado alelo do HLA. A principal vantagem desta técnica é a obtenção de resultados num curto espaço de tempo, no entanto, a identificação exponencial de novos alelos levou a que os soros não tivessem tanta especificidade como pretendido, uma vez que não detetavam pequenas diferenças nos aminoácidos das proteínas HLA. Assim, esta técnica acabou por cair em desuso (Bontadini *et al.*, 2012).

1.8.2. Genotipagem Genética

A genotipagem genética permite uma resolução alélica específica, isto é, alterações nucleotídicas de HLA definem alelos diferentes que são detetados através das técnicas baseadas na amplificação *in vitro* do gene específico, variando entre si nos métodos de detecção pós-amplificação. São designadas: PCR-SSP, PCR-SSO, rPCR-SSO, PCR-SBT e NGS (Bontadini *et al.*, 2012).

1.8.2.1. PCR-SSP (*Sequence Specific Primer*)

A técnica PCR-SSP foi desenvolvida nos anos 90 como uma alternativa à genotipagem serológica para os genes HLA-DR, e em 1995, Bunce *et al.* aplicaram-na à genotipagem dos alelos HLA-A, -B, -C, DRB1, -DRB3, -DRB4, -DRB5 e -DQB1 (Dunckley, 2012).

Esta técnica usa *primers* específicos complementares de determinada sequência alélica do HLA ou a um grupo de alelos semelhantes. Tendo em conta a especificidade dos *primers*, a incompatibilidade de uma única base irá inibir uma reação não específica, e deste modo apenas se estes forem completamente emparelháveis com a amostra teste é que se dá a amplificação desta. Para as tipagens em genes HLA-A, -B e -C, a maioria dos polimorfismos conhecidos estão nos exões 2 e 3 dos genes e desta forma os *primers* são selecionados de forma a detetar os polimorfismos nestes exões. Para os genes HLA-DR, -DQ e -DP, os *primers* são desenhados para detetar polimorfismos mais abundantes no exão 2. Este método pode ser usado para genotipagem de baixa ou alta resolução. Um *kit* para a técnica de PCR-SSP do HLA-I e HLA-II possui vários conjuntos de *primers* selecionados de forma a haver emparelhamento entre estes e a região do gene que contém o alelo que se pretende amplificar. Se os alelos estiverem presentes, irão ser detetados os amplicões destes através de uma eletroforese em gel de agarose. As reações positivas são aquelas que contêm, para além do controlo interno, bandas específicas com determinado peso molecular (Bontadini *et al.*, 2012 e Dunckley, 2012).

A grande vantagem é a rapidez de execução, sendo especialmente usada em amostras urgentes como transfusões ou transplantação de órgãos. Quando é necessária a genotipagem de muitas amostras torna-se obrigatório recorrer a outro método, devido ao elevado número de poços utilizados por amostra. Uma desvantagem manifestada ao longo dos anos, com o aumento do número de alelos HLA, é a dificuldade em distinguir determinados alelos numa única reação de PCR (*Polymerase Chain Reaction*). De forma a resolver ambiguidades, novos

conjuntos de *primers* podem ser adicionados para novos alelos, no entanto, o crescimento exponencial dos alelos incrementa as ambiguidades, tornando-se necessário recorrer a métodos mais específicos (Dunckley, 2012).

1.8.2.2. PCR-SSO (*Sequence Specific Oligonucleotide*)

A técnica PCR-SSO foi uma das primeiras a ser desenvolvida para a genotipagem do HLA com base no PCR. Para esta técnica é necessário extrair o DNA e escolher um conjunto de *primers*, marcados com biotina, que seja específico de determinado *locus*. Estes *primers* são selecionados a partir de uma região conservada de determinado *locus* e têm a capacidade de amplificar todos os alelos conhecidos desse *locus*. Para a realização desta técnica são necessárias sondas oligonucleotídicas, com cerca de 20bp, o que assegura especificidade e que a torne instável caso ocorra uma incompatibilidade. Uma sonda oligonucleotídica reconhece muitos alelos com a mesma sequência, deste modo é necessário definir um alelo ou grupo de alelos. Estas são desenhadas de forma que o polimorfismo que se pretende detetar se encontre no meio da sequência oligonucleotídica aumentando a força de ligação entre a sequência de DNA e a sequência oligonucleotídica. Assim, após amplificação, o produto de DNA é transferido para uma membrana de nylon carregada positivamente, para que o DNA amplificado, que possui carga negativa, se ligue fortemente. O DNA amplificado e biotilado hibrida com as sondas e esta só se verifica quando a sequência de DNA amplificada é complementar ao oligonucleotídeo. Assim sendo, oligonucleotídeos que não hibridam permanecem em solução e são eliminados através de sucessivas etapas de lavagem. De seguida é adicionada solução de streptavidina conjugada com uma enzima, por exemplo a fosfatase alcalina, que se liga aos produtos sonda-sequência biotilada. Por fim é adicionada a solução de substrato que catalisa a formação de um complexo colorido permitindo a identificação dos produtos de PCR ligados à sonda. As reações de cada amostra são guardadas e a genotipagem é feita recorrendo a *software* adequado. Esta técnica é muito prática para um grande volume de amostras, no entanto acarreta desvantagens para um volume baixo de amostras, pois apesar de cada indivíduo ser tipado para cada *locus*, é necessário um grande número de sondas para detetar todas as especificidades. Outro aspeto negativo é o facto de não distinguir polimorfismos *cis* e *trans*, logo não consegue resolver todas as ambiguidades (Bontadini *et al.*, 2012 e Dunckley, 2012).

1.8.2.3. *Reverse* PCR-SSO (Reverso do anterior)

É uma técnica alternativa ao PCR-SSO que usa o mesmo procedimento de amplificação, painel de sondas e suporte de hibridação, no entanto as sondas estão individualmente ligadas às membranas para cada *locus* HLA e a amplificação do DNA é feita com *primers* específicos. Tal como na técnica anterior ocorre hibridação, no entanto nesta técnica as sondas oligonucleotídicas já se encontram pré-ligadas à membrana. Os amplicões que não se ligaram são eliminados e a reação é detetada por quimiluminescência ou colorimetria. Apesar de ser adequada à genotipagem de um reduzido número de amostras, acarreta os mesmos problemas de resolução de ambiguidades (**Bontadini et al., 2012**).

O *Luminex* é uma plataforma que recorre ao método reverse PCR-SSO. Usa microesferas codificadas com fluorescência como suporte sólido para imobilizar sondas oligonucleotídicas. O DNA alvo é amplificado por PCR e hibridado no *array* com as sondas. Todo o processo ocorre num único poço numa placa de PCR com 96 poços, permitindo o processamento de 96 amostras simultaneamente. Este sistema substitui as membranas usadas em PCR-SSO por microesferas de poliestireno as quais se conseguem ligar a sondas oligonucleotídicas específicas. Neste método o DNA alvo é amplificado com *primers* específicos e biotinilados e o produto de PCR é desnaturado e hibridado com as sondas nucleotídicas. A adição da solução SAPE (Streptavidina conjugada com a R-ficoeritrina - PE) permite a sua ligação à biotina nos híbridos PCR biotinilado – sonda específica. A intensidade de fluorescência em cada microesfera é detetada no analisador de fluxo –*Luminex* (**Bontadini et al., 2012 e Dunckley, 2012**). A plataforma *Luminex*, *LABScan*, quantifica a intensidade do sinal de fluorescência originado pelo SAPE. Os sinais são processados para cada reação e a genotipagem é obtida com recurso a *software* específico (*Fusion*). Cada mistura de sonda contém um ou mais oligonucleotídeos que reagem com todos os alelos de um *locus* HLA específico, funcionando como controle da reação (**Testi e Andreani, 2015**).

Na última década, o método baseado em *Luminex* para genotipagem de baixa resolução foi bem estabelecido, dando-se particular interesse ao uso da tecnologia “*special probe*” que permite a resolução de ambiguidades *cis/trans*, caso contrário, seriam necessários mais ensaios. Para obter alta resolução foram desenvolvidas duas abordagens *Luminex*, comercialmente disponíveis, com diferenças consistindo nas etapas de amplificação e desnaturação. Um desses sistemas, o *OneLambda Labtype® High Definition* é capaz de resolver ambiguidades, muitas vezes eliminando a necessidade de testes pela tecnologia SBT (**Testi e Andreani, 2015**). Apesar do avanço tecnológico, algumas limitações vão

permanecendo como os polimorfismos fora da região amplificada ou nos locais de ligação dos *primers*, o que leva a que novas sondas sejam desenvolvidas (**Bontadini et al., 2012**).

1.8.2.4. PCR-SBT (Sequencing Based Typing)

Esta técnica consiste na adição de nucleótidos modificados, os didesoxinucleotídeos (ddNTPs) que impedem o crescimento de um fragmento de DNA quando a sua replicação pela DNA polimerase (**Kallon, 2011**).

A técnica de Sanger necessita primariamente de uma reação de PCR para amplificação do *locus* de interesse antes da etapa de sequenciação. Após a amplificação seguindo-se as etapas de purificação, os produtos de PCR em cadeia simples são hibridizados com um *primer* específico para cada locus e para cada um dos exões mais relevantes. São adicionados marcadores fluorescentes com diferentes cores para cada tipo de ddNTP permitindo a distinção das cadeias truncadas. Assim a DNA polimerase vai adicionando os nucleótidos à cadeia, até à incorporação de um ddNTP interrompendo a síntese. No final de vários ciclos é praticamente garantido que um ddNTP terá sido incorporado em todas as posições do DNA alvo, obtendo-se no mesmo tubo vários fragmentos de DNA de diferentes tamanhos, terminando em cada uma das posições nucleótidos do DNA original. Estes fragmentos são submetidos a uma eletroforese capilar, passando por um tubo muito fino (capilar) contendo uma matriz de gel. Durante este percurso o laser excita os fluorocromos dos ddNTPs, cuja fluorescência é captada numa câmara, convertendo o sinal digital num sistema de cores identificativo dos 4 nucleotídeos presentes: adenina (verde), citosina (azul), timina (vermelho) e guanina (preto) (**Bontadini et al., 2012 e Kallon, 2011**).

A interpretação dos resultados é feita num *software* específico que alinha as sequências resultantes e permite ao operador compará-las com a sequência consensus disponível numa base de dados. A partir desta comparação é possível determinar o alelo presente obtendo-se assim a genotipagem em alta resolução (**Allcock, 2012**).

Esta técnica deteta diretamente as sequências nucleotídicas permitindo uma determinação mais exata de determinado alelo e até mesmo reconhecer novos alelos. Assim, a precisão é a maior vantagem desta e para além disso é onde se atinge maior resolução alélica, constituindo assim uma das técnicas preferências na transplantação de progenitores hematopoiéticos. Em contrapartida, é uma técnica limitada pelas ambiguidades. Para ultrapassar esta limitação, quando há ambiguidades devem-se realizar amplificações adicionais com conjuntos de *primers* específicos de forma a amplificar um alelo ou um grupo

de alelos separadamente. Mais uma vez, o incremento de novos alelos nas livrarias internacionais torna necessário a sequenciação noutras exões ou intrões que nem sempre estão disponíveis num determinado *kit* (Bontadini *et al.*, 2012).

1.9. Tecnologia NGS e o HLA

O processo de identificação dos alelos HLA evoluiu desde o método serológico até ao PCR-SSO, -SSP e SBT. Nos dias de hoje, a genotipagem é bastante eficiente, estando otimizada e sendo capaz de detetar e identificar os diferentes *loci*. A genotipagem com base nos seis *loci* mais importantes da transplantação (HLA-A, -B, -C, -DRB1, -DQB1 e -DPB1) oferece os melhores resultados, com menores níveis de rejeição, tanto em órgãos sólidos como em medula óssea. Em 2005, com o lançamento do sequenciador 454/GS FLX pela RocheDiagnostics, surgiu uma nova abordagem de sequenciação, designada *Next Generation Sequencing* (NGS). Desde então deu-se uma grande evolução dos sequenciadores e consequente evolução da investigação, sendo possível a sua aplicação na rotina laboratorial (Santis *et al.*, 2013). A técnica NGS difere dos outros métodos tradicionais, pois permite a sequenciação de todo o genoma, através de reações de sequenciação a partir de um grande número de reações clonais em paralelo. Devido à extrema complexidade do HLA, o NGS constitui uma nova abordagem com estratégias de genotipagem e algoritmos de *softwares* específicos (Gabriel *et al.*, 2014).

Várias tecnologias NGS foram desenvolvidas, permitindo obter uma genotipagem HLA com elevado rendimento e alta resolução. Os níveis de expressão dos genes de HLA podem ter papéis cruciais na patogénese de várias doenças. Deste modo, é necessário detetar variantes de um único nucleótido (SNVs- *single nucleotide variants*), inserções e deleções (*Indels*) localizadas fora dos exões. A sequenciação completa do HLA, incluindo regiões reguladoras funcionais, pode levar à descoberta de novos alelos associados a patologias diversas e associados a efeitos adversos a fármacos, assim como esclarecer os níveis de expressão dos genes que afetam esses processos (Hosomichi *et al.*, 2015).

Para genotipagem NGS é necessário preparar o DNA, isolando o alvo do resto do genoma. O alvo da genotipagem pode ser o HLA por inteiro, genes específicos ou apenas certos exões. O primeiro é frequentemente usado em estudos de genética de populações, associação com doenças ou identificação de novos alelos HLA, enquanto o segundo e o terceiro são usados mais na rotina laboratorial. A técnica de PCR permite isolar a sequência alvo, através do uso de *primers* específicos, e para além disso, permite a sua amplificação, o

que possibilita o uso de pequenas quantidades de DNA. A dificuldade desta técnica consiste no desenho de *primers* que amplifiquem a grande maioria dos alelos HLA conhecidos, com boa garantia de hibridação. A maioria das bases de dados apenas possui informação sobre os alelos HLA mais comuns. Cada plataforma NGS possui uma química de sequenciação específica e dividem-se: em sequenciadores de segunda geração, ou seja, de leitura de tamanho médio, entre 250-1000bp, como o *Illumina*, *Roche* e o *Ion Torrent* da *Thermo Fisher Scientific*; e sequenciadores de terceira geração com comprimento de leitura maior que 10 kb, como o SMRT da *Pacific Biosciences* e o *Oxford Nanopore™ Technologies*. Os sequenciadores de terceira geração são capazes de analisar uma única molécula de DNA, ou seja, sem necessidade de amplificação prévia da sequência. Já no caso de sequenciadores de segunda geração as amostras têm de ser amplificadas antes da sequenciação (**Carapito et al., 2016**).

As metodologias de sequenciação NGS possuem elevado desempenho e estão cada vez mais presentes no cotidiano laboratorial. É possível observar-se ao longo dos anos um aumento na qualidade dos resultados, bem como no tamanho das sequências lidas. Adicionalmente, a diminuição na quantidade de amostra necessária, no custo por base sequenciada e a grande quantidade de dados obtidos num curto espaço de tempo têm vindo a revolucionar a pesquisa em diversas áreas (**Zolet et al., 2017**).

1.9.1. *Roche454 System*

A tecnologia *Roche454* tem por base a pirosequenciação de fragmentos de DNA inicialmente amplificados por clonagem em esfera, por PCR de emulsão (**Carapito et al., 2016**). Milhões de reações clonais de PCR de emulsão realizam-se em gotículas óleo-água emersas com os respetivos reagentes de PCR necessários e as esferas (**Gabriel et al., 2014**).

As reações de pirosequenciação ocorrem em placas de Pico Titer Plates™ com poços, sendo que cada um deles aloja apenas uma esfera. À medida que o nucleótido é incorporado pela DNA polimerase há libertação de pirofosfato, que através de reações enzimáticas converte o sinal em forma de luz observada pela câmara. A localização da esfera, a sequência e a emissão de luz é calculada para um fluxograma, que corresponde à leitura de uma sequência derivada de moléculas de DNA ligadas à esfera (**Gabriel et al., 2014**). Todo este processamento é possível ser observado no anexo C.

A grande vantagem desta plataforma é a possibilidade de leituras até 1000bp, podendo abranger regiões com dois exões. Os principais erros detetados devem-se a inserções e

deleções em consequência da presença de regiões com homopolímeros. A identificação do tamanho dos homopolímeros deve ser determinada pela intensidade da luz emitida pelo pirosequenciamento, assim sinais com intensidade muito alta ou muito baixa levam a sub ou superestimação do número de nucleótidos, ou seja, o aumento de sinal não é proporcional, levando a erros de identificação (**Kchouk et al., 2017**). Em comparação com a sequenciação de Sanger, a redução das ambiguidades é de 93,5%. O baixo rendimento e elevado custo levou a que a *Roche* tenha descontinuado os sequenciadores em 2016 (**Carapito et al., 2016**).

1.9.2. SOLiD (*Sequencing by oligonucleotide ligation and detection*)

Esta tecnologia NGS baseia-se num equipamento de sequenciação por hibridação e ligação. Aqui a amplificação é feita por *colonies*, ou seja, sequências de DNA clonadas e amplificadas a partir de uma única molécula de PCR com uma matriz de poliacrilamida imobilizada numa simples lâmina de microscópio. Tal como demonstrado no anexo D, neste método os fragmentos de interesse são amplificados na superfície de uma esfera magnética com 1 µm em PCR de emulsão. Devido à matriz de poliacrilamida a difusão dos fragmentos amplificados é retardada e assim as moléculas amplificadas permanecem muito próximas da sequência molde, ficando disposta de forma paralela na lâmina. As esferas com as amplicões são depositadas numa lâmina e de seguida há hibridização dos *primers* com a sequência adaptadora, comum a todas as sequências amplificadas, do fragmento. Assim, inicia-se então a sequenciação por ligação. Para além dos *primers*, são necessárias DNA ligases fornecidas juntamente com combinações específicas de pequenas sondas marcadas com fluorescência. Estas sondas possuem um local de ligação, a primeira base, um local de clivagem, correspondente à quinta base e quatro corantes fluorescentes ligados à última base). Assim, a ligase reconhece a sonda marcada com fluorescência que hibridou com a molécula de DNA molde e faz a ligação desta com o *primer* adjacente emitindo um sinal de fluorescência que é detetado, convertendo numa sequência nucleotídica específica. Após emissão do sinal, ocorre clivagem do fluoróforo, iniciando-se novo ciclo de ligação. Estes ciclos repetem-se a partir das sequências das sondas com tamanhos diferentes de forma a toda a sequência apresentar uma boa cobertura. Nesta metodologia, o DNA pode ser clivado em fragmentos até ~100bp e ligado aos adaptadores específicos para o PCR em emulsão ((**Liu et al., 2012** e **Zolet et al., 2017**)).

A grande vantagem desta plataforma é a elevada precisão, pois cada base é lida duas vezes, enquanto a desvantagem são leituras relativamente curtas e tempos de execução

longos. O principal tipo de erro é nas substituições e para além disso são frequentes erros devido ao ruído durante o ciclo de ligação, e consequentemente identificação errada das bases. A sequenciação *SOLiD* já foi bastante utilizada em diversos projetos: resequenciação de genomas, análises de transcritomas e sequências expressas, descoberta de miRNAs (micro RNAs) e outros pequenos RNAs, etc. Este tipo de sequenciação auxiliou na revolução das análises genéticas, através da redução de custos e aumento considerável da quantidade de dados gerados numa só reação (**Kchouk et al., 2017 e Zolet et al., 2017**).

