

MARIA DE LURDES PONTES REBELO

**APLICAÇÃO DA ANÁLISE DE SNPS AUTOSSÓMICOS À RESOLUÇÃO DE CASOS FORENSES**

Tese de Candidatura ao grau de Doutor em Ciências Forenses submetida à Faculdade de Medicina da Universidade do Porto.

**Orientador** . Professor Doutor Rui Manuel de Medeiros Melo Silva.

**Categoria** - Professor Associado Convidado com Agregação.

**Afiliação** - Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

**Co-orientador** . Professora Doutora Maria Victória Laréu

**Categoria** . Professora Catedrática.

**Afiliação** - Faculdade de Medicina da Universidade de Santiago de Compostela.

---

## **Informação Técnica**

Título:

Aplicação da análise de SNPs autossómicos à resolução de casos forenses.

Tese de Candidatura ao Grau de Doutor em Ciências Forenses

submetida à Faculdade de Medicina da Universidade do Porto.

AUTOR/EDITOR:

Maria de Lurdes Pontes Rebelo

DATA: dezembro de 2015

E-MAIL: [mlurdes.pontesrebelo@gmail.com](mailto:mlurdes.pontesrebelo@gmail.com)

---

Quando atingimos alguma maturidade, olhamos para as dificuldades que vão surgindo no nosso caminho com empenho, sem que nos impeçam de continuar nesse caminho, tornando-nos mais fortes, assentes na realidade, mas simultaneamente com um olhar permanentemente aberto à novidade do futuro

**Nomeação do Júri da Prova de Doutoramento de**

**- MARIA DE LURDES PONTES REBELO -**

**Programa Doutoral em Ciências Forenses**

Nos termos do disposto do n.º 2 do art.º 18.º do Regulamento dos Terceiros Ciclos de Estudos da Universidade do Porto, a seguir se publicita o júri de doutoramento em Ciências Forenses, da licenciada Maria de Lurdes Pontes Rebelo, nomeado por despacho vice-reitoral de 17 de novembro de 2015, com a tese "Aplicações da análise de SNPS autossómicos à resolução de casos Forenses":

Presidente: Reitor da Universidade do Porto.

VOGAIS: Doutor Ángel María Carracedo Álvarez, professor catedrático da "Facultad de Medicina da Universidad de Santiago de Compostela";

Doutor Manuel Diamantino Pires Bicho, professor catedrático da Faculdade de Medicina da Universidade de Lisboa;

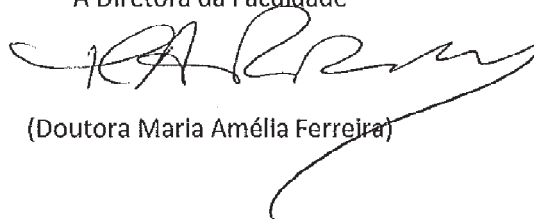
Doutora Teresa Maria Salgado de Magalhães, professora catedrática convidada da Faculdade de Medicina da Universidade do Porto;

Doutora Maria Berta de Jesus Duarte da Silva, professora associada do Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto;

Doutor Rui Manuel de Medeiros Melo Silva, investigador principal e Coordenador do Grupo de Oncologia Molecular e Patologia Viral do Centro de Investigação do IPO Porto e orientador da tese;

Divisão Académica, 9 de dezembro de 2015.

A Diretora da Faculdade



(Doutora Maria Amélia Ferreira)

---

Durante a realização deste trabalho foram efetuadas algumas publicações.

Trabalhos já publicados:

- Pontes ML, Pinheiro MF. Autosomal SNPs study of a population sample from North of Portugal and a sample of immigrants from the Eastern Europe living in Portugal. *Legal Medicine*, 16 (2014) 118. 120. (Anexo 1)
- Pontes ML, Amigo OM, Fondevila M, Porto MJ, Medeiros R. Analysis of Autosomal SNPs in Paraffin-embedded Tissue Samples: Two Cases. *J Forensic Investigation*, 2015 Vol.:3, Issue:1. (Anexo 2)
- M Lurdes Pontes & R Medeiros. Autosomal SNPs in different forensic applications. *Australian Journal of Forensic Sciences (2015)*, DOI: 10.1080/00450618.2015.1050067. (Anexo 3)
- M. Lurdes Pontes, Manuel Fondevila, Maria Victoria Laréu, Rui Medeiros. SNP Markers as Additional Information to Resolve Complex Kinship Cases. *Transfus Med Hemother* 2015;42:385. 388; DOI: 10.1159/000440832. (Anexo 4)

Trabalhos submetidos:

- Pontes ML, Carneiro de Sousa MJ, Medeiros R. SNPs and STRs in Forensic Medicine. A new strategy for kinship evaluation, submetido à revista *the American Journal of Forensic medicine and Pathology*. (Anexo 5)

---

---

## Agradecimentos

Agradeço muito especialmente ao meu orientador Professor Doutor Rui Medeiros, pelo apoio incondicional que me deu, numa situação excepcional, de que poucos seriam capazes, a escassos meses da data limite da entrega deste trabalho. Obrigada pela disponibilidade, atenção dispensada, paciência, dedicação e, sobretudo, pelo ânimo transmitido.

À Professora Doutora Maria Victória Lareu, minha co-orientadora, agradeço a sua permanente disponibilidade e apoio apesar da distância e da mudança de orientador. Agradeço ainda, por me ter acolhido em Santiago de Compostela no início deste trabalho, que ajudou a delinear conjuntamente com a minha anterior orientadora, a Professora Doutora Fátima Pinheiro.

Agradeço à Professora Doutora Teresa Magalhães, coordenadora deste Curso de Doutoramento em Ciências Forenses e que sempre me incentivou, em especial aquando da minha tomada de decisão em o frequentar.

À Professora Doutora Maria José Carneiro de Sousa que, muito amavelmente me apresentou e apoiou relativamente à escolha do meu presente orientador, muito agradeço. Foi, sem dúvida, uma ajuda que, parecendo insignificante a quem a prestou, se tornou decisiva.

Agradeço a todos os meus colegas de trabalho, anteriores e atuais, sem exceção, pelos momentos partilhados em conjunto, contrinuindo para manter sempre um bom ambiente, mesmo em momentos de muita turbulencia.

Agradeço em particular à minha mãe que, apesar de ter pouca instrução, sempre me apoiou, ajudando-me em todos os meus projetos, permitindo deste modo tornar-me uma pessoa mais realizada.

Ào Mário, meu marido, estou grata pela compreensão e encorajamento, e por todo o apoio lá em casa, durante todo o periodo do trabalho desenvolvido, com muitas ausências e rabujice à mistura. Espero poder retribuir, de algum modo, todo o teu carinho e paciência.

Aos meus filhos, Beatriz e Miguel Ângelo, que talvez sejam os que mais sofrem com esta minha vida ocupada. Obrigada por existirem!

A todos os demais, especialmente os meus amigos, agradeço o ânimo e apoio em todas as circunstâncias, sempre com Música, Fé e Esperança a colorir a nossa convivência conjunta.

---

---

# Índice Geral

<b>Lista de abreviaturas</b>	<b>1</b>
<b>Resumo</b>	<b>3</b>
<b>Abstract</b>	<b>5</b>
<b>Capítulo I</b>	
<b>1. Introdução</b>	<b>9</b>
<b>1.1. SNPs</b>	<b>10</b>
<b>1.1.1. SNPs estudados</b>	<b>14</b>
<b>1.2. STRs</b>	<b>15</b>
<b>1.3. SNPs vs.STRs</b>	<b>16</b>
<b>1.4. Metodologias de estudo</b>	<b>17</b>
<b>1.5. Aplicação Forense dos SNPs</b>	<b>25</b>
<b>Capítulo II</b>	
<b>2. Objetivo do trabalho</b>	<b>29</b>
<b>2.1. Objetivo principal</b>	<b>29</b>
<b>2.2. Objetivos Específicos</b>	<b>29</b>
<b>Capítulo III</b>	<b>31</b>
<b>3. Seleção da metodologia de estudo</b>	<b>31</b>
<b>Capítulo IV</b>	
<b>4. Aplicação em estudos populacionais</b>	<b>37</b>
<b>4.1. Sumário</b>	<b>37</b>
<b>4.1.1. Estudo populacional de uma amostra do Norte de Portugal e de uma amostra de indivíduos imigrantes do Leste europeu</b>	<b>37</b>
<b>4.1.2. Material e métodos</b>	<b>37</b>
<b>4.1.3. Resultados e discussão</b>	<b>42</b>
<b>Capitulo V</b>	
<b>5. Aplicação em amostras degradadas em condições laboratoriais controladas</b>	<b>55</b>
<b>5.1. Introdução</b>	<b>55</b>
<b>5.2. Material e métodos</b>	<b>55</b>
<b>5.3. Resultados e discussão</b>	<b>56</b>
<b>Capitulo VI</b>	
<b>6. Aplicação em casos de parentesco complexos</b>	<b>59</b>
<b>6.1. Introdução</b>	<b>59</b>
<b>6.2. Material e métodos</b>	<b>59</b>
<b>6.3. Resultados e discussão</b>	<b>61</b>

---

## Capítulo VII

<b>7. Aplicação em casos de criminalística e de identificação</b>	<b>63</b>
7.1. Introdução	63
7.2. Material e métodos	63
7.3. Resultados e discussão	64

## Capítulo VIII

<b>8. Discussão geral e conclusão</b>	<b>67</b>
8.1. Introdução	67
8.2. Discussão mediante os objectivos propostos	68
8.3. Conclusão e perspectivas futuras	71

<b>Anexos</b>	<b>73</b>
---------------	-----------

<b>Referências Bibliográficas</b>	<b>131</b>
-----------------------------------	------------

## Lista de abreviaturas

ALFRED . *Allele FREquency Database*

ASO . *Allele specific oligonucleotides* - oligonucleotidos alelo específicos

AISNPs - *Ancestry Inference SNPs* - SNPs para Inferência de ancestralidade

CAL . *Coupled Amplification and oligonucleotide Ligation*

ddNTP . di-deoxinucleótidos

DNA . ácido desoxirribonucleico

dNTP . deoxinucleótido trifosfato

DOL . *Dye-labelled Oligonucleotide Ligation*

DVI . *Disaster Victim Identification* . Identificação de vítimas de desastres de massas

FRET . *Fluorescence Resonance Energy Transfer*- Transferência de energia de ressonância fluorescente

GHEP-ISFG . Grupo que fala Espanhol e Português da Sociedade Internacional de Genética Forense

IISNPs . *Individual Identification SNPs* . SNPs para identificação individual

LCR . *Ligation Chain Reaction*

LISNPs - *Lineage Inference SNPs* - SNPs para Inferência de linhagem

MALDI-TOF . *Matrix-Assisted Laser Desorption/Ionization -Time Of Flight* . *Mass Spectrometry* - Espetrometria de massa com ionização e dessorção a laser assistida por matriz.

mtDNA . DNA mitocondrial

NGS . *Next Generation Sequencing technologies*

OLA . *Oligonucleotide ligation Assay*

PISNPs - *Phenotype Inference SNPs* - SNPs para Inferência de fenótipo

SGBF-N . Serviço de Genética e Biologia Forenses Unidade Funcional da Delegação do Norte

SNP . *Single Nucleotide Polymorphism*

STR . *Short tandem repeats*

PIC . *Polymorphic indices content*

PCR . *Polymerase Chain Reaction*

---

---

## RESUMO

A utilização de *single nucleotide polymorphisms* (SNPs) autossómicos na investigação forense, tem sido muito discutida nos últimos dez anos, em grande parte devido às vantagens que estes marcadores possuem, quando comparados com os *short tandem repeats* (STRs).

Neste trabalho foram testadas duas técnicas de genotipagem de SNPs e uma das quais foi implementada no Laboratório de Genética e Biologia Forenses, Delegação do Norte do Instituto Nacional de Medicina Legal e Ciências Forenses, IP (INMLCF,IP), tendo sido avaliada no sentido de ser validada como método a aplicar em alguns estudos forenses. Uma das metodologias foi usada para estudar uma amostra de indivíduos do norte de Portugal e compará-la com uma amostra de indivíduos imigrantes da Europa de Leste residentes em Portugal. Esta metodologia foi depois aplicada a casos concretos, com amostras reais e a casos de paternidade complexos de modo a avaliar o impacto da informação complementar obtida a partir do estudo dos SNPs autossómicos.

A primeira metodologia de genotipagem que tínhamos a intenção de validar incluía 48 marcadores SNPs autossómicos e baseava-se no sistema de tipagem ~~%Genplex SNP typing system+~~. Este objetivo foi cumprido em algumas amostras de referência de indivíduos selecionados a partir de casos de parentesco, previamente resolvidos, foi efetuada no Laboratório de Santiago de Compostela. Esta tecnologia teve de se abandonar sobretudo porque o *Geneplex kit* (Applied Biosystems) foi descontinuado.

A outra estratégia de genotipagem, consistiu na utilização de uma das mais comuns metodologias de tipagem comerciais e que se baseia numa reação de minisequenciação seguida de eletroforese e deteção da fluorescência. Trata-se de uma metodologia que utiliza o kit *SNaPshot<sup>TM</sup>* comercializado pela Applied Biosystems, criada e desenvolvida pelo *SNPforID Consortium*, que num único tubo amplifica 52 SNPs (*52-plex*) com vista à sua aplicação na investigação forense. O nosso objetivo seria utilizar esta metodologia como alternativa, uma vez que já foi validada por alguns laboratórios, tanto em testes criminais como em investigações de parentesco. A estratégia evoluiu no sentido de proceder à validação interna deste *52-plex*.

Procedeu-se a um estudo populacional, determinando-se as frequências alélicas para o *52-plex*. Foram genotipadas um total de 216 amostras que incluíram indivíduos não relacionados representativos do norte de Portugal e 85 indivíduos imigrantes da Europa de leste, na sua maioria ucranianos, também não relacionados e a residir em Portugal.

---

Procedeu-se ainda a alguns estudos de validação interna e, finalmente aplicou-se esta metodologia a amostras de casos já realizados que foram selecionados, e que se relacionavam com casos criminais, casos de investigação de parentesco complexos, e casos de identificação, como complemento à genotipagem dos marcadores STRs ou mini-STRs, habitualmente utilizados na rotina laboratorial.

A principal conclusão deste trabalho, para além de determinar as frequências destes 52 SNPs na nossa população, de fazer comparações com a população de imigrantes de leste a residir em Portugal e com outras populações, foi demonstrar que estes marcadores serão muito úteis na resolução de alguns casos específicos, inconclusivos após obtenção da informação genética a partir da análise exclusiva dos marcadores genéticos habitualmente utilizados (STRs).

---

## ABSTRACT

The use of autosomal SNPs for forensic research has been widely discussed in recent years, mainly because SNPs have important advantages compared to short tandem repeats (STRs).

In this thesis, two separate single nucleotide polymorphism (SNP) genotyping techniques were tested and one of them was set up at the Forensic Genetics Laboratory of the National Institute of Legal Medicine and Forensic Sciences, IP, North Branch pooled genotyping was evaluated as a method for forensic studies. Finally, one of the former approaches was used to study a sample from the North of Portugal and compare it with a East Europe Immigrants individuals living in Portugal and applied to concrete casework samples and complex paternity cases to monitor the impact of the information complement achieved with these markers.

The first genotyping platform that we intended to validate included 48 autosomal SNPs and was based on the Genplex SNP typing system+ was accomplished in some reference samples of individuals selected from our routine parentage testing in the Santiago de Compostela Laboratory. This technology had to be abandoned mainly because the Genplex kit (Applied Biosystems) was discontinued.

The other strategy of genotyping consisted in the use of one of the most common commercial technologies based on minisequencing reaction followed by electrophoresis and fluorescence detection, the SNaPshot™ kit from Applied Biosystems, in the core single-tube 52plex PCR created and developed by the SNPforID Consortium for forensic use. We intended to use this latter technology mainly as a point of comparison, because it was already validated in forensic and in relationship tests. This strategy was chosen as the one to be validated internally in our laboratory.

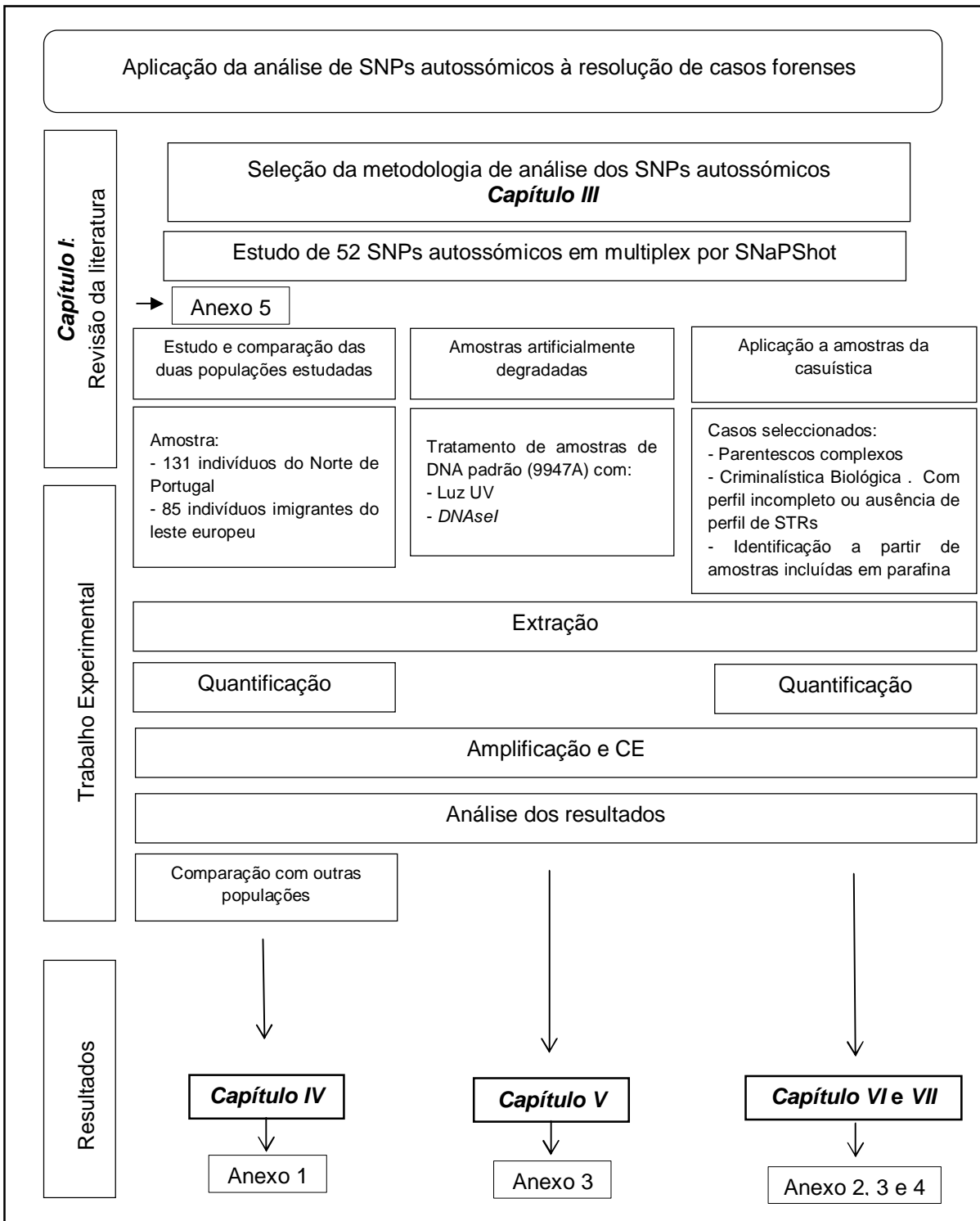
We determined the allele frequencies for the 52-plex in a population study. For that we genotyped a total of 216 samples that included non related individuals from the North of Portugal and 85 immigrant individuals from the Eastern Europe, mainly Ukrainians, equally non related and residing in Portugal. This was published in a journal.

We made some validation studies and finally we have applied this methodology in previously selected casework samples, related with criminal cases, complex paternity and identification cases, as a complement to traditional genotyping that uses STRs or mini-STRs.

The main conclusion of this work, besides determine our own frequencies, was that these markers will be very useful in our lab to solve some particular cases that are inconclusive if we take into account only the traditional markers, the STR markers.

---

A presente Tese organiza-se por capítulos. O capítulo I contempla uma introdução geral ao tema da Tese, sendo o capítulo II dedicado aos objectivos desenvolvidos ao longo do trabalho. Nos capítulos III, IV, V, VI e VII são apresentados os resultados experimentais e respetiva discussão. No capítulo VIII são apresentadas as conclusões gerais retiradas a partir dos resultados do trabalho desenvolvido bem como se perspetivam aplicações e implicações futuras.



---

Nota importante:

A estratégia descrita no projeto inicial que consistia em selecionar a metodologia mais apropriada para a caracterização de SNPs autossómicos entre duas, a minissequenciação (SNaPshot<sup>®</sup>) . metodologia proposta por Sanchez et al. (2006) e desenvolvida pelo SNPforID Consortium e o GenPlex system<sup>+</sup> (49 SNP-plex), também proposto pelo SNPforID Consortium, em 2007, e desenvolvido pela Applied Biosystems, que incluía 48 SNPs comuns aos do 52 SNP-plex anterior, bem como a amelogenina, foi alterada. Esta alteração ocorreu fundamentalmente porque o método que utilizava o sistema GenPlex<sup>+</sup> foi inviabilizado pelo facto deste *kit* ter sido descontinuado. Por isso, optou-se pela utilização exclusiva da minissequenciação.

---

# CAPÍTULO I

## 1 Introdução

O principal objetivo das perícias relacionadas com investigação de parentesco, criminalística biológica e identificação individual é a identificação genética de amostras biológicas relacionadas com este tipo de perícias. Para isso, nos últimos 20 anos, o recurso habitual é proceder à análise de vários marcadores genéticos, mais frequentemente aos STRs (*Short Tandem Repeats*), altamente polimórficos, para a obtenção de perfis de DNA aplicados às várias finalidades ligadas à identificação humana na área da Genética Forense.

Os perfis de STRs autossómicos são obtidos a partir de vestígios biológicos encontrados em cenas de crime e são posteriormente comparados com perfis de amostras de referência de indivíduos relacionados com o crime (suspeitos), conhecidos através da investigação prévia das forças policiais ou incluídos em bases de dados com fins forenses. Com o suporte estatístico adequado, um *match* de um perfil fornece uma forte evidência para uma identificação individual (com exceção dos gémeos monozigóticos . 0,35% nos países ocidentais [1]), o que não acontece no caso de uma exclusão. Perfis muito semelhantes podem indicar o envolvimento de familiares próximos e são sinónimos de interpretação estatística mais complexa.

Os STRs também são utilizados para identificação de vítimas em desastres de massa (*Disaster Victim Identification* . DVI) onde se procede a comparações com amostras *antemortem* (provenientes por exemplo de objetos de uso pessoal) ou com familiares próximos, fornecendo evidência para uma identificação positiva.

Por fim, os STRs vulgarmente são usados, com sucesso, para determinar paternidades ou outros parentescos mais ou menos complexos.

A exclusão de indivíduos relacionados com crimes ou relações de parentesco com base nos seus perfis de STRs assume igual importância.

Durante os últimos anos também existiu um grande progresso associado aos desafios técnicos envolvidos na obtenção de perfis genéticos na área da Genética Forense. Este progresso incluiu a capacidade de analisar DNA em quantidades reduzidas e/ou degradado, um aumento do poder de discriminação global e a aplicação dos perfis para procura e comparações com familiares recorrendo, por exemplo, a bases de dados genéticos.

Apesar destes avanços técnicos no que diz respeito à obtenção de perfis genéticos com base nos STR autossômicos, permanecem desafios importantes por resolver, tendo-se para tal recorrido à investigação forense na procura de outras respostas.

Atualmente existem soluções inovadoras que propiciarão a obtenção de resultados, permitindo assim uma cabal identificação genética em várias situações. Uma dessas possíveis respostas consiste em utilizar outros marcadores genéticos, concretamente os SNPs (*Single Nucleotide Polymorphisms*), com fins de identificação humana, que potencialmente irão resolver algumas das dificuldades que surgem associadas aos STRs, sobretudo como complemento ao seu estudo. Por outro lado o estudo destes marcadores alternativos também pode conduzir a novos desafios.

### 1.1. SNPs

Os SNPs são polimorfismos muito abundantes no genoma humano e terão surgido pela ocorrência de mutações pontuais ao longo da evolução. Consistem em sequências alternativas (alelos) de uma única base em determinada posição do DNA genómico (Figura 1). É normal existirem diferentes alelos em indivíduos de determinada população. Estes distinguem-se de variantes raras porque cumprem o requisito de o alelo menos abundante apresentar uma frequência mínima de 1% [2]. Podem situar-se no DNA codificante ou no DNA não codificante. Alguns autores afastam-se desta definição, mas habitualmente as variantes que são muito raras originam mutações de perda de função e habitualmente estão associadas a patologias graves (ex. Fibrose cística, distrofia muscular de Duchenne, entre outras).

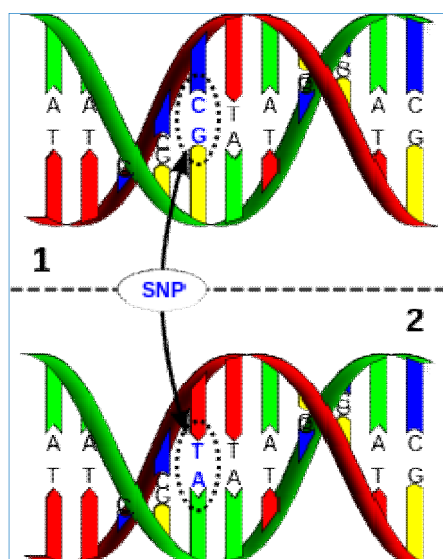


Figura 1. Representação de um SNP . a molécula 1 difere da molécula 2 apenas num par de bases em determinada posição (<http://en.wikipedia.org/wiki/File:Dna-SNP.svg>)

---

É difícil fazer uma estimativa do número de SNPs do genoma humano, mas nas bases de dados públicas e privadas o seu número ascende a 5 milhões e, destes, cerca de 4 milhões já validados, o que significa que já demonstraram ser polimórficos em, pelo menos, um ou vários grupos populacionais *major*. Apesar da sua simplicidade e do seu conteúdo polimórfico ser limitado, nos últimos anos a abundância destes polimorfismos é o principal fator que tem despertado muito interesse, essencialmente na área médica, como marcadores de doenças genéticas basicamente na identificação de genes que estão na base de doenças complexas, ou na fármaco-genómica em que as diferentes respostas a fármacos podem se cruciais para uma boa terapêutica [3].

Na Genética Populacional foram muito utilizados com a finalidade de estudar a história demográfica das populações, principalmente como marcadores de linhagem.

Nas ciências forenses, concretamente na Genética Forense, desde 2001 que se tem vindo a discutir o potencial da utilização dos SNPs na resolução de perícias, tendo a sua importância vindo a aumentar. A razão principal prende-se com o facto de estes marcadores possuírem um certo número de características que os tornam apropriados nesta área: apresentam baixas taxas de mutação, 100 000 vezes inferior à dos STRs, o que os torna muito mais interessantes para testes de paternidade, podendo substituir os STRs para este fim, logo que estejam disponíveis *kits* comerciais para o efeito [4]; são apropriados em análises que utilizam tecnologias de elevado rendimento, permitindo a automatização dos procedimentos de genotipagem, o que é vantajoso tendo em vista a criação de bases de dados. Acresce o facto de poderem ser analisados em produtos da amplificação de pequeno tamanho (fragmentos com cerca de 50bp ou menores), sendo o reduzido tamanho dos produtos amplificados um fator benéfico para o sucesso da análise de amostras degradadas, amostras em que habitualmente não se obtêm perfis de STRs ou se obtêm perfis muito incompletos, revelando-se inconclusivos. Este tipo de amostras é muito frequente em casos de criminalística ou casos de identificação a partir de restos cadavéricos. Adicionalmente, uma vez que as variações dos SNPs não envolvem repetições de sequências, evitam-se os artefactos do tipo *stutter* que complicam a interpretação dos perfis baseados em STRs, especialmente em amostras quantidades de DNA reduzidas.

Apesar da sua maior limitação ser o baixo polimorfismo, podem ser opção quando o estudo de STRs falha podendo, portanto, ser usados para muitos tipos de aplicações forenses [5]. Os SNPs permitem ultrapassar muitas das dificuldades técnicas com que os peritos forenses se deparam relativamente aos STRs, que não existem no caso dos SNPs. Assim, foram definidas quatro categorias básicas de SNPs: SNPs para identificação individual (*Individual Identification* . IISNPs), SNPs para inferência de ancestralidade (*Ancestry Inference* . AISNPs), SNPs para inferência de linhagem (*Lineage Inference* . LISNPs) e SNPs para inferência de fenótipo ou para

---

características fenotípicas (*Phenotype Inference* . PISNPs). A cada um destes propósitos forenses corresponderá uma bateria específica de SNPs e um diferente método de seleção dos SNPs a utilizar.

Os SNPs têm também algumas limitações. Obviamente, sendo os SNPs bi-alélicos, são menos polimórficos que os STR que são multi-alélicos. Por isso são menos informativos na análise de misturas que envolvem mais do que um indivíduo. No entanto, a utilização de muitos SNPs ou de SNPs tri-alélicos, em combinação com tecnologias *multiplex*, pode compensar este efeito. Estima-se que o número de SNPs necessário em Genética Forense, com o propósito de identificação genética, seja cerca de 40 vezes superior ao dos STRs (em média); por esse motivo, serão necessários cerca de 50 a 60 SNPs não ligados com uma elevada heterozigotia global, para obter um poder de discriminação semelhante ao dos sistemas STR analisados em simultâneo (*multiplex*), que habitualmente se utilizam em Genética Forense [6, 7]. Foi demonstrado que 20-50 SNPs autossómicos podem atingir probabilidades de *matching* semelhantes às obtidas a partir de 10-15 STRs [8,9].

Outro desafio consiste em desenvolver SNPs para identificação humana universal, com fiabilidade semelhante para indivíduos de diferente ancestralidade biogeográfica. Existem duas características populacionais que são importantes contemplar neste caso, os SNPs selecionáveis devem possuir um elevado grau de diversidade alélica nas populações mundiais e devem possuir um reduzido grau de diferenciação populacional entre as várias populações (10). O elevado grau de diversidade alélica permite atingir elevados probabilidades de *matching* em todas as populações. Quanto ao baixo grau de diferenciação é crucial para evitar os efeitos da substruturação quando se estimam probabilidades de *matching*.