1.9.3. *Ion Torrent PGM (Personal Genome Machine)*

A tecnologia *Ion Torrent PGM* é semelhante à tecnologia *Roche454*, no entanto não utiliza iões marcados com fluorescência nem sistemas óticos de deteção, logo apresenta como vantagem uma redução de possíveis fontes de ruído na deteção da fluorescência resultante da incorporação de um dado nucleótido (**Zolet et al., 2017**). O princípio baseia-se na medição dos iões de hidrogénio que são libertados à medida que ocorre a incorporação de nucleótidos pela DNA polimerase. Este sequenciador possui um chip semicondutor com milhões de poços e a sequenciação inicia-se com a fragmentação de uma sequência de DNA em múltiplos fragmentos. Cada poço do chip contém uma única partícula esférica onde se irá ligar um único fragmento de DNA. Para além disso será adicionado a cada um dos fragmentos, um *primer* seguindo-se a amplificação de cada um deles numa reação de PCR de emulsão, até preencher completamente a esfera. Segue-se então, um fluxo sequencial de cada um dos nucleótidos que são bombeados para o chip, e à medida que cada um é incorporado ocorre libertação de um ião de hidrogénio. Em consequência dessa libertação, dá-se um desvio no pH do poço e essa mudança é detetada e convertida num sinal digital. Esta variação do pH é proporcional ao número de nucleótidos adicionados, no entanto, tal como acontece com o sequenciador da *Roche454*, esta leitura é dificultada na presença de sequências repetitivas de DNA. Deste modo, mantém-se o problema de sequenciação de homopolímeros, em que uma alteração de sinal não pode ser encarada com precisão. Contrariamente, como vantagem, a tecnologia *Ion Torrent* é a que apresenta tempos de execução mais curtos ao nível das plataformas de segunda geração, com leituras médias de cerca de 400 bases em 3h. Para além disso, estas leituras constituem também uma característica atraente para a sequenciação HLA, uma vez que 400 bases é o comprimento correto para abranger um único exão HLA (**Carapito et al., 2016, Gabriel et al., 2014 e Kchouk et al., 2017**). Outra vantagem é a reduzida taxa de erro, sendo de 1 base por cada 100 sequenciadas. Contudo, o tamanho dos

fragmentos sequenciados e a quantidade de dados gerados não é muito grande, sendo necessária otimização desta técnica. Este sequenciador apresenta uma vasta gama de aplicações, como a sequenciação de genomas e transcritomas bacterianos, regiões alvo específicas, bem como detecção de variantes de DNA, como os SNPs (*Single Nucleotide Polymorphism*) (Zolet *et al.*, 2017).

1.9.4. *Illumina*

A tecnologia *Illumina* tem por base a sequenciação por síntese. Possui diversas plataformas, que vão desde o Miseq com tamanho de bancada para os grandes sistemas HiSeq oferecendo uma grande variedade de rendimento e produtos. Esta metodologia inicia-se pela preparação de bibliotecas (conjunto de fragmentos de DNA obtidos após a amplificação), onde o DNA é fragmentado aleatoriamente seguida de adição de adaptadores às suas extremidades. Os fragmentos ligados aos adaptadores são posteriormente amplificados por PCR, selecionados por tamanho e adicionados a uma fase sólida designada por células de fluxo (lâmina de vidro com canais de poliacrilamida) onde são capturados por complementaridade através de adaptadores específicos. Cada fragmento é amplificado em conjuntos clonais, designados *clusters* (conjuntos de genes), em ponte de fase sólida utilizando diversos *primers* imobilizados na célula de fluxo. Na amplificação por ponte, após a hibridação com o *primer* imobilizado na superfície há extensão deste fragmento pela DNA polimerase formando uma região de cadeia dupla. De seguida, esta é desnaturada e fica pronta para hibridar com outro *primer* complementar imobilizado ocorrendo assim a síntese de outra molécula. Durante a amplificação, as moléculas formam uma ponte, daí a denominação deste tipo de amplificação. Cada *cluster* irá abranger um único fragmento e corresponder a uma única leitura no final do procedimento, podendo obter-se até 1000 cópias idênticas densamente compactadas na célula de fluxo. Para assegurar que apenas as cadeias *sense* permanecem na lâmina, após a formação destes *clusters*, a sequência *antisense* de cada molécula molde é clivada e o adaptador complementar é bloqueado. Este processamento está descrito no anexo E. As sequências adaptadoras não irão apenas funcionar como ligantes entre a sequência e a célula de fluxo, mas também como *primers* para a amplificação em ponte. Assim como os outros reagentes de sequenciação também os quatro dNTPs marcados com fluorescência são adicionados e o processo de sequenciação consiste na monitorização da incorporação dos nucleótidos fluorescentes, sendo que em cada ciclo de polimerização apenas um nucleótido é incorporado em cada molécula. Este tipo de metodologia diminui a formação

de homopolímeros quando sequenciados em sequências repetitivas. Assim, a cada incorporação de um nucleótido é detetado um sinal específico de cada *cluster* pelo equipamento e o bloqueador da extremidade 3' da molécula nascente é removido para que o próximo nucleótido possa ser incorporado. Por fim, os nucleótidos não ligados são removidos por lavagem. De forma a facilitar a identificação das amostras, é frequente a introdução de pequenas sequências de DNA, designadas por “DNA *Barcoding*”, ligadas às sequências adaptadoras. Estas são incorporados durante a criação das bibliotecas e cada código corresponde especificamente a uma determinada amostra. A *Illumina* dispõe ainda de uma variedade de *kits* para diferentes comprimentos de leitura, taxas de rendimento e tempo de execução (). Esta tecnologia é a mais amplamente utilizada pela comunidade científica e a cada ano, as suas metodologias necessitam de otimizações resultando em novos *kits* e equipamentos adaptados a diferentes realidades técnicas. Atualmente existem no mercado seis equipamentos diferentes adaptados a esta metodologia (**Gabriel *et al.*, 2014 e Zolet *et al.*, 2017**).

As principais vantagens são a ausência de erros causados por homopolímeros, baixo custo e maior nível de precisão quando comparado com outras tecnologias no mercado. Como principal desvantagem, esta técnica apresenta um longo tempo de reação (**Carapito *et al.*, 2016**). A sequenciação do DNA pode englobar o genoma completo, partes do genoma, moléculas de RNA previamente convertidas em cDNA (DNA complementar), RNA total ou mRNA (RNA mensageiro), genes específicos, entre outros (**Zolet *et al.*, 2017**).

1.9.5. *Pacific Biosciences*

A tecnologia *Pacific Biosciences* usa o método SMRT (*single molecule real time sequencing*), com monitorização em tempo real da incorporação de cada um dos nucleótidos pela DNA polimerase, com sinal de fluorescência próprio (**Carapito *et al.*, 2016**). Nesta metodologia a DNA polimerase é específica (ϕ 29), pois foi necessário otimizá-la de forma a aumentar a sua afinidade pelos os nucleótidos fosfoligados. Cada nucleótido possui seis grupos fosfato ligados e o último está por sua vez ligado ao fluoróforo. A DNA polimerase encontra-se ligada a um detetor e catalisa a síntese do DNA complementar com base numa sequência molde. Este detetor possui apenas alguns nanómetros e é feito de um filme de metal depositado num extrato de vidro. Devido ao tamanho do seu orifício, o volume de reação é da ordem dos zeptolitros. Em consequência, quando um nucleótido é detetado pelo local ativo da DNA polimerase, é incorporado em apenas milésimas de segundos, seguindo-se da deteção de

um sinal de fluorescência momentâneo, que rapidamente decai com a libertação dos grupos fosfato que possuem o fluoróforo ligado (anexo F). Todos os nucleótidos marcados são libertados ao mesmo tempo na célula de sequenciação, mas apenas o nucleótido no local ativo da DNA polimerase se encontra próximo o suficiente para ser detetado, de forma que a presença dos restantes não interfere com a deteção de sinal (Zolet *et al.*, 2017).

As principais vantagens são a reduzida quantidade de reagentes e a rapidez da metodologia, já que não existem etapas de lavagem entre cada nucleótido incorporado nem de pré-amplificação. A desvantagem é a elevada taxa de erro, de 13%, dominada por erros de inserções e deleções. Esta tecnologia é usada para a sequenciação de genomas, transcritomas e epigenomas completos, conseguindo-se sequências de elevado tamanho, com uma média 1000bp podendo chegar até 10.000bp, sendo útil em aplicações relacionadas com o alinhamento de sequências de forma não ambígua e resolução de regiões complexas com sequências repetitivas, entre outros (Kchouk *et al.*, 2017 e Zolet *et al.*, 2017).

1.9.6. Oxford Nanopore Technologie

A *Oxford Nanopore Technologie* é a tecnologia mais recente e diferente das restantes, pois não depende do uso de DNA polimerases nem marcações fluorescentes, reduzindo os custos e aumentando a velocidade de processamento. Esta disponibiliza duas versões de sequenciadores: *GridION* e o *MinION*. O primeiro é um dispositivo maior, mais direcionado para sequenciação de genes completos. O segundo é uma versão melhorada, sendo um dispositivo pequeno e portátil capaz de sequenciar 1Gb de DNA. Esta tecnologia é bastante vantajosa no estudo do HLA, pois permite obter leituras longas, até 30Gb, possui baixo custo e a preparação da amostra é simples. Tem por base o uso de nanoporos, sintéticos ou biológicos os quais permitem a discriminação dos nucleótidos individualmente medindo a mudança na condutividade elétrica à medida que as moléculas de DNA atravessam o poro. É utilizada uma exonuclease I em associação com o complexo do nanoporo (possui uma molécula adaptadora, a ciclodextrina, covalentemente associada próximo ao centro do poro de forma a otimizar a discriminação das bases) permitindo a redução da velocidade com que o DNA passa através do poro, o que melhora a qualidade de captura do sinal (Zolet *et al.*, 2017 e Carapito *et al.*, 2016).

São várias as vantagens desta tecnologia: é capaz de discriminar com precisão (cerca de 90%) os dNMPs (desoxinucleosídeosmonofosfato), é capaz de conhecer formas metiladas, leituras rápidas de 400 Mb numa corrida de 48h, oferece um baixo custo e os dados são

exibidos na tela e gerados sem ter que esperar que a corrida esteja completa. Outra vantagem do *MinION* é a possibilidade de utilização em estudos locais, epidemiológicos ou de monitorização, pois é fácil a conexão USB a qualquer computador, produzindo resultados de sequenciação em tempo real à medida que as moléculas de DNA são lidas (**Zolet et al., 2017 e Carapito et al., 2016**). Esta versão pode fornecer leituras muito longas, superiores a 150kbp, no entanto produz uma elevada taxa de erro de ~ 12% distribuída em cerca de 3% de incompatibilidades, ~ 4% de inserções e ~ 5% de exclusões. A tecnologia *Oxford Nanopore Technologie* continua a evoluir. Recentemente surgiu um instrumento chamado "*PromethION*". É um sequenciador de mesa de trabalho autónomo com 48 células de fluxo individuais, cada uma com 3000 poros (equivalente a 48 *MINIONs*) operando a 500bp por segundo, o que é suficientemente poderoso para alcançar um *throughput* elevado e vantajoso para o sequenciação de grandes genomas, como o genoma humano (**Kchouk et al., 2017**).

1.9.7. Bioinformática

Apesar das particularidades de cada plataforma NGS, o pós-processamento recorre sempre à Bioinformática. Após sequenciação, o ficheiro gerado é introduzido num programa adequado sendo efetuado um pré-processamento, como remoção das sequências adaptadoras e sequências com baixa qualidade de leitura. Após alinhamento da sequência de referência com a de teste, é feita a interpretação, de acordo com a aplicação e as questões científicos. O elevado polimorfismo do HLA exige adaptações no alinhamento. Assim, as leituras de sequência são alinhadas com as sequências alélicas HLA conhecidas e disponíveis na base de dados IMGT/HLA, limitando a perda de leituras e desfasamentos. Adicionalmente, é possível diferenciar melhor os alelos e facilitar a genotipagem. Também o número de *softwares* para o HLA tem aumentado, de forma a facilitar a genotipagem (**Carapito et al., 2016**).

Capítulo 2

Material e Métodos

Este capítulo descreve a metodologia utilizada para a genotipagem HLA, de acordo com os protocolos vigentes no IPST- Porto, desde a colheita das amostras biológicas até à interpretação final dos resultados.

2.1. Colheita de amostras biológicas e triagem

Após identificação do utente, procede-se à colheita de sangue, sendo selecionados os tubos conforme os pedidos e sendo as amostras entregues nos laboratórios de destino. Em caso de amostras externas, após identificação pelo secretariado, estas são encaminhadas para os respetivos laboratórios.

No presente trabalho, foram analisadas 2054 amostras de dador CEDACE (Centro Nacional de Dadores de Células de Medula Óssea, estaminais ou de Sangue do Cordão) desde 2017 até julho de 2018 e 265 amostras de cordão desde 2013 até julho de 2018.

2.2. Extração de DNA

Para a obtenção de DNA com suficiente integridade, a amostra de sangue deve ser colhida como sangue total e armazenada com EDTA (Ácido etilenodiamina tetra-acético) ou ACD (Ácido citrato dextrose) como agentes anticoagulantes. O valor mínimo de sangue total necessário é de 200 μ L.

Nos casos em que a pureza de DNA é de extrema importância, como as amostras de rotina, deve realizar-se a extração a partir de *buffy coat*. Para tal centrifugam-se os tubos de sangue a 3500 rpm durante 10 min e a 4°C. De seguida, com uma pipeta esterilizada retira-se o anel de linfócitos para um tubo de 15 ml ao qual se adiciona RCLB (*Red Cell Lysis Buffer*: para 2 litros de solução adiciona-se; 20 ml Tris 1M pH=7.6; 10 ml de MgCl₂ 1M; 4 ml de NaCl 5M e completa-se o restante volume com H₂O) na proporção de 1:3. Volta a centrifugar-se durante 10 min a 1200 rpm e rejeita-se o sobrenadante, devendo repetir-se a

lavagem com RCLB até o *pellet* estar completamente limpo. Em seguida procede-se à extração manual ou automática.

O parâmetro da qualidade das amostras extraídas é assegurado pela quantificação e avaliação da pureza final do DNA obtido descrito em 2.3. A amostra resultante da extração deve ser guardada à temperatura de 2° a 8° C até processamento ou a -20°C, para armazenamento mais prolongado.

A extração de DNA pode ser realizada por diferentes procedimentos. O protocolo de extração manual encontra-se descrito abaixo. No entanto esta pode ser realizada por outros dois procedimentos de extração automática, descritos nos anexos G e H. São eles a extração em modo automático no *QIAcube*, cujo permite a extração de DNA a partir de um total máximo de 12 amostras; e a extração automática a partir do *QIAGEN-QIASYMPHONY DNA BloodKit* para um total máximo de 96 amostras.

Procedimento Extração Manual:

a) Identificar previamente um tubo eppendorf de 1,5 ml, pipetar 20 µl de *Protease* (*Qiagen* 20 mg/ml), 200 µl de amostra (sangue total ou *buffy-coat* previamente isolado e lavado) e adicionar 200 µl de Tampão AL. Misturar no vortex durante 15 seg e incubar a 56°C durante 10 min.

b) Após centrifugação breve para remoção de gotas na tampa, adicionar 200 µl de etanol absoluto e misturar no vortex, durante 15 seg. Após mistura, centrifugar novamente de forma breve.

c) Aplicar a mistura na microcoluna de extração inserida num tubo coletor de 2 ml e centrifugar a 8000 rpm durante 1 min. Colocar a microcoluna num tubo coletor de 2 ml limpo e descartar o tubo que contém o filtrado. Caso na microcoluna ainda permaneça líquido esta centrifugação deve ser repetida.

d) Adicionar 500 µl de Tampão AW1 e proceder como em c).

e) Adicionar 500 µl de Tampão AW2 e centrifugar a 14000 rpm durante 3 min. Caso ainda exista tampão AW2 na microcoluna, colocá-la num tubo novo coletor e centrifugar a 14000 rpm durante 1 min.

f) Colocar a microcoluna num *eppendorf* de 1,5 ml, descartar o tubo coletor com o filtrado e adicionar 200 µl de Tampão AE. Incubar à temperatura ambiente (15-25°C) durante 1 min e centrifugar a 8000 rpm, durante 1 min. **NOTA:** a incubação antes da centrifugação permite o aumento do rendimento.

g) Identificar o tubo de armazenamento, proceder à quantificação da amostra e seu armazenamento.

2.3. Quantificação e Análise da pureza do DNA

Os equipamentos Nanodrop 1000 e Nanodrop 8000 determinam a quantidade de ácidos nucleicos através da leitura das absorvâncias nos comprimentos de onda de 260 nm e 280 nm e também estabelece um grau de pureza da amostra através das razões 260/280 e 260/230. Para DNA puro, a razão 260/280 tem de apresentar valores de $\approx 1,8$. Valores inferiores a esta razão indicam a presença de proteínas, fenol ou outros contaminantes que absorvem próximo do comprimento de onda de 280nm. A razão 260/230 é considerada uma medida secundária da pureza dos ácidos nucleicos. Esta razão apresenta valores entre 1,8 e 2,2, sendo que valores inferiores desta razão indica a presença de sais e solventes orgânicos.

A amostra sujeita à quantificação deve estar o mais homogênea possível de modo a evitar desvios na quantificação. É importante que a quantidade de amostra a testar cubra uniformemente o orifício do pedestal inferior e que não haja bolhas de ar. O volume recomendado é de 1,5 a 2 μl .

Entre leituras de umas amostras para outras, é necessário proceder à limpeza dos pedestais com papel absorvente. Já no final de todas as leituras deve-se limpar os pedestais da seguinte forma: 1) Adicionar 5 μl de água ultra-pura no pedestal inferior; 2) Baixar o pedestal superior e deixar em repouso durante 2 a 3 min; 3) Secar os pedestais superiores e inferiores com papel absorvente.

A concentração ótima de DNA para cada uma das técnicas utilizadas deve ser:

- 30 a 50 ng/ μl – PCR-SSP *OLERUP*
- 20 a 60 ng/ μl - *Labtype* rSSO tecnologia *Luminex*
- 10 a 40 ng/ μl - SBT *HLAssure Texas BioGene*
- 15-30 ng/ μl - SBT *SeCore*

2.4. Genotipagem HLA Classe I e II pelo *kit OLERUP SSP*

2.4.1. Procedimento

2.4.1.1. Amplificação por PCR

a) Remover a placa de PCR contendo os *primers* liofilizados do congelador, assim como o reagente PCR Master Mix com Taq (*OLERUP SSP*).

b) Num tubo de 1,5 ml, juntar (N+2) x 3 µl de PCR Master Mix com Taq e (N+2) x 5 µl de dH₂O, sendo N o número de poços da placa de PCR. Homogeneizar a mistura no vortex e dispensar 8 µl desta em cada um dos poços, adicionando de seguida 2 µl de DNA (30ng/ µl). No último poço, que corresponde ao controlo negativo, adicionar apenas a mistura sem o DNA.

c) Tapar a placa com a película, colocá-la no termociclador iniciando o respetivo ciclo de temperaturas:

❖ 1 ciclo:	94°C	2 min.
❖ 10 ciclos:	94°C	10 seg.
	65°C	60 seg.
❖ 20 ciclos:	94°C	10 seg.
	61°C	50 seg.
	72°C	30 seg.

2.4.1.2. Pós-Amplificação

a) Preparar previamente o gel de agarose a 2% e colocá-lo na tina de eletroforese com tampão TBE (Tris/Borato/EDTA) 0,5x;

b) Com uma pipeta multicanal, depositar os produtos de PCR no gel. Por cada linha de poços no gel, reservar um poço para o marcador de peso molecular em uso no laboratório, adicionando 6 µl deste;

c) Efetuar a eletroforese a 200 volts durante 15 min.

2.4.2. Cálculo e análise de resultados

a) Colocar o gel num transiluminador UV (radiação ultravioleta) e fotografar (sistema de captação e imagem);

b) Verificar e assinalar as reações positivas, isto é, os poços que amplificam uma banda de tamanho pré-determinado pelo kit. Nas reações negativas, verificar apenas a existência da banda específica do controlo interno (concebido para amplificar e detetar uma região conservada do genoma humano. Nos casos em que os ensaios num poço são negativos, este controlo permite confirmar que a mistura de reação foi adicionada corretamente ao poço e que ocorreu a PCR).

c) Apenas as bandas com um tamanho molecular específico para cada poço, com ou sem a banda de controlo interno, devem ser consideradas positivas, podendo esse tamanho molecular relativo ser útil na interpretação dos resultados, tal como apresentado na figura 7a em resultados e discussão.

2.4.3. Validação dos resultados

- É válida uma banda positiva com o controlo interno ténue, ou sem controlo interno, porque pode existir uma competição para a ligação aos *primers*.
- A presença de bandas ténues, com o tamanho molecular específico para cada poço, pode ser considerada positiva, consoante a intensidade das restantes bandas.
- Interpretar com as tabelas específicas para cada *kit* e usar o *software SCORE*.
- Devem registar-se todos os comportamentos incorretos das bandas, como falsos positivos e falsos negativos.
- Uma separação eletroforética é considerada válida se separar e permitir definir os vários fragmentos do marcador de pesos moleculares em uso.
- O controlo negativo (sem DNA) é usado como controlo de qualidade.

2.4.4. Interpretação do PCR-SSP através do *software SCORE*

A interpretação do PCR/SSP é efetuada no programa *HELMBERG-SCORE* com o *software SCORE* de acordo com o manual de utilizador. A interpretação é feita através do registo das reações positivas na eletroforese.

2.5. Genotipagem HLA pela metodologia rSSO usando o *kit Labtype* rSSO através da tecnologia *Luminex*

2.5.1. Procedimento

2.5.1.1. Amplificação por PCR

- a) Pipetar 1 µL de DNA em cada poço da placa de PCR;
- b) Preparar a mistura de amplificação conforme descrito, sendo N o número de reações: N x 6,9 µl D-mix (*Labtype Primer set D-mix*; ref LTPDMX, *One Lambda*); N x 2 µl *primer* e N x 0.2 µl Taq polimerase. Agitar no vortex e pipetar 9 µl da mistura para cada poço.
- c) Selar e colocar a placa de PCR no termociclador e executar o programa de amplificação:

❖ 1 ciclo:	96°C	3 min.
❖ 5 ciclos:	96°C	20 seg.
	60°C	20 seg.
	72°C	20 seg.
❖ 30 ciclos:	96°C	10 seg.
	60°C	15 seg.
	72°C	20 seg.
❖ 1 ciclo	72°C	10 min.

2.5.1.2. Desnaturação/Neutralização

a) Transferir 5 μL de DNA amplificado para o poço da placa e pipetar 2,5 μL de tampão de desnaturação para cada poço, misturando bem com a pipeta em cada um dos poços. Incubar a placa 10 min entre 20 a 25 $^{\circ}\text{C}$;

b) Acrescentar 5 μL de tampão de neutralização em cada poço. Misturar com a pipeta até observar a mudança de cor. Conservar no frio.

2.5.1.3. Hibridação

A mistura de hibridação e a SAPE são sensíveis à luz, de tal modo que devem ser guardadas em ambiente escuro. Para além disso o tampão SAPE é termossensível, devendo ser refrigerado entre 2-8 $^{\circ}\text{C}$. O invólucro das esferas não deve ser retirado até à sua utilização.

a) Misturar durante 30 seg as esferas vigorosamente no vortex, de forma a suspendê-las;

b) Preparar a mistura de hibridação e proteger da luz, em que N corresponde ao número de reações: N x 17 μL de tampão de hibridação e N x 2 μL da mistura de esferas. Pipetar 19 μL da mistura de hibridação para cada poço e agitar no vortex a baixa velocidade. Incubar a placa a 60 $^{\circ}\text{C}$ durante 15 min.

c) Retirar a película de selagem e acrescentar rapidamente 75 μL de tampão de lavagem em cada poço. Centrifugar 5 min a 1300g;

d) Remover o tampão de lavagem por inversão da placa e agitar no vórtex de forma suave. Repetir a lavagem;

e) Preparar a solução SAPE de acordo com o número de reações (N): N x 0.2 μL de SAPE e N x 20 μL de tampão SAPE.

2.5.1.4. Marcação

a) Ligar o analisador *Luminex* e proceder à sua manutenção diária.

b) Pipetar 20 μL de solução SAPE para cada poço e agitar no vortex a baixa velocidade para misturar. Incubar durante 5 min a 60 $^{\circ}\text{C}$;

c) Retirar a película e acrescentar 75 μL de tampão de lavagem seguindo-se uma centrifugação de 5 min a 1300g;

d) Remover o sobrenadante por inversão da placa e pipetar 70 μL de *sheath fluid*. Misturar suavemente com a pipeta e transferir para a placa de leitura;

e) Ler as amostras. Caso não seja possível ler imediatamente, é necessário selar a placa com película e manter entre 2 a 8°C no escuro. O armazenamento prolongado das amostras, mais de 4 horas, resulta na perda do sinal.

2.5.2. Validação

A validação deste protocolo é feita mediante os resultados obtidos através dos controlos internos, controlo positivo e controlo negativo. Estes variam de acordo com o *locus* em estudo (ver anexo I) e das amostras controlo conhecidas adicionadas a cada série de ensaios. Para além disso os resultados têm de ser validados por duas pessoas.

2.5.3. Leitura e interpretação dos dados com o software HLA Fusion

O objetivo deste consiste em definir a análise da genotipagem HLA classes I e II quando são utilizados *kits* rSSO através do *software* HLA Fusion, de acordo com o manual do utilizador.

2.6. Genotipagem HLA pela metodologia PCR-SBT usando os kits HLAAssure Texas BioGene

Os *kits* HLAAssure da TexasBiogene contém *primers* de amplificação para PCR de um locus específico (LSA) para os genes HLA-A, -B, -C, -DRB1, -DQB1 e também *primers* de amplificação para PCR de um grupo específico (GSA), permitindo a separação alélica e diminuindo o número de ambiguidades. Os pares de *primers* a serem utilizados para cada grupo/alelos específicos encontram-se descritos no anexo J.

Este sistema de sequenciação apresentado foi desenhado para a amplificação dos diferentes exões no sentido *forward* e *reverse*. Para os genes HLA-A e HLA-B estão disponíveis *primers* de sequenciação que amplificam os exões 1, 2, 3, 4 e 5. No caso do HLA-C, para além destes últimos, é também possível sequenciar os exões 6 e 7. Para o HLA-DQB1 existem *primers* para a sequenciação dos exões 2 e 3 e quanto ao HLA-DRB1 há *primers* para exão 2 e ainda uma mistura que sequencia o codão 86, frequentemente usado para diminuir a frequência de ambiguidades. O anexo K apresenta a especificidade das misturas de amplificação para os diferentes *primers* de sequenciação.

2.6.1. Procedimento

No Laboratório de Genética, os pedidos para a tipagem HLA em alta resoluções são essencialmente feitos a dadores CEDACE ativados para um doente, em que a tipagem em baixa resolução já é conhecida; ou a doentes cujo estudo de família já foi realizado, mas que não têm um dador aparentado com tipagem HLA idêntica. Assim, nos casos de uma tipagem heterozigótica previamente conhecida, as amplificações são realizadas com os dois *primers* GSA ou com um *primer* GSA e outro LSA. Nos casos de tipagem homozigótica previamente determinada, é recomendável a realização das amplificações com os *primers* LSA.

2.6.1.1. Amplificação por PCR

a) Preparar um tubo de *master mix* para cada *locus* de acordo com o número de reações (N):

- ❖ Tampão AMP- (N+1) x 6 µl
- ❖ EZTaq - Classe I: (N+1) x 0.14 µl
- Classe II: (N+1) x 0.07 µl

b) Normalizar as concentrações de DNA a 40 ng/µl.

c) Adicionar à placa de PCR ou aos tubos de PCR 6 µl de *master mix*, 3 µl de DNA genómico com concentração normalizada e 3 µl de *primer* de amplificação. Executar um *short spin* à placa ou tubos de PCR.

d) Introduzir a placa ou tubos de PCR no termociclador executando o seguinte programa:

	Classe I:	Classe II:
❖ 1 ciclo:	95°C 5 min.	5 min.
❖ 36 ciclos:	93°C 30 seg.	30 seg.
	63°C 40 seg.	40 seg.
	72°C 2,5 min.	1 min.
❖ 1 ciclo:	72°C 5 min.	2,5 min.

e) Após a reação de PCR proceder à eletroforese em gel de agarose a 2%. Caso não seja possível realizar a eletroforese imediatamente, é possível guardar as reações de PCR no frigorífico 4° a 8°C, até duas semanas. **NOTA:** A eletroforese em gel de agarose não é obrigatória. No entanto deve ser executada sempre que haja suspeita de degradação do DNA ou ausência de sequencias após corrida no sequenciador.

2.6.1.2. Purificação das reações de PCR

a) Adicionar 2 µl de ExoSAP-IT a cada reação a purificar. No termociclador iniciar o programa

❖ **1 ciclo:** 37°C 30 min. (Tratamento)
80°C 15 min. (Inativação)

b) Diluir os amplicões de modo a garantir quantidade de produto de PCR apropriado para sequenciação. Na presença de uma banda nítida, adiciona-se 10 a 15 µl de água bidestilada às reações. Caso não seja possível efetuar imediatamente este passo é necessário guardar as reações de PCR purificadas a 7° - 9°C até duas semanas.

2.6.1.3. Reação de sequenciação

A reação de sequenciação é feita aos produtos de PCR purificados, utilizando *primers* de sequenciação que amplificam os exões relevantes para cada *locus*, cada qual em duas direções: *forward* e *reverse*. Para cada mistura de amplificação GSA/LSA deve consultar-se a sua especificidade para os diferentes *primers* de sequenciação.

a) Calcular o número total de reações (N) de sequenciação e determinar os volumes necessários para preparação da *master mix* BDT: N x 1 µl de BDT *Ready Reaction Mix* e N x 0,5 µl de Tampão 5x BDT;

b) Adicionar 1,5 µl de *master mix*, 2 µl de produto de PCR purificado e diluído e 2,5 µl do respetivo *primer* de sequenciação a cada poço da placa de sequenciação. Selar a placa e verificar que todo o produto a sequenciar se encontra no fundo do tubo, caso contrário, fazer um *short spin* na centrífuga de placas;

c) Introduzir a placa no termociclador e iniciar o programa:

❖ **1 ciclo:** 96°C 1 min.
❖ **25 ciclos:** 96°C 10 seg.
50°C 5 seg.
60°C 4 min.

d) Após a reação de sequenciação, retirar a placa do termociclador e proceder à purificação.

2.6.1.4. Purificação das reações por precipitação alcoólica

a) Adicionar 10 µl de água bidestilada a cada poço de reação;

b) Adicionar 2 µl de tampão NaOAc/EDTA a cada reação de sequenciação. Fazer um *short spin* à placa de sequenciação (90g durante 30 seg).

- c) Adicionar 25 µl de etanol absoluto a cada poço e misturar no vortex durante 15 seg. Centrifugar a placa à velocidade de 2000g durante 30 min.
- d) Retirar o selo autocolante e rejeitar o sobrenadante. Inverter a placa sobre papel absorvente e colocar na centrífuga de placas fazendo um *short spin* a 90g durante 30 seg.
- e) Rejeitar o papel absorvente e adicionar 50 µl de etanol a 80% (diluído com água bidestilada) a cada poço e tapar a placa com um selo autocolante. Centrifugar a placa à velocidade de 2000g durante 5 min.
- f) Retirar o selo autocolante e o sobrenadante invertendo a placa sobre papel absorvente colocado na centrífuga de placas e fazendo um *short spin* a 500 G durante 1 minuto.
- g) Rejeitar o papel absorvente e adicionar 15 µl de Formamida Hi-Di a cada poço contendo o produto de sequenciação purificado.
- h) Selar a placa com o *Plate Septa* e centrifugar de modo a que todo o líquido fique no fundo do poço e sem bolhas.
- i) Inserir a placa no termociclador e desnaturar as reações a 95°C durante 2 min. Após desnaturação, retirar a placa de 96 poços do termociclador e realizar a eletroforese capilar.

2.6.1.5. Eletroforese Capilar

Colocar a placa no Sequenciador ABI Prism® 3730 e iniciar a sequenciação capilar. A genotipagem HLA é obtida através da interpretação das sequências resultantes através do *software uTYPE*.

2.6.2. Validação

Após a sequenciação verificar a qualidade dos eletroferogramas de acordo com o *software* de análise de sequências “*Sequencing Analysis*”. Verificar que o ruído da sequência é baixo, não interferindo com a determinação do pico do nucleótido e verificar a intensidade do sinal dos diferentes picos que deve ser superior a 100 e inferior a 3000.

O *software ACCUTYPE* alinha os diferentes fragmentos exônicos, rejeita os eletroferogramas de má qualidade e determina a genotipagem.

2.7. Genotipagem HLA por Sequenciação *SeCore*

2.7.1. Procedimento

2.7.1.1. Amplificação por PCR

a) Pipetar os DNAs extraídos no fundo de cada poço da placa. No poço controlo negativo juntar água em vez do DNA e no positivo uma amostra anteriormente genotipada por alta resolução. Para tipagem HLA de classe I pipetar 5 µl de DNA de cada amostra e para HLA classe II pipetar 2 µl de DNA de cada amostra.

b) Preparar a *mastermix* de acordo com o número de reações (N):

	Classe I:	Classe II:
❖ Mix Amp-	(N+2) x 9,9 µl	(N+2) x 7,9 µl
❖ Taq-	(N+2) x 0.15 µl	(N+2) x 0.1 µl

c) Adicionar a *mastermix* aos poços da placa, sendo 10 µl para HLA classe I e 8 µl para classe II. Centrifugar em short spin e colocar a placa no termociclador iniciando o seguinte ciclo de temperaturas:

❖ 1 ciclo:	95°C	4 min.
❖ 35 ciclos:	95°C	20 seg.
	63°C	20 seg.
	72°C	40 seg.
❖ 1 ciclo:	72°C	5 min.

2.7.1.2. Purificação das reações de PCR

a) Preparar a solução de purificação com ExoSAP-IT de acordo com o número de reações (N):

	Classe I:	Classe II:
❖ ExoSAP-IT-	(N+2) x 3 µL	(N+2) x 2 µL
❖ MgCL ₂ -	(N+2) x 12 µL	(N+2) x 13 µL

b) Juntar 15 µL de ExoSAP-IT diluída a cada poço, levar ao vortex vigorosamente cerca de 5 seg e efetuar de seguida um short spin na centrífuga. Colocar a placa no termociclador e executar o programa: 37°C durante 20 min e 80°C durante 20 min.

2.7.1.3. Reação de sequenciação

a) Aos produtos de PCR de HLA classe II, diluir com 30 µl de água ultrapura. Tapar com película adesiva, vortex e centrifugar em *short spin*;

b) Transferir 2µL dos amplicões tratados com ExoSAP para as placas com 3µL de *primer* previamente distribuídos. Centrifugar em *short spin*;

c) Colocar a placa no termociclador e iniciar o programa:

❖ 25 ciclos	95°C	20 seg.
	50°C	15 seg.
	60°C	60 seg.

2.7.1.4. Purificação das reações por precipitação alcoólica

- a) Misturar no vortex a seguinte mistura, de acordo com o número de reações (N): N x 20 µL de etanol 100% e N x 1 µL de tampão PPT.
- b) Adicionar 21 µL da mistura de precipitação a cada poço e levar ao vortex +/- 1min. Centrifugar a placa 2000g durante 30min. Inverter em papel absorvente e centrifugar invertido a 100g durante 1 min;
- c) Juntar a cada poço 35 µL de etanol a 80% e centrifugar 2000g durante 5 min. Inverter em papel absorvente e centrifugar invertido 500g durante 1min;
- d) Juntar 15 µL de formamida a cada poço. Centrifugar em *short spin*;
- d) Desnaturar as amostras a 95°C durante 2 min no termociclador.

2.7.1.5. Eletroforese Capilar

Colocar a placa no Sequenciador ABI Prism® 3730 e iniciar o programa de leitura. Fazer a interpretação das sequências resultantes através do *software uTYPE*.

2.7.2. Validação

A tipagem pelo *software uTYPE* tem em consideração a qualidade das sequências obtidas no que respeita à intensidade de sinal dos picos e do ruído de fundo e ainda ao espaçamento uniforme entre os picos, permitindo a correta determinação das bases. Assim, no caso de uma sequência com uma intensidade de sinal de ruído muito elevado ou no caso de uma sequência com uma intensidade de sinal muito baixo (normalmente abaixo de 100), o próprio *software* desativa a sequência e assume-a como UN (não definida), não permitindo a sua ativação e análise. A verificação manual de toda a sequência deve ser feita de modo a corrigir alguma base incorretamente determinada, assim como todas as bases de pior qualidade assinaladas pelo programa.

Como controlo de qualidade em cada corrida insere-se uma amostra conhecida e previamente genotipada.

2.7.3. Interpretação de sequências pelo software *uTYPE*

O software *uTYPE* HLA é o programa utilizado para a análise de sequências HLA obtidas com os kits *SeCore* SBT da *One Lambda*, permitindo deste modo a genotipagem das diferentes amostras testadas por comparação com a base de dados IMGT/HLA. Este software permite a análise de todos os *loci* HLA nomeadamente HLA-A, -B, -C, -DRB1, -DRB3, -DRB4, -DRB5, -DQA1, -DQB1, -DPA1 e -DPB1. Qualquer ficheiro de dados de sequenciação obtido nos instrumentos da *Applied Biosystems* pode ser importado e analisado no *uTYPE*, após a análise automática no software *Sequencing Analysis*.

2.8. Cálculo das frequências alélicas

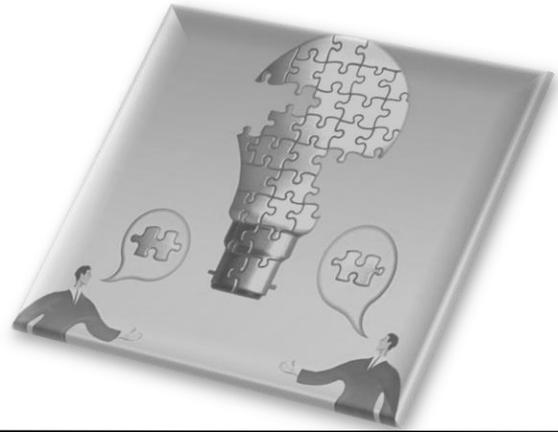
Nos resultados serão apresentadas as frequências alélicas, em baixa resolução, para a população do Norte de Portugal a partir das genotipagens das amostras dos dadores CEDACE. A tabela 1, em resultados e discussão, apresenta o cálculo das frequências alélicas para um total de 2054 amostras de sangue a partir de dadores de medula óssea genotipados desde 2017 até julho de 2018, enquanto a tabela 2 demonstra as frequências obtidas a partir de 265 amostras de sangue do cordão genotipados desde 2013 até julho de 2018.