As bases dos dados públicas, como é exemplo o *HapMap Project*, são úteis mas não incluem todas as populações. Mais recentemente foi desenvolvida uma outra base de dados que inclui dados de frequências alélicas, denominada ALFRED (*Allele FREquency Database*), exatamente com o objetivo de tornar as frequências deste tipo de marcadores disponíveis à comunidade científica.

Nos últimos tempos têm sido desenvolvidos muitos *multiplex*, mas estes não estão disponíveis comercialmente. Em princípio um elevado número de SNPs, como os que estão incluídos nos *microarrays* disponíveis, que contêm centenas de milhares de marcadores, são ideais para esta finalidade de identificação individual universal. No entanto, os *microarrays* atualmente disponíveis requerem quantidades substanciais de DNA genómico, pouco fragmentado, requisito que frequentemente é impossível de concretizar nos casos criminais.

---

Para além de todos os aspetos relacionados com a biologia e a técnica, outro aspeto importante torna o uso geral potencial dos SNPs autossómicos para investigar evidências relativas a cenas de crime, pouco aliciante ou popular: todas as bases de dados de DNA dos diferentes países se baseiam em STRs. Adicionalmente, em muitos países a legislação obriga a que não se possam reter amostras relativas a agressores condenados. Teriam portanto de ser construídas novas bases de dados que incluíssem perfis baseados nestes marcadores, o que não é uma opção muito sensata. No entanto, países que possam arquivar amostras ou que ainda não possuam bases de dados, podem considerar os SNPs para fins de identificação criminal, logo que existam *kits* comerciais disponíveis para esse fim.

Uma exceção importante, onde não serão necessárias bases de dados pré-existentes, é a questão da identificação de pessoas desaparecidas, que consta das perícias de DVI. Nestes casos a maioria das amostras *ante mortem* das vítimas, ou das amostras de referência de familiares próximos das vítimas, são colhidas *de novo* e processadas conjuntamente com os restos cadavéricos das vítimas, pelo que a substituição de STRs por SNPs é mais razoável e pode ser mais vantajosa, uma vez que, frequentemente o DNA dos restos cadavéricos está mais degradado.

A genotipagem de SNPs funciona como suplemento da obtida com recurso aos *kits* comerciais para STRs convencionais, correntemente utilizados com fins de identificação humana. Estes métodos analíticos que se baseiam na tipagem de STRs são utilizados para obter um elevado poder de discriminação. Nos casos em que é necessário um poder de discriminação acrescido, para indivíduos proximamente relacionados, pode ser necessário incluir STRs adicionais. Em alternativa, pode-se recorrer à caracterização de SNPs autossómicos [4,11]. Phillips C e col. E Børsting C e col. Evidenciaram que estes marcadores genéticos serão muito úteis no estudo de casos de imigração (menos frequentes no nosso país), na reconstrução de *pedigrees* complexos e na análise de famílias incompletas.

Os custos envolvidos na análise de SNPs autossómicos ainda não estão completamente calculados, embora se pense que, utilizando algumas técnicas de elevado rendimento, possa ser muito reduzido e claramente inferior ao dos STRs, usando os *kits* comerciais.

A juntar a estes fatores existe a experiência acumulada pelos peritos ao longo dos últimos anos com os STRs, que lhes permite, por exemplo, conhecer já, as mutações que flanqueiam os polimorfismos, e que podem constituir um problema ou desvantagem relativamente aos SNPs.

Em relação aos SNPs é necessária uma extensa validação em grupos populacionais para um número elevado de marcadores [8,12]. Com esta finalidade foi criado um consórcio, o SNPforID

---

*Consortium*, com o objetivo de desenvolver um conjunto de marcadores com utilidade forense. Assim, em 2006 foi publicado um artigo que descrevia um *multiplex* que permite estudar 52 SNPs em simultâneo, amplificados numa única reação de PCR (*polymerase chain reaction*) e posteriormente caracterizados através da eletroforese dos produtos de duas reações de SnaPshot<sup>®</sup>, também em *multiplex* [8]. No referido trabalho é descrita a forma como foram escolhidos os SNPs com interesse forense, isto é, SNPs com elevado poder de discriminação: localizados em cromossomas distintos, se possível em braços distintos, para haver pouca probabilidade de recombinação durante a meiose, com distribuição dos dois alelos possíveis equiprovável e altamente polimórficos.

O *SNPforID Consortium*, desenvolveu e validou duas metodologias de análise simultânea de vários SNPs autossómicos: o método anteriormente referido que tipa 52 SNPs [8], muito polimórficos em Europeus, Asiáticos e Africanos e, posteriormente, em 2007, o *GenPlex system*, que caracterizava 48 desses SNPs iniciais e que incluía também o estudo da Amelogenina [13]. Esta última metodologia revelou-se mais sensível do que a primeira, uma vez que utilizava sondas específicas dos alelos e *locus*. Uma das dificuldades relativamente a esta última metodologia residia na necessidade de efetuar múltiplas pipetagens, que poderiam ser obviadas recorrendo à utilização de *robots* para esse fim [14] [15].

### **1.1.1. SNPs estudados**

Os SNPs estudados neste trabalho são os descritos por Sanchez JJ *et al.* [8]. Habitualmente estes marcadores são analisados utilizando duas reações de extensão de base única (SBE), denominadas por Auto 1 e Auto 2. Estes SNPs foram escolhidos tendo em conta alguns critérios já referidos anteriormente, de modo a que fosse possível utilizá-los com fins de identificação em Genética forense, e foi demonstrado que estavam em equilíbrio de *linkage*, ou seja, espera-se que sejam herdados independentemente, tornando possível a chamada regra do produto nos cálculos estatísticos associados à sua utilização, nomeadamente nos cálculos de índices de paternidade (IP).

### **1.2. STRs: Short Tandem Repeats**

Desde a última metade dos anos 1980, aquando da introdução das tecnologias que envolvem o estudo do DNA nos laboratórios de Genética Forense, que tem havido uma evolução das metodologias mais aplicadas nestes laboratórios. A tecnologia atualmente mais aplicada nesta área é a análise de STR utilizando a PCR (Figura 2). As razões que justificam a seleção de *loci* específicos dos muitos disponíveis para laboratórios de genética forense, são importantes para assegurar que os *loci* tenham um elevado poder de discriminação, sejam estáveis em evidências

forenses e que haja um painel consensual de alguns *loci* considerados mínimos, para usar em bases de dados de DNA.

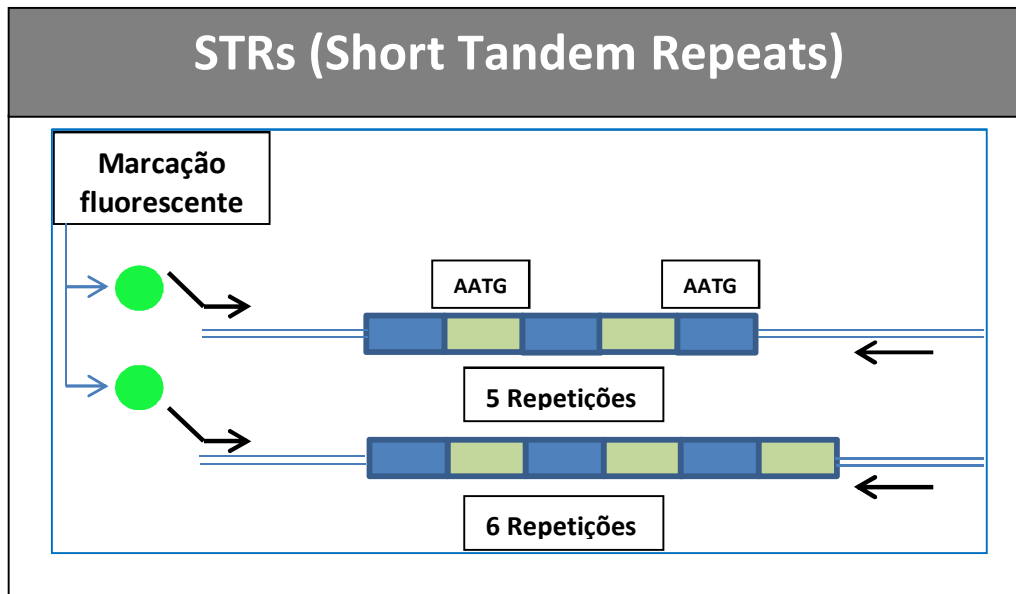


Figura 2. A análise de STRs recorre à marcação fluorescente dos primers.

Os tamanhos dos fragmentos amplificados (*amplicons*) dos STRs que habitualmente se utilizam nesta área e que estão incluídos em *kits* comerciais, podem atingir até 500bp. Isto pode originar falhas na genotipagem quando se trabalha com DNA muito fragmentado proveniente de amostras degradadas, como é o caso das amostras que normalmente se encontram em cenas de crime ou em casos em que se procede à identificação a partir de restos cadavéricos em avançado estado de decomposição ou de vítimas de catástrofes (DVI).

Com a introdução relativamente recente de *kits* comerciais com os chamados mini - STR, que estudam o mesmo tipo de marcadores (STRs) recorrendo a um novo desenho dos *primers* utilizados para amplificar o DNA inicial, originam-se fragmentos de menor tamanho (abaixo dos 200bp), havendo assim a expectativa de continuarem a ser os marcadores de eleição para a resolução da maior parte dos casos que são requisitados aos Laboratórios de Genética Forense (Figura 3).

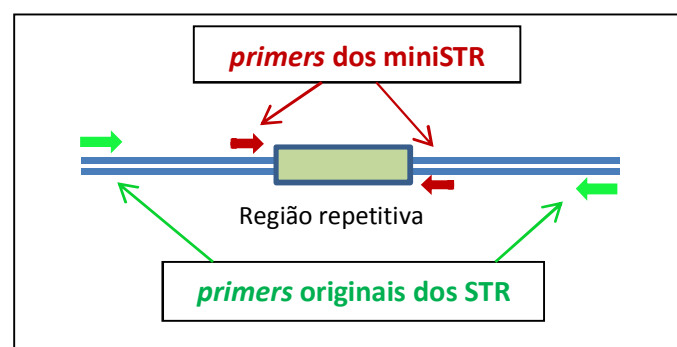


Figura 3. Utilizando diferentes primers os STR obtêm-se os chamados miniSTR

---

Apesar da evolução que existiu em torno de novos *kits* disponíveis, que caracterizam este tipo de marcadores, com o desenvolvimento de novos tampões pelos laboratórios comerciais, que permitem obviar alguns problemas de inibição da PCR aliado ao maior número de marcadores analisados em simultâneo, alguns casos permanecem sem uma resposta satisfatória. Como exemplo do que foi referido, temos os casos de paternidade que envolvem o estudo de familiares do pretense pai, frequentemente um irmão deste, resultando na obtenção de Índices de Paternidade pouco elevados e pouco discriminativos. Existem outros casos de difícil resposta, mais raros, em que obtemos exclusões ou incompatibilidades de transmissão alélica.

O mesmo acontece relativamente a casos de Criminalística Forense, que envolvem frequentemente amostras muito degradadas ou em quantidades muito vestigiais, que tornam a perícia não conclusiva, mesmo recorrendo ao estudo dos miniSTRs.

### **1.3. SNPs versus STRs**

Até que ponto os SNPs substituirão os STRs como primeiro método de análise na área forense não parece ser uma opção no momento. No entanto, não há dúvida de que estes marcadores são muito úteis para algumas aplicações específicas como, por exemplo, na definição de haplogrupos do cromossoma Y [16] e do mtDNA [17], para análise da origem populacional de uma amostra ou para determinar algumas características ligadas aos aspetos fenotípicos ou a características físicas [18]. A standardização e validação inter - laboratorial serão a chave para a sua cada vez maior utilização na área forense [19].

Não existe um método ideal para a análise de SNPs e a seleção de um método apropriado dependerá das necessidades do utilizador. Para a Genética Forense essas necessidades incluem uma capacidade de análise de vários sistemas em simultâneo e elevada precisão, enquanto para os projetos de genotipagem em larga escala, para fins clínicos ou criação de bases de dados, o elevado rendimento e os baixos custos assumem um papel essencial [15].

Como já foi referido, em Genética Forense, desde 2001 que se tem vindo a discutir o seu potencial. Neste sentido foi criado o *SNPforID Consortium*, com o objetivo de desenvolver um conjunto de marcadores com utilidade forense. Assim em 2006 foi publicado um artigo que descrevia um *multiplex* que permite estudar 52 SNPs em simultâneo. Mais recentemente foram descritos outros *multiplex* com finalidade semelhante [20].

---

## 1.4. Metodologias de estudo de SNPs

Para compreender cada tecnologia disponível é importante distinguir entre reações de discriminação alélica, tipos de ensaio e métodos de detecção.

Os produtos de reações de discriminação alélica podem ser detetados por mais de um método e o mesmo método de detecção pode analisar produtos obtidos a partir de ensaios diferentes.

A maior parte dos ensaios para genotipagem de SNPs podem-se dividir em quatro grupos, baseados no mecanismo molecular: hibridação alélico-específica, extensão de *primers*, ligação oligonucleotídica e corte invasivo.

Existem vários métodos de detecção para analisar cada tipo de reação . fluorescência, luminescência, medição de massa, etc.

Os tipos de ensaio podem ainda ser divididos em duas grandes categorias tendo em conta a detecção dos SNPs: reações homogêneas, quando estas ocorrem em solução e reações que ocorrem com recurso a suportes sólidos, como sejam, lâminas de vidro, um *chip*, uma pérola-(*bead*), etc.

Em geral as reações homogêneas são mais fáceis de automatizar, uma vez que não envolvem passos de separação ou purificação posteriores à reação de discriminação. No entanto este tipo de reação tem limitada capacidade de *multiplex*.

Pelo contrário, reações em suporte sólido têm mais possibilidades de *multiplex* mas são necessárias mais manipulações.

### 1.4.1. Ensaios que usam a hibridação alelo-específica

A hibridação alelo-específica, também conhecida como hibridação alelo-específica oligonucleotídica (ASO), baseia-se na distinção de dois DNAs alvo que diferem apenas numa posição nucleotídica, por hibridação. São desenhadas duas sondas (*probes*) específicas normalmente com a base polimórfica situada numa posição central da sequência da sonda. Em condições de ensaio otimizadas, apenas as sondas que hibridam perfeitamente com o alvo são estáveis e os híbridos com um desemparelhamento ou *mismatch* são instáveis (Figura 4).

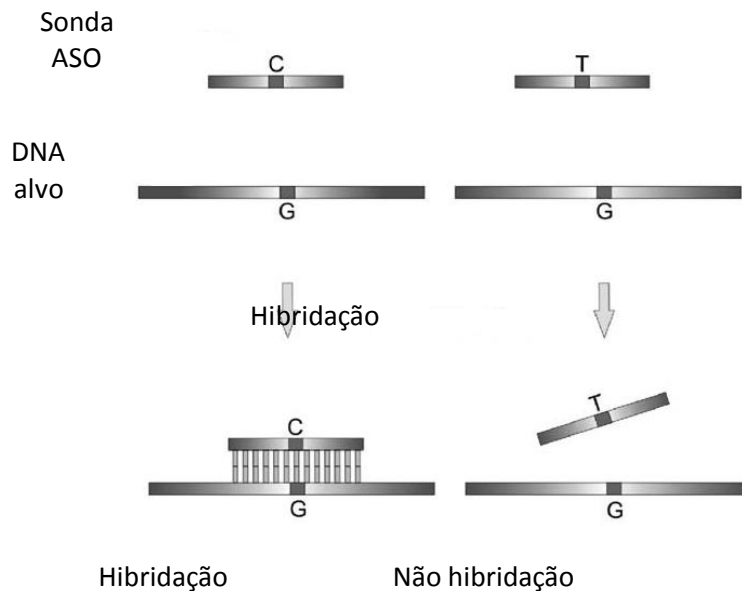


Figura 4. Hibridação com oligonucleótidos alelo específicos (ASO): duas sondas ASO são hibridadas com o DNA alvo que contém o SNP. Sob condições de reação otimizadas, apenas os híbridos sonda-alvo perfeitamente complementares são estáveis. (*Forensic Sci Int.* 154(2005) 181-194)

Este tipo de sondas ASO, com formato de *reverse dot-blot*, foi usado para detetar os primeiros polimorfismos analisados por PCR na área da Genética Forense (sendo ainda utilizadas em alguns laboratórios), embora tenham sido progressivamente substituídas pela análise de STRs. Para rentabilizar todas as vantagens destes novos formatos de sondas na genotipagem de SNPs é necessário utilizar métodos de detecção que permitam uma elevada precisão, elevada sensibilidade e alto rendimento.

#### 1.4.1.1. Hibridação homogénea usando a tecnologia FRET

A transferência de energia de ressonância fluorescente (FRET) ocorre quando dois marcadores fluorescentes estão muito próximo um do outro e o espectro de emissão de um deles se sobrepõe ao espectro de excitação do outro fluoróforo. Os métodos que utilizam FRET combinam a discriminação alélica que usa sondas ASO com reações PCR em tempo real e quantificação. Por isso, além das sondas para genotipagem são necessários dois *primers*. O incremento de fluorescência pode ser medido em tempo-real durante a PCR ou quando a PCR termina. Existem múltiplas variações baseadas no mesmo princípio.

Em Genética Forense por vezes recorre-se a esta metodologia, em conjugação com a chamada TaqMan<sup>®</sup> assay (Applied Biosystems), que explora a função 5' exonucleotídica da Taq polimerase. Neste ensaio TaqMan<sup>®</sup> a Taq polimerase que desloca e corta as sondas hibridadas com o DNA alvo, gerando um sinal fluorescente. São necessárias duas sondas TaqMan<sup>®</sup> que diferem no local polimórfico; uma sonda é complementar ao alelo correspondente ao alelo ancestral (*wild-type*) e a outra ao alelo variante ou alternativo. Estas sondas possuem diferentes marcações fluorescentes na extremidade 5' e um *quencher* (supressor) ligado à extremidade 3'

Quando as sondas estão intactas, o *quencher* interage com o fluoróforo por FRET, absorvendo a sua fluorescência. Na fase de extensão, o fluoróforo na extremidade 5q é cortado pela polimerase, através da sua atividade 5qexonucleotídica, levando a um aumento de fluorescência do fluoróforo *reporter* (repórter) (Figura 5). As sondas que apresentam desemparelhamento sofrem deslocamento sem se fragmentarem.

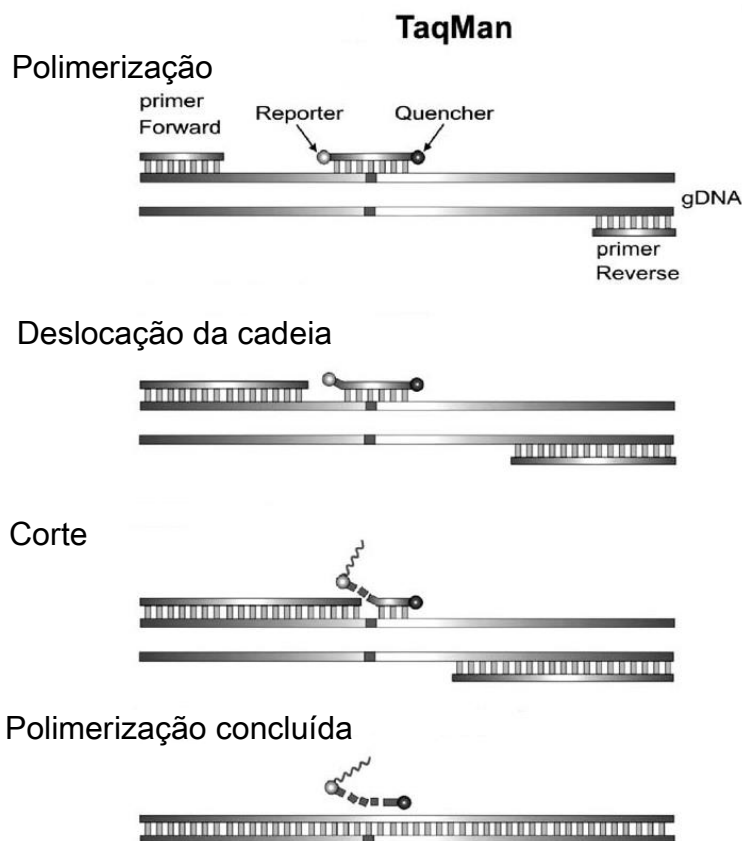


Figura 5. Ensaio TaqMan; Este ensaio baseia-se na atividade de nuclease 5qda Taq polimerase. Quando as sondas estão intactas, o quencher interage com o fluoróforo por FRET, captando a sua fluorescência. No passo de extensão, o marcador fluorescente da extremidade 5q é cortado por ação da actividade da 5q nuclease da Taq polimerase, originando um aumento de fluorescência do marcador reporter. Sondas não complementares são deslocadas sem sofrer fragmentação. (*Forensic Sci Int.* 154(2005) 181-194)

A genotipagem de determinada amostra obtém-se por medição da intensidade do sinal dos diferentes fluoróforos. A deteção destes fluoróforos depende da capacidade do equipamento que se utiliza e pode ser feita num espectrómetro de luminescência ou num equipamento de PCR em tempo real (normalmente deteta poucos fluoróforos, tendo como consequência uma capacidade de multiplex limitada).

No *LightCycler® Real-Time PCR System* (Roche) também se utilizam duas sondas marcadas com fluoróforos. A sonda 1 está marcada com fluoresceína na sua extremidade 3q e a sonda 2 possui outra marcação (LC Red) na extremidade 5q. As sequências dos dois oligonucleótidos são seleccionadas para que hibridem adjacentes uma à outra no DNA alvo. Quando os oligonucleótidos hibridam, os dois fluoróforos ficam posicionados de forma muito próxima. O primeiro (fluoresceína), é excitado e emite luz fluorescente verde que, por sua vez, vai excitar o

---

fluoróforo adjacente (LCRed) emitindo ele próprio e de seguida, luz fluorescente. Esta transferência de energia, chamada transferência de energia de ressonância por fluorescência ou FRET (*fluorescence resonance energy transfer*), é altamente dependente do espaçamento entre as duas moléculas fluorescentes. Apenas se ambas estiverem a uma distância entre 1-5 nucleótidos é que há transferência de energia com eficiência suficientemente elevada. O aumento de fluorescência, emitida pelo fluoróforo LCRed filtrada e medida pelo sistema ótico do equipamento do *LightCycler*<sup>®</sup> é proporcional ao aumento de DNA gerado pela PCR. Uma vez que o LC Red apenas emite um sinal quando ambos os nucleótidos estão hibridados, a medição de fluorescência realiza-se após o passo de ligação dos *primers* ou *annealing*. Uma das sondas fica numa posição central de acordo com a base polimórfica e a outra fica numa posição adjacente para permitir o FRET. Um único desemparelhamento ou *mismatch* pode reduzir significativamente a temperatura de fusão ou de *melting* do oligonucleótido que depende fundamentalmente do comprimento deste oligonucleótido e da posição relativa do não emparelhamento.

A redução da temperatura de *melting* pode observar-se realizando uma análise de curvas de *melting*. Mais do que um SNP pode ser genotipado em simultâneo usando este método, combinando diferentes marcadores fluorescentes com sondas com diferentes temperaturas de *melting*. A primeira versão do *LightCycler*<sup>®</sup> apenas detetava duas marcações fluorescentes, mas a nova versão do instrumento, permite usar 4 marcações. Foi descrito um 4-plex de Y-SNPs [21]. Neste trabalho discute-se a vantagem desta tecnologia e que consiste na elevada sensibilidade.

#### **1.4.2. Ensaios que utilizam a extensão dos primers**

Esta metodologia baseia-se na capacidade que a DNA polimerase possui de incorporar desoxiribonucleótidos complementares à sequência do DNA alvo. Existem muitas variantes desta reação, mas podem-se dividir em dois principais tipos:

- A minisequenciação, ou extensão dos *primers* em apenas um nucleótido, em que a base polimórfica é detetada por adição de um dideoxinucleótido (ddNTP), complementar à base em questão, pela DNA polimerase. Existe ainda a chamada extensão alelo-específica, em que a DNA polimerase apenas amplifica se o *primer* tiver uma complementaridade perfeita com o DNA molde;
- Outra alternativa é a chamada pirosequenciação, que também se baseia na reação da DNA polimerase.

Descreve-se um pouco a reação de minisequenciação.

### 1.4.2.1. Minisequenciação

Na reação de minisequenciação existe um *primer* que se liga ao seu DNA alvo imediatamente adjacente ao SNP sofre extensão de apenas um nucleótido, complementar ao local polimórfico, por ação da DNA polimerase (Figura 6).

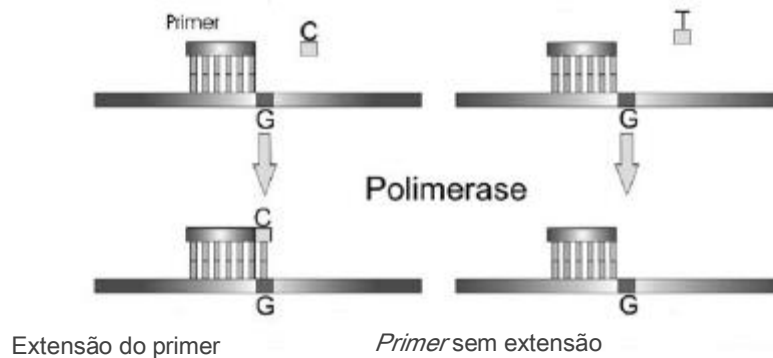


Figura 6. Reação de extensão dos primers . minisequenciação. Um primer liga-se ao seu DNA alvo imediatamente a montante ao SNP e sofre extensão de apenas um nucleótido complementar à base polimórfica. (*Forensic Sci Int.* 154(2005) 181-194).

Este método baseia-se na precisão com que as DNA polimerases fazem a incorporação dos nucleótidos.

A detecção dos SNPs no genoma humano, e também noutros genomas de grande dimensão, requer uma amplificação prévia da região que flanqueia o local da variação, por PCR. Trata-se de uma necessidade frequente em muitas das tecnologias usadas com a finalidade de genotipar SNPs. Antes da minisequenciação propriamente dita, de modo a obter produtos específicos a partir da extensão dos *primers*, é necessário remover o excesso de reagentes que foram utilizados na PCR anterior, tais como *primers* e os dNTPs (deoxinucleótidos trifosfato).

Existem ainda diferentes tecnologias para analisar estes produtos de extensão. Na reação de minisequenciação podem-se utilizar nucleótidos marcados ou não marcados, ddNTP combinados com dNTP ou apenas ddNTP, dependendo do método que vai depois ser usado para a deteção dos produtos.

#### 1.4.2.1.1. Extensão alelo-específica

A extensão alelo-específica baseia-se na alteração da eficiência de extensão por parte da DNA polimerase quando os *primers* nas suas extremidades 3' são ou não compatíveis com o DNA alvo. A DNA polimerase apenas efetua a extensão dos *primers* quando a sua extremidade 3' é perfeitamente complementar ao DNA alvo (Figura 7). São necessários dois *primers*, um para cada um dos alelos de determinado SNP.

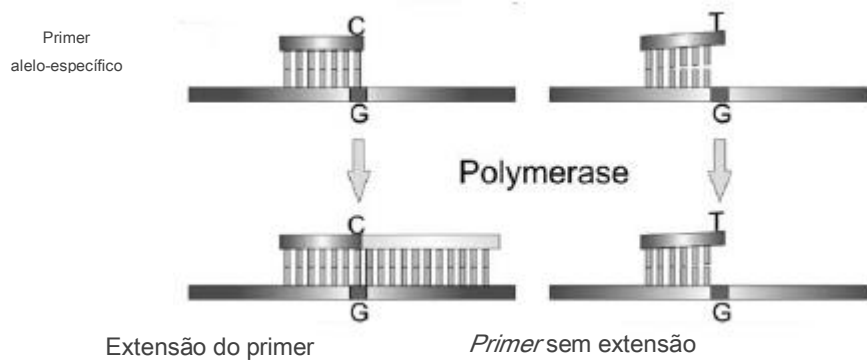


Figura 7. Reação de extensão dos primers . extensão alelo-específica. A extremidade 3' dos primers é complementar a cada alelo do SNP. (*Forensic Sci Int.* 154(2005) 181-194).

A detecção diferencial do *primer* de que é composto o produto, determina o genótipo de uma amostra. O produto desta reação pode ser detetado em *microarrays* usando nucleótidos marcados com fluorescência [22] (Figura 8).

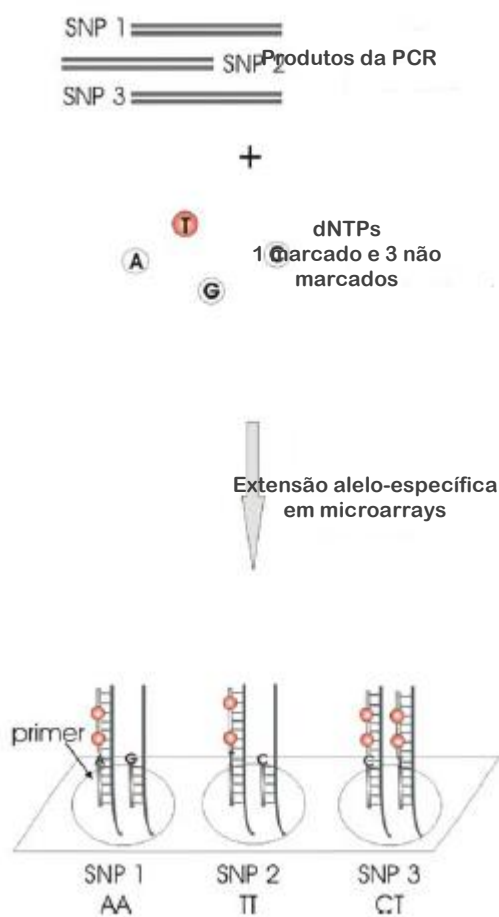


Figura 8. Exemplo de tecnologia de tipagem de SNPs envolvendo microarrays. Extensão alelo-específica . os primers alelo-específicos estão ligadas ao microarray. Quando há complementaridade perfeita com o DNA alvo, os primers sofrem extensão. A extremidade 3' dos primers é complementar a cada alelo do SNP. (*Genome Res.* 10 (2000) 1031. 1042).

---

Outra variante desta reação consiste na utilização habitual de um *primer* reverso, conjuntamente com os *primers* alelo-específicos: nos casos em que o *primer* complementar permite a amplificação de um alelo específico numa amostra. Neste caso a reação denomina-se por PCR alelo-específica. A deteção dos produtos de PCR apropriados permite a genotipagem das amostras. A utilização de *primers* marcados permite a identificação específica de cada produto de PCR, utilizando vários métodos de deteção.

A capacidade de *multiplex* depende também da tecnologia usada na deteção. As tecnologias mais comuns para analisar produtos de minisequenciação são a eletroforese e deteção de fluorescência, MALDI-TOF (*matrix assisted laser desorption/ionization time-of-flight*), espectrometria de massa e *microarrays*.

#### 1.4.3. Reação de ligação alelo-específica

A DNA ligase é altamente específica na reparação de interrupções (cortes) no DNA (falta de uma ligação fosfodiéster). Landergren et al. [23] descreveram o ensaio de ligação oligonucleotídica ou OLA (*oligonucleotide ligation assay*) como método de tipagem de SNP baseado na capacidade da ligase ligar covalentemente dois oligonucleótidos quando hibridam perto um do outro num DNA molde. O ensaio OLA necessita que sejam concebidas três sondas sejam, uma habitual/normal e duas alelo específicas. A sonda normal liga-se ao DNA imediatamente a jusante do SNP. Uma das sondas alélicas tem na extremidade 3'q nucleótido complementar a um dos alelos, com a outra sonda alélica complementar ao alelo alternativo. Estas duas sondas alélicas competem aquando do *annealing* com o DNA alvo adjacente à sonda normal. Isto faz com que se gere uma região de dupla cadeia contendo uma falha correspondente ao local do alelo. Apenas a sonda alelica perfeitamente complementar com o alvo vai ser ligada à sonda comum pela DNA ligase (Figura 9).

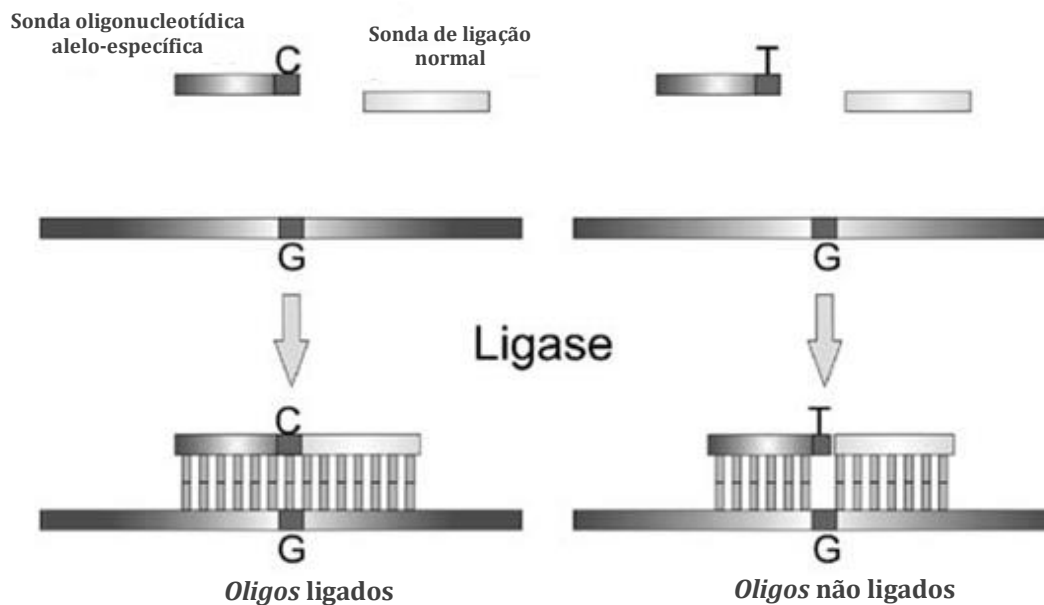


Figura 9. Reação de discriminação alélica - Ensaio de ligação oligonucleotídica (OLA): por cada SNP são necessárias duas sondas alelo-específicas e uma sonda de ligação normal. A sonda de ligação normal hibrida num local adjacente ao da sonda alelo-específica. Quando existe complementaridade perfeita da sonda alelo-específica, a ligase une as duas sondas a alelo-específica e a sonda normal. (*Forensic Sci Int.* 154(2005) 181-194).

A utilização de uma DNA ligase termo-estável permite a repetição de ciclos térmicos, resultando num aumento linear de produtos de ligação. Se forem usadas ambas as cadeias de DNA genómico como alvo para hibridação oligonucleotídica, o aumento de produtos de ligação pode ser exponencial. Esta reação denomina-se reação em cadeia da ligase (*ligase chain reaction - LCR*) [24]. São usados dois conjuntos de oligonucleótidos, complementares a cada cadeia alvo. Os produtos de ligação resultantes da primeira reação de ligação poderão ser os alvos do próximo ciclo de ligação. Têm sido desenvolvidos diferentes formatos de ensaios para deteção de produtos de ligação. O uso da biotina na sonda normal e um grupo *reporter* nas sondas alelo-específicas permite a captura e deteção do produto [23]. Outros ensaios substituem a biotina por modificadores de mobilidade sendo as sondas alelo-específicas marcadas com diferentes fluoróforos, permitindo assim discriminar os produtos de ligação por tamanho e por cor [25]. A utilização de modificadores de mobilidade e de fluoróforos permite fazer OLA em *multiplex* em que os produtos são separados por eletroforese sob condições desnaturantes com deteção fluorescente. Os modificadores de mobilidade permitem uma regulação precisa da mobilidade de cada produto de ligação, independentemente do tamanho do oligonucleótido. Esta estratégia foi denominada de separação *sequencecoded* (*sequencecoded separation*) por Grossman et al. [25]. Neste ensaio é realizada uma PCR antes da reação OLA. No entanto, os procedimentos associados de amplificação e de ligação oligonucleotídica denominada CAL (*coupled amplification and oligonucleotide ligation*), combinam a amplificação por *multiplex* e a genotipagem dos SNPs

---

numa única reação. Para isso são utilizados *primers* da PCR com temperaturas de *melting* elevadas e sondas de ligação com temperaturas *melting* inferiores. Todos os reagentes são colocados em simultâneo num único tubo, minimizando o trabalho manual e permitindo a fácil automação. Durante a primeira fase de reação os *primers* da PCR ligam-se ao DNA alvo e a amplificação ocorre a uma temperatura superior à temperatura de *melting* das sondas de ligação dos oligonucleótidos. Na segunda fase a temperatura é reduzida, permitindo que as sondas hibridem e que a ligação ocorra. Os produtos de ligação são detetados usando sequenciadores automáticos.

Uma variante do ensaio CAL é a *dye-labelled oligonucleotide ligation* (DOL) [26]. Neste caso os produtos de ligação são detetados por monitorização da FRET em tempo real.

## 1.5. Aplicação Forense dos SNPs

Têm sido desenvolvidas muitas tecnologias de genotipagem de SNPs nos últimos anos. Como resultado disso existem vários protocolos disponíveis para os investigadores, no entanto nenhum deles cumpre todos os objetivos pretendidos. Convém ter em conta diversos aspetos quando se opta por uma tecnologia com propósitos forenses, tais como a sensibilidade, a reprodutibilidade, a precisão, a capacidade de *multiplex* e a capacidade de rendimento. Também é importante ter em conta a flexibilidade da tecnologia, o tempo de análise e os custos envolvidos, no que respeita quer aos equipamentos quer aos reagentes por genótipo.

A capacidade de reposta pretendida depende de cada aplicação. Algumas aplicações permitem estudar um número reduzido de marcadores numa amostragem elevada, outras estudam um número elevado de marcadores em poucas amostras, outras ainda, permitem estudar muitos marcadores em grandes amostragens. Para aplicações forenses há necessidade de um rendimento global intermédio no caso das paternidades e casos criminais, mas no que se refere à implementação de bases de dados genéticas seria necessário uma metodologia com elevado rendimento. Um passo limitante em qualquer destas aplicações é a quantidade inicial de DNA necessária á genotipagem. Algumas técnicas procedem à genotipagem direta do SNP no DNA genómico sem necessidade de uma PCR prévia, como é exemplo o ensaio SNPlex. Nestes casos, tem de existir um melhor ajuste da quantidade mínima de DNA inicial a usar, mas será sempre superior do que nos casos em que se procede a uma PCR prévia.

Para as restantes tecnologias que requerem uma PCR antes de proceder à reação para a discriminação alélica, é essencial a capacidade de realizar a PCR em *multiplex*, não só porque geralmente a quantidade de DNA disponível para análise nos casos criminais é muito reduzida, mas também sob o ponto de vista do rendimento global. Em casos críticos, alguns autores

---

sugerem que seria aconselhável proceder à amplificação total do genoma antes da análise dos marcadores escolhidos, para superar a pouca quantidade de DNA disponível.

Outro aspeto muito importante na Genética Forense é a análise de misturas. Devido à natureza bi-alélica da maioria dos SNPs será difícil detetar quando estamos perante uma mistura em determinada amostra, usando estes marcadores. Por isso, a possibilidade de quantificar cada um dos alelos numa amostra pode ajudar na determinação da contribuição de cada componente num perfil de mistura. Algumas tecnologias como é o caso da espectrometria de massa, permitem algum tipo de quantificação. Esta característica tem sido usada de forma rotineira para fazer estimativas de frequências alélicas em conjuntos de amostras, mas poderia ser uma característica útil na Genética Forense.

Em tecnologias que se baseiam em hibridação homogénea com deteção FRET como é o caso do ensaio TaqMan<sup>®</sup>, as reações de PCR e da discriminação alélica são realizadas na mesma reação. Esta vantagem evita passos adicionais, favorecendo a automação e o rendimento global do processo, especialmente quando se usam os equipamentos de elevado rendimento para os ensaios TaqMan<sup>®</sup>. A principal desvantagem destes ensaios consiste na reduzida capacidade de *multiplex*. Como consequência, estas tecnologias são uma boa opção para validar SNPs selecionáveis e para construir bases de dados criminais, mas não para ser usadas como tecnologia de rotina nos casos criminais.

As tecnologias de minissequenciação são, de momento, os métodos preferidos nos laboratórios de Genética Forense, especialmente o SNaPshot<sup>™</sup>, uma vez que a sua deteção se realiza num equipamento de eletroforese capilar automático, que já faz parte do equipamento habitual dos laboratórios para análise de STRs. Foram desenvolvidos vários projetos, em diferentes laboratórios, relacionados com a análise de SNPs com recurso a esta tecnologia, relacionadas Y-SNPs e SNPs do DNA mitocondrial [27,28,29]. A capacidade de *multiplex* responde às necessidades forenses mas é necessário um trabalho de otimização de concentrações, desenho de *primers* e reação de minissequenciação para conjuntos de múltiplos SNPs.

Outros métodos como matrizes de tipagem (*microarrays*) e MALDI-TOF parecem ser apropriados para fins forenses, tal como descrevem alguns autores [30,31]. No entanto, a maior desvantagem destas tecnologias é o facto de serem necessários equipamentos específicos que teriam de ser adquiridos pelos laboratórios, em contraste com o SNaPshot<sup>™</sup>. Acresce ainda que a capacidade de fazer *multiplex* no caso da MALDI-TOF é menor do que com as tecnologias de minissequenciação. O maior problema da deteção em *microarrays* é a dificuldade na reprodutibilidade e na validação.

A piro-sequenciação é muito limitada no que se refere à possibilidade de desenvolver reações em *multiplex*, além disso tem limitações na automação, uma vez que envolve muitos passos antes da

---

deteção. A maior vantagem é a possibilidade de quantificar a contribuição de cada alelo, uma característica muito útil para análise de perfis de mistura.

A possibilidade de multiplex de várias tecnologias baseadas em OLA é muito elevada, mesmo acima dos requisitos forenses. O maior problema destas tecnologias é o facto de as reações de discriminação alélica se realizarem diretamente no DNA genómico. Trata-se de uma grande vantagem quando existe quantidade suficiente de DNA para análise mas não nos casos criminais. Todas estas tecnologias têm sido descritas nos últimos anos e é necessário desenvolver mais esforços para a sua utilização com propósitos forenses.

As tecnologias que cumprem os critérios necessários para aplicações forenses têm de ser bem exploradas para que as suas limitações possam ser bem conhecidas e para aferir da possibilidade de serem usadas na genética forense. É difícil encontrar uma única tecnologia que sirva para responder a todas as necessidades e, portanto, não é uma tarefa fácil definir a melhor tecnologia para genotipar SNPs em Genética Forense.

Acresce dizer que existiu um progresso tecnológico muito rápido, devido ao facto de as companhias biotecnológicas estarem a desenvolver novas estratégias de tipagem de SNPs, tornando-se difícil escolher o melhor método para aplicações específicas.

Provavelmente, poderão ser usadas diferentes tecnologias no trabalho de rotina, distinguindo-se aquele onde a quantidade e integridade do DNA é mais crítica, do que o que envolve testes de paternidade ou criação de bases de dados criminais de DNA. É ainda possível diferenciar entre SNPs autossómicos e Y-SNPs e SNPs do DNA mitocondrial, tendo em conta o número de marcadores que é necessário analisar por amostra e a estratégia para a análise. Por isso, as necessidades de *multiplex* não são as mesmas para os diferentes tipos de aplicações bem como a tecnologia mais apropriada pode ser distinta para diferentes tipos de SNPs.

Em geral, todas as tecnologias que possuem capacidade limitada de *multiplex* devem ser excluídas da rotina forense; no entanto poderão ser úteis para estimar frequências alélicas de determinados SNPs selecionados e também para criação de bases de dados genéticas criminais.

Em resumo, das diversas metodologias existentes e que permitem a caracterização de SNPs, na área da Genética Forense será importante selecionar aquelas que possibilitem a caracterização de um elevado número de SNPs a partir de uma quantidade inicial de amostra relativamente reduzida. Será ainda necessário ter em conta os equipamentos de que habitualmente dispõem os laboratórios de Genética Forense.

Como já foi anteriormente referido, o *SNPforID consortium*, um grupo de investigadores europeus, criado em 2003, com o objetivo de estudar SNPs capazes de ser suficientemente informativos

---

para a resolução de casos forenses, identificou um conjunto de 52 SNPs que podem ser caracterizados usando métodos padronizados de minissequenciação [7]. Este *multiplex* baseia-se na detecção por SBE (Single Base Extension), usando o SNaPshot<sup>®</sup> (kit comercializado pela Applied Biosystems).

A Applied Biosystems (AB) também desenvolveu um sistema de caracterização simultânea de SNPs (*multiplex*), fundamentado na informação fornecida pelo *SNPforID consortium*, que incluía 48 dos 52 SNPs autossômicos e a amelogenina, baseado no sistema GenPlex<sup>®</sup>, que é uma modificação do sistema SNPLEX<sup>®</sup> (AB).

---

## CAPÍTULO II

### 2. Objetivo do trabalho

#### 2.4. Objetivo principal

O objetivo principal deste trabalho consistiu em fazer uma validação interna de uma metodologia de análise de SNPs autossómicos por minissequenciação (SNaPshot<sup>®</sup>) metodologia proposta por Sanchez et al. (2006) e desenvolvida pelo *SNPforID Consortium*, que consiste na caracterização de 52 SNPs autossómicos amplificados numa única reação de PCR, seguida de duas reações de SBE (Single Base Extension) [7] para análise de casos de parentesco, complexos, com o intuito de aumentar poder informativo.

Um objetivo adicional inclui ainda a aplicação desta metodologia à análise de amostras forenses, com particular aplicação ao estudo de amostras "difíceis", relativas a perícias do âmbito da criminalística biológica; ou seja, as que pela quantidade ou qualidade do seu material genético não proporcionaram resultados, através do estudo dos STRs.

#### 2.2. Objetivos específicos:

Para cumprir o objetivo geral, teríamos como objetivos específicos:

- Efetuar o estudo populacional de uma amostra da população do Norte de Portugal e de um grupo populacional de indivíduos do leste europeu a residir em Portugal, maioritariamente Ucrânios.
- Validação primária da metodologia selecionada, através da comparação dos resultados obtidos no âmbito da realização de um exercício colaborativo proposto pelo GHEP-ISFG (Grupo de Línguas Espanhola e Portuguesa da *International Society for Forensic Genetics*).
- Aplicar o estudo de SNPs autossómicos na população do Norte de Portugal a perícias forenses já concluídas, designadamente casos complexos de investigação de parentesco, casos de criminalística biológica e casos de identificação genética individual.

- 
- Estudo de amostras seguras provenientes de casos de parentesco complexos já concluídos nos quais, com recurso aos STRs, apresentando reduzidos valores de IP, e com interesse complementar a análise convencional (estudo de STRs).
  - Estudo de amostras de casos de criminalística biológica e de amostras degradadas em condições laboratoriais controladas.

---

## CAPÍTULO III

### 3. Seleção da metodologia de análise dos SNPs autossómicos

#### MATERIAL E MÉTODOS

##### **Amostras**

- Amostra de alguns casos complexos de parentesco já realizados na Delegação do Norte do INML relativamente aos quais se obtiveram reduzidos índices de paternidade ( $IP < 10\ 000$ ); complementar com amostras de indivíduos não relacionados entre si representativos da nossa população (Norte de Portugal).
- Amostras obtidas a partir de casos de criminalística biológica e de identificação genética individual, já concluídos em que a abordagem tradicional (STRs) não resultou ou a partir das quais não foi possível obter um perfil completo, relativamente às quais se pretenda complementar a análise convencional (estudo de STRs);

##### **Colheita**

- Amostras de saliva foram colhidas a indivíduos não relacionados do norte de Portugal envolvidos em casos de investigação de parentesco complexos, após consentimento informado, as quais foram caracterizadas para 49 marcadores autossómicos.

As amostras respeitantes à nossa população (norte de Portugal) foram previamente estudadas com os *kits* PowerPlex16 (Promega) e Identifiler (Applied Biosystems) verificando-se deste modo a unicidade de cada amostra.

Todas as amostras, cerca de 80 no total, foram sujeitas a uma codificação para que fosse garantida a confidencialidade relativamente à identidade dos indivíduos analisados neste estudo.

##### **Extração**

Os extratos de DNA foram obtidos por recurso a métodos considerados tradicionais como é o caso do método de Chelex modificado, com fenol-clorofórmio- álcool isoamílico e ou recorrendo à utilização de bio-robots como é o caso do EZ1 (Qiagen) ou utilizando o kit PrepFiler no AutoMate Expressi Forensic DNA Extraction System da Applied Biosystems® (AB).

---

## **Quantificação**

As concentrações de DNA nos extratos foram determinadas recorrendo à metodologia de real-time PCR usando o Quantifiler™ Human DNA Quantification Kit (AB) e o ABI PRISM® 7500 Sequence Detection System (AB). As concentrações finais variaram entre 0.20 e 10 ng/μl tendo as amostras sido quantificadas em duplicado e as concentrações ajustadas a 1ng/μl.

## **Marcadores escolhidos**

Os marcadores escolhidos neste estudo são SNPs bi-alélicos e foram descritos em Sanchez et al. (2006). Estes SNPs estão posicionados nas porções distais dos braços dos cromossomas, exceto o do único SNP localizado no cromossoma 19. Por isso, são estudados pelo menos dois SNPs na maioria dos cromossomas, mas em alguns deles são estudados quatro, nos cromossomas 1, 9, 10, 13, 15, 21, e 22 [8]. Nos cromossomas mais pequenos como o 21 e o 22, o número médio de eventos de recombinação por meiose é um [13,14], como consequência, dois dos três SNPs nestes cromossomas vão situar-se do mesmo lado, por recombinação, e os descendentes herdam dois *loci* ligados e um *locus* independente dos pais. No entanto pode-se calcular o Índice de Paternidade (PI) tal como calcularíamos se os *loci* fossem independentes, porque estes estão em equilíbrio na população [8] e porque os pais, numa paternidade típica, são desconhecidos.

## **Amplificação**

Para cada amostra 52 ou 48 SNPs, comuns às duas metodologias, foram amplificados numa única reação de PCR e caracterizados recorrendo ao %Genplex system+ (metodologia abandonada) ou por SNaPshot™ em duas reações SBE.

## **Í Genplex systemí**

O ensaio do %Genplex system+ envolveu vários passos, realizados em placas %microtite+ de 96 poços (MTP plates):

- **Reação de PCR** em que se pré-amplificam os 48 SNPs, usando os mesmos primers descritos para o 52 SNP-plex, de modo a originar produtos com 59-115 bp de tamanho, incluindo também um par de %primers+ para a amelogenina, originando um produto com 170bp, usando produtos incluídos num kit desenvolvido pela Applied Biosystems (AB).

- **Limpeza pós-PCR** em que se removem as bases não incorporadas e os %primers+ com ExoSAP-IT1.

---

- **OLA**: ligação de oligonucleótidos específicos de locus (LSOs) marcados com Biotina que se ligam à sequência imediatamente após o SNP e oligonucleótidos específicos dos alelos (ASOs) que hibridam directamente aos alelos e à sequência a montante do SNP. Cada ASO é identificado por uma sequência 3' que inclui complementar a uma sonda de detecção ZipChute®.

- **Ligação dos produtos OLA marcados com Biotina a uma placa MTP de captura**, cujos poços possuem Estreptavidina, placa esta que funciona como uma fase sólida para a hibridação com as sondas de detecção ZipChute®.

- **Hibridação das sondas ZipChute® com os produtos OLA capturados**, cada uma das quais com cor e mobilidade determinadas pelas propriedades químicas de modificadores de mobilidade e pelos corantes FAM e /dR6G.

- **Eluição seguida de eletroforese capilar das sondas ZipChute®**, após mistura com uma solução de carga contendo um padrão de pesos moleculares marcado, o  $\lambda$ DZ® size standard.

A eletroforese deve incluir controlos negativos e vários ladder alélicos. Os produtos são separados utilizando o polímero POP-7i durante 15 minutos.

Todos os reagentes necessários estão incluídos no SNPlexi system (AB).

A análise dos produtos do GenPlex e SNPlexi por electroforese capilar requer um sequenciador 3130 ou 3730 Genetic Analyzer e o GeneMapper v.4.0, como software de análise.

Esta experiência, que decorreu em Santiago de Compostela e utilizou o sistema GenPlexi, envolveu cerca de 80 amostras, resultando daí a caracterização da maior parte dessas amostras. Seria necessário realizar um número mais elevado de ensaios no sentido de adquirir experiência relativamente a esta metodologia de análise. Entretanto, a circunstância do GenPlex ter sido descontinuado, inviabilizou a prossecução do estudo com este kit.

### **SnaPshot!**

Procedeu-se à caracterização dos SNPs autossómicos desenvolvidos e validados pelo consórcio SNPforID usando a tecnologia de SnaPshoti.

O ensaio 52 SNP-plex envolve a amplificação simultânea dos 52 SNPs autossómicos. Depois da remoção do excesso de primers e nucleótidos não incorporados, segue-se a inclusão dos produtos SBE resultantes de dois multiplexes por SnaPshoti (AB), analisados por CE

---

(eletroforese capilar), usando o POP-4 como polímero, após tratamento com SAP (Shrimp Alkaline Phosphatase).

**Amplificação por PCR:** 52 SNPs autossômicos foram amplificados em simultâneo por PCR (polymerase chain reaction), num volume final de 12.5  $\mu$ l contendo DNA de concentração variável, de 0.40 to 3 ng. As condições de amplificação foram conforme já descrito [2]. Depois da desnaturação a 95°C durante 10 minutos, a PCR decorreu em 35 ciclos de 95°C 30s, 60°C 50s e 65°C 45s, com uma extensão final de 6min a 65°C.

**Limpeza pós-PCR:** Após a PCR é necessário proceder à remoção das bases não incorporadas e dos primers, com ExoSAP-IT. Adiciona-se de 2  $\mu$ L ExoSAP IT a 5  $\mu$ L de produto de PCR e incuba-se 45 minutos a 37°C. A inativação da enzima faz-se por incubação a 85°C durante 15 minutos.

**Reação SBE:** Realizaram-se duas reações single base extension (SBE) como também foi descrito pelos mesmos autores, com ligeiras modificações [2] e incluíram 2.5  $\mu$ l de Snapshot reaction mix (AB), 1.5  $\mu$ l de SBE primer mix (0.01. 0.27 M) e 2  $\mu$ l de produto de PCR purificado com ExoI-SAP (Exonuclease I/Shrimp Alkaline Phosphatase), totalizando 6  $\mu$ l. Colocaram-se no termociclador a amplificar durante 30 ciclos de 96°C durante 11segundos, 55°C durante 5 segundos e 60°C durante 30 segundos.

**Limpeza pós-SBE:** Os nucleotidos em excesso foram removidos juntando à mistura no tubo de reação (SBE mix) 1  $\mu$ l de SAP (1 U/  $\mu$ l) e procedendo a duas incubações a 37°C durante 45 minutos, seguida de 15 minutos a 85°C.

**Genotipagem:** A separação dos produtos de SBE realizou-se por recurso à eletroforese capilar, usando um sequenciador ABI Prism 3130xl Genetic Analyzer (AB), com arrays de capilares de 36 cm e usando POP-4 como polímero (AB).

**Análise dos produtos de SBE:** Foi levada a cabo recorrendo ao software GeneMapper® ID-X. As atribuições alélicas foram realizadas manualmente.

### **Validação inicial**

Procedeu-se a uma validação primária desta segunda metodologia, ou seja do 52-plex. Esta validação consistiu em comparar os resultados consenso dos laboratórios forenses participantes num exercício internacional promovido pelo Grupo Português e espanhol da Sociedade Internacional de Genética Forense (GHEP-ISFG) com os resultados obtidos pelo laboratório onde

---

este estudo foi desenvolvido (Serviço de Genética e Biologia Forense da Delegação do Norte do Instituto Nacional de Medicina Legal, I.P. . SGBF-N).

Foram também usados alguns controlos positivos de DNA (9947<sup>A</sup> e 007), incluídos nos kits comerciais e disponíveis para tipagem de STRs autossómicos e os seus resultados comparados com os publicados e estabelecidos para vários controlos positivos de DNA [32].

Aconteceu ainda uma participação num segundo exercício colaborativo promovido pelo GHEP-ISFG, que envolveu vários Laboratórios de Genética Forense. Esta participação constituiu um marco fundamental na seleção final do método de caracterização de SNPs e funcionou como motivação adicional para a conclusão do trabalho.

Como complemento da validação primária já referida procedeu-se à caracterização de SNPs autossómicos em amostras laboratorialmente degradadas. Foram estudadas amostras artificialmente degradadas no laboratório, usando diferentes condições de exposição a raios de UV e tratamento com DnaseI, de modo a obter fragmentos de DNA de tamanho reduzido, descrita num dos capítulos seguintes.

A estratégia descrita no projeto inicial foi alterada fundamentalmente porque o método que utilizava o sistema %GenPlex+foi inviabilizado pelo facto deste kit ter sido descontinuado.

Por isso, neste trabalho, optou-se pela utilização exclusiva da minissequenciação.

---

---

## CAPÍTULO IV

### 4. Aplicação do 52-plex em estudos populacionais

#### 4.1. Estudo populacional de uma amostra do Norte de Portugal e de uma amostra de indivíduos imigrantes do leste europeu a residir em Portugal

##### 4.1.1. Sumário

Neste estudo foram utilizadas amostras conhecidas como amostras seguras, também denominadas de *amostras de referência*, provenientes de alguns casos de parentesco.

Foi realizada a extração de DNA, quantificação e caracterização de SNPs relativamente a 131 amostras de indivíduos do Norte de Portugal, selecionadas a partir de casos de investigação de parentesco, bem como de 85 amostras de indivíduos imigrantes do leste Europeu, maioritariamente Ucrânianos residentes em Portugal, utilizando a metodologia selecionada descrita por Sanchez J. e col. (2006) [7], correspondendo aos marcadores que integram o 52-Plex, que foi objeto de publicação numa revista científica [33]. O estudo populacional da amostra do Norte de Portugal (n=131) serviu para calcular os valores de heterozigotia e de outros parâmetros estatísticos de interesse forense. Este trabalho incluiu também o estudo comparativo com a amostra do grupo populacional de indivíduos imigrantes do leste europeu a residir em Portugal, bem como com outros grupos populacionais já publicados.

##### 4.1.2. Material e métodos

###### **Colheita**

Foram colhidas amostras de saliva a 216 indivíduos não relacionados, de 2 grupos populacionais: norte de Portugal (131) e imigrantes do Leste Europeu, sobretudo Ucrânianos (79) e Russos (6), totalizando 85 indivíduos, a residir em Portugal, após consentimento informado, as quais foram caracterizadas para 52 marcadores autossómicos.

As amostras respeitantes à nossa população (norte de Portugal) foram previamente estudadas com os *kits* PowerPlex16 (Promega) e Identifiler (Applied Biosystems) verificando-se deste modo a unicidade de cada amostra.

Todas as amostras foram sujeitas a uma codificação para que fosse garantida a confidencialidade relativamente à identidade dos indivíduos analisados neste estudo.

---

## **Extração**

Os extratos de DNA foram obtidos por recurso a métodos considerados tradicionais como é o caso do método de Chelex modificado, com fenol-clorofórmio- álcool isoamílico e ou recorrendo à utilização de bio-robots como é o caso do EZ1 (Qiagen) ou utilizando o kit PrepFiler no AutoMate Expressi Forensic DNA Extraction System da Applied Biosystems® (AB).

## **Quantificação**

As concentrações de DNA nos extratos foram determinadas recorrendo à metodologia de real-time PCR usando o Quantifiler™ Human DNA Quantification Kit (AB) e o ABI PRISM® 7500 Sequence Detection System (AB). As concentrações finais variaram entre 0.20 e 10 ng/µl tendo as amostras sido quantificadas em duplicado e as concentrações ajustadas a 1ng/µl.

## **Marcadores escolhidos**

As características dos SNPs estudados neste trabalho, descritos por Sanchez JJ e *col.* [7], encontram-se resumidas na Tabela 1.