Para o cálculo das frequências alélicas foi utilizada a fórmula $FA(\%) = \frac{n}{2N} \times 100$, em que n corresponde ao somatório de um determinado alelo e N ao número total de amostras genotipadas para dado gene HLA.

Genotipagem do *Human Leucocyte Antigen* (HLA):
Técnicas atuais e aplicabilidade da abordagem de *Next Generation Sequencing*

Capítulo 3

Resultados e Discussão



Há cerca de 60 anos atrás foram descobertos os primeiros antígenos HLA (**Dausset, 1981**). Desde então foi possível constatar que esta região do genoma humano é altamente polimórfica. Inicialmente, apenas uma pequena parte do polimorfismo HLA foi descrita, consistindo em algumas dezenas de alelos para os genes HLA-A, -B e -DRB1. Com o decorrer do tempo e avanço da tecnologia, atingiu-se um número considerável de alelos com distintas frequências, criando desafios na estimativa de distâncias genéticas. Em paralelo, tornou-se visível um padrão de diversidade HLA, sob a forma de uma elevada heterozigotia na maioria das populações e elevado número de polimorfismos não sinónimos dentro dos *loci* HLA. Estes resultados evidenciaram que a função imunológica do *cluster* de genes HLA é alvo de uma seleção equilibrada (**Meyer e Thomson, 2001**). Para além disso, foram sendo descobertas associações tanto positivas como negativas entre o HLA e doenças, com resultados em muitos casos heterogêneos entre populações. Por último, a descoberta de interações entre o HLA e outras moléculas altamente polimórficas do sistema imunológico, como o KIR (*Killer-cell immunoglobulin-like receptors*), levou à possibilidade de coevolução imunomediada entre o HLA e outras famílias de genes. Devido à complexidade da região HLA, a caracterização dos seus polimorfismos torna-se uma tarefa árdua. Felizmente, a constante atualização das bases de dados e dos repositórios de definições alélicas, a criação de ferramentas bioestatísticas e bioinformáticas e o avanço tecnológico, simplificam a análise destes dados. Como resultado, observa-se um grande progresso na caracterização do HLA (**Sanchez-Mazas e Meyer, 2014**).

O principal objetivo do presente trabalho foi a aprendizagem e aplicação das técnicas de rotina para a genotipagem HLA. Esta aprendizagem ocorreu no laboratório de Genética Molecular do Instituto Português do Sangue e Transplantação- unidade de transplantação do Porto. Foi efetuada uma extensa revisão sobre o *cluster* de genes HLA, incluindo a sua caracterização e importante relevância na resposta imunitária, bem como a sua associação

com doenças e revisão das técnicas para a sua genotipagem na rotina laboratorial. É dado especial enfoque à aplicabilidade do NGS nos laboratórios de genotipagem HLA.

Tendo em conta a base de dados deste Instituto foi realizado um estudo comparativo das frequências alélicas HLA entre os dadores voluntários de medula óssea e as amostras do banco de cordão umbilical e também entre outros estudos realizados a nível nacional. Também é abordada a interpretação e discussão dos resultados em aplicações biomédicas específicas, nomeadamente decisão sobre a prescrição de Abacavir após a genotipagem para HLA-B57 e a genotipagem de HLA-B27 no diagnóstico clínico de espondilite anquilosante.

Na tabela 1 são apresentadas as frequências alélicas na região Norte de Portugal, em baixa resolução, para dadores de medula óssea genotipados desde 2017 até julho de 2018.

Tabela 1. Frequências alélicas, em baixa resolução, das amostras de dadores de medula óssea do Norte de Portugal, genotipados desde 2017 até julho de 2018.

Dadores de Medula (N=2054)											
HLA-A	Nº de ALELOS	FA (%)	HLA-B	Nº de ALELOS	FA (%)	HLA-C	Nº de ALELOS	FA (%)	HLA-DRB1	Nº ALELOS	FA (%)
01	435	10,59	07	284	6,91	01	109	2,65	01	508	12,37
02	1102	26,83	08	282	6,86	02	264	6,43	03	410	9,98
03	439	10,69	13	64	1,56	03	285	6,94	04	521	12,68
11	256	6,23	14	301	7,33	04	623	15,17	07	743	18,09
23	181	4,41	15	245	5,96	05	286	6,96	08	180	4,38
24	472	11,49	18	205	4,99	06	356	8,67	09	38	0,93
25	57	1,39	27	123	2,99	07	1013	24,66	10	56	1,36
26	122	2,97	35	451	10,98	08	265	6,45	11	411	10,00
29	220	5,36	37	47	1,14	12	272	6,62	12	58	1,41
30	119	2,90	38	99	2,41	14	123	2,99	13	625	15,21
31	102	2,48	39	75	1,83	15	219	5,33	14	98	2,39
32	176	4,28	40	156	3,80	16	248	6,04	15	362	8,81
33	137	3,33	41	43	1,05	17	44	1,07	16	98	2,39
34	22	0,54	42	9	0,22	18	1	0,02	Total	4108	100,0
36	2	0,05	44	589	14,34	Total	4108	100,0			
66	27	0,66	45	49	1,19						
68	206	5,01	47	11	0,27						
69	23	0,56	48	4	0,10						
74	5	0,12	49	142	3,46						
80	5	0,12	50	147	3,58						
Total	4108	100,0	51	456	11,10						
			52	25	0,61						
			53	38	0,93						
			55	42	1,02						
			56	17	0,41						
			57	123	2,99						
			58	76	1,85						
			67	2	0,05						
			73	3	0,07						
			Total	4108	100,0						

O sangue do cordão umbilical contém células-tronco hematopoiéticas e pode ser usado como uma alternativa para o transplante de medula óssea quando doadores adultos adequados não estão disponíveis. Este material biológico é uma fonte de DNA facilmente disponível, rico em células estaminais hematopoiéticas. O Instituto Português do Sangue e da Transplantação tem como atividade a criopreservação das células do cordão umbilical, e antes do processamento destas amostras é retirada uma amostra representativa do saco, contendo o sangue do cordão umbilical. No instituto o DNA das amostras é extraído a partir de sangue total. A tabela 2 apresenta as frequências alélicas para amostras de sangue do cordão doadas e genotipadas desde 2013 até julho de 2018 na região Norte de Portugal.

Tabela 2. Frequências alélicas em baixa resolução para as amostras doadas de sangue do cordão genotipadas desde 2013 até julho de 2018, na região Norte de Portugal.

Amostras de sangue do cordão (N=265)											
HLA-A	Nº de ALELOS	FA (%)	HLA-B	Nº de ALELOS	FA (%)	HLA-C	Nº de ALELOS	FA (%)	HLA-DRB1	Nº de ALELOS	FA (%)
01	63	11,89	07	33	6,23	01	10	1,89	01	73	13,77
02	146	27,55	08	38	7,17	02	24	4,53	03	55	10,38
03	60	11,32	13	7	1,32	03	24	4,53	04	61	11,51
11	41	7,74	14	39	7,36	04	92	17,36	07	88	16,60
23	33	6,23	15	33	6,23	05	30	5,66	08	22	4,15
24	40	7,55	18	27	5,09	06	50	9,43	09	8	1,51
25	3	0,57	27	16	3,02	07	136	25,66	10	12	2,26
26	13	2,45	35	68	12,83	08	37	6,98	11	58	10,94
29	30	5,66	37	10	1,89	12	30	5,66	12	8	1,51
30	27	5,09	38	8	1,51	14	18	3,40	13	77	14,53
31	9	1,70	39	9	1,70	15	28	5,28	14	13	2,45
32	18	3,40	40	10	1,89	16	41	7,74	15	43	8,11
33	18	3,40	41	6	1,13	17	9	1,70	16	12	2,26
34	2	0,38	42	4	0,75	18	1	0,19	Total	530	100,0
36	4	0,75	44	74	13,96	Total	530	100,0			
66	3	0,57	45	11	2,08						
68	17	3,21	47	1	0,19						
69	1	0,19	49	18	3,40						
74	1	0,19	50	20	3,77						
80	1	0,19	51	53	10,00						
Total	530	100,0	52	5	0,94						
			53	10	1,89						
			55	3	0,57						
			56	1	0,19						
			57	16	3,02						
			58	9	1,70						
			81	1	0,19						
			Total	530	100,0						

Tendo em conta os resultados obtidos é possível observar que em ambos os grupos, para o gene HLA-A, o alelo mais frequente é o HLA-A*02, tendo-se registado uma frequência de 26,83% para as amostras de medula óssea e 27,55% para as amostras de cordão. Já o menos frequente relativamente às amostras de medula é o alelo HLA-A*36 com uma frequência de 0,05% e para as amostras de cordão verificam-se três alelos com frequência baixa de 0,19%, sendo o HLA-A*69; -A*74 e -A*80. Para o gene HLA-B, o alelo mais frequente para ambos os grupos, medula e cordão, é o alelo HLA-B*44, com frequências de 14,34% e 13,96%, respetivamente. O alelo menos frequente corresponde ao HLA-B*67 nas amostras de medula, com percentagem de 0,05%, e aos alelos HLA-B*47, -B*56 e -B*81, todos com frequências alélicas de 0,19% nas amostras de cordão. Quanto ao gene HLA-C, o alelo mais frequente e o menos frequente é o mesmo em ambos os grupos, sendo o HLA-C*07 o mais frequente e o HLA-C*18 o menos frequente. As percentagens correspondem a 24,66% e 0,02%, respetivamente para as amostras de medula, e para as amostras de cordão correspondem a 25,66% e 0,19%. Por último, temos as frequências alélicas relativas aos alelos do gene HLA-DRB1. Neste caso o alelo HLA-DRB1*07 mostrou ser o mais frequente nas amostras de medula, com valores de 18,09% e o menos frequente com valores de 0,93% corresponde ao HLA-DRB1*09. Nas amostras de cordão com frequência alélica de 16,60% temos o alelo mais frequente o HLA-DRB1*07, e quanto aos alelos menos frequentes temos o HLA-DRB1*09 e -DR*12, ambos com o mesmo valor de 1,51%.

É possível observar-se que o gene mais polimórfico é o HLA-B com 29 alelos diferentes identificados nos doadores de medula e 27 nas amostras de cordão. De seguida temos o gene HLA-A, com 20 alelos identificados para cada conjunto de dados. Por último, para o gene HLA-C e HLA-DRB1 foram reconhecidos 14 e 13 alelos, respetivamente para ambos os conjuntos de dados. Estes valores entram em concordância com os valores obtidos da base de dados internacional - IMGT/HLA- já que de acordo com esta, também o HLA-B é o gene mais polimórfico com 6537 alelos conhecidos, seguido do HLA-A com 5266 alelos, o HLA-C com 5140 alelos e o HLA-DRB1 com 2581 alelos (dados atualizados em julho de 2019). Nesta base de dados é também possível perceber-se a evolução do número de alelos identificados nos últimos anos. Esta é alvo de uma atualização constante de três em três meses em que se regista um aumento constante do número de alelos identificados. Os últimos dados apresentados no mês de julho de 2019 apresentam uma identificação de 23907 alelos, um aumento de 1379 alelos face aos 22528 alelos identificados até abril de 2019.

Os *loci* HLA são o *cluster* de genes mais polimórficos em humanos. Desde a descoberta do sistema HLA, a aplicação clínica mais importante deste, tem sido a seleção de dadores adequados para transplante ou transfusão. Apesar da expansão dos registos de dadores, continua a ser um desafio, a compatibilidade HLA (*full match*) entre um dador e um recetor, devido à grande diversidade de alelos e haplótipos HLA. Doentes que são portadores dos alelos HLA mais comuns em haplótipos conservados têm maior probabilidade de encontrar dadores não relacionados do que aqueles com alelos raros ou haplótipos menos conservados. A mesma origem étnica é outro fator importante que aumenta a probabilidade de identificar um dador compatível (Lima e Alves, 2013).

Num estudo de Lima e Alves (2013) foi feita a análise das frequências alélicas e dos haplótipos a partir de quase 38.000 dadores voluntários de medula óssea recrutados no Norte de Portugal. Neste estudo os alelos mais frequentes são os mesmos encontrados no presente estudo, sendo os alelos HLA-A*02, -B*44, -C*07 e -DRB1*07, mantendo-se assim uma concordância dos resultados, apesar do tempo decorrido entre os dois estudos.

Em São João *et al.* (2013) foi realizado uma análise das frequências alélicas do HLA de 1021 indivíduos selecionados aleatoriamente a partir do registo do banco português de dadores voluntários de medula óssea. Os alelos mais frequentes obtidos foram o HLA-A*02 (34,3%), o HLA-B*44 (16%) e o HLA-DRB1*13 (18,3%). Os resultados obtidos por estes autores são concordantes com o nosso, à exceção do *locus* DRB1, no entanto seguido deste apresenta-se o -DRB1*07 com frequência de 15,9% o que o torna um dos alelos HLA-DRB1 mais frequente. Neste estudo não foram considerados os alelos de HLA-C. De acordo com estes autores estas frequências são concordantes com alelos clássicos de populações europeias caucasianas, logo também os nossos resultados estão de acordo com este tipo de população.

Por fim, num estudo mais recente de Bettencourt *et al.* (2016) referente à população açoriana, também foram registadas as frequências alélicas para os *loci* HLA-A, -B, -C e -DRB1 num total de 130 indivíduos. Os alelos mais frequentes obtidos foram o HLA-A*02 (53,6%), o HLA-B*35 (28,2%), o HLA-C*07 (40,4%) e por fim o HLA-DRB1*07 (31,6%). Neste estudo, o único alelo que não está em concordância com os resultados aqui expressos é o HLA-B*35, que no nosso estudo corresponde ao segundo alelo mais frequente. Este é um alelo também comum a populações europeias, tal como é possível observar-se na tabela 3 cujo alelo é o mais frequente em estudos de populações na França, Espanha, Inglaterra, Sérvia, Itália e Grécia. A caracterização da composição dos alelos HLA na população portuguesa é importante para perceber mecanismos relacionados com várias doenças

autoimunes e para organizar registos de dadores voluntários de medula óssea e cordão. Estes dados constituem uma contribuição importante para estudos antropológicos e de associação de doenças envolvendo a população portuguesa (**Lima e Alves, 2013**).

Ao longo da tabela 3 estão sumarizados os alelos mais frequentes em dadores medula óssea e amostras de sangue do cordão para os *loci* HLA-A, -B, -C e -DRB1 a partir da análise na base de dados The Allele Frequency Net Database (<http://www.allelefrequencies.net/>). Esta base de dados permite selecionar cada *locus* individualmente e o tipo de estudo pretendido, compilando vários estudos de diferentes populações. Para além disso apresenta as frequências obtidas para cada alelo. Como parâmetro de escolha foram selecionados apenas estudos de genotipagem HLA publicados na literatura.

Tabela 3. Frequências alélicas para os *loci* HLA-A, -B, -C e -DRB1 em estudos de genotipagem de dadores de medula óssea em diferentes populações (<http://www.allelefrequencies.net/>).

População	HLA-A	HLA-B	HLA-C	HLA-DRB1
Brasil	02	07		04
República Checa	02	07	07	03
Inglaterra	02	35	12	
França	02	44;35	07	13
Alemanha	02	07	07	07
Grécia	02	35	07	11
Irlanda do Norte	02	07	07	04
Itália	02	35	07	11
Polónia	02	07	07	07
Rússia	02	44	07	13
Sérvia	02	35	07	11
Espanha	02	14; 35	07	01
Suíça	02	07	07	07
País de Gales	02	44		04

Como já referido anteriormente a genotipagem HLA é um grande desafio devido aos elevados polimorfismos deste. Inicialmente a genotipagem era feita através de métodos serológicos, onde eram detetadas variantes proteicas na superfície das células usando anticorpos específicos para antigénios. No entanto esta não era completamente satisfatória a nível clínico na transplantação de medula, principalmente em casos de dadores não relacionados. Deste modo, foi necessário desenvolver métodos de genotipagem que identificassem e combinassem dador-paciente a nível alélico do HLA, sendo estes: PCR-SSO, PCR-SSP e PCR-SBT. Estes métodos de genotipagem expressam resultados com diferentes níveis de ambiguidade, dependendo do tipo de metodologia e *kit* usado (**Paunic et al., 2012**). As figuras seguintes ilustram um exemplo para cada metodologia de genotipagem HLA utilizada no Laboratório de Genética Molecular do IPST- transplantação Porto.