Tabela 1. Listagem dos SNPs autossômicos estudados, separados conforme as duas reações SBE usadas na sua caracterização: Auto 1 e Auto 2. A cada SNP com um número de referência específico na base de dados NCBI (NCBI rs No. . *NCBI reference SNP ID number*) é atribuído um código de 1 a 54.

<b>Código do marcador</b>	<b>Cromossoma</b>	<b>NCBI rs No.</b>	<b>Marker code</b>	<b>Cromossoma</b>	<b>NCBI rs No.</b>
<b>Auto1</b>			<b>Auto2</b>		
<b>A1</b>	1	rs1490413	<b>A24</b>	21	rs2831700
<b>A2</b>	2	rs876724	<b>A25</b>	14	rs873196
<b>A3</b>	3	rs1357617	<b>A26</b>	16	rs1382387
<b>A4</b>	4	rs2046361	<b>A27</b>	12	rs2111980
<b>A5</b>	5	rs717302	<b>A28</b>	8	rs2056277
<b>A6</b>	6	rs1029047	<b>A29</b>	18	rs1024116
<b>A7</b>	7	rs917118	<b>A30</b>	6	rs727811
<b>A8</b>	8	rs763869	<b>A32</b>	1	rs1413212
<b>A9</b>	9	rs1015250	<b>A33</b>	17	rs938283
<b>A10</b>	10	rs735155	<b>A34</b>	4	rs1979255
<b>A11</b>	11	rs901398	<b>A35</b>	9	rs1463729
<b>A12</b>	12	rs2107612	<b>A36</b>	11	rs2076848
<b>A13</b>	13	rs1886510	<b>A37</b>	3	rs1355366
<b>A14</b>	14	rs1454361	<b>A38</b>	2	rs907100
<b>A15</b>	15	rs2016276	<b>A39</b>	13	rs354439
<b>A16</b>	16	rs729172	<b>A40</b>	22	rs2040411
<b>A17</b>	17	rs740910	<b>A41</b>	7	rs737681
<b>A18</b>	18	rs1493232	<b>A42</b>	21	rs2830795
<b>A19</b>	19	rs719366	<b>A43</b>	5	rs251934
<b>A20</b>	20	rs1031825	<b>A44</b>	21	rs914165
<b>A21</b>	21	rs722098	<b>A45</b>	1	rs10495407
<b>A22</b>	22	rs733164	<b>A46</b>	9	rs1360288
<b>A23</b>	10	rs826472	<b>A48</b>	10	rs964681
			<b>A49</b>	20	rs1005533
			<b>A50</b>	15	rs8037429
			<b>A51</b>	1	rs891700
			<b>A52</b>	13	rs1335873
			<b>A53</b>	22	rs1028528
			<b>A54</b>	15	rs1528460

Novos métodos de genotipagem estão a ser continuamente desenvolvidos e é muitas vezes difícil decidir qual a melhor metodologia, de entre as disponíveis.

O Genplex (AB) é um novo sistema de genotipagem baseado numa amplificação inicial, seguida de um ensaio de ligação de oligos (OLA). A OLA consiste na hibridação e posterior ligação de

---

oligonucleótidos específicos para alelos e locus (ASOs e LSOs) a produtos de PCR. Seguidamente, os produtos assim ligados são imobilizados em placas %microtitre+ e hibridados a nucleótidos %reporter+ (ZipChute<sup>®</sup> probes). Estas %probes+ são depois eluídas e detetadas por eletroforese capilar (CE).

Pretendia-se neste trabalho aplicar este %Genplex system+ a amostras provenientes de casos de parentesco complexo, nos quais se obtiveram reduzidos índices de paternidade, amostras degradadas laboratorialmente em condições controladas e a amostras de casos de criminalística biológica, relativamente às quais tenha falhado a análise convencional por estudo de STRs. Pretendia-se, ainda, confrontar os resultados obtidos deste sistema com os do anterior, que permite a genotipagem de 52 SNPs.

O consórcio SNPforID desenvolveu um 52 SNP-plex numa única reacção multiplex e subsequente detecção de SBE, usando a tecnologia de SNaPshot<sup>™</sup> (AB). O sistema requer equipamento que habitualmente se encontra em laboratórios forenses. A Applied Biosystems (AB) desenvolveu o Genplex que inclui a amelogenina e 48 SNPs dos 52 SNP-plex anteriormente descrito e validado pelo consórcio SNPforID (criado em 2002) [13].

Como foi já referido, optou-se pelo ensaio 52 SNP-plex que envolve a amplificação simultânea dos 52 SNPs autossómicos. Depois da remoção do excesso de %primers+ e nucleótidos não incorporados, segue-se a detecção dos produtos SBE resultantes de dois multiplexes por SNaPshot<sup>™</sup> (AB), analisados por eletroforese capilar (CE), usando o POP-4 como polímero, após tratamento com SAP (Shrimp Alkaline Phosphatase).

### **Tratamento de resultados e análise estatística**

As frequências alélicas, esperadas e observadas, as heterozigotias esperadas e observadas ( $H_e$  e  $H_o$ , respetivamente), o equilíbrio de Hardy-Weinberg (HWE) e os testes de desequilíbrio de ligação e outros parâmetros como FIS (coeficiente de inbreeding) e  $F_{ST}$  (distância de Slatkin) e análise da variância molecular (AMOVA) foram calculados utilizando o software Arlequin ver 3.1 [34].

A probabilidade de matching combinada ( $pM$ ), o poder de exclusão (PE) e o poder de discriminação (PD), foram calculados manualmente, usando uma folha Excel caseira.

A Probabilidade de *Match* ou *Matching probability*, também conhecida por probabilidade de *matching* ( $pM$ ), corresponde ao número de indivíduos que pode se estudado antes de encontrar o mesmo padrão de DNA num indivíduo escolhido ao acaso [35].

---

É representada pela seguinte equação:

$$pM = \sum_{i=a}^n \sum_{j=1}^n P_{ij}^2$$

Onde  $i$  e  $j$  representam as frequências de todos os alelos possíveis de  $a$  a  $n$ ;

$P_{ij}$  representa as frequências de todos os genótipos possíveis.

A probabilidade de *match* combinada para mais do que um *locus* corresponde ao produto das probabilidades de *match* individuais, para cada *locus*, assumindo que não estão em *linkage*.

O poder de exclusão (PE) [34] é definido como a fração de indivíduos que possui um perfil de DNA que é distinto do de um indivíduo escolhido de forma aleatória num caso típico de paternidade. O seu valor varia para cada caso estudado. A média, para um dado *locus*, é representada pela seguinte equação:

$$PE = h^2(1 - 2hH^2)$$

O PE para vários *loci* é representado vulgarmente pela seguinte equação:

$$PE_{total} = 1 - \prod_{i=1}^n (1 - PE_i)$$

O poder de discriminação corresponde a um menos  $pM$ . O poder de discriminação combinado para múltiplos *loci* pode-se calcular pela seguinte equação [36]:

$$Pd_{combinado} = 1 - \prod_{i=1}^n (1 - P_{di})$$

Nas comparações realizadas entre os resultados para a população do norte de Portugal com os obtidos para o grupo populacional de indivíduos imigrantes de leste a residir em Portugal, bem como com outros grupos populacionais já publicados, e por consulta do *SNPforID browser* [37] utilizou-se o software estatístico já referido.

As comparações efetuadas incluíram os resultados publicados para 49 destes 52 SNPs para uma amostra populacional da Eslovénia [38]. Outras comparações incluíram a população do NW Espanha (NW Spain) e da Rússia (*SNPforID browser*) [37] escolhidas sobretudo pela sua relativa proximidade geográfica das amostras populacionais aqui estudadas.

### 4.1.3. Resultados e discussão

Em face da limitação da utilização de apenas uma metodologia de análise, procedeu-se ao estudo populacional de uma amostra da população do Norte de Portugal e de um grupo populacional de indivíduos imigrantes do leste europeu a residir em Portugal, maioritariamente

ucranianos. Para isso, foram caracterizados 52 SNPs autossômicos de 216 amostras, 131 das quais relativas a indivíduos do Norte de Portugal e 85 relativas aos indivíduos imigrantes do leste Europeu, residentes em Portugal. Estes 52 marcadores são bi-alélicos e de pequeno tamanho. Para cada amostra, os 52 SNPs foram genotipados recorrendo a uma única PCR inicial, seguida de duas reações SBE, correspondendo aos dois conjuntos de marcadores que integram o 52-*plex*, também conhecidos por Auto1 e Auto2, do qual se dá um exemplo (Figura 10).

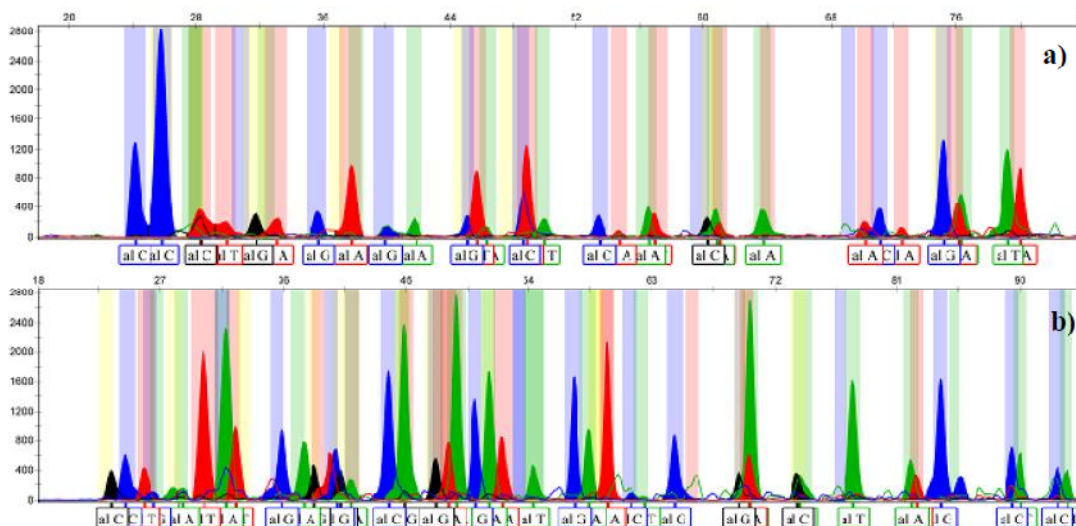


Fig.10 . Caracterização do 52-plex numa amostra: a) Auto 1; b) Auto 2.

As amostras correspondentes aos indivíduos representativos do norte de Portugal foram selecionadas a partir de casos de parentesco complexos já concluídos. No caso dos indivíduos imigrantes, as amostras (zaragatoas bucais) foram colhidas após consentimento prévio e informado. Este trabalho foi objeto de publicação numa revista internacional indexada de referência, da área da Medicina Legal e Forense [33] (Anexo 1).

Para o cálculo dos valores de heterozigotia e de outros parâmetros estatísticos de interesse forense, este estudo populacional, foi complementado com o estudo de amostras de referência relativas a outros casos de parentesco já genotipadas para os marcadores convencionais (STRs) e posteriormente caracterizadas para estes marcadores, de modo a que a amostra fosse representativa do norte de Portugal.

Compararam-se estes resultados com os obtidos para o grupo populacional de indivíduos de leste a residir em Portugal, bem como com outros grupos populacionais já publicados.

Como já foi referido, foi genotipado um total de 216 amostras com o 52SNP*plex* recorrendo ao *kit* SNaPshot<sup>TM</sup>. Dos 262 perfis gerados a partir de 131 amostras do norte de Portugal, genotipadas em duplicado, 240 (91.6%) permitiram obter perfis completos (Tabela 1A e 1B . Anexo 6).

No caso das amostras da população de imigrantes, constituída por 85 indivíduos, obtiveram-se perfis completos em 164 de 170 (96.5%) (Tabela 2A e 2B . Anexo 7).

---

Tipicamente o número de amostras necessárias para se fazer uma estimativa robusta das frequências alélicas em loci que possuem de 5-15 alelos é de 100-150 amostras para cada grupo populacional. Uma vez que se estudaram marcadores bi-alélicos, que apenas podem apresentar 3 tipos de genótipo, um número inferior de amostras será suficiente desde que se utilize uma frequência alélica mínima de cerca de  $5/2N$ , para que seja um valor conservador, ou seja, em caucasianos deve ser usada uma frequência mínima de 0.034 ( $N=74$ ).

Nas tabelas seguintes são apresentadas as frequências alélicas observadas para cada um dos grupos populacionais estudados (Tabela 3 e Tabela 4). Para testar se as populações estudadas estavam em Equilíbrio Hardy-Weinberg efetuaram-se 104 testes ( $52 \times 2$  populações). Como tem sido descrito é normal que cerca de 5% das comparações efetuadas se desviem do equilíbrio Hardy-Weinberg [39].

Tabela 3. Frequências alélicas para uma população do norte de Portugal (n=131)

Auto1						Auto2					
SNP	Alelos	Frequências	SNP	Alelos	Frequências	SNP	Alelos	Frequências	SNP	Alelos	Frequências
A1	C	0.51145	A13	A	0.51527	A24	C	0.41794	A40	A	0.64695
	T	0.48855		G	0.48473		T	0.58206		G	0.35305
A2	C	0.77481	A14	A	0.62977	A25	A	0.47901	A41	C	0.54580
	T	0.22519		T	0.37023		G	0.52099		T	0.45420
A3	A	0.36260	A15	C	0.20992	A26	G	0.40076	A42	A	0.70992
	T	0.63741		T	0.79008		T	0.59921		G	0.29008
A4	A	0.39122	A16	G	0.52863	A27	A	0.66031	A43	A	0.59924
	T	0.60878		T	0.47137		G	0.33969		G	0.40076
A5	A	0.44657	A17	A	0.62977	A28	A	0.25191	A44	C	0.55344
	G	0.55344		G	0.37023		G	0.74809		T	0.44657
A6	A	0.43893	A18	G	0.12595	A29	A	0.50763	A45	C	0.56870
	T	0.56107		T	0.87405		G	0.49237		T	0.43130
A7	A	0.22328	A19	C	0.43893	A30	A	0.52672	A46	A	0.50382
	G	0.77672		T	0.56107		C	0.47328		G	0.49618
A8	C	0.53435	A20	G	0.82061	A32	C	0.65649	A48	C	0.29389
	T	0.46565		T	0.17939		T	0.34351		T	0.70611
A9	C	0.79962	A21	A	0.64695	A33	C	0.23283	A49	A	0.30153
	G	0.20038		G	0.35305		T	0.76718		G	0.69847
A10	C	0.44847	A22	A	0.11069	A34	C	0.33206	A50	A	0.55725
	T	0.55153		G	0.88931		G	0.66794		G	0.46183
A11	C	0.23283	A23	A	0.38740	A35	A	0.61450	A51	A	0.45611
	T	0.76718		G	0.61260		G	0.38550		G	0.54389
A12	A	0.55344				A36	A	0.45802	A52	A	0.62977
	G	0.44657					T	0.54199		T	0.37023
						A37	C	0.55344	A53	A	0.79008
							T	0.45130		G	0.20992
						A38	C	0.83206	A54	A	0.51527
							G	0.16794		G	0.50763
						A39	A	0.50382			
							T	0.49618			

Tabela 4. Frequências alélicas de uma população de imigrantes do leste europeu a residir em Portugal (n=85).

Auto1						Auto2					
SNP	Alelos	Frequências	SNP	Alelos	Frequências	SNP	Alelos	Frequências	SNP	Alelos	Frequências
A1	C	0.50000	A13	A	0.41176	A24	C	0.43530	A40	A	0.68824
	T	0.50000		G	0.58824		T	0.56471		G	0.31177
A2	C	0.76471	A14	A	0.54706	A25	A	0.39413	A41	C	0.66471
	T	0.23529		T	0.45294		G	0.60588		T	0.33530
A3	A	0.25294	A15	C	0.20588	A26	G	0.42941	A42	A	0.77647
	T	0.74706		T	0.79412		T	0.57059		G	0.22353
A4	A	0.31471	A16	G	0.58529	A27	A	0.67647	A43	A	0.60588
	T	0.68529		T	0.41471		G	0.32353		G	0.39412
A5	A	0.47647	A17	A	0.71765	A28	A	0.29412	A44	C	0.60000
	G	0.52353		G	0.28235		G	0.70588		T	0.40000
A6	A	0.34118	A18	G	0.27647	A29	A	0.58235	A45	C	0.67647
	T	0.65882		T	0.72353		G	0.41765		T	0.32353
A7	A	0.21765	A19	C	0.35294	A30	A	0.48235	A46	A	0.42353
	G	0.78235		T	0.64706		C	0.51765		G	0.57647
A8	C	0.51765	A20	G	0.80000	A32	C	0.65294	A48	C	0.40588
	T	0.48235		T	0.20000		T	0.34706		T	0.59412
A9	C	0.84118	A21	A	0.71177	A33	C	0.16471	A49	A	0.48824
	G	0.15882		G	0.28824		T	0.83530		G	0.50588
A10	C	0.49412	A22	A	0.23824	A34	C	0.43530	A50	A	0.53529
	T	0.50588		G	0.76176		G	0.56471		G	0.46471
A11	C	0.25882	A23	A	0.37059	A35	A	0.59412	A51	A	0.56471
	T	0.74118		G	0.62941		G	0.40588		G	0.48824
A12	A	0.68824				A36	A	0.34706	A52	A	0.69412
	G	0.31176					T	0.65294		T	0.30588
						A37	C	0.47647	A53	A	0.80588
							T	0.52353		G	0.19412
						A38	C	0.83530	A54	A	0.37941
							G	0.16471		G	0.62059
						A39	A	0.45882			
							T	0.54118			

Para a amostra da população do norte de Portugal calculou-se o índice de diversidade génica médio para todos os *loci* correspondendo a  $0.4229 \pm 0.2072$ , a partir de dados de 49 em 52

---

marcadores (isto é, os que tinham menos de 0.05% de dados em falta) e  $0.4287 \pm 0.2099$  com os dados obtidos a partir de todos os 52 *loci* para a amostra populacional de imigrantes (maioritariamente Ucrânicos e alguns Russos).

O SNP autossômico A22 (rs733164) apresentava o menor valor de índice de conteúdo polimórfico (PIC) (0.1585) (Tabela 5).

Tabela 5. Parâmetros estatísticos para cada marcador estudado na população do Norte de Portugal.

Auto 1																								
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A23
<b>pM</b>	0,488	0,492	0,380	0,374	0,527	0,394	0,499	0,523	0,518	0,515	0,495	0,340	0,422	0,568	0,517	0,590	0,555	0,659	0,389	0,646	0,434	0,744	0,367	0,367
<b>PD</b>	0,512	0,508	0,620	0,626	0,473	0,606	0,501	0,477	0,482	0,485	0,505	0,660	0,578	0,432	0,483	0,410	0,445	0,341	0,611	0,354	0,566	0,256	0,633	0,633
<b>PIC</b>	0,374	0,283	0,354	0,356	0,367	0,371	0,279	0,374	0,266	0,371	0,289	0,372	0,375	0,349	0,269	0,369	0,353	0,182	0,371	0,224	0,352	0,159	0,360	0,360
<b>Pexcl</b>	0,360	0,082	0,052	0,108	0,394	0,206	0,082	0,409	0,070	0,390	0,128	0,089	0,263	0,405	0,106	0,490	0,398	0,029	0,198	0,011	0,213	0,008	0,106	0,106
<b>IP</b>	1,443	0,759	0,685	0,820	1,549	1,042	0,759	1,600	0,730	1,537	0,865	0,775	1,179	1,588	0,814	1,909	1,561	0,623	1,024	0,567	1,059	0,553	0,816	0,816
Auto 2																								
	A24	A25	A26	A27	A28	A29	A30	A32	A33	A34	A35	A36	A37	A38	A39	A40	A41	A42	A43	A44	A45	A46	A48	A49
<b>pM</b>	0,484	0,351	0,381	0,433	0,466	0,453	0,431	0,470	0,485	0,428	0,409	0,515	0,397	0,563	0,490	0,423	0,468	0,426	0,584	0,382	0,479	0,412	0,430	0,434
<b>PD</b>	0,516	0,649	0,619	0,567	0,534	0,547	0,569	0,530	0,515	0,572	0,591	0,485	0,603	0,437	0,510	0,577	0,532	0,574	0,416	0,618	0,521	0,588	0,570	0,566
<b>PIC</b>	0,368	0,374	0,366	0,350	0,302	0,375	0,374	0,348	0,292	0,344	0,361	0,373	0,372	0,235	0,375	0,354	0,371	0,325	0,365	0,372	0,371	0,375	0,328	0,330
<b>Pexcl</b>	0,336	0,138	0,164	0,205	0,104	0,310	0,274	0,260	0,114	0,178	0,198	0,393	0,216	0,061	0,364	0,202	0,319	0,087	0,471	0,188	0,336	0,248	0,124	0,038
<b>IP</b>	1,372	0,889	0,947	1,040	0,810	1,300	1,208	1,173	0,833	0,977	1,024	1,548	1,066	0,709	1,455	1,032	1,327	0,771	1,829	1,000	1,372	1,143	0,855	0,648

Legenda: pM- probabilidade de *Matching*; PD-Probabilidade de discriminação; PIC . Índice de conteúdo polimórfico; Pexcl-Probabilidade de exclusão; IP- Índice de paternidade.



O valor médio de heterozigotia observado para os 52 SNPs do *52plex* foi  $0.4954 \pm 0.1663$  para a população do norte de Portugal (Tabela 6) Tal como seria de esperar, uma vez que os SNPs autossómicos incluídos neste sistema são os mais informativos dos actualmente descritos, os valores de heterozigotia na nossa população são elevados e estão de acordo com os descritos em estudos efetuados em populações do mesmo grupo populacional [4].

Tabela 6. Dados relativos à população do Norte de Portugal

Locus	Obs.Het.	Exp.Het.	P-value	s.d.
A1	0.6797	0.5017	0.000	1.000E-05
A2	0.3438	0.3431	1.000	0.000E+00
A3	0.2764	0.4614	0.000	0.000E+00
A4	0.4634	0.4760	0.849	3.400E-04
A5	0.7638	0.4959	0.000	0.000E+00
A6	0.5200	0.4938	0.588	5.200E-04
A7	0.3167	0.3213	1.000	0.000E+00
A8	0.6953	0.4995	0.000	0.000E+00
A9	0.3228	0.3126	1.000	0.000E+00
A10	0.6880	0.4966	0.000	1.000E-05
A11	0.4219	0.3518	0.023	1.500E-04
A12	0.3548	0.4956	0.002	4.000E-05
A13	0.5760	0.5012	0.108	3.000E-04
A14	0.7323	0.4660	0.000	0.000E+00
A15	0.3858	0.3222	0.025	1.500E-04
A16	0.8095	0.5000	0.000	0.000E+00
A17	0.6875	0.4665	0.000	0.000E+00
A18	0.1969	0.2031	0.663	4.800E-04
A19	0.5156	0.4941	0.719	4.800E-04
A20	0.1177	0.2578	0.000	0.000E+00
A21	0.5397	0.4563	0.050	2.200E-04
A22	0.0880	0.1677	0.000	1.000E-05
A23	0.3871	0.4726	0.055	2.300E-04
A24	0.6279	0.4874	0.001	3.000E-05
A25	0.4419	0.5009	0.217	3.700E-04

A26	0.4720	0.4803	0.853	3.500E-04
A27	0.5349	0.4487	0.031	1.700E-04
A28	0.3828	0.3725	0.816	3.900E-04
A29	0.6308	0.5018	0.005	7.000E-05
A30	0.5859	0.5005	0.074	2.600E-04
A32	0.5892	0.4512	0.000	2.000E-05
A33	0.4000	0.3564	0.218	3.900E-04
A34	0.4884	0.4436	0.317	4.600E-04
A35	0.5118	0.4740	0.449	4.900E-04
A36	0.6846	0.4983	0.000	0.000E+00
A37	0.5385	0.4961	0.380	5.300E-04
A38	0.2946	0.2736	0.523	5.000E-04
A39	0.6797	0.5019	0.000	1.000E-05
A40	0.5156	0.4577	0.177	3.900E-04
A41	0.6462	0.4977	0.001	3.000E-05
A42	0.3516	0.4093	0.129	3.200E-04
A43	0.7344	0.4813	0.000	0.000E+00
A44	0.5077	0.4961	0.859	3.600E-04
A45	0.6279	0.4922	0.002	4.000E-05
A46	0.6142	0.5020	0.013	1.100E-04
A48	0.4154	0.4153	1.000	0.000E+00
A49	0.2284	0.4178	0.000	0.000E+00
A50	0.5440	0.4948	0.279	4.700E-04
A51	0.5736	0.4983	0.109	2.900E-04
A52	0.5231	0.4676	0.189	4.000E-04
A53	0.3846	0.3304	0.065	2.500E-04
A54	0.3496	0.5019	0.001	3.000E-05
Média	0.4954	0.4386		
s.d.	0.1663	0.0847		
Diversidade génica média calculada a partir dos 52 loci			0.4229+/-0.2072	

O valor de  $F_{ST}$  global, calculado para as duas populações foi de 0.00905 e o valor de FIS global para a população do norte de Portugal foi de -0.0913, refletindo a aleatoriedade da amostragem (Tabela 7).

Tabela 7. Índices de fixação médios F-estatísticos sobre todos os loci  
(F- Statistics over all loci Fixation Indices)

Índices de Fixação F- Estatísticos	
<b>FIS</b>	-0.09133
<b>FST</b>	0.00905
<b>FIT</b>	-0.08145

Legenda: FIS- coeficiente de *inbreeding*;  $F_{ST}$ - distância de Slatkin ; FIT- Coeficiente de inbreeding total

O valor de  $F_{ST}$  linearizado obtido a partir das comparações efetuadas entre as duas amostras populacionais estudadas foi de 0.0085, demonstrando que existem diferenças significativas entre elas (nível de significância = 0.05) (Tabelas 8 e 9).

Tabela 8 . Matriz de distâncias genéticas entre duas amostras populacionais:  
1- Norte de Portugal; 2- Imigrantes do leste europeu

Matriz de FSTs linearizados de Slatkin como $t/M=FST/(1-FST)$ ( $M=N$ para dados haploides. $M=2N$ para dados diploides data)		
Amostra populacional	1	2
1	0.00000	
2	0.00846	0.00000

Tabela 9. Matriz de significância . valores de P para os Fst  
(Significance Level=0.0500; Number of permutations : 110)

Amostra populacional	1	2
1		+
2	+	

Decidiu-se incluir as poucas amostras de indivíduos russos no total dos 85 indivíduos imigrantes estudados, uma vez que após efetuar comparações, incluindo ou excluindo as 9 amostras que lhes correspondiam, os resultados não variavam uns dos outros de forma significativa.

Compararam-se estes resultados com os obtidos para o grupo populacional de indivíduos de leste a residir em Portugal, bem como com outros grupos populacionais já publicados.

As comparações, realizadas entre os resultados obtidos para estas populações, com os publicados relativos a uma população Eslovénia [38], para 49 dos 52 SNPs estudados, demonstraram que todas as três populações comparadas apresentavam diferenças significativas entre elas e, que a amostra populacional do norte de Portugal se afastava mais das restantes duas populações comparadas.

Realizaram-se também comparações destas amostras populacionais com as populações do NW-Espanha e da Rússia, disponíveis *online* (*SNPforID browser*) [37] escolhidas sobretudo devido à sua proximidade geográfica relativa, As populações da Rússia e Eslovenia não apresentavam diferenças significativas, mas, como seria de esperar, eram diferentes das restantes populações. As comparações efetuadas, entre a amostragem relativa ao norte de Portugal (no presente trabalho) com as amostras relativas à população do NW-Espanha, demonstrou que, estas duas populações, apresentam diferenças significativas (dados não apresentados), Este facto pode ser devido ao grande poder discriminativo de alguns dos 52 SNPs incluídos no 52-SNP-plex que foram estudados, uma vez que são marcadores usados para inferir a ancestralidade, permitindo distinguir entre indivíduos relativamente idênticos uns aos outros, ao contrário do que acontece quando a análise se baseia apenas no estudo de STRs [40], No entanto, os resultados da análise de AMOVA revelaram que, a maior parte da variação molecular se deveu à variação intra-populacional e não à variação inter-populacional.

Usando os SNPs estudados no presente trabalho, obteve-se uma probabilidade de *match* (pM) média de  $3,0621 \times 10^{-18}$  para a população do norte de Portugal correspondendo a um poder de discriminação (PD) comparável ao que se obtém quando se usam alguns dos *kits* de STRs disponíveis por rotina nos laboratórios de Genética Forense, A probabilidade média de exclusão obtida foi de 99,9997% (Tabela 10).

Tabela 10. Parâmetros estatísticos com interesse forense.

	<b>Norte de Portugal</b>	<b>Imigrantes de leste</b>
<b>pM</b>	3.062E-18	1.613E-19
<b>PD</b>	0.99999999999999990	0.99999999999999990
<b>PIC</b>	1.476E-25	2.378E-25
<b>Pexcl.</b>	0.999997	0.99995

Legenda: pM . probabilidade de matching; PD - Poder de Discriminação; PIC- Conteúdo polimórfico; Pexcl.- Probabilidade de exclusão.

---

## CAPÍTULO V

### 5. Caracterização de SNPs autossómicos em amostras degradadas em condições laboratoriais controladas

#### 5.1. Introdução

A validação primária da metodologia selecionada neste trabalho, isto é, do 52SNP-*plex*, consistiu em participar em exercícios colaborativos, promovidos pelo GHEP-ISFG, que envolveu vários Laboratórios de Genética Forense. Nesta validação compararam-se os resultados consensos dos laboratórios forenses participantes com os resultados obtidos pelo laboratório onde este estudo foi desenvolvido (SGBF-N). Esta participação constituiu um marco fundamental na seleção do método de caracterização dos SNPs autossómicos e funcionou como motivação adicional para a conclusão do trabalho prático realizado.

Além destes exercícios colaborativos de intercâmbio de resultados, foram também comparados os resultados destas experiências com os publicados e estabelecidos para os controlos positivos de DNA, concretamente o 9947A e o 007, incluídos nos *kits* comerciais e disponíveis para genotipagem de STRs autossómicos, que serviram como controlos positivos das experiências realizadas [32]. Relativamente aos DNAs padrão utilizados, foram obtidos os resultados esperados para todos os marcadores relativos ao GenPlex™ referidos no artigo, existindo concordância de resultados em 49 dos 52 SNPs.

Procedeu-se ainda ao estudo de amostras degradadas em condições laboratoriais controladas de modo a obter fragmentos de DNA de tamanho reduzido com o objetivo de que, numa fase posterior, se aplicasse esta metodologia a amostras provenientes de perícias relativas a situações reais, já concluídas, provenientes de casos em que a análise dos STRs foi insuficiente ou não conclusiva.

#### 5.2. Material e métodos

Controlo de qualidade: Como já foi referido, a metodologia do 52-SNP *plex* foi validada, localmente e por participação num exercício colaborativo inter-laboratorial, promovido pelo grupo GHEP-ISFG, antes de procedermos às genotipagens populacionais. Esta validação constituiu numa validação primária da metodologia seleccionada, através da comparação dos resultados obtidos no âmbito da realização de um exercício colaborativo proposto pelo GHEP-ISFG (Grupo de Línguas Espanhola e Portuguesa da International Society for Forensic Genetics) e consistiu

---

em comparar os resultados consenso dos laboratórios forenses participantes com os resultados obtidos pelo laboratório onde este estudo foi desenvolvido (SGBF-N).

Neste trabalho foram também realizadas comparações dos resultados com os publicados e estabelecidos para os controlos positivos de DNA (9947A e 007) incluídos nos kits comerciais e disponíveis para tipagem de STRs autossómicos.

O passo seguinte consistiu em submeter réplicas do DNA padrão (9947A) fornecido nos kits comerciais juntamente com os reagentes necessários à amplificação dos marcadores STRs, a uma concentração de 1ng/ul. a uma degradação artificial controlada no laboratório, usando diferentes condições de exposição a raios de UV (525nm) numa câmara %workstation+destinada à esterilização do material (The Clean Spot Work Station . Coy - Michigan) e tratamento com DNaseI (U/ul), durante os mesmos períodos de tempo de exposição aos raios UV (1. 2. 5.10 e 15 min).

Procedeu-se à amplificação do 52-plex para todas as amostras tratadas bem como para o DNA padrão não tratado.

### **5.3. Resultados e discussão**

Relativamente a estas amostras, obteve-se sempre um perfil completo, qualquer que fosse o período de exposição do DNA padrão (9947A) aos raios UV (1. 2. 5. 10 e 15 min). O mesmo aconteceu para as amostras submetidas a digestão com DNaseI para todos os períodos testados (Figura 11). Os resultados para o DNA padrão podem ser validados por comparação com os já descritos [32]. Este resultado era esperado. como tem sido descrito por outros investigadores [41].

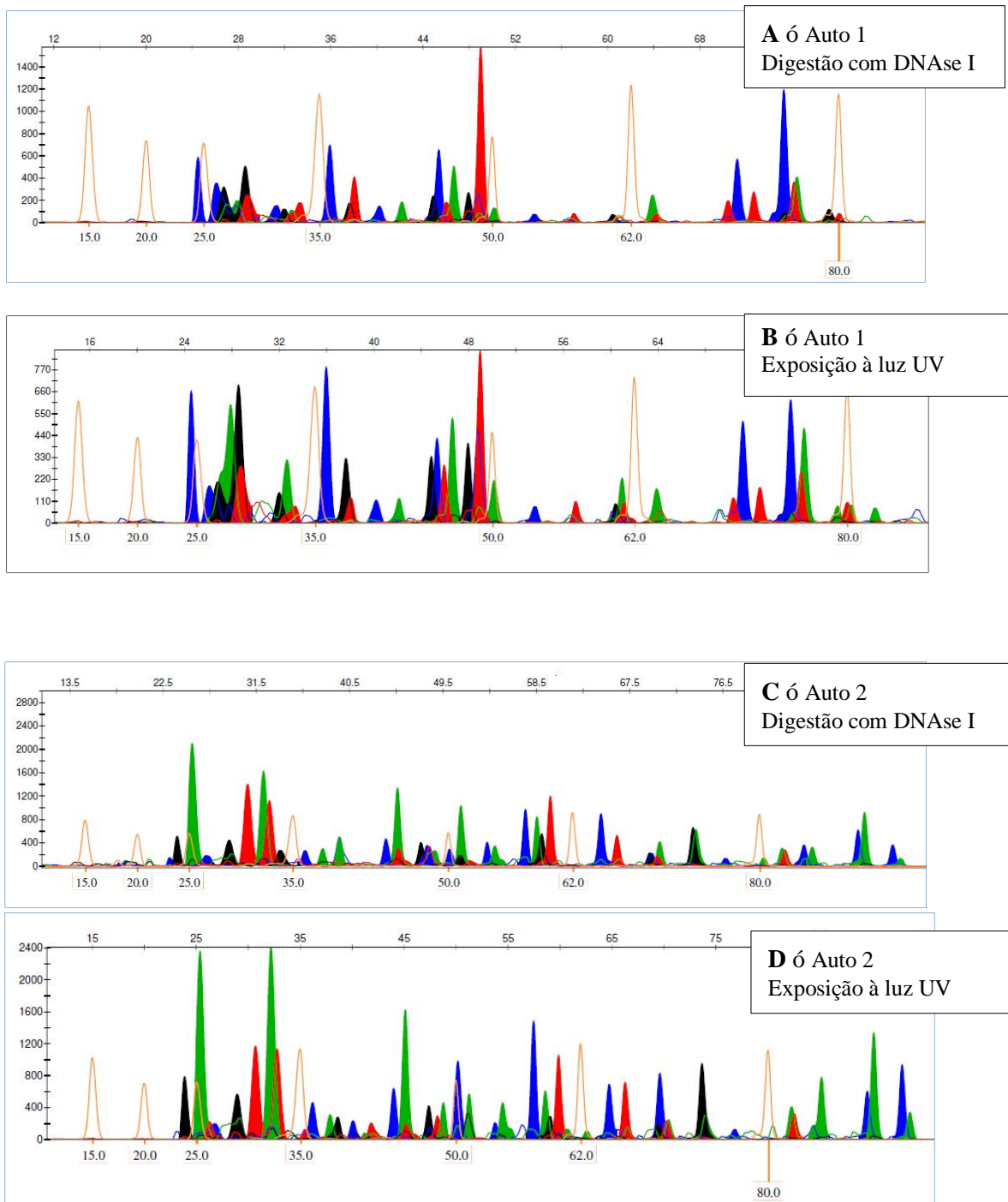


Fig. 11. DNA padrão. 9947A. artificialmente degradado: A- Resultados da tipagem para o Auto 1 após 15 minutos de digestão com DNase I; B- Resultados da tipagem para o Auto 1 após 15 minutos de exposição à luz UV; C- Resultados da tipagem para o Auto 2 após 15 minutos de digestão com DNase I; D . Resultados da tipagem para o Auto 2 após 15 minutos de exposição à luz UV.



---

## CAPÍTULO VI

### 6. Caracterização de SNPs autossômicos em casos de parentesco complexos

#### 6.1. Introdução

A utilização de SNPs autossômicos na investigação em Genética Forense, tem sido muito discutida nos últimos anos, sobretudo porque os SNPs possuem algumas vantagens quando comparados com os STRs [42]. Quando combinados com os marcadores clássicos ou habituais, os SNPs autossômicos aumentam o poder de discriminação e fornecem dados adicionais para os testes de paternidade [4,7].

Embora menos polimórficos do que os STRs (multialélicos), os SNPs (bialélicos) podem permitir a distinção entre indivíduos não relacionados ou relacionados entre si, desde que se utilize um elevado número destes SNPs. Um grupo de investigadores internacional, o *SNPforID consortium* propôs e desenvolveu a construção de sistemas *multiplex* robustos com fins de identificação, que permitem a tipagem simultânea de elevado número de SNPs, tal como o *52 SNP-plex* ou o *GenPlex System*. O seu estudo pode complementar a informação em casos complexos de identificação e de parentesco.

Após realização do estudo populacional prévio, aplicou-se a metodologia de genotipagem de SNPs autossômicos seleccionada, isto é, o *52 SNP-plex*, a perícias forenses já concluídas. Estas perícias incluíram 14 casos complexos de investigação de parentesco, casos de criminalística biológica e casos de identificação genética individual (por exemplo os que envolvem restos cadavéricos).

#### 6.2. Material e métodos

Fizeram-se estudos de amostras seguras provenientes de casos de parentesco complexos em que, com recurso aos STRs, se obtiveram valores reduzidos de  $IP(\leq 10\ 000)$ . Estes casos consistem sobretudo em investigações de parentesco em que não se dispõe do pretense pai mas sim de familiares seus, como por exemplo, irmãos, filhos biológicos comprovados, pais ou outros. Também foram incluídos no estudo alguns trios (Suposto Pai, Filho e Mãe) nos quais foram observadas uma ou mais incompatibilidades de transmissão alélica aquando do estudo dos STRs.

---

Os marcadores escolhidos neste estudo, uma vez mais, foram os 52 SNPs autossômicos, bi-aliélicos, descritos em Sanchez et al. (2006). Estes SNPs estão posicionados nas porções distais dos braços dos cromossomas, exceto o do único SNP localizado no cromossoma 19. Por isso, são estudados pelo menos dois SNPs na maioria dos cromossomas, mas em alguns deles são estudados quatro, nos cromossomas 1, 9, 10, 13, 15, 21, e 22 [8]. Nos cromossomas mais pequenos como o 21 e o 22, o número médio de eventos de recombinação por meiose é um [13,14], como consequência, dois dos três SNPs nestes cromossomas vão situar-se do mesmo lado, por recombinação. e os descendentes herdam dois *loci* ligados e um *locus* independente dos pais. No entanto podemos calcular o Índice de Paternidade (PI) tal como calcularíamos se os *loci* fossem independentes, porque estes estão em equilíbrio na população [8] e porque os pais, numa paternidade típica, são desconhecidos.

Relativamente ao tratamento dos resultados obtidos e da sua análise estatística, foi utilizado o *software* de análise de *pedigree* Familias [34], assumindo independência dos *loci* estudados. Para o cálculo dos PI todos os genótipos de STR e de SNP foram comparados nos indivíduos testados usando os dados de frequências aliélicas de cada SNP obtidos no estudo populacional da amostra da população do norte de Portugal estudada (n=131) [33]. Em todos os casos, para o cálculo do índice de paternidade e respetiva conversão do valor obtido numa probabilidade percentual (W), foi atribuído um valor *a priori* igual para todas as hipóteses testadas. No caso de existirem apenas duas hipóteses (a maioria), usou-se o valor *a priori* de 0.5. O programa Familias sugere o parentesco mais provável dados os genótipos inseridos (obtidos por genotipagem) para os indivíduos testados, calculando as probabilidades de diferentes pedigrees possíveis previamente indicados.

Quando se observam exclusões de segunda ordem e mutações do tipo *single step mutations*, o *software* Familias consegue valorizá-las por introdução de taxas de mutação para os loci em causa de modo a obter uma probabilidade para as relações definidas. Quando lidamos com STRs podemos inserir valores de *m* descritos, por exemplo, na STRbase [43], com valores que podem ser variáveis conforme o marcador em causa, mas também se pode usar um valor médio, como é exemplo de  $m = 0.0016$ , conforme sugerem Phillips C. e col. (2005) [13]. No caso dos SNPs, foi utilizada uma taxa de mutação universal de  $m = 2.5 \times 10^{-8}$  [13,44].

Os cálculos foram realizados tendo assumido que podem existir alelos silenciosos. Para loci onde previamente estes alelos silenciosos ainda não tinham sido observados, a frequência dos alelos silenciosos calculou-se segundo a equação  $1/(n + 1)$ , em que *n* corresponde ao número total de alelos observados.

### 6.3. Resultados e discussão

Relativamente aos resultados da aplicação do estudo do 52-plex em amostras seguras provenientes de casos de parentesco complexos em que, com recurso aos STRs (marcadores convencionais), se obtiveram reduzidos valores de IP muitos deles  $\leq 10\ 000$ , que consistiram sobretudo em investigações de parentesco em que não se dispõe do pretense pai mas sim de familiares seus, como por exemplo, irmãos, filhos biológicos comprovados, pais ou outros, ou casos que envolvem trios (Suposto Pai, Filho e Mãe) nos quais foram observadas uma ou mais incompatibilidades de transmissão alélica aquando do estudo dos STRs (ex. família F5), que se encontram resumidos na Tabela 12.

Em alguns casos, os SNPs não foram muito informativos, com valores de  $IP < 20$ . Este resultado era esperado, uma vez que se trata de casos de difícil resolução que surgem na nossa rotina nos quais é vulgar existir falta de informação, que seja suficiente para investigações de parentesco, como é o caso em que se verificam duas incompatibilidades de transmissão alélica na análise de STRs (família 5- F5) e noutras em que dispomos apenas de familiares afastados ou por linha lateral do filho investigado, como é o caso de avós, tios ou irmãos (F6, F7, F10 e F11).

Tabela 12. Impacto do estudo dos SNPs em casos seleccionados de parentesco complexos.

Família	IP (STR)	W	IP (SNPs)	IP Total	W	
F1	1.25E+04	0.99992	1.15E+03	1.44E+07	0.99999993	
F2	2.20E+03	0.9995	1.29E+02	2.84E+05	0.999992948	1 incompatibilidade
F3	6.13E+03	0.9998	9.71E+05	5.95E+09	0.999999998	1 incompatibilidade
F4	5.50E+00	0.85	1.79E+00	9.80E+00	0.907791609	
F5	2.96E+03	0.99996	1.18E+01	3.50E+04	0.999971395	2 incompatibilidades
F6	2.20E+00	0.6825	2.25E+00	4.80E+00	0.82869379	ISP+F+M
F7	1.20E+03	0.9992	4.12E+00	4.92E+03	0.99979693	PSP+SISP+F+M
F8	4.24E+03	0.9998	1.50E+01	6.37E+04	0.99998429	SP+F+M 2 incompatibilidades
F9	9.97E+03	0.9999	1.10E+00	1.10E+04	0.999908789	SP+F+M 1 incompatibilidade
F10	2.41E+02	0.9959	7.31E+00	1.76E+03	0.999431915	2 FSP(M1) +F(M)
F11	2.66E+03	0.9996	1.68E+00	4.46E+03	0.999775887	MSP+F+M
F12	1.29E+04	0.99992	5.67E+00	7.32E+04	0.999986346	SP+F
F19	1.38E+04	0.99993	2.23E+00	3.07E+04	0.999967388	
F21	1.55E+03	0.9994	9.43E+02	1.46E+06	0.999999317	MSP+F+M

Em todos os casos, dependendo do número de SNPs para os quais se obtiveram resultados, o valor final de IP sofreu um aumento significativo, pelo menos 5 vezes superior, realçando-se assim a importância que estes marcadores podem ter como marcadores adicionais nestes casos

---

de parentesco mais complexos. Como exemplo referem-se os casos assinalados na Tabela como sendo correspondentes às famílias F2 e F21.

O Caso F3 merece uma atenção especial uma vez que apresentava uma incompatibilidade no sistema D16S539, provavelmente uma mutação *single step*, no qual a aplicação da tipagem dos SNPs aumentou o IP em duas ordens de magnitude. Estes resultados podem ser explicados pelo facto de, a aplicação dos sistemas *multiplex*, como é o caso do *52-plex*, permitir um poder de discriminação elevado nas populações europeias, por vezes superior ao que se obtém a partir dos STRs incluídos nos *kits* de identificação habitualmente utilizados [8].

O resumo dos resultados dos casos de paternidade complexa analisados apresentados na Tabela 12, evidenciam que a análise de SNPs permitiu uma melhoria significativa na probabilidade da relação de parentesco testada. As conclusões baseadas apenas nos perfis obtidos a partir dos marcadores convencionais, ou seja, nos STRs, eram ambíguas em alguns destes casos, com valores de LR  $\leq 10\ 000$ . Em todos os casos as ambiguidades ficam resolvidas com a informação obtida a partir do *52-plex* de SNPs autossómicos. Como já foi descrito por outros autores a informação adicional dos 52 SNPs remove o elemento de dúvida envolvido nestes.

Tal como tem sido descrito por vários autores [4,5], estes casos servem para ilustrar claramente que ao adicionar a informação dos 52SNPs se remove o elemento de dúvida que está presente na interpretação deste tipo de testes de parentesco, mesmo após a caracterização extensiva de marcadores STR e MiniSTR. Uma característica que favorece o uso deste tipo de marcadores para esta finalidade, ou seja, para teste de parentesco, é a baixa taxa de mutação, comparativamente aos STRs, reduzindo, deste modo, o risco de exclusões ambíguas que podem surgir por ocorrência de mutações. No entanto, como foi anteriormente referido, os resultados obtidos têm de ser cuidadosamente interpretados, especialmente se familiares próximos estejam envolvidos como potencial suposto pai [4].

---

## CAPÍTULO VII

### 7. Caracterização de SNPs autossômicos em casos de criminalística biológica e de identificação genética individual

#### 7.1. Introdução

A utilização de SNPs autossômicos na investigação em Genética Forense, em combinados com os marcadores clássicos ou habituais, aumentam o poder de discriminação e pode constituir uma vantagem na análise de amostras muito degradadas [45].

Para que se possa obter informação a partir de casos de rotina, especialmente os relacionados com amostras de DNA degradado ou presente em quantidades reduzidas, é necessária utilização de sistemas *multiplex* robustos, que permitem a tipagem simultânea de elevado número de SNPs, tal como o 52 SNP-*plex*, uma vez que são menos polimórficos que os STRs. O reduzido tamanho dos produtos de PCR destes SNPs autossômicos em particular, pode ser útil na análise de variadas amostras de DNA degradado frequentemente associadas a casos de criminalística biológica e complementar a informação em casos complexos de identificação e de parentesco, como foi já evidenciado.

A fase final deste trabalho consistiu em aplicar estes marcadores a casos concretos em que as amostras escolhidas eram de especial dificuldade. Nesta fase, como exemplo de aplicação, foram estudados dois casos de identificação, em tudo semelhantes, em que se analisaram células incluídas em blocos de parafina de dois indivíduos nos quais foi diagnosticado um cancro, para posterior comparação com amostras de referência, confirmando ou não a sua proveniência (confirmação de identidade).

Procedeu-se à seleção de algumas amostras referentes a casos já concluídos em que, relativamente aos quais, seria interessante complementar a informação obtida através da análise de STRs.

#### 7.2. Material e Métodos

Foram selecionados dois casos de criminalística biológica, previamente estudados, nos quais se obteve um perfil genético incompleto com recurso aos STRs ou Mini-STRs, num deles, ou ausente no restante. Foram ainda estudados dois casos de identificação, em tudo semelhantes, que envolviam amostras obtidas a partir de células tumorais incluídas em blocos de parafina (Ex.

Caso 1 - Figura 12). Nestes dois últimos casos também foram obtidos perfis incompletos com recurso à tipagem convencional de STRs ou Mini-STRs e o seu interesse residia sobretudo no facto de serem casos relativamente invulgares nos laboratórios de Genética Forense, para fins de identificação individual.

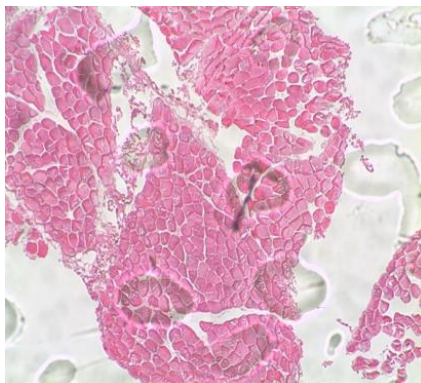


Figura 12 . Lâmina correspondente a células incluídas num bloco de parafina com diagnóstico positivo de cancro (Caso 2), a partir das quais se procedeu à extração de DNA e obtenção do perfil de SNPs autossómicos.

### 7.3. Resultados e discussão

Relativamente aos dois casos que envolviam amostras incluídas em parafina, apresenta-se a genotipagem incompleta obtida com recurso aos STRs convencionais e complementada a *posteriori* com o estudo de mini-STRs (Tabela 13).

Tabela 13 . Resultados para os STRs/mini-STRs (Caso 1)

Marcador STR	Corte de tecido incluído em parafina	Zaragatoa bucal de referência
D8S1179	13	13
D3S1358	15.16	15.16
D19S433	14.16.2	14.16.2
D21S11	30.32.2	30.32.2
D7S820	11	11
CSF1PO	11.12	11.12
D13S317	12.13	12.13
D16S539	12	12
D2S1338	20.25	20.25
D18S51	12.18	12.18
FGA	24	24
Amelogenina	XY	XY

Quando se estudaram os 52- SNPs, num dos casos foi conseguido um perfil completo (Caso 2 - Figura 13).

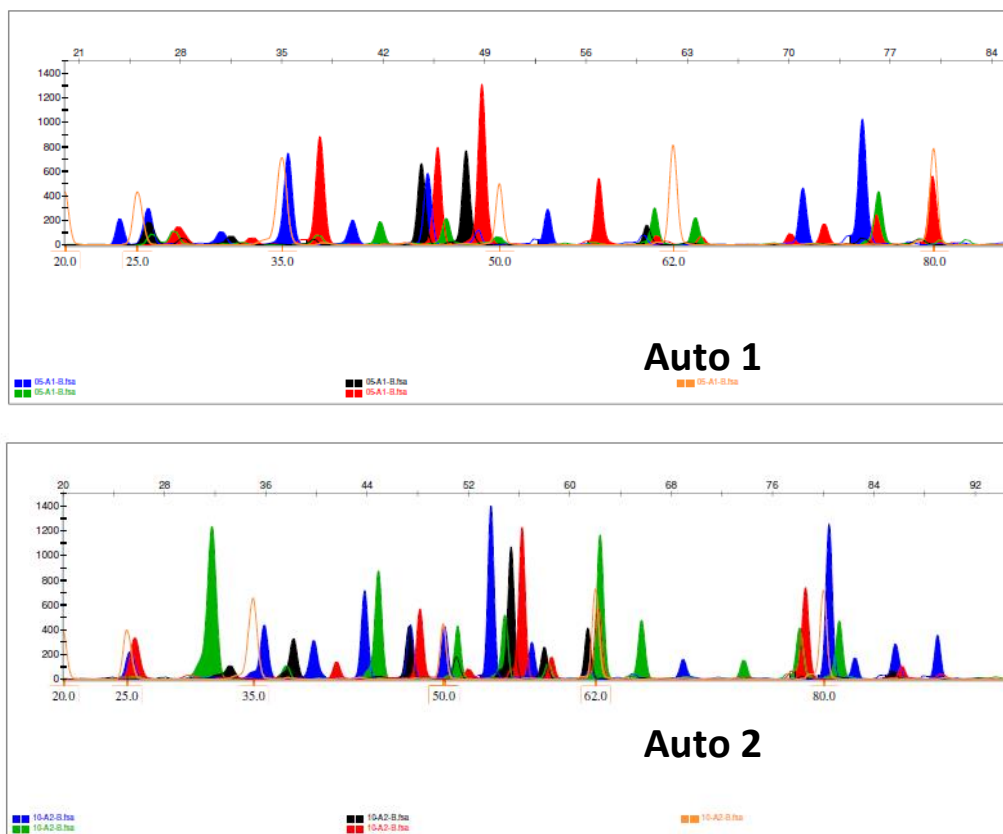


Fig. 13 . Electroferograma correspondente à genotipagem dos produtos de SBE em dois multiplex, Auto 1 e Auto 2, do 52-plex (Caso 2)

No segundo caso conseguiu-se resultados para 48 dos 52 SNPs do *multiplex*.

O impacto do estudo dos SNPs nestes dois casos encontra-se traduzido na Tabela 14.

Tabela 14. Cálculo dos *Likelihood ratio* para os casos estudados

Casos estudados	Likelihood Ratio (LR)		
	STRs	SNPs	Total
<b>Caso 1</b>	1.338E+15	9.046E+26	1.211E+42
<b>Caso 2</b>	7.354E+19	4.682E+27	3.443E+47

Este trabalho foi apresentado numa reunião nacional da medicina legal e ciências forenses e posteriormente submetido para publicação numa revista internacional (Anexo 2).

Foram também realizadas, a título de exemplo, as caracterizações de dois vestígios correspondentes a dois casos criminais já concluídos em que, com recurso aos STRs

convencionais (*kit* Identifiler), num deles foi obtido um perfil incompleto e no restante não se conseguiu obtenção de qualquer perfil genético (Figura 14).

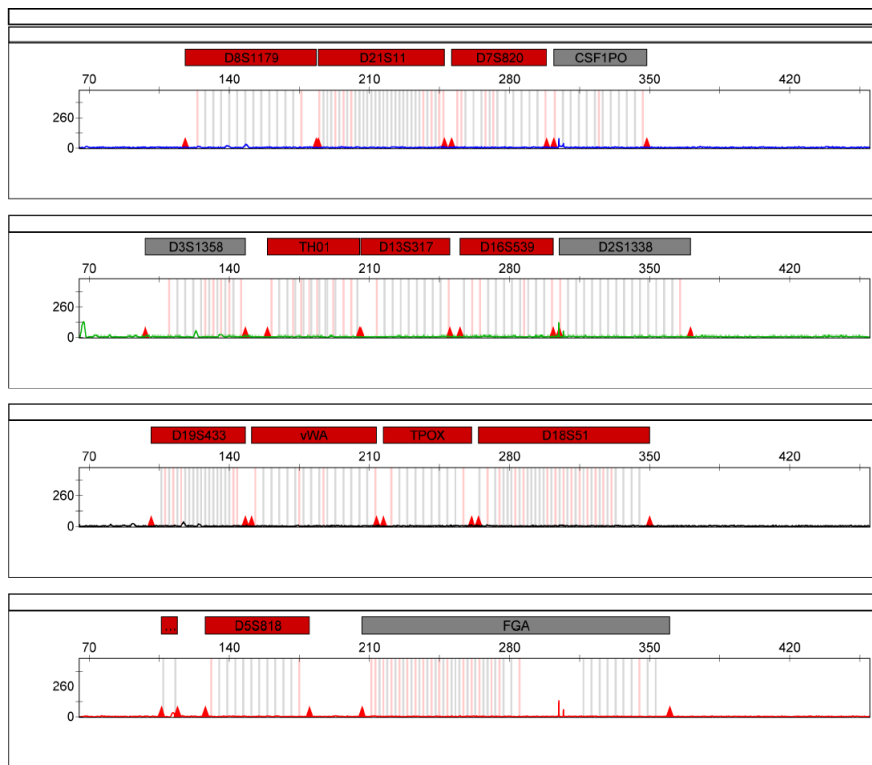


Figura 14. Eletroforetograma após amplificação do *kit* Identifiler de uma amostra única relativa a um caso criminal.

Sendo este o único vestígio disponível para investigação, em cada caso, seria interessante verificar sobre a possibilidade de se obter o perfil para os SNPs autossómicos. De facto, nas duas amostras estudadas correspondentes aos casos referidos, foi possível obter informação para todos os SNPs do 52-SNP-*plex* estudado (Figura15).

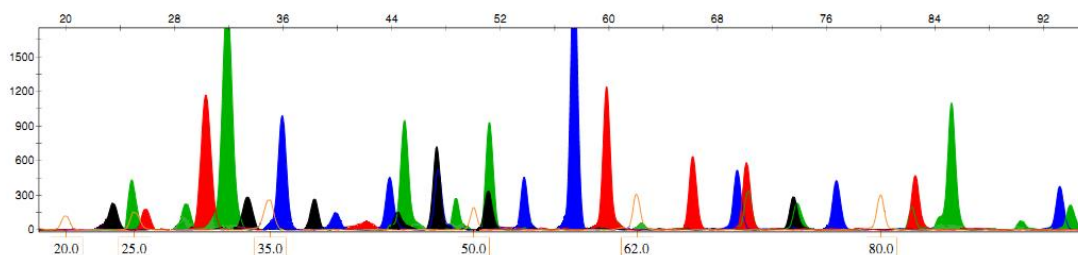


Fig. 15 . Genotipagem do Auto 2 de uma amostra relativa a um caso criminal sem resultados após análise de STRs.

---

## CAPITULO VIII

### 8. Discussão Geral e conclusão

#### 8.1. Introdução

A utilização de SNPs autossômicos na investigação em Genética Forense tem sido muito discutida nos últimos anos, sobretudo porque os SNPs possuem algumas vantagens quando comparados com os STRs [42]. Quando combinados com os marcadores clássicos ou habituais, os SNPs autossômicos aumentam o poder de discriminação e fornecem dados adicionais para os testes de paternidade [4,7]. Adicionalmente o seu estudo pode constituir uma vantagem na análise de amostras muito degradadas [42].

Embora menos polimórficos do que os STRs (multialélicos), os SNPs (bialélicos) podem permitir a distinção entre indivíduos não relacionados ou relacionados entre si, desde que se utilize um elevado número destes SNPs. No entanto, para que se possa obter informação a partir de casos de rotina, especialmente os relacionados com amostras de DNA degradado ou presente em quantidades reduzidas, foi necessária a construção de sistemas *multiplex* robustos, que permitem a tipagem simultânea de elevado número de SNPs, tal como o 52 SNP-plex ou o *GenPlex System*, propostos pelo *SNPforID consortium*. O reduzido tamanho dos produtos de PCR destes SNPs autossômicos em particular, pode ser útil na análise de variadas amostras de DNA degradado frequentemente associadas a casos de criminalística biológica e complementar a informação em casos complexos de identificação e de parentesco.

O desenvolvimento deste tipo de multiplexes maiores e robustos, conforme demonstrado por Sanchez et al. [10], que incluem estes e outros SNPs, foi necessário para conseguir obter dados genéticos a partir de amostras de DNA degradado ou em baixo número de cópias, presente de forma rotineira em evidências forenses [11]. O pequeno tamanho dos produtos de PCR destes SNPs autossômicos (ver Tabela 1) permitiu obter informação a partir de amostras de DNA severamente degradado (ex. Blocos) e complementou os resultados de alguns casos de identificação, criminais e de parentesco.

---

## 8.2. Discussão mediante os objectivos propostos

Neste trabalho conseguimos atingir os objetivos propostos, no relatório apresentado em 2013, depois de reformular os objetivos iniciais.

Esses objetivos incluíam:

- **A validação primária da metodologia seleccionada**, isto é, do 52SNP-*p/lex*, consistiu em participar em exercícios colaborativos, promovidos pelo GHEP-ISFG, que envolveu vários Laboratórios de Genética Forense. Esta validação consistiu em comparar os resultados consenso dos laboratórios forenses participantes com os resultados obtidos pelo laboratório onde este estudo está a ser desenvolvido (SGBF-N). Esta participação constituiu um marco fundamental na selecção do método de caracterização dos SNPs autossómicos e funcionou como motivação adicional para a conclusão do trabalho prático realizado.