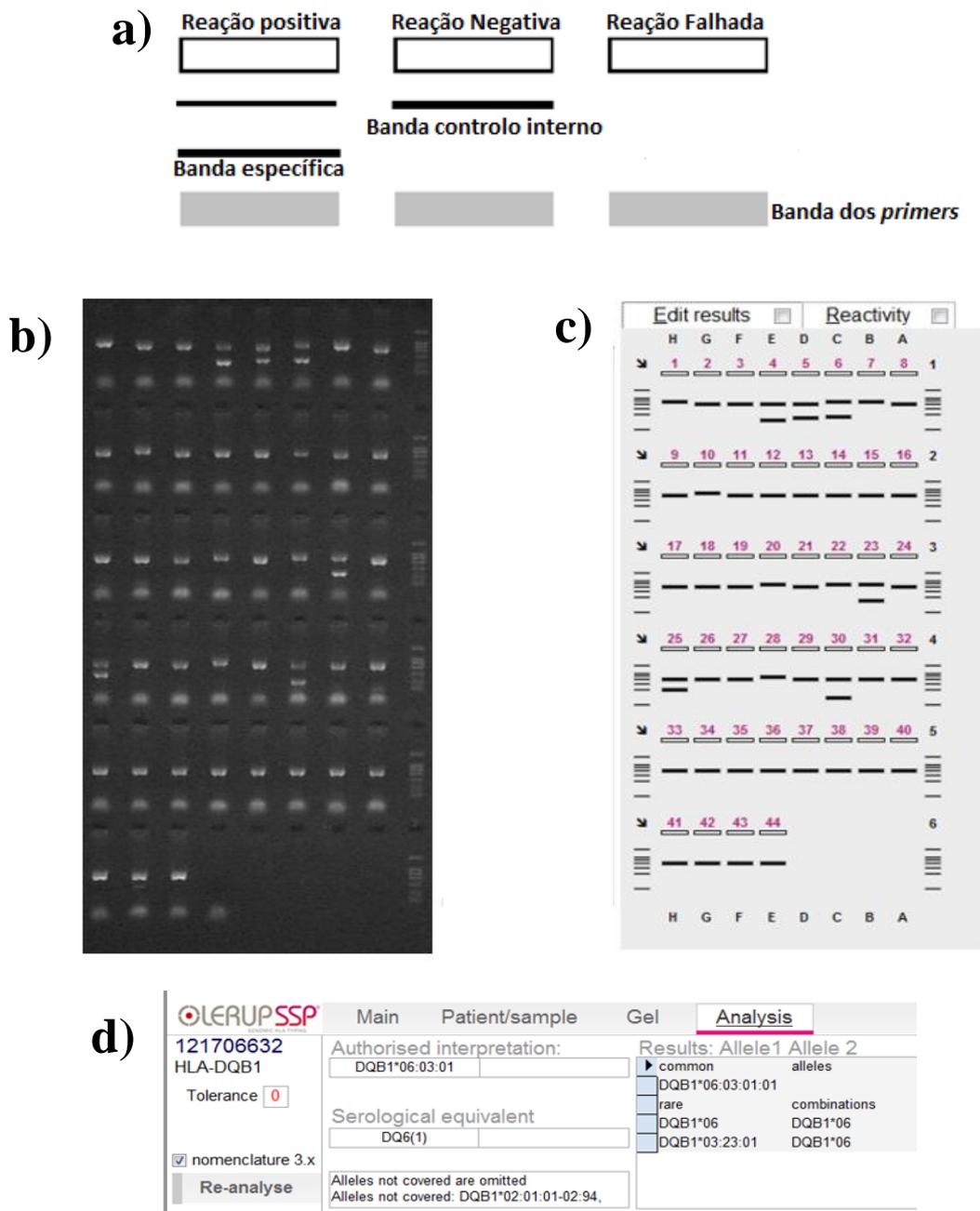
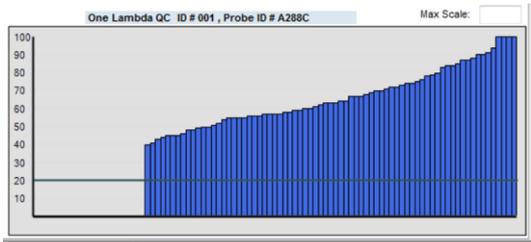


Figura 7. Interpretação pelo *software SCORE* no programa *HELMBERG-SCORE*, para a genotipagem HLA Classe I e II pelo kit *OLERUP SSP*. a) Interpretação universal da presença ou ausência de produtos específicos de PCR. As reações positivas correspondem aos poços que amplificam uma banda específica de tamanho pré-determinado pelo kit. Nas reações negativas verifica-se apenas a existência banda específica do controlo interno. Numa reação em que não haja amplificação não há amplificação de nenhuma banda. b) Gel de agarose para a eletroforese da nossa amostra. c) Corresponde ao resultado da eletroforese obtido em b) dada pelo operador ao *software SCORE*. Aqui o operador seleciona as bandas positivas obtidas na eletroforese, de forma a que o *software* possa dar um resultado final. d) Resultado final da genotipagem HLA-DQB1*06 para a nossa amostra, ou seja, temos a presença do alelo DQB1*06:03:01

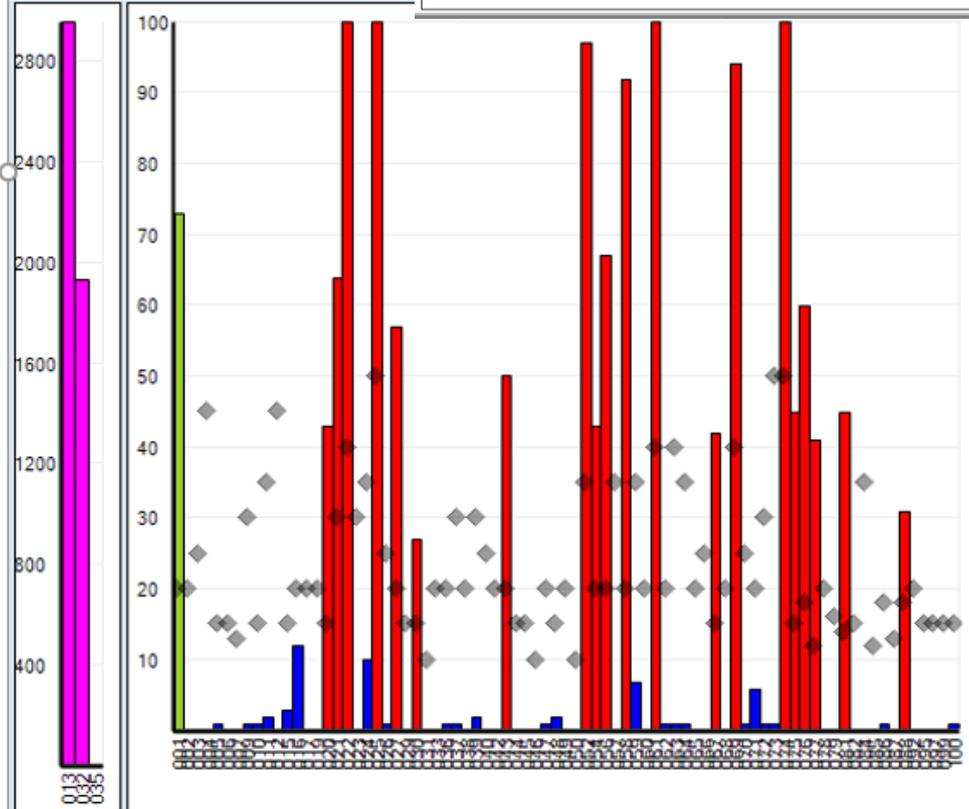
a)



b)



c)



d)

Pairs	Force	Type/SubType	Match	Sero
Local Demographic Data [CWD_TDX_3.29.1]				

Group 1: frequent on both alleles				

A*01:01:01:01	A*26:01:01:01 (G1)			
A*01:01:01:01	A*26:15 (G1)			
A*01:01:01:01	A*26:17 (G1)			
A*01:04N	A*26:01:01:01 (G1)			
A*01:04N	A*26:15 (G1)			
A*01:04N	A*26:17 (G1)			

e)

Possible Allele Code
A*01:AYFPX A*26:AYFPY
AYFPX=01:01:01:04N/01:09/01:11N/01:15N/01:16N/01:17/01:18N/01:22N
AYFPY=26:01:26:10/26:15/26:17/26:20/26:23/26:24/26:25N/26:26/26:32/

Figura 8. Histogramas da interpretação da tecnologia *Luminex* através do software *HLA Fusion* da *One Lambda*. a) Histograma que revela o comportamento da *bead* selecionada, neste caso a *bead* 001. Aqui, as barras estão acima do *cut-off* revelando um comportamento positivo ao longo das amostras. b) Apresenta o comportamento das amostras que foram genotipadas, sendo que a vermelha se encontra a nossa amostra. c) Este histograma corresponde ao resultado da amostra para cada *bead*. No canto esquerdo temos três barras que correspondem aos controles, positivo e negativo: as barras rosa são os controles positivos, neste caso a *bead* 013 e 032; a barra incolor é o controle negativo (*bead* 035). As barras azuis e vermelhas revelam o comportamento de cada *bead* para a amostra. Os diamantes a cinza são o valor de *cut-off* para cada *bead*, deste modo se a fluorescência for inferior a este *cut-off* o resultado é negativo (barras azuis), caso ocorra o contrário a *bead* tem

um comportamento positivo (barras vermelhas). A barra verde corresponde à *bead* selecionada e está a ser avaliada no histograma b). d) Lista de todos os possíveis alelos como resultado da genotipagem. O primeiro corresponde à genotipagem mais correta para a amostra e as seguintes por ordem decrescente como possíveis hipóteses de genotipagem. e) Resultado final da genotipagem, em baixa resolução, corresponde aos alelos HLA-A*01; HLA-A*26.

a)

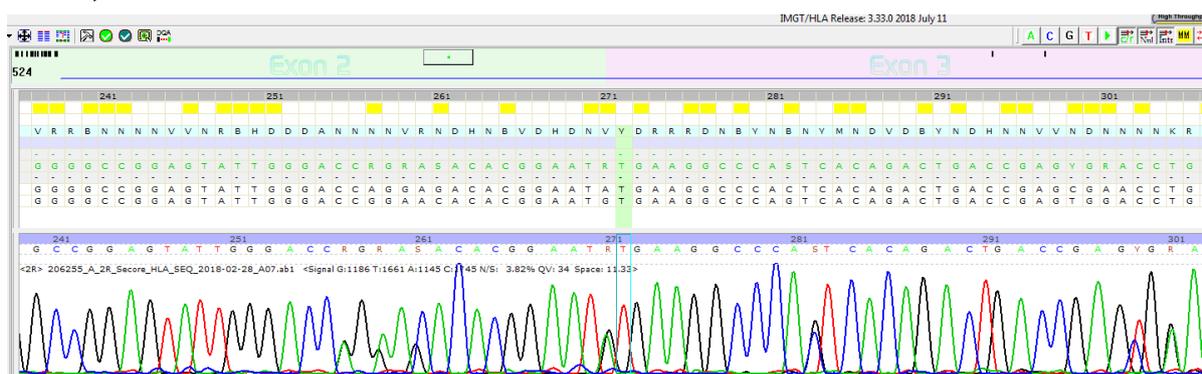


b)

Allele 1	Allele 2	Coverage	MM	MM1	MM2	MM...	MM...	MM...
A*01:01:01...	A*02:05:0...		0	0	0	0	0	0
A*01:01:01...	A*02:05:0...		0	0	0	0	0	0
A*01:01:01...	A*02:05:0...		0	0	0	0	0	0
A*01:01:01...	A*02:05:0...		0	0	0	0	0	0
A*01:01:01...	A*02:05:0...		0	0	0	0	0	0
A*01:234	A*02:05:0...		0	0	0	0	0	0
A*01:234	A*02:05:0...		0	0	0	0	0	0
A*01:01:01...	A*02:05:0...		0	1	0	0	0	0
A*01:221	A*02:172		0	1	1	0	0	0
A*01:01:68	A*02:05:03		0	1	1	1	1	1
A*01:01:21	A*02:05:05		0	1	1	1	1	1

Figura 9. Genotipagem em alta resolução obtida pelos kits *HLAssure Texas BioGene* e interpretação pelo *software ACCUTYPE*. A tipagem em baixa resolução (HLA-A*01, *02) já era conhecida pelo que se procedeu à sequenciação por separação alélica com *primers* diferentes (A-01 e A-03) a) Eletroferogramas das sequências *Foward* e *Reverse* para cada conjunto de *primers*. Foram utilizados os conjuntos de *primers* A-01 e A-03, de forma a amplificar o alelo A*01 e A*02, respetivamente. A sequência a preto assinalada como “Reference” corresponde à sequência da livraria HLA-A. As sequências a rosa e azul por baixo da referência correspondem às respetivas sequências dos alelos apresentados (A*01:01:01:10 e 02:05:01:01). A sequência a preto logo por baixo representa a combinação destes dois alelos. Cada eletroferograma dispõe também da sua leitura e o resultado final da combinação das sequências dos eletroferogramas é o que está representado a azul em “consensus”. b) O resultado final foi a genotipagem HLA-A*01:01; HLA-A*02:05, com nenhum *mismatch*. Ao longo da lista estão representados todos os possíveis alelos para esta genotipagem por ordem decrescente de possibilidade.

a)



b)

MM	(7)	Allele	Info
0	A*01:01:01:01	A*69:01:01:01	
0*	A*01:04:01:01N	A*69:01:01:01	
1	A*01:03:01:01	A*69:01:01:01	
1	A*01:09:01:01	A*69:01:01:01	
1	A*01:17	A*69:01:01:01	
1	A*01:25	A*69:01:01:01	
2	A*36:04	A*69:01:01:01	

Figura 10. Genotipagem HLA pelos kits *SeCore* e interpretação no *software uTYPE*. a) Eletroferograma para a sequência genotipada. A sequência a preto corresponde à base da livraria para o gene HLA-A, as duas sequências juntas a preto correspondem às leituras de bases para os dois alelos e a sequência a verde, a combinação destes dois alelos. b) Resultado final da genotipagem, sendo os alelos HLA-A*01:01:01:01 e HLA-A*69:01:01:01, com zero *mismatch*.

No laboratório de genética molecular do IPST-transplantação Porto a genotipagem HLA ainda não é realizada por NGS, mas sim pelos métodos mais tradicionais. A genotipagem HLA neste laboratório aplica-se a amostras de doadores de medula óssea e banco do cordão, bem como aos doadores e recetores candidatos a transplantes e pesquisa de alelos HLA associados com doenças. Na rotina laboratorial e diária todas as análises que chegam ao laboratório são prescritas por um médico com determinada finalidade:

- Genotipagem HLA-B*27 para confirmar o diagnóstico de espondilite anquilosante. Durante o período de estágio no laboratório (setembro de 2017 a julho de 2018) foram requisitados 10 pedidos de pesquisa HLA-B*27. A metodologia usada é a PCR-SSO usando o kit *Labtype* rSSO através da tecnologia *Luminex*. Foram obtidos dois resultados positivos. Nestes casos, esta genotipagem permite um diagnóstico confirmativo para a doença levando a um início do tratamento direcionado para a doença;

- O pedido de genotipagem HLA-B*57:01 inicia com a metodologia de PCR-SSO usando o kit *Labtype* rSSO através da tecnologia *Luminex* e nos casos de HLA-B*57 positivos é confirmado o subtipo B*57:01 por PCR-SBT. No mesmo período referido anteriormente, foram requisitados 236 pedidos de pesquisa para este alelo HLA, com treze resultados positivos. A genotipagem confirmatória deste alelo informa o médico que terá de prescrever

um antirretroviral que não seja o abacavir, prevenindo que o doente desenvolva uma reação de hipersensibilidade a este medicamento frequentemente utilizado como antirretroviral;

- A nível de transplantação hematopoiética as requisições solicitadas englobam a genotipagem para definição alélica dos *loci* HLA-A, -B, -C, -DRB1, -DQA1, -DQB1, -DPA1 e -DPB1 do doente conforme solicitação do médico. Um possível dador pode ser um familiar do doente, HLA compatível, ou na ausência deste, um dador inscrito no banco de dadores de medula óssea. No primeiro caso podem ser requisitados pelo médico estudos da família sendo realizada a genotipagem dos *loci* HLA-A, -B, -C, -DRB1 e -DQB1 de familiares diretos pelas técnicas de genotipagem HLA em baixa resolução como o PCR-SSO ou PCR-SSP. Nos casos de incompatibilidade HLA dentro da família, a pesquisa de dadores inscritos no banco de dadores de medula óssea implica uma genotipagem do doente e do seu dador em alta resolução pelas técnicas de PCR-SBT. A nível de transplantação de órgãos sólidos, os dadores e recetores são estudados em baixa resolução usando tanto as técnicas de PCR-SSO como PCR-SSP, sendo esta mais utilizada no caso de um dador cadáver, permitindo uma rápida definição alélica para os *loci* HLA-A, -B, -C, -DRB1 e -DQB1 e a nível laboratorial é destacada como uma Urgência, tornando-se prioridade sobre as restantes análises de rotina.

A plataforma <http://ipst.pt/> fornece dados estatísticos sobre a atividade nacional na doação e transplantação de órgãos relativamente ao primeiro trimestre de 2019. Relativamente à tipologia dos dadores temos 1 dador sequencial (recetor de um transplante de órgão (fígado), cujo órgão nativo pode ser considerado para transplantação noutra doente, 14 dadores vivos (pessoa que doou um rim ou porção de fígado, em vida), 6 dadores em paragem circulatória (dador falecido a quem foi declarada a morte com base em critérios circulatórios, verificando-se a cessação irreversível das funções cardiocirculatórias) e 74 dadores em morte cerebral (dador falecido a quem foi declarada a morte com base em critérios neurológicos, verificando-se a cessação irreversível das funções do tronco cerebral). Foram colhidos 230 órgãos para transplante e foram transplantados 212, dando uma taxa de utilização de 92% dos órgãos. Dos transplantes realizados 8 foram cardíacos, 133 renais, 55 hepáticos, 6 pulmonares e 10 pancreáticos. Relativamente aos dadores CEDACE, a plataforma apenas fornece dados até 2015, onde foi atingido 385323 genotipagens de dadores de medula óssea e 2415 dádivas de sangue do cordão umbilical.

A viabilidade da genotipagem HLA por NGS tem sido demonstrada por vários estudos, uma vez que esta tecnologia caracteriza completamente o *cluster* de genes HLA. Este desenvolvimento tem sido um marco no campo imunogenético permitindo a caracterização

conjunta de vários genes polimórficos, que são fundamentais em muitos processos imunológicos. Com o passar dos anos torna-se visível a disponibilidade de produtos comerciais para genotipagem HLA por NGS, devido ao crescente número de relatos na literatura descrevendo os seus benefícios, tornando-se a sua utilização uma realidade presente.

A tabela 4 apresenta as vantagens *versus* desvantagens que levaram a evolução a partir da serologia para a genotipagem genética.

Tabela 4. Vantagens e desvantagens que levaram à evolução da genotipagem serológica para a genotipagem genética.

Genotipagem Serológica	VS	Genotipagem Genética
<ul style="list-style-type: none"> ☺ Resultados rápidos ☹ Baixa especificidade ☹ Caiu em desuso 		<ul style="list-style-type: none"> ☺ Mais flexível ☺ Mais robusta ☺ Resultados mais específicos

A tabela 5 apresenta uma lista de vantagens e desvantagens, com base em diferentes estudos retratados ao longo desta revisão, comparando as ambiguidades entre os métodos tradicionais e a tecnologia NGS.

Tabela 5. Resumo das vantagens e desvantagens para os métodos convencionais e para a tecnologia NGS.

Genotipagem Genética				
PCR-SSP	PCR-SSO	rPCR-SSO	PCR-SBT	NGS
<ul style="list-style-type: none"> ☺ Elevada especificidade dos <i>primers</i> ☺ Rápida ☹ Pequeno número de amostras ☹ Desenvolvimento constante de novos conjuntos de <i>primers</i> 	<ul style="list-style-type: none"> ☺ Mais prática ☺ Grande volume de amostras ☹ Não distingue polimorfismos cis e trans 	<ul style="list-style-type: none"> ☺ Reduzido número de amostras ☺ Resolução de ambiguidades cis e trans 	<ul style="list-style-type: none"> ☺ Maior exatidão ☺ Maior precisão ☺ Maior resolução alélica ☹ Resolução de ambiguidades com amplificações adicionais 	<ul style="list-style-type: none"> ☺ Sequenciação de todo o genoma ☺ Elevado rendimento ☺ Alta resolução

Através da entropia de Shannon (mede o grau médio de incerteza), **Paunic et al. (2012)** demonstraram que a genotipagem SBT é a que apresenta níveis de entropia menores e consequentemente menor ambiguidade. Quanto à genotipagem SSO, esta apresentou mais nível de ambiguidade, logo apresenta uma menor cobertura que o SBT. Quanto à genotipagem serológica, que também foi avaliada neste estudo, é a que apresenta maior nível de entropia e menor resolução. Apesar de a metodologia SBT apresentar resultados com menor ambiguidade, estes autores afirmam que alguns laboratórios dão preferência ao método SSO, devido ao custo mais baixo e mais fácil configuração e manuseamento.

Num estudo de **Mellet et al. (2015)**, foi feita a genotipagem em alta e baixa resolução de 20 amostras provenientes banco de dados de medula óssea da África do Sul. Tendo em

consideração a grande diversidade genética desta população, reduzindo deste modo a probabilidade de encontrar um par dador-recetor 100% compatível na transplantação hematopoiética, tornou-se expectável que a tecnologia NGS seria uma mais-valia na determinação alélica, principalmente nas genotipagens com ambiguidades. Assim, fez-se uma comparação dos resultados obtidos pelas técnicas convencionais, neste caso SBT e SSP, com os resultados obtidos pela tecnologia NGS, mais especificamente a plataforma *Roche454*. Os resultados apresentaram-se muito semelhantes entre si com uma concordância de 99% em 10 amostras e 98% nas restantes 10. A discordância foi observada apenas em quatro amostras, sendo que estas discrepâncias foram resultado de genotipagens ambíguas, em que a genotipagem foi atribuída mediante a probabilidade de alelos mais comuns. Neste estudo concluiu-se que a resolução alélica do NGS não apresentava um incremento suficiente de forma a justificar economicamente a sua implementação. No entanto há que ter em consideração o número de amostras reduzido deste estudo.