- Além destes exercícios colaborativos de intercâmbio de resultados, foram também comparados os resultados destas experiências com os publicados e estabelecidos para os controlos positivos de DNA, concretamente o 9947A e o 007, incluídos nos kits comerciais e disponíveis para genotipagem de STRs autossómicos, que serviram como controlos positivos das experiências realizadas [32]. Relativamente aos DNAs padrão utilizados, foram obtidos os resultados esperados para todos os marcadores relativos ao GenPlex™ referidos no artigo, existindo concordância de resultados em 49 dos 52 SNPs.

Como resultado deste trabalho cumpriu-se o objetivo principal: aplicar os SNPs autossómicos aqui estudados, incluídos no 52-*p/lex*, aos casos solicitados no Serviço de Genética e Biologia Forenses do INMLCF, IP, quando necessário.

- **Estudos populacionais nas amostragens estudadas** de indivíduos representativos do norte de Portugal (131) e para indivíduos imigrantes do leste Europeu (85), cujos dados foram publicados (Anexo 1).

Tal como seria de esperar, uma vez que os SNPs autossómicos incluídos neste sistema são os mais informativos dos atualmente descritos, os valores de heterozigosidade na nossa população são elevados e estão de acordo com os descritos em estudos efetuados em populações do mesmo grupo populacional.

- 
- **Estudo de amostras degradadas em condições laboratoriais controladas** de modo a que, numa fase posterior, se aplicasse esta metodologia a amostras provenientes de perícias relativas a situações reais, já concluídas, provenientes de casos em que a análise dos STRs foi insuficiente ou não conclusiva.

Relativamente a estas amostras, obteve-se sempre um perfil completo, qualquer que fosse o período de exposição do DNA padrão (9947A ou 007) aos raios UV (0. 1. 5. 10 e 15 min). O mesmo aconteceu para as amostras submetidas a digestão com DNaseI para todos os períodos testados. Este resultado era esperado, como tem sido descrito por outros investigadores, tendo sido aceite para publicação (Anexo 3).

- **Aplicação a alguns casos de parentesco complexo** já concluídos, seleccionados daqueles em que, com recurso aos STRs, se obtiveram reduzidos valores de IP ( $\leq 10\ 000$ ). Estes casos consistem sobretudo em investigações de parentesco em que não se dispõe do pretense pai mas sim de familiares seus, como por exemplo, irmãos, filhos biológicos comprovados, pais ou outros.

Os resultados obtidos para os casos de parentesco complexo estudados e apresentados, quando baseados apenas nos perfis obtidos a partir dos marcadores STR e MiniSTR, não eram suficientemente conclusivos. Em todos os casos foi possível diminuir as ambiguidades existentes possibilitando a resolução dos casos com sucesso, por inclusão da análise de SNPs. Estes dados constam da publicação já referida (Anexo 4).

- **Aplicação do 52 SNP-plex, a exemplos de perícias forenses já concluídas.** Estas perícias incluíram além dos casos complexos de investigação de parentesco (por exemplo os que envolvem restos cadavéricos), casos de criminalística biológica e casos de identificação genética individual.

- Mais uma vez comprovou-se que, a genotipagem destes SNPs autossómicos, funcionam bem como marcadores para amostras mais difíceis, frequentemente associadas a casos criminais ou a casos de identificação genética individual (Anexo 2), a partir de vestígios em pior estado de conservação ou em avançado estado de degradação.

- Finalmente, é de referir que se efetuaram alguns testes de sensibilidade, através da quantificação prévia do DNA extraído das amostras estudadas e das experiências efetuadas com

---

DNA padrão de diferentes concentrações. Concluiu-se que conseguimos caracterizar amostras com uma concentração até 0.5ng/ul.

Nota: O estudo inicial levado a cabo com o *GenPlex system* em algumas amostras, cerca de 80, das quais apenas tivemos resultados para cerca de 60. teria sido muito relevante. A relevância desta validação seria fundamental para que a aplicação do estudo deste tipo de marcadores se tornasse mais fácil e acessível a um maior número de laboratórios que atualmente estão envolvidos na resolução de casos de identificação humana.

---

### 8.3. Conclusão e perspectivas futuras

Durante os últimos anos existiu um grande progresso associado aos desafios técnicos envolvidos na obtenção de perfis genéticos na área da Genética Forense. Estes sugerem que as investigações forenses melhorem substancialmente num futuro próximo. No entanto mantêm-se alguns desafios.

Um exemplo é a necessidade óbvia de métodos que permitam identificar múltiplos dadores numa amostra de mistura, especialmente se estes forem do mesmo género. Nestes casos o uso da *next-generation sequencing technologies* (NGS) a partir de moléculas únicas, especialmente as que envolvem PCR, poderá ser a solução.

Estas tecnologias também poderão ser a solução para analisar quantidades exíguas de DNA degradado, característica importante quando se analisam muitas das perícias forenses e nos casos de DVI. Estas novas tecnologias têm a desvantagem de utilizar outro tipo de equipamentos e abordagem técnica que implicarão uma avaliação inicial dos resultados de trabalhos de investigação nesta área específica. bem como um investimento a nível monetário que poderá ser importante.

Alguns destes marcadores, nomeadamente os que permitem estudar a ancestralidade, também poderão ser utilizados com outras finalidades, conjuntamente com os STRs, nomeadamente ver a influência desta ancestralidade biológica/populacional nalgumas características que condicionam por exemplo a predisposição para determinados doenças, como é o caso de alguns tipos de cancro.



---

# Anexos



Contents lists available at ScienceDirect

Legal Medicine

journal homepage: [www.elsevier.com/locate/legalmed](http://www.elsevier.com/locate/legalmed)

## Announcement of Population Data

## Autosomal SNPs study of a population sample from North of Portugal and a sample of immigrants from the Eastern Europe living in Portugal

M. Lurdes Pontes<sup>a,b,\*</sup>, M.F. Pinheiro<sup>a,b,c,d</sup><sup>a</sup> INMLC, LP – North Branch, Portugal<sup>b</sup> CENFOR – Forensic Science Center, Portugal<sup>c</sup> Faculty of Medicine, University of Porto, Portugal<sup>d</sup> Biomedical Sciences Institute “Abel Salazar”, University of Porto, Portugal

## ARTICLE INFO

## Article history:

Received 12 August 2013

Received in revised form 2 December 2013

Accepted 2 December 2013

Available online 12 December 2013

## Keywords:

Forensic and population genetics

Autosomal single nucleotide

polymorphisms

52-SNP multiplex

North of Portugal

## ABSTRACT

The use of autosomal single nucleotide polymorphisms (SNPs) for forensic research has been widely discussed in recent years, mainly because SNPs have important advantages compared to short tandem repeats (STRs).

In this study a total of 131 non related individuals from the North of Portugal and 85 immigrant individuals from the Eastern Europe, mainly Ukrainians, equally non related and residing in Portugal, were typed for 52 loci included in the in the SNP for ID 52plex with the SNaPshot™ assay.

© 2013 Elsevier Ireland Ltd. All rights reserved.

**Population:** In this study a total of 216 samples were collected, mostly buccal swabs, 131 from North of Portugal individuals and 85 from volunteer immigrants from the Eastern Europe, mainly Ukrainians (76 samples) and some Russians (9 samples), after informed consent, anonymized and used for the population study. The two sample populations were compared.

**Extraction:** DNA was extracted using a slightly modified version of the Chelex extraction method [1] and using the bio-robot AutoMate Express™ Forensic DNA Extraction System (Applied Biosystems® – AB).

**Quantification:** DNA concentrations were determined by real-time PCR using the Quantifiler™ Human DNA Quantification Kit (AB) and the ABI PRISM® 7500 Sequence Detection System (AB). The final concentrations were between 0.20 and 10 ng/μl.

**PCR amplification:** A total of 52 autosomal SNPs were amplified by polymerase chain reaction (PCR) in 12.5 μl reactions containing DNA ranging from 0.40 to 3 ng. The PCR conditions were performed as previously described [2].

**SBE reaction:** Two single base extension (SBE) reactions were performed as previously described with slight modifications [2] and included 2.5 μl SNaPshot reaction mix (Applied Biosystems – AB), 1.5 μl SBE primer mix (0.01–0.27 μM) and 2 μl of purified PCR product (with Exonuclease I/Shrimp Alkaline Phosphatase – Exo-SAP),

totalizing 6 μl. Excess nucleotides were then removed by the addition of 1 μl SAP (1 U/μl) to the SBE mix.

**Typing:** SBE products separation was performed by capillary electrophoresis, using an ABI Prism 3130xl Genetic Analyzer (AB), with 36 cm capillary arrays and POP-4 polymer (AB). Analysis was made using GeneMapper® ID-X. Allele calls were made manually. All samples were analyzed at least twice and the results from the two experiments were compared. If the results differed the sample was reanalyzed up to three times.

**Quality Control:** The 52plex SNP assay was validated, locally and through inter-laboratory collaborative exercises, prior to sample populations typing, promoted by the GHEP-ISRG group. The recommendations of the International Society of Forensic Genetics on the analysis of DNA polymorphisms were strictly followed which included the use of recommended nomenclature and guidelines regarding quality control and statistical calculations [3].

**Results:** Allele frequencies for the 52 SNPs are shown (Table 1 and 2). All the genotype distributions matched Hardy-Weinberg equilibrium (HWE) expectations, after Bonferroni corrections.

**Analysis of data:** Allele frequencies, expected and observed heterozygosities (He and Ho, respectively), HWE and linkage disequilibrium tests, FIS and FST values and AMOVA (analysis of molecular variance) analysis were calculated using Arlequin ver 3.1 [4] (Supplementary material). The combined matching probability (pM), power of exclusion (PE) and power of discrimination (PD), were calculated using an in house excel spreadsheet.

\* Corresponding author at: INMLC, LP – North Branch, Portugal. Tel.: +351 22 207 3850.

E-mail address: [lrebelo@dpinmlmj.pt](mailto:lrebelo@dpinmlmj.pt) (M. Lurdes Pontes).

**Access of data:** Available upon request to mlurdes.pontesrebelo@gmail.com.

**Other remarks:** The use of autosomal SNPs for forensic research has been widely discussed in recent years, mainly because SNPs have important advantages compared to short tandem repeats (STRs) [5]. In combination with standard markers, autosomal SNPs improve the power of discrimination and provide potential supplementary data for paternity testing [6,7]. Moreover their study may constitute an improvement in the analysis of degraded samples [8].

A total of 216 samples were typed with the 52plex SNP assay using the SNaPshot<sup>®</sup> kit. Of the 262 profiles generated from the 131 samples from the north of Portugal typed in duplicate, 240 (91.6%) were complete profiles and 164 in 170 (96.5%) for the immigrant population sample constituted by 85 individuals.

The average gene diversity index for all loci was 0.4229 ± 0.2072 for the north of Portugal sample, was obtained with data from 49 in 52 user loci (less than 0.05% missing data) and 0.4287 ± 0.2099 with data obtained from 52 loci for the immigrant population sample (Ukrainians and Russians).

The autosomal SNP A22 (rs733164) showed the lowest (0.1585) PIC value (data not shown). The mean value of the observed heterozygosity for the 52 SNPs in the 52plex was 0.4954 ± 0.1663 for the north of Portugal population (Supplementary material).

The overall *F<sub>ST</sub>* value calculated for the two populations was 0.00905 and the overall *F<sub>IS</sub>* for the north of Portugal population was -0.0913 reflecting the random sampling.

The linearized *F<sub>ST</sub>* value of 0.0085 achieved from comparisons between the two studied sample populations showed that there are significant differences between them (significance level = 0.05). We have decided to include the few Russian samples in the total 85

studied immigrant individuals, because the comparisons made, including or excluding these 9 samples, did not differ significantly between each other.

When we compared these populations with the published results from 49 of these 52 SNPs for a Slovenian population [9], results showed that all the three tested populations were significantly different between each other and the North of Portugal population more apart from the other two compared populations. We also compared these population samples with the ones from NW-Spain and from Russia (SNPforID browser) [10] mainly based on their relative geographic proximity. The compared populations of Russia and Slovenia were not significantly different, but were different from the other three populations.

The comparisons of the North of Portugal with the NW-Spain population samples showed that these two population samples were significantly different (data not shown). This may be due to the discrimination power of some of the studied 52 SNPs (ancestry informative) that are able to distinguish between relatively similar individuals when the analysis is based exclusively in STRs [11]. Although, AMOVA results revealed that, most of the molecular variation was due to variation within populations rather than among them.

The combined mean match probability using the studied SNPs was  $3.0621 \times 10^{-18}$  for the North of Portugal population, corresponding to a combined power of discrimination comparable to the one achieved with some of the routinely used STR Kits in forensics. The mean exclusion probability was 99.9997%.

In conclusion, the typing results revealed that the 52plex SNP assay is robust and sufficiently sensitive to be used for forensic purposes.

**Table 1**  
Allele frequencies for a North of Portugal population sample (n = 131).

Autol						Auto2					
AutSNP	Alleles	Frequencies	AutSNP	Alleles	Frequencies	AutSNP	Alleles	Frequencies	AutSNP	Alleles	Frequencies
A1	C	0.51145	A13	A	0.51527	A24	C	0.41794	A40	A	0.64695
	T	0.48855		G	0.48473		T	0.58206		G	0.35305
A2	C	0.77481	A14	A	0.62977	A25	A	0.47901	A41	C	0.54580
	T	0.22519		T	0.37023		G	0.52069		T	0.45420
A3	A	0.36260	A15	C	0.20992	A26	G	0.40076	A42	A	0.70992
	T	0.63741		T	0.79008		T	0.59921		G	0.29008
A4	A	0.39122	A16	G	0.52863	A27	A	0.60031	A43	A	0.59924
	T	0.60878		T	0.47137		G	0.39969		G	0.40076
A5	A	0.44657	A17	A	0.62977	A28	A	0.25191	A44	C	0.55344
	G	0.55344		G	0.37023		G	0.74809		T	0.44657
A6	A	0.43893	A18	G	0.12595	A29	A	0.50763	A45	C	0.56870
	T	0.56107		T	0.87405		G	0.49237		T	0.43130
A7	A	0.22328	A19	C	0.43893	A30	A	0.52672	A46	A	0.50382
	G	0.77672		T	0.56107		C	0.47328		G	0.49618
A8	C	0.53435	A20	G	0.82061	A32	C	0.65649	A48	C	0.29389
	T	0.46565		T	0.17939		T	0.34351		T	0.70611
A9	C	0.79962	A21	A	0.64695	A33	C	0.23283	A49	A	0.30153
	G	0.20038		G	0.35305		T	0.76718		G	0.69847
A10	C	0.44847	A22	A	0.11069	A34	C	0.33206	A50	A	0.55725
	T	0.55153		G	0.88931		G	0.66794		G	0.46183
A11	C	0.23283	A23	A	0.38740	A35	A	0.61450	A51	A	0.45611
	T	0.76718		G	0.61260		G	0.38550		G	0.54389
A12	A	0.55344				A36	A	0.45802	A52	A	0.62977
	G	0.44657					T	0.54199		T	0.37023
						A37	C	0.55344	A53	A	0.79008
							T	0.45130		G	0.20992
						A38	C	0.83206	A54	A	0.51527
							G	0.16794		G	0.50763
						A39	A	0.50382			
							T	0.49618			

**Table 2**  
Allele frequencies for a population of Eastern Europe immigrants living in Portugal (n=85).

Autol						Aut02					
AutSNP	Alleles	Frequencies	AutSNP	Alleles	Frequencies	AutSNP	Alleles	Frequencies	AutSNP	Alleles	Frequencies
A1	C	0.50000	A3	A	0.41176	A24	C	0.43530	A40	A	0.68824
	T	0.50000		C	0.58824		T	0.56471		G	0.31177
A2	C	0.76471	A4	A	0.54706	A25	A	0.39413	A41	C	0.66471
	T	0.23529		T	0.45294		G	0.60588		T	0.33530
A3	A	0.25294	A5	C	0.20588	A26	G	0.42941	A42	A	0.77647
	T	0.74706		T	0.79412		T	0.57059		G	0.22353
A4	A	0.31471	A6	C	0.58529	A27	A	0.67647	A43	A	0.60588
	T	0.68529		T	0.41471		G	0.32353		G	0.39412
A5	A	0.47647	A7	A	0.71765	A28	A	0.29412	A44	C	0.60000
	G	0.52353		G	0.28235		G	0.70588		T	0.40000
A6	A	0.34118	A18	C	0.27647	A29	A	0.58235	A45	C	0.67647
	T	0.65882		T	0.72353		G	0.41765		T	0.32353
A7	A	0.21765	A9	C	0.35294	A30	A	0.48235	A46	A	0.42353
	G	0.78235		T	0.64706		C	0.51765		G	0.57647
A8	C	0.51765	A20	C	0.80000	A32	C	0.65294	A48	C	0.40588
	T	0.48235		T	0.20000		T	0.34706		T	0.59412
A9	C	0.64118	A21	A	0.71177	A33	C	0.16471	A49	A	0.48824
	G	0.15882		G	0.28824		T	0.83530		G	0.50588
A10	C	0.49412	A22	A	0.23824	A34	C	0.43530	A50	A	0.53529
	T	0.50588		G	0.76176		G	0.56471		G	0.46471
A11	C	0.25882	A23	A	0.37059	A35	A	0.50412	A51	A	0.56471
	T	0.74118		C	0.62941		G	0.40588		G	0.48824
A12	A	0.68824				A36	A	0.34706	A52	A	0.69412
	G	0.31176					T	0.65294		T	0.30588
						A37	C	0.47647	A53	A	0.80588
							T	0.52353		G	0.19412
						A38	C	0.83530	A54	A	0.37941
							G	0.16471		G	0.62059
						A39	A	0.45882			
							T	0.54118			

**Acknowledgements**

We thank Mykola Stasyuk for helping us in the collection of samples from immigrant individuals Ukrainian and Russian, living in Portugal. We also thank to Prof. María Victoria Lareu, Christopher Phillips, Olalla M. Amigo and Manuel Fondevila from the Institute of Legal Medicine of the University of Santiago de Compostela, for their advising and training.

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.legalmed.2013.12.002>.

**References**

[1] Walsh PS, Metzger DA, Higuchi R. CHEX® 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biochem Biophys Res Commun* 1991;105:06–13.

[2] Sanchez JJ, Phillips C, Barsting C, Balogh K, Bogus M, Fondevila M, et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 2006;27:1713–24.

[3] Schneider PM. Scientific standards for studies in forensic genetics. *Forensic Sci Int* 2007;165:238–43.

[4] Excoffier L, Laval G, Schneider S. Arlequin (version 3.5): a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010;10:564–7.

[5] Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med* 2001;114:204–10.

[6] Amorim A, Pereira L. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Sci Int* 2005;150:17–21.

[7] Phillips C, Fondevila M, Garcia-Magarinos M, Rodriguez A, Salas A, Carracedo A, et al. Resolving relationship testes that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Sci Int Genet* 2008;2:198–204.

[8] Butler JM, Coble MD, Vallone PM. STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol* 2007;2007(3):200–5.

[9] Dobnic K, Barsting C, Roclenbauer E, Tomas C, Morling N. Typing of 49 autosomal SNPs by SNaPshot® in the Slovenian population. *Forensic Sci Int Genet* 2010;4:e125–7.

[10] SNPflorID browser; 2013 [accessed 25.07.13].

[11] Phillips C, Salas A, Sanchez J, Fondevila M, Gómez-Tato A, Álvarez-Díaz J, et al. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Sci Int Genet* 2007;1:273–80.

Avens Publishing Group

J Forensic Investigation

February 2015 Vol.:3, Issue:1

© All rights are reserved by Pontes et al.

# Analysis of Autosomal SNPs in Paraffin-embedded Tissue Samples: Two Cases

**Keywords:** Forensic genetics; Autosomal single nucleotide polymorphisms; S2-SNP-multiplex; Casework; Paraffin-embedded tissue samples

## Abstract

Despite the numerous STR loci available for analysis of cases submitted to forensic laboratories, namely with the use of commercial kits that amplify mini - short tandem repeats (STRs), some cases remain "unsolved" i.e. fail to yield a STR DNA profiling result or result in likelihood ratio calculations that are not sufficiently high. Autosomal single nucleotide polymorphisms (SNPs) may constitute a potential advantage to solve some of these difficult cases, as has already been described by others.

In this work we present two revisited cases of our routine work, where the type of samples available for analysis was histological blocks of paraffin-embedded tissue samples from a man (Case 1- "lymph node") and from a woman (Case 2). The application of autosomal SNPs in the analysis of paraffin-embedded tissues constitutes one of the possible approaches that can be used in our laboratory in some "difficult" cases, and is part of a validation process of the described methodology.

Following extraction the two sample DNAs had already been genotyped with the routine commercial STR or mini-STR kits, like AmpFLSTR® Identifier®, MiniFiler™ and NGM™ PCR Amplification Kits and compared to their respective reference samples. These samples were then typed for the 52 loci included in the in the SNP for ID 52 plex using a SNaPshot™ assay previously described by Sanchez et al. in order to test and validate this multiplex in real casework [1].

The study showed that autosomal SNP can be used as additional markers to the profiling of difficult samples such as paraffin embedded tissue.

## Introduction

Currently forensic laboratories solve almost all the requested cases through the traditional use of commercial kits that amplify short tandem repeats (STRs) or mini-STRs. In spite of this, there are still some "unsolved" or challenging cases that are occasionally encountered in forensic laboratories like paraffin-embedded tissue. "Unsolved" or challenging cases can be paternity cases, usually where the father is deceased and other relatives are disposable to investigate the paternity, that often lead to low paternity indexes after STR profiling. Paraffin embedded samples are not usual in forensic investigations. But these types of samples due to their high level of degraded sample DNAs, may also lead to low Likelihood Ratios (LR) after conventional STR analysis.

The use of autosomal SNPs for forensic research has been widely discussed in recent years. SNPs have important advantages compared to short tandem repeats (STRs) [2] and in combination with standard markers, improve the power of discrimination and provide potential supplementary data for paternity testing [3,4]. Moreover their study may constitute an improvement in the analysis of degraded samples [5].

Open Access

Case Report



## Journal of Forensic Investigation

M Lurdes Pontes<sup>1,2\*</sup>, Olalla Maroñas Amigo<sup>3</sup>,  
Manuel Fondevila<sup>4</sup>, M. João Porto<sup>1,2</sup> and R.  
Medeiros<sup>4,5,6</sup>

<sup>1</sup>Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal

<sup>2</sup>CENICFOR- Forensic Science Center, Portugal

<sup>3</sup>Institute of Legal Medicine, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

<sup>4</sup>Molecular Oncology group, Portuguese Institute of Oncology, Portugal

<sup>5</sup>ICBAS, Abel Salazar Institute for the Biomedical Sciences, University of Porto, Largo Professor Abel Salazar, 4099-003, Porto, Portugal

<sup>6</sup>LPCC, Research Department-Portuguese League Against Cancer (NRNorte), Estrada Interior da Circunvalação 6637, 4200 Porto, Portugal

### Address for Correspondence

M Lurdes Pontes, Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal, E-mail: mlurdes.pontes@icml.com

Submission: 13 January, 2015

Accepted: 07 February, 2015

Published: 11 February, 2015

Reviewed & Approved by: Prof. Shawn L. Parsons, Director, National Forensic Autopsy and Tissue Recovery Services, Overland Park, KS, USA

As it was already said, paraffin embedded samples are not usual in forensic investigations. Here, the only two interesting cases involving paraffin-embedded tissue samples we have had over the years are presented. These cases were previously resolved using the traditional STR markers, but they were revisited in order to access the relevance of studying autosomal SNPs. This was achieved by calculating the LR before and after adding the information of the analyzed SNPs. The use of SNPs for challenging forensic sample type's is documented by others, as is the recovery and analysis of DNA from formalin-fixed paraffin-embedded (FFPE) tissues but mostly in medical context and using other analysis methodologies. Equally, the combining of statistical data from STR and SNP loci has been demonstrated before.

In spite of these two cases the STR results alone would have provided sufficient certainty of source identification to be used without the additional SNP work, our aim was to validate the use of the SNP for ID 52 plex and demonstrate the increment of the calculated LRs after their use as additional markers to the profiling of difficult samples such as paraffin embedded tissue, in order to apply their use in similar cases that may be encountered in our forensic laboratory in the future.

## Materials and Methods

### Samples

In this study histological material from two persons with a histological tumor - positive diagnostic was available. The aim of the requested examinations was to confirm the identity of the persons involved, supposedly the donors of the paraffin-embedded samples,

ISSN: 2330-0396

who questioned the results of the histology laboratory biopsies observations. For that, a DNA profile had to be obtained from cuts of the histological blocks of paraffin-embedded tissue samples (Figure 1); then it was necessary to compare the obtained profiles with the ones of the reference samples from the donors of the histological samples, respectively a man and a woman.

### Extraction

DNA of reference samples (buccal swabs) was extracted using the slightly modified Chelex® method [6] and the DNA from the casework samples was extracted using the EZ1 DNA Tissue Kit together with the EZ1 DNA Forensic Card protocols with a BioRobot® EZ1 (Qiagen) [7].

### Quantification

DNA concentrations were determined by real-time PCR using the Quantifiler™ Human DNA Quantification Kit (AB) and the ABI PRISM® 7500 Sequence Detection System (AB). The final concentrations of the casework samples were 0.02 and 0.05 ng/µl, respectively.

### PCR amplification

The STR and mini-STR amplification of the extracted samples was achieved following the manufacturer protocols for the kits Identifier® and Minifiler™ or NGM™ respectively.

A total of 52 autosomal SNPs were amplified by polymerase chain reaction (PCR) in 12.5 µl reactions containing DNA, in the case of these particular casework samples we adjusted the final volume to 25 µl in order to have a higher input of DNA sample. The PCR conditions were performed as previously described [1]. When using the problem samples of cases 1 and 2, we used a DNA input of -0.10 and -0.25 ng respectively.

### SBE reaction

Two single base extension (SBE) reactions were performed (Auto 1 and Auto 2) as previously described with slight modifications [1] and included 2.5 µl SNaPshot reaction mix (Applied Biosystems - AB), 1.5 µl SBE primer mix (0.01–0.27 µM) and 2 µl of purified PCR product (with Exonuclease I/Shrimp Alkaline Phosphatase - ExoI-

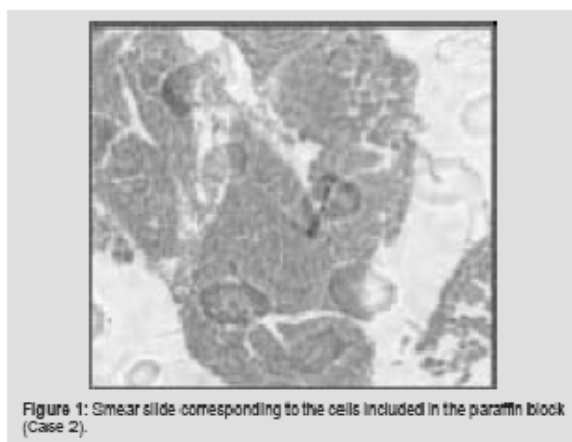


Figure 1: Smear slide corresponding to the cells included in the paraffin block (Case 2).

SAP), totaling 6 µl. Excess nucleotides were then removed by the addition of 1 µl SAP (1 U/µl) to the SBE mix. The multiplexes were balanced in the laboratory in order to validate the described method in our particular conditions.

### Typing

STR and SBE products separation was performed by capillary electrophoresis, using an ABI Prism 3130 xl Genetic Analyzer (AB), with 36 cm capillary arrays and POP-4 polymer (Applied Biosystems, Foster City, CA, USA). Analysis was made using GeneMapper® ID-X. For the autosomal SNPs allele calls were made manually.

### Quality control

The 52 plex SNP assay was validated, locally and through inter-laboratory collaborative exercises, promoted by the GHEP-ISFG group.

### Analysis of data

The LR's were calculated using an in house excel spread sheet, simply by multiplying the individual LR's from each of the SNP markers, using the previous published frequencies for those markers in a population sample of the North of Portugal, assuming that all the markers are independent [8].

### Results

The results obtained for the STRs for Case 1 are presented in Table 1. These were only possible for 11 and 14 STR markers, respectively for case 1 and 2. Concerning SNPs, as it was expected, the results were possible for all the 52 typed SNPs in one of the cases (Case 1) and for 50 SNPs in the other (Case 2). The results for Case 2 are shown (Figure 2).

Table 1: STRs/mini-STRs results (Case 1).

STR marker	Paraffin embedded tissue cut	Reference Buccal swab
D6S1179	13	13
D3S1358	15,16	15,16
D19S433	14,16,2	14,16,2
D21S11	30,32,2	30,32,2
D7S820	11	11
C8F1PO	11,12	11,12
D13S317	12,13	12,13
D16S539	12	12
D2S1338	20,25	20,25
D18S51	12,18	12,18
FGA	24	24
Amelogenin	XY	XY

Table 2: Likelihood ratio calculation results.