Num outro estudo desenvolvido por **Gowda et al. (2016)** com 1020 indivíduos da população indiana, a genotipagem HLA pelos métodos tradicionais de baixa resolução como o SSP apresentou-se pouco eficiente em transplantes de órgãos sólidos e medula óssea. Neste estudo foram selecionadas 80 amostras das 1020 cuja genotipagem era a correspondente aos alelos mais frequentes dessa população. As 80 amostras foram sequenciadas por NGS, usando a plataforma *Roche*, e 8 destas foram sequenciadas na sua totalidade usando o sequenciador de terceira geração, o *PacBio*. Os resultados obtidos pela tecnologia NGS demonstraram 95% de concordância com os dados obtidos pelas técnicas convencionais de SSP. No entanto as limitações desta técnica foram essencialmente as suas leituras curtas, focando-se apenas a alguns exões perdendo-se informação de outras regiões intrónicas e exónicas. Para além disso, acarreta mais erros em torno de sequências homopolimórficas. Com a tecnologia *PacBio*, que recorre a leituras mais longas envolvendo toda a região codificante e não codificante, foi possível a descoberta de três novos alelos. Neste estudo concluiu-se que tanto as técnicas convencionais como a tecnologia *Roche* apenas conseguem obter informações proteicas do HLA dos exões estudados, o que difere da plataforma *PacBio* que fornece informações completas sobre o HLA incluindo regiões codificantes e não codificantes, como intrões e UTRs. Para além disso é uma técnica que define o haplótipo HLA completo, tornando-a crucial na transplantação e estudos haplotípicos para a deteção de novos alelos na população Indiana. A única contrapartida desta tecnologia é o seu custo mais elevado por amostra em

comparação com as outras técnicas. No entanto, os autores consideram que este custo irá diminuir com o desenvolvimento de novas tecnologias.

Várias tecnologias NGS estão disponíveis no mercado, não havendo, no entanto, prevalência de uma única metodologia em relação a outra. É necessário identificar a combinação certa de protocolos e plataformas de sequenciação que serão utilizados para diferentes fins e necessidades (Monos e Maiers, 2015). A tabela 6 demonstra características de cada uma das plataformas NGS, bom como os seus pontos fortes e fracos.

Tabela 6. Visão geral sobre as plataformas NGS usadas em genotipagem HLA e comparação de suas características (Adaptado de Carapito *et al.*, 2016).

Plataformas	<i>Roche GS FLX</i>	<i>Ion Torrent PGM</i>	<i>Illumina MiSeq</i>	<i>Pacific Bioscience</i>	<i>Oxford Nanopore Minlon</i>
Preparação da amostra	PCR emulsão	PCR emulsão	Pontes de amplificação	Sem amplificação	Sem amplificação
Química de sequenciação	Pirosequenciação	lões semi-condutores	Sonda terminadoras reversíveis	Sequenciação de uma molécula única em tempo real	Nanoporos
Comprimento de leitura máximo	1000 b	400 b	2 x 300 b	40 Kb	50 Kb
Rendimento máximo por corrida	700 Mb	2 Gb	15 Gb	260 Mb	400 Mb
Tempo médio	23h	3 h	56 h	3 h	48h
Precisão média	99,9%	99%	99,9%	87%	72 – 80%
Tipo erros primários	Indels	Indels	Substituições	Inserções	Indels
Pontos fortes	-Longas leituras -Elevado rendimento	-Tempos de sequenciação rápidos	-Elevado rendimento -Baixos custos -Sequenciação final emparelhado	-Longas leituras -Sequências consensos circulares -Baixos tempos de sequenciação	-Longas leituras -Baixo custo
Pontos fracos	-Erros com homopolímeros -Elevado custo	-Erros com homopolímeros	-Tempo de sequenciação elevado	-Elevada taxa de erro -Baixo rendimento	-Elevada taxa de erro -Baixo rendimento

As melhorias proporcionadas pela tecnologia NGS relativamente aos problemas enfrentados pela genotipagem HLA por métodos tradicionais são muito significativas, pois formam a base para a caracterização precisa e completa dos genes HLA. Isso, por sua vez, provará ser um desenvolvimento que influenciará vários campos interdependentes, como autoimunidade, transplantes, cancro, desenvolvimento de vacinas, farmacogenómica, doenças infecciosas e outras que nem podemos suspeitar atualmente (Duke *et al.*, 2016).

Desde 2015 que alguns laboratórios, noutros países, que possuem programas de transplante de células hematopoiéticas e de órgãos sólidos, usam o NGS para fornecer

genotipagens de HLA para ambos os programas (**Edgerly e Weimer, 2018**). A introdução da tecnologia NGS nos laboratórios traz inúmeros benefícios:

1) O NGS é uma tecnologia relevante na seleção de doadores não só para o transplante de medula, mas também de órgãos sólidos, já que permite uma maior compreensão das incompatibilidades HLA (**Monos e Maiers, 2015**). As ambiguidades do genótipo HLA geralmente requerem testes adicionais significativos para garantir a correspondência precisa entre dador e paciente. No entanto, muitas destas ambiguidades por vezes não são resolvidas devido à presença de alelos raros ou envolvem outras áreas com efeitos pouco claros sobre os desfechos clínicos. A tecnologia NGS tem o potencial de superar essas limitações através do processamento paralelo de milhões de sequências para genotipagem HLA de alta resolução. A natureza desta tecnologia facilita a sequenciação de genes completos, permitindo a avaliação de polimorfismos fora dos exões tradicionalmente examinados. É importante ressaltar que a natureza encadeada das sequências de DNA geradas pela tecnologia NGS permite a determinação para locais heterozigóticos, reduzindo uma fonte significativa de ambiguidades. Estes benefícios melhoram a precisão da genotipagem do HLA, facilitam a identificação de alelos raros ou novos e minimizam a possibilidade de complicações clínicas da rejeição do transplante ou doença do enxerto contra o hospedeiro (**Allen et al., 2018 e Seshasubramanian et al., 2019**).

2) Adicionalmente, permite expandir os nossos horizontes em relação à relevância biológica das sequências HLA tanto as codificantes como as não codificantes (**Monos e Maiers, 2015**). A tecnologia NGS ao fornecer informações completas sobre a sequência de todos os *loci* HLA numa única etapa, permite analisar a variabilidade em regiões não codificantes e de mutações sinónimas representando uma oportunidade de melhoria para a correspondência de doadores e recetores no transplante de medula óssea. No entanto, avaliar os efeitos clínicos da compatibilidade em regiões não codificantes dos genes HLA exigirá analisar um maior número de amostras, com informação multicentro e, de preferência, internacional. (**Allen et al., 2018 e Seshasubramanian et al., 2019**).

3) Para além destas vantagens, temos o seu custo. Em 2015, **Monos and Maiers** afirmaram que este tem vindo a decair com o aumento de produtos no mercado, a sua utilização em grande escala bem como a otimização dos protocolos aliada com a automatização resulta na simplicidade e poupança, tornando-a deste modo, uma técnica de genotipagem bastante competitiva relativamente às tecnologias tradicionais. Um estudo mais recente, **Seshasubramanian et al., (2019)**, também caracteriza a tecnologia NGS como um

método económico e preciso, principalmente após a primeira etapa de aquisição do equipamento e sobretudo quando o número de amostras processado é maior.

Para além das vantagens da aplicação clínica do NGS, um estudo de **Sanchez-Mazas e Nunes (2019)** apresenta quatro vantagens destacam-se para o uso do NGS de forma a se perceber a diversidade populacional do HLA. Como primeira abordagem o uso do NGS possibilita uma maior identificação de alelos HLA associados a doenças. Esta tecnologia ao reduzir substancialmente as ambiguidades e melhorar as análises estatísticas fazendo menos suposições, permite a obtenção de resultados mais claros para as associações de doenças com determinados alelos HLA. Segundamente, disponibiliza novas formas de detetar assinaturas demográficas na variação populacional, ou seja, é possível determinar diferenças de nucleótidos entre pares de sequências em uma dada população. As técnicas tradicionais de genotipagem HLA realizadas em centenas de populações de todo o mundo contribuíram para distribuições de frequências alélicas úteis de forma a descrever a variação genética HLA dentro e através dos diferentes continentes e para confirmar o papel principal das migrações humanas na formação dos padrões genéticos de HLA observados. Nestes estudos, no entanto, apenas uma parte limitada da informação relativa a cada população foi levada em conta e como a variação molecular também é encontrada fora dos exões 2 e 3, uma determinação mais detalhada da variação de nucleótidos em todos os genes HLA usando tecnologias NGS permitiria aperfeiçoar os resultados fornecidos por este tipo de abordagem. Esta medida torna-se útil para detetar assinaturas de eventos demográficos e/ou pressões seletivas, como aquelas que afetam os genes HLA nas populações. Os efeitos da demografia ou da seleção natural sobre os padrões de diversidade de HLA seriam mais fáceis de separar comparando-se a diversidade de nucleótidos de diferentes regiões codificantes com as regiões não codificantes que não seriam alvos diretos de seleção ou que seriam submetidas a tipos de pressões seletivas. A geração de novos dados do NGS para o enorme conjunto de amostras populacionais que já foram genotipadas para HLA com técnicas tradicionais constituía claramente uma excelente estratégia para melhorar o conhecimento sobre a história das populações. Para além disso, permite decifrar os processos evolutivos que moldaram os padrões de diversidade do HLA. A maior proporção de variação molecular de HLA foi relatada dentro dos exões HLA que codificam a região peptídica de ligação ao antigénio, sendo exão 2 e 3 para classe I e exão 2 para genes de classe II, aplicando-se a genotipagem a esses exões durante várias décadas. No entanto, os trabalhos publicados ao longo dos anos apenas se aplicam algumas amostras incapazes de cobrir a enorme diversidade da região HLA

descrita hoje dentro e entre diferentes populações em todo o mundo, assim em dados populacionais extensos podem-se produzir resultados substancialmente diferentes. Assim, os resultados obtidos até agora demonstram que as análises de NGS aplicadas aos dados de HLA são mais informativas para compreender os processos evolutivos que moldaram os padrões de diversidade de HLA observados em humanos. Por último, fornece dados promissores para entender os mecanismos envolvidos no processo imunológico. Vários estudos demonstram que os dados da sequência HLA podem ser altamente informativos para compreender os mecanismos envolvidos no processo imunológico. Estudos recentes sugerem relações de cooperação importantes envolvidos no processo imunológico entre alelos de *loci* distintos. De forma a compreender o modelo de cooperação entre alelos de *loci* distintos é necessário estender as genotipagens além dos exões onde a variabilidade alélica é frequentemente relatada, ou seja, nas regiões não codificantes. Estas podem ter envolvidos mecanismos de regulação, como níveis de expressão gênica e é aqui que a tecnologia NGS pode fornecer informações essenciais.

No geral, estas vantagens têm o potencial de fornecer informações completas, precisas e possibilitando descobertas que não eram possíveis até hoje. No entanto, a tecnologia NGS também acarreta alguns problemas, como o tempo necessário para gerar os dados informáticos. Enquanto as tecnologias mais convencionais são capazes de processar os seus dados em intervalos de algumas horas, o NGS despende alguns dias no processamento da informação. Outro desafio que esta tecnologia enfrenta surge no desenvolvimento de programas de *software* que apresentem resultados de genotipagem com um elevado grau de confiança e reprodutíveis. Tendo em conta que esta tecnologia gera uma infinidade de dados numa única sequenciação, o programa de software deve permitir a avaliação de muitos parâmetros de qualidade de modo que, em casos desafiantes como na identificação de novos alelos, o operador possa examinar os resultados obtidos em detalhe e possa relatar a genotipagem com precisão. Para tal são necessários estudos extensivos e completos, protocolos bem validados, simples e rentáveis de forma a promover o potencial científico desta nova informação derivada da caracterização detalhada dos genes (**Ghandi et al., 2016 e Monos e Maiers, 2015**). Por último, temos as limitações bioinformáticas, essencialmente devidas à sensibilidade na deteção de mutações, dificuldade em sequenciar certas áreas do genoma, dificuldade em interpretar raras e novas mutações e na capacidade limitada para detetar variações genéticas (**Yohe e Thyagarajan, 2017**).

Para além de se avaliar as vantagens e desvantagens da tecnologia NGS **Ghandi et al., (2016)** dão especial atenção a diferentes parâmetros que constituem a análise pré-analítica, analítica e pós-analítica necessária para a prática da genotipagem HLA por NGS em laboratórios de diagnóstico clínico. O sistema pré-analítico inclui as etapas de seleção das amostras, extração de DNA, controlo de qualidade e quantidade de DNA, reação de amplificação por PCR, quantificação e fragmentação dos amplicões, ligação do adaptador, e seleção do tamanho de sequenciação. Neste sistema é necessário ter em atenção a qualidade do DNA obtido pode limitar o método usado, logo tanto a quantidade como a qualidade do DNA devem ser dentro do intervalo sugerido para um desempenho ideal de PCR, o que é crítico para as etapas seguintes. Outra consideração importante é a capacidade de fornecer uma biblioteca que satisfaça muitas métricas pós-sequenciação, ou seja, alguns protocolos comerciais possuem conjuntos de *primers* multiplex para a amplificação combinada de múltiplos genes, simplificando o protocolo e reduzindo o número total de reações de PCR. No entanto, em consequência da complexidade do sistema, podem ocorrer consequências imprevisíveis, logo, os programas de *software* para analisar os dados do HLA são de extrema relevância. Relativamente à fragmentação do DNA em partes menores é necessária nas tecnologias *Illumina* e *Ion Torrent*, pois não podem sequenciar fragmentos maiores do que cerca de 1200 bases e 600 a 700 bases, respetivamente. Ao contrário dos sequenciadores de terceira geração, como o *PacBio* ou o *Minion*, a fragmentação não é necessária. Segundo estes autores é preferível a sequenciação de pequenos fragmentos, já que fornecem resultados com melhor qualidade e maior eficiência. No entanto, contrariamente a estes autores, **Weimer et al. (2016)** demonstraram que grandes fragmentos também podem gerar resultados eficazes e de alta qualidade, atribuindo as falhas na genotipagem à fraca qualidade da amostra e manuseio inadequado das esferas.

Quanto aos sistemas analíticos, a bioinformática tem um papel crucial para obter genotipagens precisas. Esta caracteriza-se por um passo inicial de remoção das sequências do adaptador e leituras de baixa qualidade, seguindo-se o alinhamento com uma sequência de referência e por fim a anotação das variantes. Devido ao elevado polimorfismo desta família de genes, uma única sequência de referência não é viável, logo os resultados obtidos são alinhados com todos os alelos HLA conhecidos e disponíveis no banco de dados IMGT/HLA. O programa de *software* deve ser flexível, de forma a que as suas ferramentas de análise forneçam informações relevantes e ajudem os analistas a avaliar a qualidade dos resultados obtidos. Por fim, ao nível dos sistemas pós-analíticos, estes devem estabelecer a simplicidade

e robustez do protocolo. Protocolos mais simples muito provavelmente são mais eficientes e rápidos, associados com menores taxas de erro, e conseqüentemente melhores tempos de resposta. A robustez do protocolo também é relevante porque ter resultados reprodutíveis é de extrema importância para os laboratórios em consequência do polimorfismo associado ao HLA o que pode levar com variações imprevisíveis. Relativamente ao tempo de geração de resultados são necessárias novas ferramentas para ajustar os protocolos atuais de modo a alcançar tempos de resposta mais curtos. Assim, tendo em conta todos estes parâmetros é necessário avaliar e escolher o melhor método para as necessidades de cada laboratório (**Ghandi et al., 2016**).

Relativamente à aplicabilidade clínica, nos últimos anos, vários estudos têm vindo a demonstrar a capacidade da tecnologia NGS em melhorar a genotipagem HLA. Em 2014, **Zhou et al.** demonstraram que através do uso da tecnologia NGS foi possível obter genotipagens sem ambiguidades em 97,7% dos resultados para o HLA-A, -B, -C e DQB1. A média para os alelos DRB1 foi de apenas 42,3%, no entanto os autores acreditavam que esta baixa taxa de sucesso se devia a terem genotipado apenas o exão 2. Mais tarde, **Allen et al. (2018)** determinou uma taxa de análise sem ambiguidades de 97,0% para a tecnologia NGS, resultado significativamente maior do que a dos métodos padrão atuais. Num estudo de **Seshasubramanian et al. (2019)** onde as amostras de dadores partiram de sangue do cordão, foi obtida uma precisão de 98,1% na genotipagem por NGS. Outro parâmetro importante avaliado neste estudo é a viabilidade do DNA obtido no pós-processamento de amostras de sangue do cordão ser adequado para genotipagem NGS. Noutros dois estudos, **Baier et al. (2018)** e **Smith et al. (2019)**, a aplicabilidade da tecnologia NGS em transplantação foi o enfoque dos estudos. Em ambos, a tecnologia NGS apresentou taxas de erros mais baixas, menores ambiguidades, deteção de novos alelos, menores custos, bem como o uso da genotipagem de alta resolução por NGS na rotina laboratorial para candidatos a transplante de órgãos sólidos fornece informações valiosas, como a identificação de alelos raros que podem ter um impacto devastador na sobrevida do enxerto. Para além disso, o NGS parece facilitar a monitorização pós-transplante para o desenvolvimento de anticorpos específicos do dador e facilita a genotipagem inicial para o registro de dadores de medula óssea.

A aplicação clínica do NGS não se limita só ao diagnóstico. Apesar de este presente trabalho não abordar este tipo de aplicação da tecnologia NGS é importante destaca-la brevemente. A evolução desta tecnologia, tem permitido detetar SNVs, rearranjos estruturais e monitorizar o DNA do tumor circulante, bem como analisar áreas do genoma desafiadora,

identificar mutações alvo para direcionar a terapia de tumores e identificar populações de alto risco para certos tipos de cancro hereditário (**Qin, 2019 e Yohe e Thyagarajan, 2017**).

Concluindo, as tecnologias NGS revolucionaram a pesquisa biológica e transformaram profundamente os campos da patologia diagnóstica e da medicina clínica. No futuro, o uso de NGS em laboratórios clínicos certamente aumentará como tecnologia e bioinformática, a fim de ultrapassar as limitações atuais e melhorar a qualidade dos resultados. O grande desafio para os laboratórios clínicos será o de realizar a abordagem mais adequada dos testes NGS, tendo em conta a relevância clínica, custo-efetividade e atendimento clínico do paciente (**Resta and Ferrari, 2018**).

A complexidade do sistema HLA e sua aplicação em transplantes continuará a impulsionar o desenvolvimento de novas tecnologias. Como já referido, inicialmente acreditava-se que apenas os exões da estrutura do gene HLA eram importantes, no entanto o conhecimento sobre o gene HLA na expressão dos alelos ou outros *loci* está a evoluir. A genotipagem de HLA via NGS forneceu informações sobre a importância potencial no grau de polimorfismo nos transplantes, em particular em transplantes de medula óssea. Deste modo, os laboratórios que suportam transplantes de medula óssea e órgãos sólidos, são potenciais candidatos à utilização desta tecnologia, uma vez que potencia a interpretação das especificidades entre o dador e o recetor. Para além disso, é expectável que o risco de transplante com incompatibilidades HLA mudará à medida que as tecnologias avançam e resultados mais detalhados podem ser produzidos em prazos curtos, de forma que a seleção de dadores seja mais rápida e mais específica (**Holdsworth et al., 2018**).

Também em consequência da elevada complexidade do *cluster* de genes HLA e a sua grande densidade de marcadores associados a diferentes doenças e características, ainda muito pouco é conhecido sobre os elementos genómicos exatos responsáveis pelos diferentes fenótipos. É de prever, que a sua detalhada caracterização revelará muitos segredos que podem eventualmente ajudar-nos a compreender o seu papel complexo e importante na saúde e doença (**Monos e Maiers, 2015**). Apesar do grande avanço tecnológico dos métodos tradicionais para a genotipagem por NGS é importante salvaguardar que o mundo HLA não deve se desviar de várias décadas de tecnologias tradicionais que deram origem a milhares de genótipos usados em muitos campos de pesquisa relacionados à imunogenética. Os vários tipos de informação “antiga” e “nova” - podem ser combinados para alcançar novos objetivos, já que a maior parte do nosso conhecimento sobre a diversidade HLA foi obtida antes do desenvolvimento de estratégias de sequenciação de alto rendimento, o que indica que técnicas

de sequenciação tradicionais ou consideradas mais limitadas possuem um grande potencial de histocompatibilidade tanto quanto função imune. É importante considerar que as diferentes tecnologias ainda desempenham um papel complementar na pesquisa de HLA. O NGS traz claramente dados novos, altamente relevantes e originais, mas isso não deve se traduzir em ignorar outras abordagens que ainda podem ser usadas por pequenos laboratórios e que podem fornecer resultados satisfatórios para alguns estudos (**Sanchez-Mazas e Nunes, 2019**).

Como qualquer tecnologia, o NGS na prática laboratorial evoluiu e irá continuar a evoluir (**Yohe e Thyagarajan, 2017**).

Genotipagem do *Human Leucocyte Antigen* (HLA):
Técnicas atuais e aplicabilidade da abordagem de *Next Generation Sequencing*

Capítulo 4

Conclusão



Os inúmeros estudos das últimas décadas permitiram a compreensão de parte das funções das moléculas de MHC classe I e de classe II na imunologia e na genética (**Rock, 2016**). As moléculas MHC regulam as respostas imunológicas específicas da apresentação do antígeno, ligando os péptidos derivados de agentes patogénicos e, em seguida, apresentando-os às células T-CD4⁺ ou CD8⁺, via MHC-II ou MHC-I, respetivamente. Para além destas funções o *cluster* de genes HLA possui outras funções importantes, em que certos alelos HLA são os principais determinantes genéticos da suscetibilidade, mas também proteção, a muitas doenças autoimunes ou reações de hipersensibilidade a fármacos. Para além destas funções, as moléculas de HLA desempenham um papel fundamental no controle da rejeição do transplante e parecem ter um papel importante no cancro à medida que se avança com abordagens individualizadas para prevenção de cancro e imunoterapia.

A genotipagem HLA é extremamente importante na clínica laboratorial. Tal como referido ao longo desta dissertação, os métodos PCR-SSP, PCR-SSO e PCR-SBT são amplamente utilizados na rotina laboratorial. O uso mais recente do NGS aumentou o entusiasmo a nível clínico em comparação com os métodos tradicionais, já que possui maior resolução e rendimento, menos ambiguidades e têm cada vez mais um menor custo associado.

O avanço na genotipagem HLA por NGS tem-se mostrado bastante promissor, pois permite a caracterização precisa e completa do *cluster* de genes HLA. Tendo em conta que a região HLA tem forte associação com doenças é expectável que ainda não se conheça todas as suas funcionalidades dos seus genes. A chegada das tecnologias de sequenciação via NGS, tem permitido uma maior caracterização dos genes HLA levando a mais perguntas da comunidade científica. O futuro do HLA poderá passar por transformar toda a abordagem, ou seja, se há alguma doença que é encontrada para ser associada com um SNP dentro do HLA há razões para procurar uma base imunológica possível dessa doença, independentemente do fenótipo aparente e qualquer não relação percebida dessa doença com a resposta imune.

O cálculo das frequências alélicas para as amostras de doadores de medula óssea voluntários e amostras de cordão a nível do Norte de Portugal são concordantes com as frequências alélicas obtidas por outros autores relativamente à população Portuguesa. Os alelos mais frequentes foram o HLA-A*02, HLA-B*44, HLA-C*07 e HLA-DRB1*07, cujos são comuns a populações europeias.

Concluindo, a sequenciação através da tecnologia NGS é uma opção viável para ser realizada em todo o tipo de laboratórios que tenham programas de transplante. O método é reprodutível, rentável, preciso, eficiente e fornece informações de sequenciação mais completas sobre os genes de HLA que as metodologias tradicionais. A genotipagem HLA bem-sucedida por NGS requer a combinação e moldagem de vários elementos, incluindo a plataforma de sequenciação, o protocolo de preparação da amostra e a análise bioinformática para fornecer resultados de genotipagem precisos. Esta dissertação fornece uma breve descrição dos métodos para genotipagem HLA e o principal elemento em estudo para um maior sucesso na genotipagem do HLA, será o NGS. É de esperar que a caracterização completa da região HLA leve a uma melhor compreensão dos seus papéis na saúde e na doença, bem como na genética das populações, transplantes e suscetibilidade a determinados fármacos.

Referências

- Allen**, E.S.; Yang, B.; Garrett, J.; Ball, E.D.; Maiers, M. e Morris, G.P. (2018) Improved accuracy of clinical HLA genotyping by next-generation DNA sequencing affects unrelated donor search results for hematopoietic stem cell transplantation. *Human Immunology*. **79**:848–854
- Allcock**, R.J.N. (2012). The Major Histocompatibility Complex: A Paradigm for Studies of the Human Genome. Em: Christiansen, F.T. e Tait, B.D. Immunogenetics: Methods and Applications in Clinical Practice. *Methods in Molecular Biology*. **882**: 1-6
- Anaya**, J.M.; Shoenfeld, Y.; Villarraga, A.R.; Levy, R.A.; e Cervera, R. (2013) Autoimmunity: From bench to Bedside. *El Rosario University Press*. 1-855
- Baier**, D.M.; Hofmann, J.A.; Fischer, H.; Rall, G.; Stolze, J.; Ruhner, K.; Lange, V.; Sauter, K. e Schmidt, A.H. (2018). Very low error rates of NGS-based HLA typing at stem cell donor recruitment question the need for a standard confirmatory typing step before donor work-up. *Springer Nature*. 1-3
- Bettencourt**, B.F.; Santos, M.R.; Pereira, J.; Amaro, B.; Fialho, R.; Meneses, R.; Couto, A.R. e Armas, J.B. (2016). HLA-A, -B, -C, -DQA1, -DQB1, -DRB1, -E, -F and -G genotyping of 130 individuals from Terceira Island, Azores, Portugal. *Human Immunology*. 1-2
- Bodis**, G.; Toth, V.; e Schwarting, A. (2018). Role of human leukocyte antigens (HLA) in autoimmune diseases. *Rheumatology and Therapy*. **5**(1): 5-20
- Bontadini**, A. (2012). HLA techniques: typing and antibody detection in the laboratory of immunogenetics. *Methods*. **56**(4): 471-476
- Carapito**, R.; Radosavljevic, M.; e Bahram S. (2016) Next-generation sequencing of the HLA locus: methods and impacts on HLA typing population genetics and disease association studies. *Human Immunology*. **77**(11): 1016-1023
- Choo**, S.Y. (2007). The HLA System: Genetics, Immunology, Clinical Testing and Clinical Implications. *Yonsei Medical Journal*. **48**(1): 11-23
- Dashti**, N.; Mahmoudi, M.; Aslani, S.; e Jamshidi, A. (2018) HLA-B*27 subtypes and their implications in the pathogenesis of ankylosing spondylitis. *Gene*. **670**: 15-21
- Dean**, L. (2015) Abacavir Therapy and HLA-B*57:01 Genotype. *Medical Genetics Summaries*. 1-12
- Duke**, J.L.; Lind, C.; Mackiewicz, K.; Ferriola, D.; Papazoglou, A.; Gaseiewski, A.; Heron, S.; Huynh, A.; McLaughlin, I., Rogers, M.; Slavich, L.; Walker, R.; e Monos, D.S. (2016). Determining performance characteristics of an NGS-based HLA typing method for clinical applications. *HLA Immune Response Genetics*. **87**(3): 141-152
- Dunckley**, H. (2012). HLA typing by SSO and SSP methods. Em: Christiansen, F.T. e Tait, B.D. Immunogenetics: Methods and Applications in Clinical Practice, Methods in Molecular Biology. *Springer Science + Business Media New York*. **882**: 9-25
- Edgerly**, C.H. e Weimer, E.T. (2018). The Past, Present, and Future of HLA typing in Transplantation. *Methods in Molecular Biology*. **1802**: 1-10
- Gabriel**, C.; Fürst, D.; Faé, I.; Wenda, S.; Zollikofer, C.; Mytilineos, J.; e Fischer, G.F. (2014). HLA typing by next-generation sequencing – getting closer to reality. *Tissue Antigens*. **83**(2): 65-75
- Gandhi**, M.J.; Ferriola, D.; Huang, Y.; Duke, J.L.; e Monos, D.; (2016). Targeted Next-Generation Sequencing for Human Leukocyte Antigen Typing in a Clinical Laboratory: Metrics of Relevance and Considerations for Its Successful Implementation. *College of American Pathologists*. **41**(6):806-812

Gowda, M.; Ambardar, S.; Dighe, N.; Manjunath, A.; Shankaralingu, C.; Hirannaiah, P.; Harting, J.; Ranade, S.; Jagannathan, L.; e Krishna, S. (2016) Comparative Analyses of Low, Medium and High-Resolution HLA Typing Technologies for Human Populations. *Journal of Clinical & Cellular Immunology*. **7**(2): 1-8

Harrison, L.C. (2018). Type 1 Diabetes. Clinical Immunology 5th Edition. *Elsevier*. 957-966

Hermann, C.; Trowsdale, J.; e Boyle, H. (2015). TAPBPR: a new player in the MHC class I presentation pathway. *Tissue antigens*. **85**(3): 155-167

Holdsworth, , .; Diviney, M. e Hart, C. (2018) Next Generation Sequencing and Histocompatibility in Hematopoietic Stem Cell Transplantation. *Revista de Medicina de Hematologia e Transfusão de Sangue*. **28**(3):275-279

Hosomichi, K.; Shiina, T.; Tajima, A.; e Inoue, I. (2015). The impact of next-generation sequencing technologies on HLA research. *Journal of Human Genetics*. **60**(11): 665-673

Howell, W.M. (2014). HLA and disease: guilt by association. *International Journal of Immunogenetics*. **41**(1): 1-12

Janeway, C.A.; Travers, P.; Walport, M.; e Shlomchik, M. (2001). Immunobiology: the immune system in health and disease. 5ª Edição. *Garland Science Publishing*.

Kallon, D.M. (2011). Histocompatibility & Immunogenetics: A collection of brief revision notes.e-book em: <https://histocompatibilityandimmunogenetics.com/>.

Kchouk, M.; Gibrat, J.F. e Elloumi, M. (2017). Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine*. **9**(3):1-8

Kennedy, A.E.; Ozbek, U.; e Dorak, M.T. (2017). What has GWAS done for HLA and disease associations? *International journal of immunogenetics*.**44**(5):195-211

Lima, B.A e Alves, H. (2013). HLA-A, -C, -B, AND -DRB1 allelic and haplotypic diversity in bone marrow volunteer donors from northern Portugal. *Organs, Tissues & Cells*. **16**:19-26

Liu, L.; Li, Y.; Li, S.; Hu, N.; He, Y.; Pong, R.; Lin, D.; Lu, L.; e Law, M. (2012). Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*. 1-11

Mahdi, B.M. (2013). A glow of HLA typing in organ transplantation. *Clinical and translational medicine*. **2**(6): 1-5

Mannon, P.J. (2018) Immunological Diseases of the Gastrointestinal Tract. Clinical Immunology 5th Edition. *Elsevier*. 1005-1019

Medini, D.; Serruto, D.; Parkhill, J.; Relman, D.A.; Donati, C.; Moxon, R.; Falkow, S.M e Rappuoli, R. (2008). Microbiology in the post-genomic era. *Nature Reviews Microbiology*. **6**(6): 419-430

Mellet, J.; Gray, C.M.; e Pepper, M.S. (2015). HLA typing: Conventional techniques v. next generation sequencing. *South African Medical Journal*. **106**(1): 88-91

Metzker, M. (2010). Sequencing technologies – the next generation. *Nature Reviews Genetics*. **11**(1): 31-46

Meyer, D. e Thomson, G. (2001). How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics*. **65**(1):1-26

Monos, D. e Maiers, M.J. (2015). Progressing towards the complete and thorough characterization of the HLA genes by NGS (or single-molecule DNA sequencing): Consequences, opportunities and challenges. *Human Immunology*. **76**(12): 883-886

Monos, D.S. e Winchesterm R.J. (2018) The Major Histocompatibility Complex. Clinical Immunology 5th Edition. *Elsevier*. 79-92

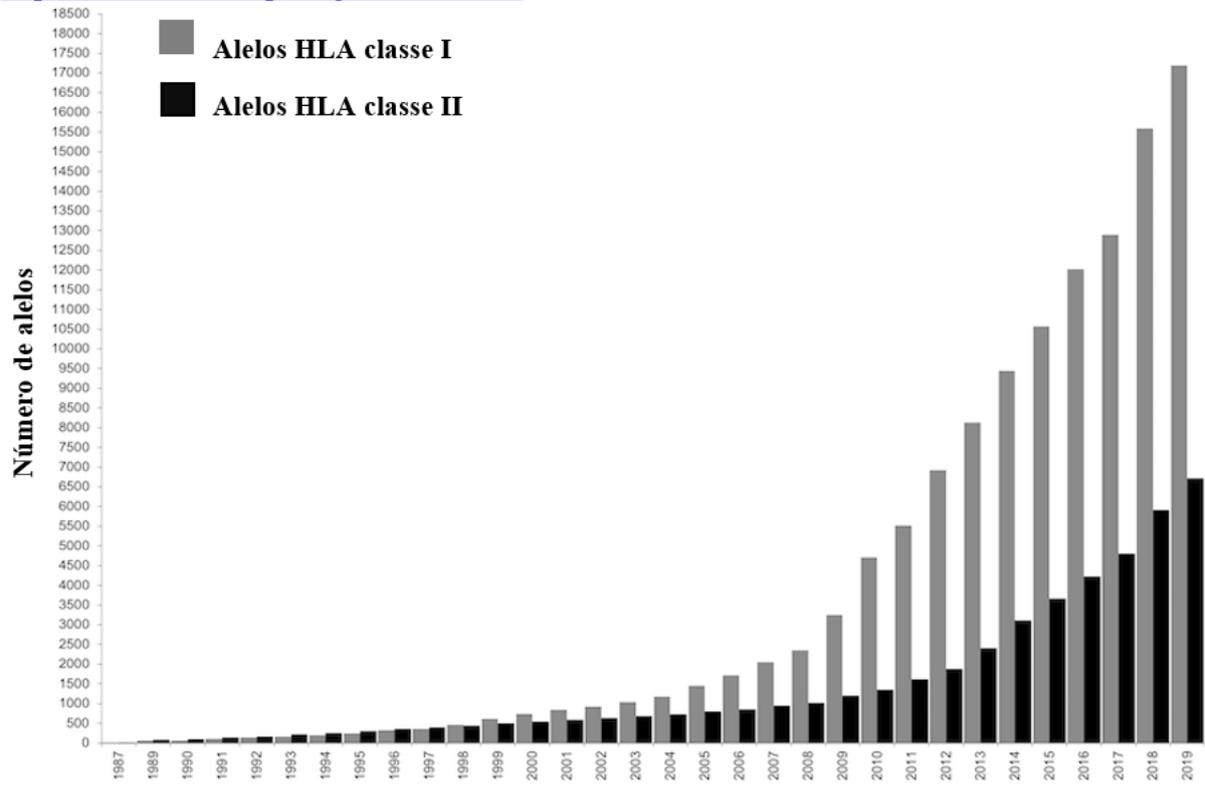
- Paunic, V.;** Gragert, L.; Madbouly, A.; Freeman, J.; e Maiers, M. (2012). Measuring Ambiguity in HLA Typing Methods. *Plos Ones*. **7**(8): 1-10
- Qin, D.** (2019). Next-generation sequencing and its clinical application. *Cancer Biology & Medicine*. **16**(1):4-9
- Resta, C.D. e Ferrari, M.** (2018). Next generation sequencing: from research area to clinical practice. *Journal of the International Federation of Clinical Chemistry*. 215-220
- Robinson, J.;** Sooemally, A.R.; Hayhurst, J.D.; e Marsh, S.G.E. (2016). The IPD-IMGT/HLA Database - New developments in reporting HLA variation. *Human Immunology*. **77**(3): 233-237
- Robinson, P.C. e Brown, M.A.** (2013). Genetics of ankylosing spondylitis. *Molecular Immunology*. **57**(1):2-11
- Rock, K.L.,** Reits, E.; e Neefies, J. (2016). Present Yourself! By MHC Class I and MHC class II molecules. *Trends in Immunology*. **37**(11): 724-737
- Sanchez-Mazas, A. e Meyer, D.** (2014). The Relevance of HLA Sequencing in Population Genetics Studies. *Journal of Immunology Research*. 1-12
- Sanchez-Mazas, A. e Nunes, J.M.** (2019). Does NGS typing highlight our understanding of HLA population diversity? Some good reasons to say yes and a few to say be careful. *Elsevier*. **80**:62-66
- Santis, D.;** Dinauer, D.; Duke J.; Erlich, H.A.; Holcomb, C.L.; Lind, C.; Mackiewicz K.; Monos D.; Moudgil A.; Norman P.; Parham P.; Sasson A.; e Allcock R.J.N. (2013). 16th IHIW: Review of HLA typing by NGS. *International Journal of Immunogenetics*. **40**(1): 72-76
- São João, R.;** papoila, A.L.; Ligeiro, D.; e Trindade, H. (2013) HLA Allele and Haplotype Frequencies of the Portuguese BoneMarrow Donors Registry. *Springer*. 417-426
- Seshasubramanian, V.;** Venugopal, M.; Kanna, A.D.S; Naganathan, C.; Manisekar, N.K.; Kumar, Y.N.; Narayan, S. e Periathiruvadi, S. (2019). Application of high-throughput next-generation sequencing for HLA typing of DNA extracted from postprocessing cord blood units. *Wiley: HLA Immune Response Genetics*. 1-6
- Shien, M.;** Chitnis, N. e Monos, D. (2018). Human Leukocyte Antigen and Disease Associations a Broader Perspective. *Clinical Chemistry and Laboratory Medicine*. 1-15
- Silva, M.E.R.;** Mory, D.; e Davini, E. (2008). Marcadores Genéticos e Auto-Imunes do Diabetes Melito Tipo 1: da Teoria para a Prática. *Archives of Endocrinology and Metabolism*. **52**(2): 166-180
- Smith, A.;** Pereira, S.; Jaramillo, A.; Stoll, S.T.; Khan, F.M.; Berka, N.; Mostafa, A.A.; Pando, M.J.; Usenko, C.Y.; Bettinotti, M.P.; Pyol, C.; Nelso, W.C.; Willis, A.; Askar, M.; e Geraghtyl, D.E. (2019). Comparison of sequence-specific oligonucleotide probe vs next generation sequencing for HLA-A, B, C, DRB1, DRB3/B4/B5, DQA1, DQB1, DPA1, and DPB1 typing: Toward single-pass high-resolution HLA typing in support of solid organ and hematopoietic cell transplant programs. *Wiley: HLA Immune Response Genetics*. 1-10
- Testi, M. e Andreani, M.** (2015). *Luminex*-Based Methods in High-resolution HLA typing. Em: Peter B. Molecular Typing of Blood Cell Antigens. *Methods in Molecular Biology*. **1310**: 231-245
- Torres, M.A. e Moraes, M.E.H.** (2011). Nomenclatura dos fatores do Sistema HLA. *Einstein*. **9**(2): 249-251
- Weimer, E.T.;** Montgomery, M.; Petraroia, R.; Crawford, J.; e Schmitz, J.L. (2016). Performance Characteristics and Validation of Next-Generation Sequencing for Human Leucocyte Antigen Typing. *The Journal of Molecular Diagnostics*. **18**(5): 668-675
- Yohe, S. e Thyagarajan, B.** (2017). Review of clinical next-generation sequencing. *Archives of Pathology & Laboratory Medicine*. **141**(11): 1544-1557