Studied cases	Likelihood Ratio (LR)		
	STRs	SNPs	Total
Case 1	1.338E+15	9.046E+26	1.211E+42
Case 2	7.354E+19	4.682E+27	3.443E+47

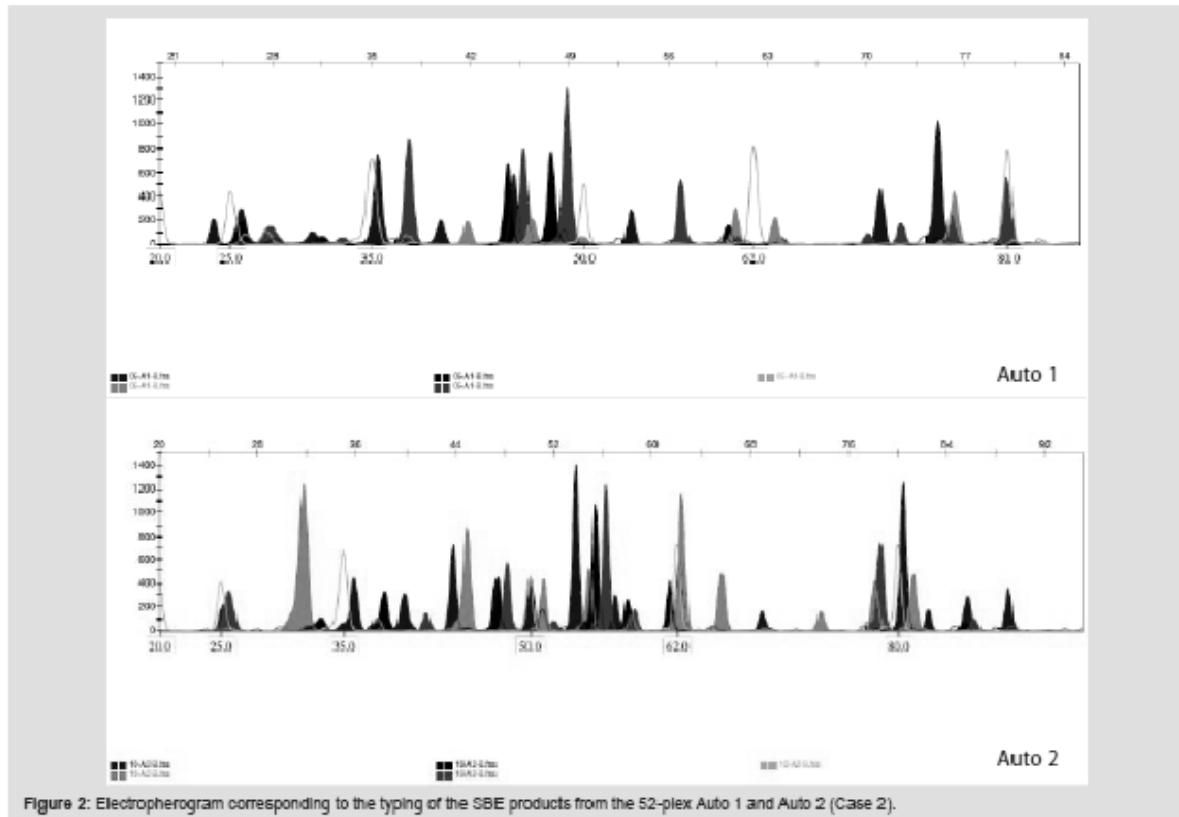


Figure 2: Electropherogram corresponding to the typing of the SBE products from the 52-plex Auto 1 and Auto 2 (Case 2).

The impact of studying SNPs in the overall Likelihood Ratio (LR) is presented by calculation of this parameter before and after adding the additional information given by the typed SNPs (Table 2), assuming all the studied loci are independent. Combining of statistical data from STR and SNP loci has been demonstrated before, but it is still not known if all SNP loci are independent of each other or of the STR loci used in this study. It is only known that this set of 52 SNP markers was carefully tailored in order to be sufficiently apart from each other guaranteeing this independence as reported by Sanchez et al. [1].

Before this example of autosomal SNP application some validation work was made in our laboratory, namely using internal controls of known standard DNAs and through inter-laboratory collaborative exercises, promoted by the GHEP-ISFG group. We also used allele frequencies for the 52-SNPs on our reference population – north of Portugal population [8].

### Discussion and Conclusion

The novelty of this work consisted in the application of the 52-SNP multiplex previously described [1] for study paraffin-embedded tissue samples, traditionally not usual in forensic cases. In the medical field there are many studies that involve paraffin-embedded samples. Another study looked at paraffin-embedded tissue for identification

of the tissue source using a SNP multiplex, but with a more recent methodology [9].

The typing results and the performed LR calculations revealed that the 52 plex SNP assay may constitute a robust and sufficiently sensitive tool to complement the results from the traditional markers (STR and mini-STRs) in difficult forensic casework, which involves samples with material particularly degraded, like paraffin-embedded tissue samples, old cadaveric remains like bones as it was already reported [10], and other low-template DNA and highly degraded DNA from crime case samples involved in routine casework [11,12].

### References

1. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, et al. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27: 1713-1724.
2. Gill P (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med* 114: 204-210.
3. Amorim A, Pereira L (2005) Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Sci Int* 150: 17-21.
4. Phillips C, Fondevila M, Garcia-Magarinos M, Rodriguez A, Salas A, et al. (2008) Resolving relationship testes that show ambiguous STR results: using autosomal SNPs as supplementary markers. *Forensic Sci Int Genet* 2: 198-204.
5. Butler JM, Coble MD, Vallone PM (2007) STRs vs. SNPs: thoughts on the

ISSN: 2330-0396

- future of forensic DNA testing. *Forensic Sci Med Pathol* 3: 200-205.
6. Walsh PS, Metzger DA, Higuchi R (1991) CHELEX 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 54: 134-139.
  7. (2004) EZ1 DNA Handbook, Qiagen.
  8. Lurdes Pontes ML, Pinheiro MF (2014) Autosomal SNPs study of a population sample from North of Portugal and a sample of immigrants from the Eastern Europe living in Portugal. *Leg Med (Tokyo)* 16: 115-120.
  9. Miller JK, Buchner N, Timms L, Tam S, Luo X, et al. (2014) Use of Sequenom sample ID Plus® SNP genotyping in identification of FFPE tumor samples. *PLoS One* 9: e88163.
  10. Fondevila M, Phillips C, Navaran N, Fernandez L, Cenezo M, et al. (2008) Case report: Identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. *Forensic Sci Int Genet* 2: 212-218.
  11. Børsting C, Mogensen HS, Mørling N (2013) Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples. *Forensic Sci Int Genet* 7: 345-352.
  12. Freije-Aradas A, Fondevila M, Kriegel AK, Phillips C, Gill P, et al. (2012) A new SNP assay for identification of highly degraded human DNA. *Forensic Sci Int Genet* 6: 341-349.

#### Acknowledgements

We thank Prof. María Victoria Laneu from the Institute of Legal Medicine of the University of Santiago de Compostela, Galicia, Spain, for her support and advising.

## Autosomal SNPs in different forensic applications

M. Lurdes Pontes<sup>a,b\*</sup>  and R. Medeiros<sup>c,d,e</sup> 

<sup>a</sup>*North Delegation, Instituto Nacional de Medicina Legal e Ciências Forenses, LP, Porto, Portugal;* <sup>b</sup>*INMLCF, CENCIFOR- Forensic Science Center, IP, Coimbra, Portugal;* <sup>c</sup>*Molecular Oncology group, Portuguese Institute of Oncology, Porto, Portugal;* <sup>d</sup>*ICBAS, Abel Salazar Institute for the Biomedical Sciences, University of Porto, Porto, Portugal;* <sup>e</sup>*LPCC, Research Department-Portuguese League Against Cancer (NRNorte), Porto, Portugal*

(Received 6 March 2015; accepted 28 April 2015)

The aim of this work was to test a 52 SNP-plex in our laboratory in order to apply it in casework cases, namely in kinship involving relatives of the true father leading to low Parentage Index (PI) values, and in unusual cases such as paraffin-embedded tissue samples or in other cases involving severely degraded samples related to some identification cases. For this, a standard DNA was used, artificially degraded by digestion with DNase I and by different UV exposure times. STR previously analysed cases were revisited and studied. All the samples were typed for the SNP for ID 52plex using the SNaPshot™ assay. In all the cases, a SNP profile was achieved. In two cases involving paraffin-embedded tissue samples, it was only possible to obtain results for 48% (11 out of 23) and 61% (14 out of 23) STR and mini-STR markers, respectively. Concerning SNPs, results were possible for all the SNPs (52) in one of the cases and for 50 SNPs (96%) in the other. The PI or LR calculated values suffered a minimum of a fourfold increment. Autosomal SNPs can be used as additional markers to the profiling of 'difficult' cases that emerge in the routine of many laboratories.

**Keywords:** legal medicine; forensic genetics; autosomal SNPs; SNP multiplex; casework samples

### Introduction

Currently, although almost all cases that are sent to forensic laboratories have a satisfactory response through the use of the available commercial kits that amplify short tandem repeats (STRs), mini-STRs or both, some cases remain challenging, and additional markers may be useful. Examples of these 'unsolved' or challenging cases can be complex paternity cases, usually when the father is deceased or close relatives of the biological father are investigated. These cases often lead to low paternity indexes after STR profiling. Furthermore, casework that involves the analysis of very low quantities or severely degraded samples also tends to lead to low Likelihood Ratios (LR). Moreover, unusual kinds of samples sent to the forensic laboratories, such as paraffin-embedded samples, due to their high level of degraded sample DNAs, may lead to low Likelihood Ratios after conventional STR analysis. The genotyping of autosomal SNPs in laboratories of forensic research constitutes one of the possible approaches that can be used and, specifically in the authors' laboratory, that deals with

\*Corresponding author. Email: [bebelo@dpinml.mj.pt](mailto:bebelo@dpinml.mj.pt)

all kinds of forensic investigations and which has been widely discussed in recent years. SNPs have important advantages compared with short tandem repeats (STRs)<sup>1</sup> and, in combination with standard markers, they improve the power of discrimination and provide potential supplementary data for paternity testing<sup>2,3</sup>. Moreover, their study may constitute an improvement in the analysis of degraded samples<sup>4</sup>.

Before applying these markers to the routine casework, it is necessary to do some validation work in forensic laboratories, especially when accreditation issues are taken into consideration. For that reason, normally known standard DNAs (ex: 9947A – Applied Biosystems) available in the commercial kits are used in the daily routine of forensic genetics laboratories. Our proposed strategy involved the use of replicates of this DNA that were digested with DNase I at different time intervals. Another set of replicates were submitted to different times of UV light exposures. With these different treatments, it was intended to mimic the degradation and other problems that arise when real casework samples are analysed. The DNA samples were then typed for the 52 loci included in the SNP for ID 52plex using a SNaPshot assay previously described by Sanchez et al. (2005).

This typing methodology was also applied to some revisited cases that had previously been studied using the traditional STR markers. This was valued by calculating the LR before and after adding the information of the analysed SNPs. The use of SNPs for challenging forensic sample types is documented by others, as is the recovery and analysis of DNA from formalin-fixed paraffin-embedded (FFPE) tissues, but mostly in the medical context and using other analysis methodologies<sup>5</sup>. Equally, the combining of statistical data from STR and SNP loci has been demonstrated before. The 52plex SNP assay was also validated through inter-laboratory collaborative exercises, promoted by the GHEP-ISFG group<sup>6</sup>. The SNP for ID 52plex was used for testing in these revisited cases. In most of these cases the STR results alone provided sufficient information to make the report.

The aim of this work was to validate the use of the SNP for ID 52plex and demonstrate the increment of the calculated LRs after their use as additional markers to those profiling of cases that need this additional information, and it may be very useful to have as alternative to STRs or mini-STRs in our forensic laboratory applying it in future casework.

## Material and methods

### *Samples*

The validation experiments included a known standard DNA available on the commercial STR kits commonly used for routine testing, the 9947A (applied Biosystems), for which the profile has already been published by others. Replicates were made of this standard DNA (2 ng/ $\mu$ l) that was digested with DNase I (0,5U), in a total volume of a 10  $\mu$ l solution including the DNase digestion Buffer (1x), during different intervals: 0, 1, 2, 5, 10 and 15 min, totalling 12 sample tubes. The digestion was stopped by placing the sample tubes on ice. Another set, like this replica, with an additional 12 sample tubes, was exposed to UV lights (254 nm) in a workstation (The Clean Spot Work Station, Coy, Michigan) during the same time intervals. The samples were then stored at  $-20^{\circ}\text{C}$ , until genotyping. As well as these replicas, the profile for the standard 9947A was obtained. The total number of samples analysed was 22.

Samples from selected cases were: Case 1 – a complex paternity case where a paternal grandfather and an uncle instead of the alleged father, already deceased, the

mother and the child; Case 2 – a complex paternity case, similar to the above, where we had only the alleged paternal grandmother, the mother and the child; Case 3 and Case 4 – cases that involved histological material, paraffin-embedded samples, from two persons, a man and a woman, with histological tumours (positively diagnosed), who questioned the results of the histology laboratory biopsy observations. Case 5 and Case 6 – these cases involved trace blood samples related to two different crime scenes, sent by the investigating police forces that, after analysis, gave no STR or mini-STR profiles.

The selected samples for study were from different types of cases, representing the variability of casework samples that can be analysed by Forensic Genetics laboratories.

#### *Extraction*

The DNA of reference samples (buccal swabs) was extracted using the slightly modified Chelex method<sup>7</sup> and the DNA from the casework samples and from cuts of the histological blocks of the paraffin-embedded tissue samples was extracted using the EZ1 DNA Tissue Kit together with the EZ1 DNA Forensic Card protocols with a BioRobot EZ1 (Qiagen)<sup>8</sup>.

#### *Quantification*

DNA concentrations were determined by real-time PCR using the Quantifiler Human DNA Quantification Kit (AB) and the ABI PRISM 7500 Sequence Detection System (AB). The final concentrations of the casework samples varied between 0.02 and 0.05 ng/ $\mu$ l.

#### *PCR amplification*

The STR and mini-STR amplification of the extracted samples was achieved following the manufacturer protocols for the Identifiler and Minifiler or NGM kits. A total of 52 autosomal SNPs were amplified by polymerase chain reaction (PCR) in 12.5  $\mu$ l reactions containing DNA. In the case of casework samples with low DNA concentration, the final volume was adjusted to 25  $\mu$ l in order to have a higher input of DNA template. The PCR conditions were performed as previously described<sup>9</sup>. In the case of the paraffin-embedded samples, the DNA input was ~0.10 and ~0.25 ng respectively.

#### *SBE reaction*

Two single base extension (SBE) reactions were performed (Auto 1 and Auto 2) as previously described with slight modifications<sup>9</sup> and included 2.5  $\mu$ l SNaPshot reaction mix (Applied Biosystems – AB), 1.5  $\mu$ l SBE primer mix (0.01–0.27  $\mu$ M) and 2  $\mu$ l of purified PCR product (with Exonuclease I/Shrimp Alkaline Phosphatase – ExoI-SAP), totalling 6  $\mu$ l. Excess nucleotides were then removed by the addition of 1  $\mu$ l SAP (1 U/ $\mu$ l) to the SBE mix. The multiplexes were balanced in the laboratory in order to validate the described method in our particular conditions.

#### *Typing*

STR and SBE products separation was performed by capillary electrophoresis, using an ABI Prism 3130xl Genetic Analyzer (AB), with 36 cm capillary arrays and POP-4

polymer (Applied Biosystems, Foster City, CA, USA). Analysis was made using GeneMapper ID-X. The allele calls were made manually.

### **Quality control**

The 52plex SNP assay was validated, locally and through inter-laboratory collaborative exercises, promoted by the GHEP-ISFG group.

### **Analysis of data**

The PIs of the complex paternity case were calculated using the Familias<sup>10</sup> software and the LR<sub>s</sub> from the paraffin embedded cases were calculated using an in-house excel spreadsheet, simply by multiplying the individual LR<sub>s</sub> from each of the SNP markers, using the previous published frequencies for those markers in a population sample of the North of Portugal, assuming that all the markers are independent<sup>11</sup>.

### **Results**

Among all the samples analysed of the standard DNA (1 ng/μl) without treatment and of the replicates of this standard DNA, it was possible to obtain complete profiles after treatment with DNase I or the exposure to UV lights (254 nm) during 1, 2, 5, 10 or 15 min (Figure 1).

The results obtained for the STRs for one of the cases (Case 3), involving paraffin-embedded samples, are presented in Table 1. In both of the two such cases that were sent to our lab, it was only possible to obtain results for 48% (11 out of 23) and 61% (14 out of 23) STR and/or mini-STR markers, respectively. Concerning SNPs, as expected, the results were possible for all the 52 typed SNPs in one of the cases (Figure 2) and for 50 (96%) SNPs in the other. For the two samples (Case 5 and Case 6) with no previous results (no profiles achieved) with the conventional STR typing (Figure 3) it was also possible to achieve an incomplete profile after 52 SNP-plex profiling (Figure 4).

Following extraction, the paternity or the casework sample DNAs used in this work had already been genotyped using the routine commercial STR or mini-STR kits, such as AmpFLSTR Identifiler, MiniFiler and NGM PCR amplification kits and for each case a report was produced with the results and respective conclusions. These samples were then typed for the 52 loci included in the SNPforID 52plex using a SNaPshot™ assay previously described by Sanchez et al.<sup>9</sup> in order to test and internally validate this multiplex for posterior application in real casework. The impact of studying SNPs in the overall Paternity Index or in the Likelihood Ratio (LR) is presented by calculation of this parameter before and after adding the additional information given by the typed SNPs (Table 2), for the paternity cases (Case 1 and Case 2) and Likelihood Ratios (LR), for the two reported cases of tissue paraffin-embedded samples, assuming all the studied loci as independent. The final values of PI, considering the joint STR and SNP information, had a minimum of a fourfold increase. For the LR calculations this increase was much higher than in the paternity cases.

### **Discussion and conclusion**

This work consisted of an internal validation study using a known standard DNA and an application of the 52-SNP multiplex previously described<sup>9</sup> to different types of case

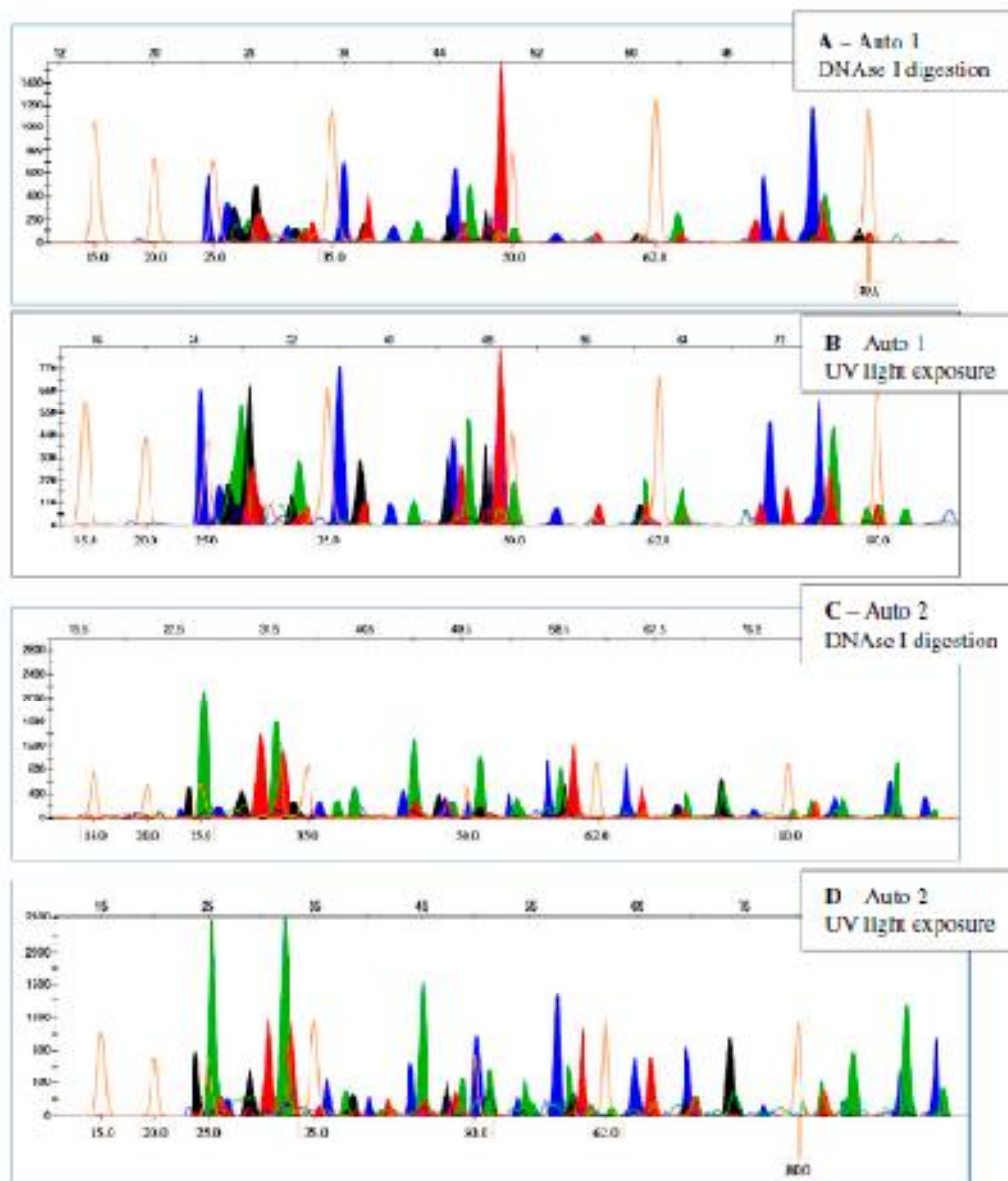


Figure 1 Artificially 'degraded' standard DNA 9947a: A – Auto 1 typing results, after 15 min digestion with DNase I, B – Auto 1 typing results, after 15 min exposure to UV light (254 nm), C – Auto 2 typing results, after 15 min digestion with DNase I; D – Auto 2 typing results, after 15 min exposure to UV light (254 nm).

work samples, corresponding to the possible cases that can be requested to forensic genetic laboratories related to some legal problems that need to be solved. The validation work described above, using the internal control of known standard DNA and also some validation through inter-laboratory collaborative exercises, promoted by the GHEP-ISFG group and on the present examples of autosomal SNP applications, will enable the application of this SNP 52-plex in our future routine casework. This set of 52 SNP markers was carefully tailored in order to be sufficiently apart from each other to guarantee this independence as reported by Sanchez et al.<sup>2</sup>, but it is still not known

Table 1. STR results for the samples of one of the two reported paraffin-embedded cases (Case 3).

Studied markers	Paraffin-embedded sample	Reference buccal swab
D8S1179	13	13
D3S1358	15,16	15,16
D19S433	14,16.2	14,16.2
D21S11	30,32.2	30,32.2
D7S820	11	11
CSF1PO	11,12	11,12
D13S317	12,13	12,13
D16S539	12	12
D2S1338	20,25	20,25
D18S51	12,18	12,18
FGA	24	24
Amelogenin	XY	XY

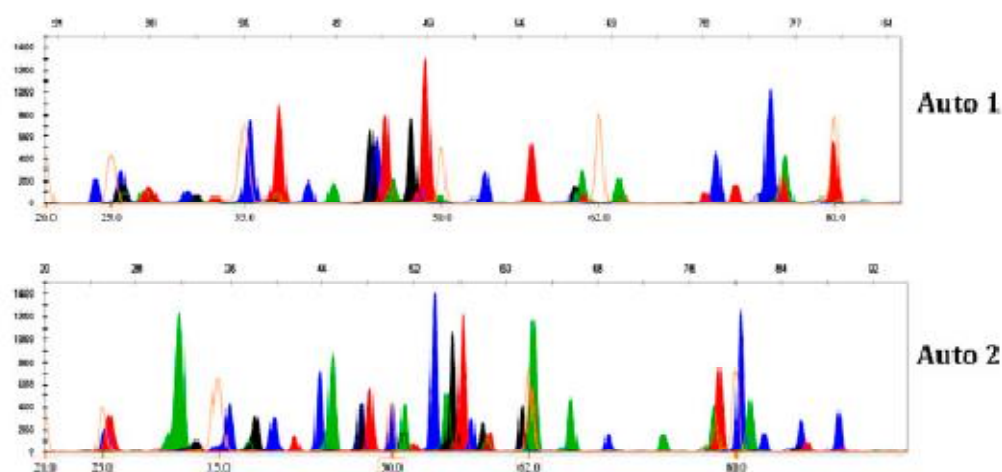


Figure 2. Electropherogram corresponding to the typing of the SBE products from the SNP 52-plex Auto 1 and Auto 2 for a case sample involving a paraffin-embedded tissue (Case 3).

if all SNP loci are independent of each other or of the STR loci used in this study. For the calculations, it was possible to achieve a minimum of a fourfold increase of the PIs after adding the information from SNP typing. The allele frequencies for the 52-SNPs enabling the calculations were previously determined on our reference population – the population of the north of Portugal<sup>11</sup>.

The two studied specific cases reported here are very unusual in the forensic context, where histological blocks of paraffin-embedded tissue samples were the only samples available for analysis. The purpose of using these two particular cases was to serve as concrete examples of ‘difficult’ cases that can be requested from forensic laboratories, where the 52 SNP-plex may constitute a possible methodology to apply. These samples are more common in the medical field where many studies that involve paraffin-embedded samples have already been reported. At least one other study has looked at paraffin-embedded tissue for identification of the tissue source using a SNP multiplex, but with a more recent methodology<sup>12</sup>.

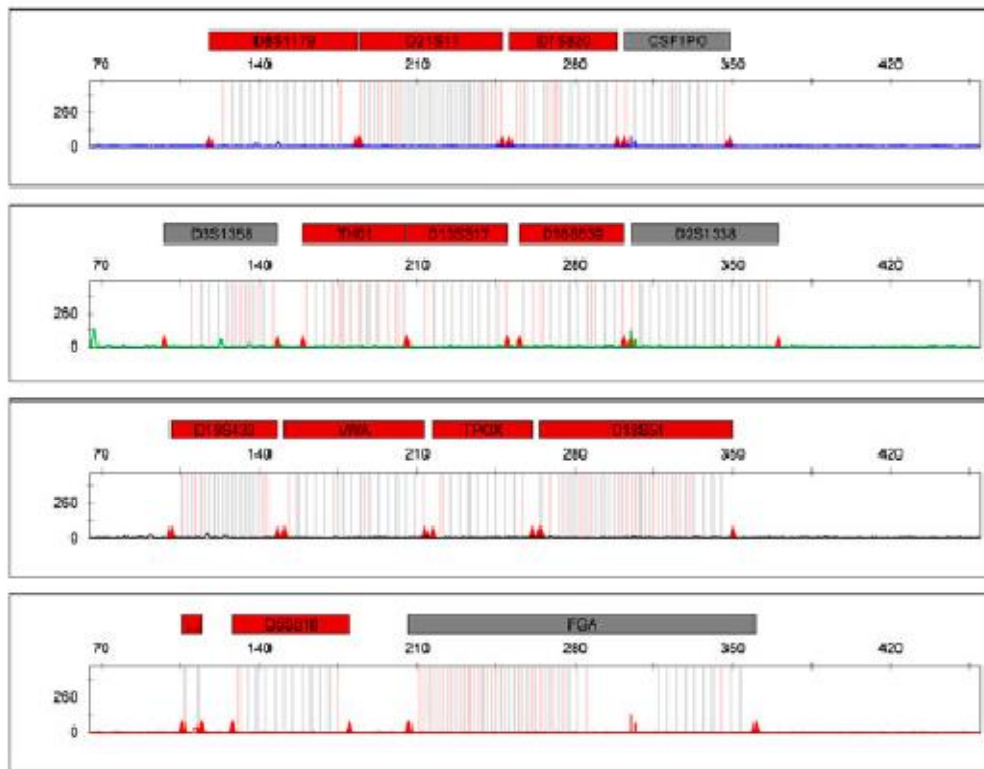


Figure 3. Electropherogram obtained from amplification of the sample using the Identifier kit – Case 6.

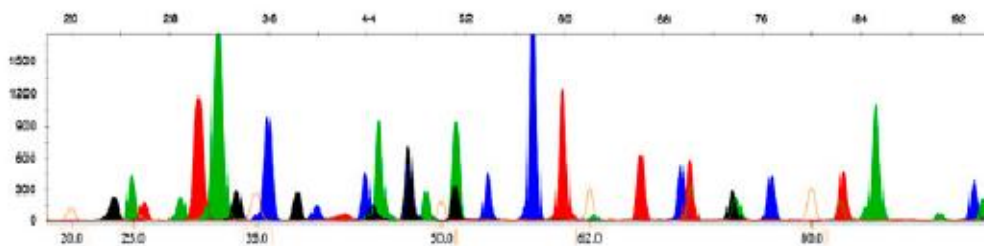


Figure 4. Auto 2 Typing of a casework sample (Case 6) that gave no profile with STR analysis.

Table 2. Complex studied families and paraffin-embedded cases: Comparison of the calculated Paternity Index and Likelihood ratio values using STR, SNPs or a combined STR/SNP.

Studied cases	Paternity Index(PI)/Likelihood Ratio (LR)		
	STRs	SNPs	Total
Complex paternity (GF+U+C+M) (Case 1)	1.195E+03	4.120E+00	IP:4.923E+03
Complex paternity (GM+C+M) (Case 2)	1.553E+03	9.431E+02	IP:1.465E+06
Paraffin embedded tissue (Case 3)	1.338E+15	9.046E+26	LR:1.211E+42
Paraffin embedded tissue (Case 4)	7.354E+19	4.682E+27	LR: 3.443E+47

Footnote: GF – Alleged paternal grandfather; U – Alleged paternal Uncle; C – Child; M – Mother; GM – Alleged paternal grandmother.

The internal validation work, the typing results for some revisited case examples involving complex paternity testing and samples from casework that gave incomplete or no STR profiles and the performed PI or LR calculations presented here, showed that the 52plex SNP assay may constitute a robust and sufficiently sensitive tool to complement the results from the traditional markers (STR and mini-STRs) in difficult forensic cases. Combining statistical data from STR and SNP loci has been demonstrated before. It is a methodology that may be applied in cases that involve samples with particularly degraded material, such as paraffin-embedded tissue samples, old cadaveric remains such as bones, as has already been reported<sup>13</sup>, and other low-template DNA and highly degraded DNA from crime case samples involved in routine casework<sup>14,15</sup>.

The study showed that autosomal SNPs can be used as additional markers to the profiling of various types of samples that forensic laboratories are requested to analyse, which would not be resolved based only on STR, in order to explain the facts according with the alternative hypotheses that can emerge from each particular case, enabling a satisfactory answer to a large number of casework.

#### Acknowledgements

The authors thank Professor María Victoria Lareu from the Institute of Legal Medicine of the University of Santiago de Compostela, Galicia, Spain, for her support and advice.

#### Disclosure statement

No potential conflict of interest was reported by the authors.