Zhou, M.; Gao, D.; Chai, X.; Liu, J.; Lan, Z.; Liu, Q.; Yang, F.; Guo, Y.; Fang, J.; Yang, L.; Du, D.; Chen, L.; Yang, X.; Zhang, M.; Zeng, H.; Lu, J.; Chen, H.; Zhang, X.; Wu, S.; Han, Y.; Tan, J.; Cheng, Z.; Huang, C.; e Wang, W. (2014). Application of high-throughput, high-resolution and cost-effective next generation sequencing-based large-scale HLA typing in donor registry. *Tissue Antigens*. **85**(1): 20-28

Zolet, C.T.A.; Turchetto, C.; Guzman, F.; Silva, G.A.; Ludwig, S.; e Veto, N. (2017). Capítulo 8: Polimorfismo de Nucleotídeo único (SNP): metodologias de identificação, análise e aplicações. 1-179

Anexos

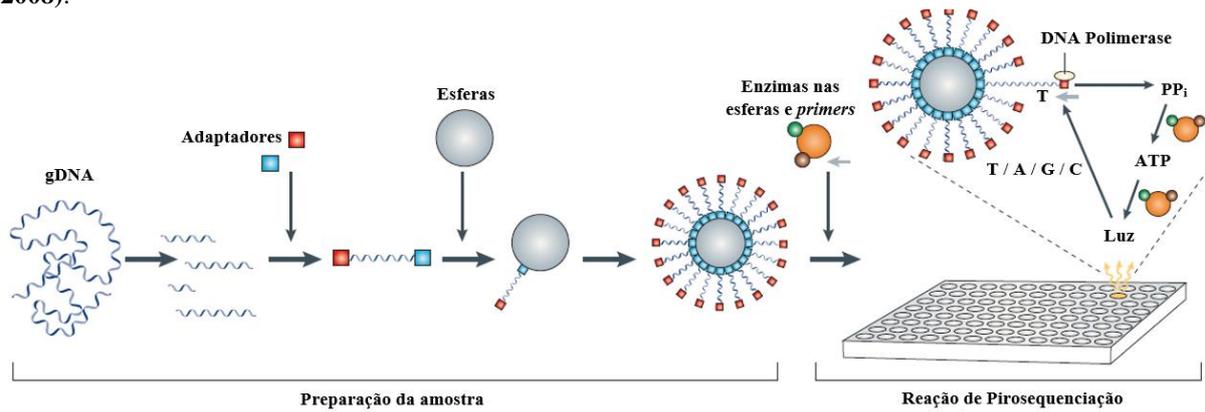
Anexo A- Número de alelos HLA identificados entre o ano de 1987 e julho de 2019 (Adaptado de <https://www.ebi.ac.uk/ipd/imgt/hla/intro.html>, acessado em setembro de 2018).



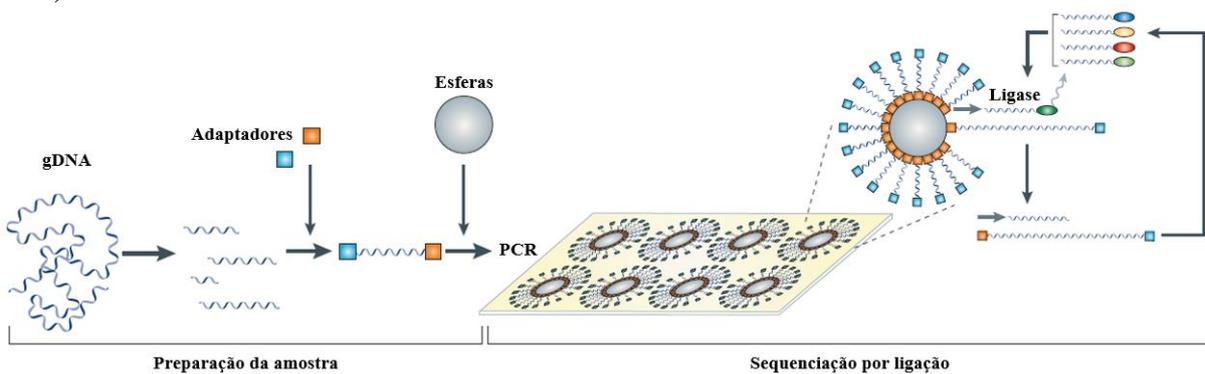
Anexo B- Haplótipos HLA-DR e -DQ associados tanto à suscetibilidade como à proteção da doença (adaptado de Anaya *et al.*, 2013).

DRB1	DQA1	DQB1
Elevado Risco		
0401, 0403, 0405	0301	0302
0301	0501	0201
Risco Moderado		
0801	0401	0402
0404	0301	0302
0101	0101	0501
0901	0301	0303
Proteção Moderada		
0403	0301	0302
0701	0201	0201
1101	0501	0301
Elevada Proteção		
1501	0102	0602
1401	0101	0503
0701	0201	0303

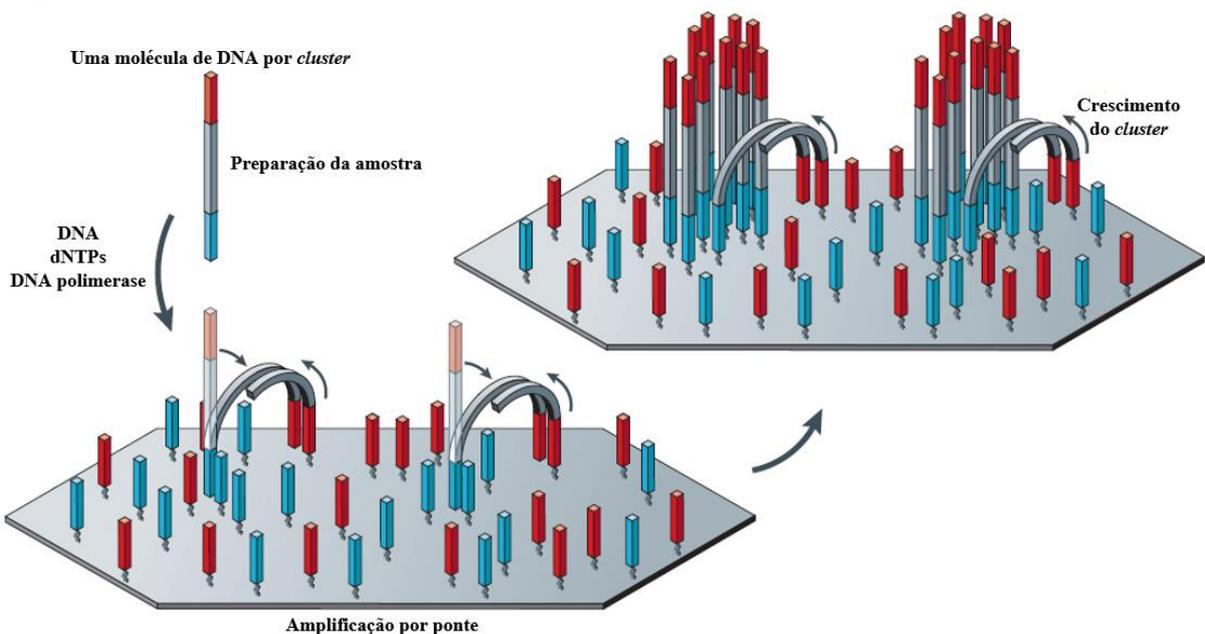
Anexo C- Base da sequenciação por pirosequenciação da *Roche454 System* (adaptado de Medini *et al.*, 2008).

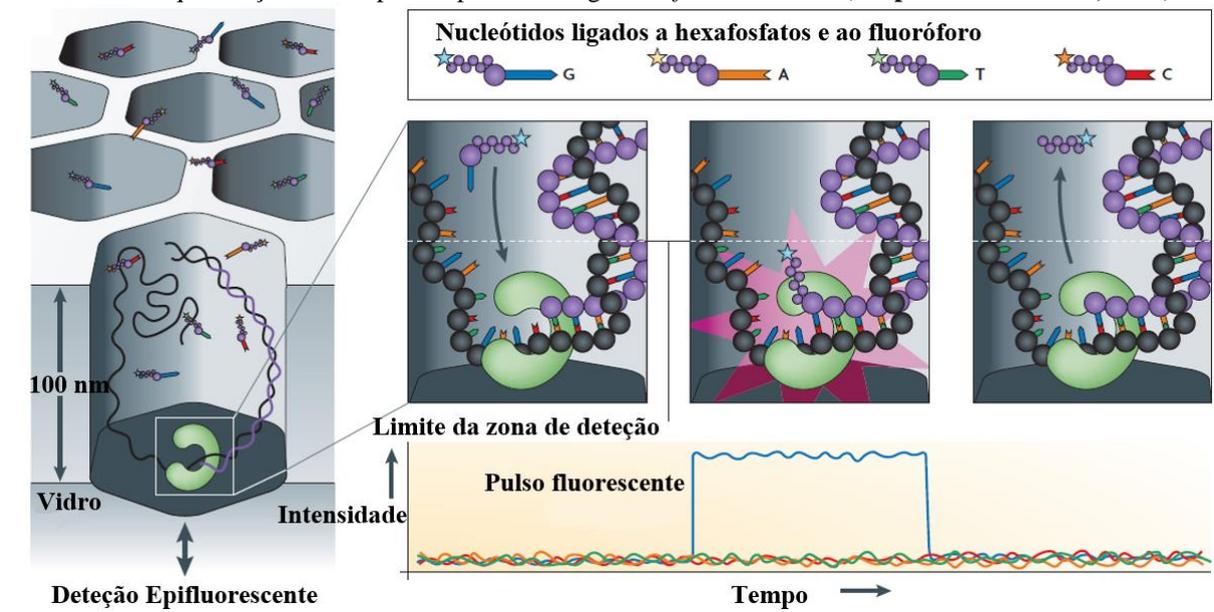


Anexo D- Sequenciação através da metodologia *SOLiD* da *Applied Biosystems* (adaptado de Medini *et al.*, 2008).



Anexo E- Amplificação por ponte através da sequenciação por síntese da *Illumina* (adaptado de Metzker, 2009).



Anexo F- Sequenciação em tempo real pela tecnologia *Pacific Biosciences* (adaptado de Metzker, 2009).**Anexo G-** Extração em modo automático no *QIAcube*

- Selecionar a opção DNA do menu, o *kit – QIAamp DNA Blood Mini*, o tipo de amostra usado e o volume de eluição da amostra pretendido, 100 ou 200 μ l. Pressionar “start” e seguir os passos indicados pelo equipamento.
- Colocar um *ependorf* de 1,5 ml na posição A da grelha para reagentes extra, com o volume de protease necessário consoante o número de amostras a extrair. Preparar e colocar os adaptadores de rotor da centrífuga conforme o número de amostras a extrair.
- Colocar os tubos com as amostras (tubos de 2 ml, com rosca e fundo cónico) nas posições corretas no agitador e peças de borracha no lugar das tampas desses tubos. Fechar a porta do equipamento e pressionar “start”.
- Quando a corrida terminar, selecionar “OK” e seguir os passos de limpeza indicados no aparelho.
- Transferir os ácidos nucleicos eluídos para um tubo de armazenamento devidamente identificado e proceder à quantificação da amostra e seu armazenamento.

NOTA: É possível parar uma corrida a meio, devido a uma emergência, pressionando “cancel”. Se uma corrida é parada, não pode ser reiniciada, tendo de ser terminada manualmente.

Anexo H- Extração por *QIAGEN-QIASYMPHONY DNA Blood Kit*

- Efetuar o login, para que todas as gavetas do aparelho sejam desbloqueadas e efetuar os passos de preparação do aparelho para a extração.
- Colocar os suportes carregados com as amostras no equipamento começando pela posição 1. Selecionar as amostras e definir o protocolo a utilizar. Selecionar a posição de eluição e o volume de eluição e pressionar o botão “run” para iniciar a extração.

NOTA: A ordem em que os suportes são carregados determina a posição dos DNAs eluídos no suporte de eluição, pelo que a posição das amostras nos suportes e a posição dos tubos de eluição devem ser correspondentes.

- Abrir a gaveta “eluate” e retirar a placa com os tubos de eluição e quantificar as amostras. Terminada a extração é necessário realizar a limpeza do equipamento. Quantificar o DNA no Nanodrop.

Anexo I- Tabela de validação de resultados para cada locus HLA em estudo.

	Esferas			
	13	32	34	35
Locus A	Controlo positivo (exão 2)	Controlo positivo (exão 3)	-----	Controlo negativo
Locus B	Controlo positivo (exão 2)	Controlo positivo (exão 3)	-----	Controlo negativo
Locus C	Controlo positivo (exão 2)	Controlo positivo (exão 3)	-----	Controlo negativo
Locus DRB1	-----	-----	Controlo positivo (exão 2)	Controlo negativo
Locus DQB1	-----	-----	Controlo positivo (exão 2)	Controlo negativo

Anexo J- Pares de *primers* para cada grupo e alelos específicos.

Mistura de amplificação	Grupo alélico	Exclui alelos	Tamanho do amplicão (bp)
A-01	01, 11, 36	-	2400
A-02	11	-	2300
A-03	02	-	2000
A-04	03	-	2500
A-05	23	23:17	2300
A-06	23, 24	24:33	2900
A-07	29, 32, 74	32:04	2300
A-08	30	-	2800
A-09	31, 33	-	2600
A-10	25, 26, 34, 43, 66	-	2400
A-11	68, 69, 34:01, 66:02/03	34:05	2400
A-12	80:01	-	2800
A-L	Todos os alelos A	-	3000

Genotipagem do *Human Leucocyte Antigen* (HLA):

Técnicas atuais e aplicabilidade da abordagem de *Next Generation Sequencing*

B-01	07, 48, 81	48:02	2400
B-02	08, 42, ?35:87	-	2300
B-03	13, 57	-	2400
B-04	14, 38, 39, 67	67:02	2300
B-05	14	-	2200
B-06	15, 46	15:42	2300
B-07	18, 37	-	2700
B-08	27, 40, 47, ?35:42	40:01/48/55, 27:05:04	2700
B-09	15:42, 35, 51, 52, 53, 58, 78, 56:06	51:42, 35:87	2200
B-10	35, 53, 57	-	3300
B-11	40	40:02/03/04/06/08/40/55	2400
B-12	41	-	2400
B-13	44, 51:42, 83:01	44:27, ?44:15/18	2600
B-14	45, 49, 50	-	2700
B-15	54, 55, 56, 59, 82	?56:06, 59:02/03/04	2300
B-L	Todos os alelos B	-	3000
C-01	1	-	2300
C-02	02, 15	-	3200
C-03	15	-	2300
C-04	03	-	2900
C-05	04	-	2900
C-06	05, 08	-	3200
C-07	05, 06, 08, 12, 15	15:43	2100
C-08	07	-	2900
C-09	14	-	2600
C-10	16	-	3100
C-11	17	-	3100
C-12	18	-	2500
C-L	Todos os alelos C	-	3100
DRB-01	01	-	430
DRB-02	15, 16	-	730
DRB-03	03, 14:02/03/06, 13:15, 14:13/19/20	03:07, 14:01/04/05/10/18/22/54/84, ?03:17, 14:17	760
DRB-04	04	-	550
DRB-05	07	-	400
DRB-06	08, 13:17	08:06	860
DRB-07	09	-	440
DRB-08	10	-	460
DRB-09	12	12:01	590
DRB-10	03, 08:06, 11, 12:01, 12:17, 13, 14	13:17	470
DRB-11	03:07, 13:01/02/36, 13:34, 14:17/21/33	13:03/04/05/12/13	460
DRB-12	11:01, 11:01:08, 12:17, 14:01/04/05/10/18/54/84, 11:13/17, 13:08/83	14:02/03/06/13/17/19/22/33, 14:13/17/19/33	470
DRB-L1	Todos os alelos DRB1 (exão 1)	-	260
DRB-L2	Todos os alelos DRB1 (exão 2)	-	460-740
DRB-L3	03, 08, 11 a 14 (exão 3)	12:17	1000
DQB-01	02	-	480
DQB-02	03	03:02/03, ?03:05-08/11	530
DQB-03	03	03:01/04/09/19, ?03:10/12/13	525
DQB-04	04	-	525
DQB-05	05	-	480
DQB-06	06	-	540
DQB-L2	Todos os alelos DQB1 (exão2)	-	510
DQB-L3	Todos os alelos DQB1 (exão 3)	-	400

Anexo K- Especificidade das misturas de amplificação para os diferentes *primers* de sequenciação.

Mistura de amplificação	Gene (exão)	<i>Primers</i> de sequenciação (F-forward; R-reverse)
A-01, 04, 05, 07 - 12, L	HLA-A (exão 1)	A1F
A-01 - 12, L	HLA-A (exão 2, 3, 4)	A2F, A2R, A3F, A3R, A4F, A4R
A-01, 02, 04, 07 - 10, 12, L	HLA-A (exão 5)	A5F
B-01, 03, 04, 07 - 12, 14, L	HLA-B (exão 1)	B1F
B-01 - 15, L	HLA-B (exão 2, 3, 4)	B2F, B2R, B3F, B3R, B4F, B4R
B-01 - 08, 11 - 15, L	HLA-B (exão 5)	B5R
C-02 - 04, 06 - 12, L	HLA-C (exão 1)	C1R
C-01 - 12, L	HLA-C (exão 2, 3, 4)	C2F, C2R, C3F, C3R, C4F, C4R
C-01, 02, 04 - 06, 08 - 12, L	HLA-C (exão 5)	C5R
C-02, 04, 05, 06, 08, 10, 11, L	HLA-C (exão 6, 7)	C6F, C7R
DRB-L1	HLA-DRB1 (exão 1)	DRB1F, DRB1R
DRB-L2	HLA-DRB1 (exão 2)	DRB2F, DRB2R, DRB2R86 (codão86)
DRB-L3	HLA-DRB1 (exão 3)	DRB3F
DRB-01 - 12	HLA-DRB1 (exão 2)	DRB2F, DRB2R
DRB-01 - 12, L	HLA-DRB1 (exão 2)	DRB2R-86 (GTG)
DQB-01 - 06, L2	HLA-DQB1 (exão 2)	DQB2F, DQB2R
DQB-L3	HLA-DQB1 (exão 3)	DQB3F, DQB3R