#### ORCID

M. Lurdes Pontes  <http://orcid.org/0000-0003-0673-1296>

R. Medeiros  <http://orcid.org/0000-0003-3010-8373>

#### References

1. Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med.* 2001;114:204–210.
2. Amorim A, Pereira L. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Sci Int.* 2005;150:17–21.
3. Phillips C, Fondevila M, García-Magarinos M, Rodríguez A, Salas A, Carracedo A, Laréu MV. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Sci Int Genet.* 2008;2:198–204.
4. Butler JM, Coble MD, Vallone PM. STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol.* 2007;3:200–205.
5. Baak-Pablo R, Dezentje V, Henk-Jan Guchelaar, van der Straaten T. Genotyping of DNA samples isolated from formalin-fixed paraffin-embedded tissues using preamplification. *J Mol Diagn.* 2010;12(6):746–749.
6. [www.gep-is.fg.org/en/working-commissions/snps-forensic-genetics/](http://www.gep-is.fg.org/en/working-commissions/snps-forensic-genetics/)
7. Walsh PS, Metzger DA, Higuchi R. CHELEX<sup>®</sup> 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques.* 1991;10:506–513.
8. EZ1 DNA handbook. For automated purification of DNA from blood, tissue, and other samples using the BioRobot<sup>®</sup> EZ1 workstation. 2nd ed. Qiagen; 2004.
9. Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis.* 2006;27:1713–1724.

10. Egeland T, Mostad PF, Mevag B, Steinsen M. Beyond traditional paternity and identification cases: selecting the most probable pedigree. *Forensic Sci Int*. 2000;110:47–59.
11. Pontes ML, Pinheiro MF. Autosomal SNPs study of a population sample from North of Portugal and a sample of immigrants from Eastern Europe living in Portugal. *Legal Med*. 2014;16:118–120.
12. Miller JK, Buchner N, Timms L, Tam S, Lao X, et al. Use of Sequenom sample ID Plus<sup>®</sup> SNP genotyping in identification of FFPE tumor samples. *PLoS One*. 2014;9(2):e88163.
13. Fondevila M, Phillips C, Naveran N, Fernandez L, Salas A, et al. Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. *Forensic Sci Int Genet*. 2008;3:212–218.
14. Børsing C, Mogensen HS, Morling N. Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples. *Forensic Sci Int Genet*. 2013;7:345–352.
15. Freire-Aradas A, Fondevila M, Kriegl AK, Phillips C, Gil P et al. A new SNP assay for identification of highly degraded human DNA. *Forensic Sci Int Genet*. 2012;6:341–349.

## SNP Markers as Additional Information to Resolve Complex Kinship Cases

M. Lurdes Pontes<sup>a,b</sup> Manuel Fondevila<sup>c</sup> Maria Victoria Laréu<sup>c</sup> Rui Medeiros<sup>d,e,f</sup>

<sup>a</sup>Laboratório de Genética e Biologia Forenses, National Institute of Legal Medicine and Forensic Sciences, North Delegation, Porto, Portugal;

<sup>b</sup>CENCIFOR- Forensic Science Center, Porto, Portugal;

<sup>c</sup>Instituto de Ciencias Forenses, Universidade de Santiago de Compostela, Santiago de Compostela, Spain;

<sup>d</sup>Molecular Oncology, Portuguese Institute of Oncology, Porto, Portugal;

<sup>e</sup>ICBAS, Abel Salazar Institute for the Biomedical Sciences, University of Porto, Porto, Portugal;

<sup>f</sup>LPCC, Research Department-Portuguese League Against Cancer (NRNorte), Porto, Portugal

### Keywords

Forensic genetics · Paternity testing · STRs · SNP markers

### Summary

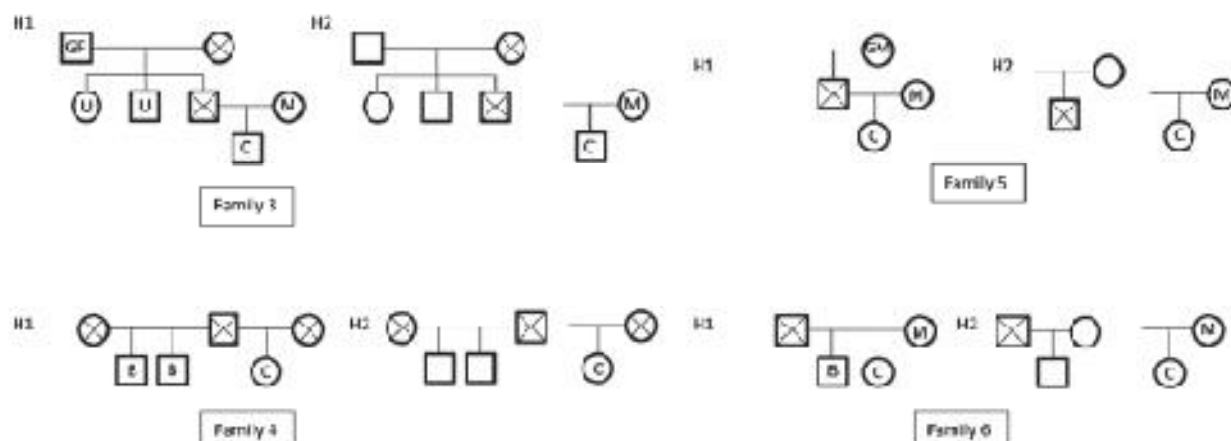
**Background:** DNA profiling with sets of highly polymorphic autosomal short tandem repeat (STR) markers has been applied in various aspects of human identification in forensic casework for nearly 20 years. However, in some cases of complex kinship investigation, the information provided by the conventionally used STR markers is not enough, often resulting in low likelihood ratio (LR) calculations. In these cases, it becomes necessary to increment the number of loci under analysis to reach adequate LRs. Recently, it has been proposed that single nucleotide polymorphisms (SNPs) could be used as a supportive tool to STR typing, eventually even replacing the methods/markers now employed. **Methods:** In this work, we describe the results obtained in 7 revised complex paternity cases when applying a battery of STRs, as well as 52 human identification SNPs (SNPforID 52plex identification panel) using a SNaPshot methodology followed by capillary electrophoresis. **Results:** Our results show that the analysis of SNPs, as complement to STR typing in forensic casework applications, would at least increase by a factor of 4 total PI values and correspondent Essen-Möller's W value. **Conclusions:** We demonstrated that SNP genotyping could be a key complement to STR information in challenging casework of disputed paternity, such as close relative individualization or complex pedigrees subject to endogamous relations.

© 2015 S. Karger GmbH, Freiburg

### Introduction

In forensic genetics, DNA samples are analyzed through the comparison of a combination of variants from a particular group of polymorphisms, resulting in a highly unlikely probability of being reproduced in a second individual within the population. In fact, although more than 99% of the genome is invariable across the human population, polymorphic variations in DNA sequence can be used to both differentiate and correlate individuals. This DNA feature allows the resolution of many problems related with paternity testing and identification cases, including disaster victim identification (DVI). Moreover, it can be used to solve some criminal cases.

Short tandem repeats (STRs) have been the first-choice genetic markers of forensic scientists all over the world, since their use enables a satisfactory answer for almost all the requested cases, mainly due to their high degree of polymorphism and consequent discrimination power. Lately, however, another type of genetic markers called SNPs (single nucleotide polymorphisms), which consists in DNA sequence variations that results from single base changes in the genome sequence, has been considered [1, 2] as additional informative markers, providing reliable likelihood ratios (LRs). Some reports go even further, considering that SNPs will replace STRs in forensic investigation [3]. In spite of almost all SNPs are bi-allelic, and consequently less polymorphic than multi-allelic STRs, they have several known advantages over STRs: SNPs are more stable genetic markers, with low mutation rates and so less likely changes over generations which is crucial concerning paternity cases. They also will probably allow a cheaper, easier and faster analysis and demand a much lower DNA consumption due to the high degree of multiplexing and automation of the genotyping methodologies [4, 5]. Moreover, as single base markers, they can be typed with ex-



**Fig. 1.** Pedigrees of complex studied families: family 3, 4, 5, and 6.

tremely short amplicons (around 50–90 bp) and since the typing reaction is independent from PCR amplification, a large number of markers can be amplified with smaller amplicon lengths without any effect over genotyping [3]. This last property becomes critical when dealing with degraded DNA samples which display highly fragmented DNA. However, SNP analysis also presents some limitations. In fact, using statistic simulations to compare STRs and SNPs effectiveness in kinship studies, it has been reported that the possibility of inconclusive results is much higher when using only SNPs [2, 6]. Furthermore, these authors discuss the validity of using exclusively SNP polymorphisms in routine paternity investigations. It is generally accepted that SNPs are individually less informative than STRs; however, their simplicity allows a multiplexing level so high that a single SNP typing multiplex could pile up a number of markers large enough to compensate this limitation [7].

In this work we present some paternity testing results, using conventional STR markers and the 52 SNP multiplex reported by the SNPforID consortium, that was locally validated using the SNaPshot<sup>®</sup> reaction mix (Applied Biosystems, Walham, MA, USA) and capillary electrophoresis.

## Material and Methods

### Samples and Pedigree Description

The samples were selected from 7 complex families that were already typed using conventional STR markers included in the Identifiler (Applied Biosystems) and in the PowerPlex 16 (Promega, Madison, WI, USA) kits. The first three reported families concerned trios that had shown incompatibilities in one or two STR, suspected to be single step mutations (Family 1 (F1)), Family 2 (F2) and Family 7 (F7), leading to reduced paternity indexes. In the other studied families instead of the alleged father we had other relatives from the paternal line: the alleged paternal grandfather and alleged uncle (Family 3 (F3)), two confirmed sons of the alleged father from a different mother (Family 4 (F4)), the alleged paternal grandfather (Family 5 (F5)) and a single confirmed son of the alleged father (Family 6 (F6)) (Fig. 1). All these families were analyzed after typing with all markers included in the above mentioned commercial kits, totaling 17 STRs.

### DNA Extraction and Quantification

DNA was extracted using a slightly modified version of the Chelex extraction method [8]. Quantification was achieved by real-time PCR using the Quantifiler<sup>™</sup> Human DNA Quantification Kit (Applied Biosystems) and the ABI PRISM<sup>®</sup> 7500 Sequence Detection System (Applied Biosystems).

### PCR and Genotyping

A total of 52 autosomal SNPs were amplified by PCR in 12.5  $\mu$ l reactions containing DNA ranging from 0.40 to 3 ng. The PCR procedures were performed as previously described [7]. Two single base extension (SBE) reactions were performed as previously described with slight modifications [7]: 2.5  $\mu$ l SNaPshot reaction mix (Applied Biosystems), 1.5  $\mu$ l SBE primer mix (0.06–0.27  $\mu$ mol/l) and 2  $\mu$ l of purified PCR product (with exonuclease I / shrimp alkaline phosphatase (ExoI-SAP)), 6  $\mu$ l in total. Excess nucleobases were then removed by the addition of 1  $\mu$ l SAP (1 U/ $\mu$ l) to the SBE mix. SBE products separation was performed by capillary electrophoresis, using an ABI Prism 3130xl Genetic Analyzer (Applied Biosystems), with 36 cm capillary arrays and POP-4 polymer (Applied Biosystems). Analysis was made using GeneMapper ID-X. Allele calls were made manually.

### Statistical Analysis

After applying the SNPforID 52-plex assay, calculations of paternity indexes (PI) were made using the software Familias [9] considering the information of STRs and of the 52 SNPs as stand-alone assays. The two assays were joined on a single genetic profile (table 1), using the previously published frequencies for those markers in a population sample of the North of Portugal [10] and assuming that all markers are independent.

Three of the studied cases involved incompatibilities, two (F1, F7) with one suspected single step mutation and the other one (F2) with two suspected single step mutations. In Familias software it is possible to take the mutation rates (mr) into consideration for those loci in order to obtain a value of probability for the alleged kinship relation. In the case of STRs we have inserted mr values from the STRbase with different values according to the marker with the observed incompatibility. The STRbase website lists a wide range of observed STR mutation rates ([www.csl.lisid.gov.si/tech/strbase/mutation.htm](http://www.csl.lisid.gov.si/tech/strbase/mutation.htm)) based on data from the most extensive recent AAFBI survey on typing trios with standard STRs ([www.aafbi.org/ia/publications/Documents/patstrp03.pdf](http://www.aafbi.org/ia/publications/Documents/patstrp03.pdf)). A universal mutation rate of  $mr = 2.5 \times 10^{-6}$  was used for all SNPs [11, 12]. Silent alleles were also assumed. The silent allele frequencies were calculated according the equation  $1 / (n + 1)$ , where n corresponds to the total number of observed alleles.

**Table 1.** Complex studied families: comparison of the calculated PI values using STR, SNPs, or a combined STR/SNP

Family	PI (STRs)	W	PI (SNPs)	Total PI (STRs + SNPs)	Total W	
F1	2,198.0	0.9995	$1.29 \times 10^2$	$2.83 \times 10^5$	0.9999929475	1 Mut D2S1338
F2	2,955.0	0.9996	$1.18 \times 10^1$	$3.49 \times 10^4$	0.9999713948	2 Mut FGA and SE33
F3	1,195.0	0.9992	4.12	$4.92 \times 10^5$	0.9997969296	GF + 2U + C + M
F4	241.0	0.9959	7.31	$1.75 \times 10^5$	0.9994319150	2 S (M1) + C (M2)
F5	1553.0	0.9994	$9.43 \times 10^2$	$1.46 \times 10^6$	0.9999993171	GM + C + M
F6	3,865.0	0.9997	$1.76 \times 10^1$	$6.81 \times 10^4$	0.9999853163	Brother + C + M
F7	6,133.0	0.9998	$9.71 \times 10^5$	$5.95 \times 10^6$	0.9999999998	1Mut D16S539

PI = Paternity index; W = Essen-Möller's probability value; GF = alleged paternal grandfather; U = alleged paternal uncle; C = child; M = mother; S = son of the alleged father; GM = alleged paternal grandmother; Mut = inconsistency in STR transmission.

#### Quality Control

The recommendations of the International Society of Forensic Genetics (ISFG) on the analysis of DNA polymorphisms were strictly followed which included the use of recommended nomenclature and guidelines regarding quality control and statistical calculations [13]. The SNPforID 52plex assay was validated locally and through inter-laboratory collaborative exercises, prior to sample typing, promoted by the Spanish and Portuguese Working Group from ISFG (GHEP-ISFG group).

## Results and Discussion

In table 1 the studied families and the respective PI calculations with STRs, SNPs, and both markers are resumed. Concerning genetic data interpretation, STR PI results range from 241 to 6,133, SNP PI results range from 4 to 971,396, whereas using the combined STR/SNP PI results range from 1,759 to 5,957,575,011. Our results are in agreement with previous reports from other authors stating that in most families the PI value obtained analyzing just SNPs was lower than with STRs analysis [2]. These authors propose that the use of SNPs, in some cases, will lead to a PI lower than the one obtained with STR information, returning an inconclusive statistic calculation [2]. However, our results show that, by gathering the information of these two types of markers together, it was possible to achieve an increased PI in all revised cases. In 4 out of the 7 cases, SNPs were not very informative, with  $PI < 20$ . This was expected, since these are the most difficult cases that can emerge from the routine work in which it is very common to lack sufficient information for a kinship investigation, since in one case there are two incompatibilities in STR analysis (F2) and in the others we dispose of far relatives of the child, as is the case of grandparents, uncles or brothers (F3, F4, and F6). In view of these data it is demonstrated that studying SNPs as a complement to STRs resulted in an at least fourfold increased total PI values, while in the remaining cases the PI improvement is several magnitude orders greater. Special attention is given to the case F7 displaying a mutation on D16S539 across generations in which the application of SNP typing and ignoring STR data increase the PI by two magnitude orders. These results may be explained by the fact that the applied SNP typing multiplex, SNPforID 52plex, in European populations provides a higher statistical power than the 15 human identity STRs contained on the employed amplification kit [7]. Furthermore, since the mutation rate of SNPs is on average five magnitude

orders lower than that of STRs, it is highly unlikely that any parent-to-child mutation will be spotted. Moreover, the combined STR/SNP increased PI results, leading to a strongly supported hypothesis of paternity and an accurate conclusion. Similar results when applied to complex casework resolution have been obtained with SNP typing by other authors [6, 13–15] – hardly a surprise, since this methodology accomplishes the single-tube-reaction study of a great number of markers. Additionally, these markers show an extremely low mutation rate, and those on the SNPforID 52plex have been selected to cover all chromosomes with an ample distribution across the genome in order to ensure the discovery of recombination-born differences even among members of complex inbred pedigrees [6].

Human identity testing using DNA has undergone an evolution from multilocus restriction fragment length polymorphism (RFLP) and less polymorphic single-locus RFLP probes to even less polymorphic STR systems, mostly to achieve laboratory automation [3]. Recently, SNPs offered a further possibility for this purpose. Future identity typing laboratories will undoubtedly utilize simpler test systems involving SNPs in order to more easily automate the laboratory process. The SNPforID 52plex assay was developed by the SNP consortium for identification purposes. Many authors have confirmed the utility of these markers; namely in severely degraded samples like bones and in other applications [16, 17].

## Conclusion

After revisiting some complex paternity cases and complementing the information achieved by the study of the conventional STR markers, by using the 52 SNP markers included in the SNPforID 52plex assay developed by the SNPforID Consortium, we found greatly increased PI values and a significantly reduced uncertainty in all cases investigated. SNP typing represents an extremely useful tool to complement STR information in deficient cases of disputed paternity regardless of the cause of the deficiency. However, even if the methodology used in this study is perfectly fitting for forensic casework requirements, it has its drawbacks. SNaPshot is a laborious technique if it is accomplished without some degree of automation, and, above all, the data analysis process requires a significant level of expertise to be properly held. Moreover, it is very difficult to resolve mixtures when using these markers [18]. In fact, mixture interpretation is a major drawback or limitation of foren-

sic SNP typing 'with the currently available typing techniques' [3]. This remains true even if better technologies allow for enhancing mixture resolution with SNPs.

New promising technologies are being developed that hopefully would lead to the construction of SNP typing forensic kits in the near future that could provide a more user-friendly alternative to the current SNP typing methods. With regard to findings like those

presented in this paper, we sincerely hope that, after implementing these more user-friendly SNP typing technologies, these markers will be included in the everyday forensic toolbox.

## Disclosure Statement

The authors declare no conflict of interest.

## References

- 1 Butler JM, Coble MD, Vallone PM: STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol* 2007,3:200-206.
- 2 Amorim A, Pereira L: Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Sci Int* 2005,150:17-21.
- 3 Kayser M, de Knijff P: Improving human kinship through advances in genetics, genomics and molecular biology. *Nat Rev Genet* 2011,12:179-191.
- 4 Sobrino B, Brian M, Carracedo A: SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 2005,154:181-194.
- 5 Pui-Yan K, Xiangting C: Detection of single nucleotide polymorphisms. *Curr Issues Mol Biol* 2003,5:B-60.
- 6 Børsting C, Mikkelsen M, Mørding N: Kinship analysis with diallelic SNPs - experiences with the SNPforeID multiplex in an ISO17025 accredited laboratory. *Transfus Med Hemother* 2012,39:195-201.
- 7 Sanchez JJ, Phillips C, Børsting C, et al: A multiplex assay with 57 single nucleotide polymorphisms for human identification. *Electrophoresis* 2006,27:1713-1714.
- 8 Walsh TG, Meier DA, Higuchi R: CHELSEA® 108 as a medium for simple extraction of DNA for PCR-based typing. *Forensic Sci Int* 1991,50:295-312.
- 9 Egeblad T, Meisner W, Mengler R, Sorenson M: Beyond traditional paternity and identification cases: selecting the most probable pedigree. *Forensic Sci Int* 2006,165:47-59.
- 10 Pereira ML, Pinheiro MF: Autosomal SNP study of a population sample from North of Portugal and a sample of immigrants from the Eastern Europe living in Portugal. *Legal Med* 2014;36:118-123.
- 11 Phillips C, Wang H, Hillard D, Fondevis M, Heston G, Hyland P, Muirgrave-Brown E, Proll C, Ramos-Late E, Sobrino B, Turida M, Syndercombe Court D, Carracedo A, Schneider PM: SNPforeID Consortium: Evaluation of the Genolex SNP typing system and a 19plex forensic marker panel. *Forensic Sci Int Genet* 2007,1:80-105.
- 12 Børsting C, Sanchez JJ, Birk AH, Bruun FQ, Hüllerberg C, Hansen AJ, Hansen HE, Simonsen BT, Mørding N: Comparison of paternity indices based on typing 15 STRs, 7 VNTRs and 52 SNPs in 50 Danish mother-child father trios. *Prog Forensic Genet* 2006,1:146-158.
- 13 Phillips C, Fondevis M, Garcia Magarinos M, Rodriguez A, Salas A, Carracedo A, Lareu MV: Resolving relationship issues that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Sci Int Genet* 2006,2:195-204.
- 14 Thoren J, Martinez M, Brotons-Aranda A, Fondevis M, Carracedo A, Porras L, Guzmán L, Ueng STB, Min ISTB and SNP markers to solve complex cases of kinship analysis. *Forensic Sci Int Genet* 2013,7:103-107.
- 15 Lindner I, Warmo-Schwartz N von, Meier P, Fimmers R, Böttner A: Usefulness of SNPs as supplementary markers in a paternity case with 3 genetic incompatibilities at autosomal and 1 chromosomal loci. *Transfus Med Hemother* 2014;41:17-21.
- 16 Fondevis M, Phillips C, Navean N, Fernandez L, Salas A, Carracedo A, Lareu MV: Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. *Forensic Sci Int Genet* 2008,2:212-218.
- 17 Børsting C, Mørding N, Mørding N: Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples. *Forensic Sci Int Genet* 2011;7:344-347.
- 18 Butler JM, Coble MD, Vallone PM: STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol* 2007,3:200-205.



**Submission  
Confirmation**

Thank you for submitting your manuscript to *Australian Journal of Forensic Sciences*.

Manuscript ID: TAJF-2015-0070

Title: SNPs and STRs in Forensic Medicine. A new strategy for kinship evaluation

Pontes, Lurdes

Authors: Carneiro de Scusa, Maria José  
Medeiros, Rui

Date Submitted: 05-May-2015

 [Print](#)  [Return to Dashboard](#)

---

## **SNPs and STRs in Forensic Medicine. A new strategy for kinship evaluation**

Lurdes Pontes<sup>1,2</sup>, M. J. Carneiro de Sousa<sup>1,3</sup>, R. Medeiros<sup>3,4,5</sup>

<sup>1</sup>Laboratório de Genética e Biologia Forenses, National Institute of Legal Medicine and Forensic Sciences, North Delegation

<sup>2</sup>CENCIFOR- Forensic Science Center, Portugal

<sup>3</sup>ICBAS, Abel Salazar Institute for the Biomedical Sciences, University of Porto, Largo Professor Abel Salazar, 4099-003, Porto, Portugal

<sup>4</sup>Molecular Oncology group, Portuguese Institute of Oncology, 4200-072 Porto, Portugal

<sup>5</sup>LPCC, Research Department-Portuguese League Against Cancer (NRNorte), Estrada Interior da Circunvalação 6657, 4200 Porto, Portugal

---

## Abstract

The emerging technologies for accessing and exploiting genomic information may be a useful tool for the development of novel diagnostic products and strategies. These technologies include the analyses of single nucleotide polymorphisms (SNPs) that have attracted much interest in recent years applied to various scientific areas. The abundance of SNPs, despite their simplicity and rather limited polymorphic content, is the main reason for their study in the medical field, since they have been extensively used as markers to identify genes that underlie complex diseases and also to realize the potential of pharmacogenomics in relation to different drug responses. Additionally SNPs have been shown to be very useful in forensic genetics resolving all kinds of legal problems, namely crime cases, disaster victim identification and paternity and kinship investigation testing. The low mutation rate of SNPs, estimated to be 100,000 times lower than that of short tandem repeats (STRs), makes these markers very suitable for relationship testing. In the great majority of the cases, analyses with the widely used sets of STR markers provide powerful statistical evidence but some of them remain with ambiguous results. Those include cases that show some problems inherent to the use of STRs like mutations or cases with more complex pedigrees to analyze than simple paternity. Until the development of the SNPforID multiplex assay by the SNP for ID Consortium, there was no real alternative to the commonly used STR kits in kinship. At this time several forensic laboratories are using this methodology especially as a complement to the study of STRs in some of their casework cases.

This paper intends to make a data update on the joint use of STRs and autosomal SNPs in the evaluation and kinship calculation, one of the strategies currently used for this purpose, namely reviewing and comparing results published by various working groups.

## Keywords

Relationship testing, Supplementary markers, Single nucleotide polymorphisms, SNPforID, Mutation

## Introduction

Single nucleotide polymorphisms (SNPs) are the class of polymorphisms more abundant in the human genome and are estimated to be 10<sup>6</sup>11 million with an average of 1 SNP per 275 base pairs (bp) once in every 300 nucleotides on average, which means there are roughly 10 million SNPs in the human genome [1, 2]. Most SNPs are bi-allelic, and the average mutation rate at a particular base pair position is extremely low (approximately 10<sup>67</sup> to 10<sup>69</sup>). For these reasons, SNPs have been widely used to trace the demographic history of human populations and for association studies between SNP markers and some traits and diseases [3-6]

---

In the last decade SNPs have been given special attention in forensic medicine, tending to continue, since they are used in genetic research concerning ethnic origin, physical traits [7, 8], molecular pathology, among others SNPs are already used by some laboratories in routine casework investigations and for human identification purposes and relationship testing, and have proved to be very valuable markers [9, 10]. In these laboratories SNPs are mainly used as supplementary markers, but there are situations where SNPs may be the best and often the only solution to choice [9, 11]. If the DNA is highly degraded, frequently STR analysis fails and there is a much higher chance of obtaining results with SNP markers [11, 12]. This is due to the fact that SNPs may be amplified on very short amplicons (<100 bp). Furthermore, as STR kits that amplify 15617 STRs are commonly used in routine work, one mutation is expected to occur in one of the loci typed with these kits in approximately 3% of all meioses [1]. In contrast, mutations occur 100,000 times less frequently in SNPs than in STRs, and inconsistencies caused by mutations in SNP loci are extremely rare. [13]. The 1000 Genomes Project showed that rare variants account for a large majority of the existing genetic diversity between individuals as well as within populations [14]. SNPs as other structural variants such as deletions, duplications, insertions among others, larger than SNPs, are classified into common and rare based on minor allele frequencies (MAF). A widely used cut-off for defining rare SNPs being a MAF of less than 0.05 [3]. But in forensic genetics it is considered that in a rare SNP the MAF is greater than 1% [2]. However, this cut-off does not have any special biological relevance. Although differences in SNP allele frequencies might be influenced by various demographic factors like selection and population size, time is the major determinant in the rise or fall of allele frequencies [15]. The relative phenotypic importance of common and rare SNPs is highly debated and while most of the Mendelian traits and deleterious mutations have been shown to be rare, several studies suggest that some continuous traits, like height, blood pressure and others, in which the distribution of phenotypes in the population varies along a continuum, might well be explained in terms of common SNPs [5]. Some studies have also shown that SNPs and structural variants are often  $\varnothing$ population-specific meaning that they have been found to occur only in a single population. These SNPs might be important in assigning characteristic phenotypes and disease susceptibility/protection to a population [6].

In paternity testing the genetic profiles of the individuals are used to compare the relative likelihood ratios (LR) of two probabilities: on the numerator the probability of observed genotypes, given the tested man is the father, and on the denominator the probability of observed genotypes, given a random man (being unrelated) is the father. In the great majority of the cases, analyses with the widely used sets of short tandem repeat markers (STRs) provide powerful statistical evidence, favoring one of the alternative hypotheses. Nevertheless, there are some situations where the final statistical result is ambiguous, namely if the calculated LR is  $\leq 10\ 000$ . If one or two genetic inconsistencies are detected between a child and the alleged parents in a relationship case, it can be difficult to draw a conclusion because the inconsistencies may be

---

mutations or indicate that one of the alleged parents may not be the true biological parent. When the alleged father shows incompatible genotypes at a few loci, a very high paternity index may be achieved in the remaining systems. Other ambiguous situations emerge if there are deficient cases families instead of the alleged father we have other relatives from the paternal line to analyse, like his alleged brother, father, son or other. In the first type of cases, the possibility that the alleged father is actually a close relative of the real one (son, father or brother) can reasonably be raised. Another situation can occur if partial profiles are obtained when highly degraded DNA is analyzed, for instance in identification cases arising from skeletonized remains. In all cases, when the statistical evidence obtained is considered as insufficient, one common practice is to extend the set of analyzed markers [16-18]. Lineage markers are also an option, but sometimes are not sufficient to reach to a concluding result [19]. In this context, bi-allelic markers such as SNPs have been suggested as markers of choice by many authors, as they are more stable markers with lower mutation rate than STRs.

There are two kinds of binary markers available with nearly identical characteristics: single nucleotide polymorphisms (SNPs) and insertion/deletion polymorphisms (indels). The indels have a typing system that uses directly a PCR amplification followed by capillary electrophoresis detection, thus reducing tube transfers to one, using the same methodology as STR kits with a direct relationship between signal strength and the DNA input. The only difference is that forensic SNP typing with SNaPshot primer extension methodology employs two reactions.

There have been developed various autosomal SNP multiplexes by some working groups, for different applications: ancestry determination [8], human identification [20], and phenotypic studies [21]. The SNPforID consortium ([www.snpforid.org](http://www.snpforid.org)) developed a SNP typing assay for human identification purposes that made it possible to amplify 52 SNPs from diminutive amounts of genomic DNA [13-22]. All SNP loci were amplified in one PCR reaction and the SNPs were analyzed by two single base extension (SBE) reactions and capillary electrophoresis. The SBE products were analyzed by Capillary electrophoresis.

The selection of the SNP loci in the SNPforID multiplex was made based on some main criteria and included SNPs with high polymorphism in most human populations (heterozygosity > 0.32), SNPs that were not in linkage disequilibrium, SNPs not associated with cellular functions or with STR loci commonly used in forensic genetics investigations and finally all SNPs could be amplified by PCR in one reaction [22].

There are some published scientific papers that present some simulation results concerning the use of autosomal SNPs as supplementary markers or as sole markers in kinship testing [9, 23-24]. However, there are very few reported concrete cases, with the calculated values of LR or Paternity Indexes (PI), if we have a paternity investigation, before and after the use of SNPs analyzed [19, 25, 26] and some of them studied another class of SNPs, the indels [27].

---

In this study, we sought to review and compare the results of employing a set of autosomal SNPs to complement the results obtained by laboratories that use routine batteries of STRs in paternity investigations where the final statistical result is ambiguous, and where the hypothesis of the alleged father may be, in fact, a close relative of the biological father.

## Material and methods

A revision of the literature concerning the use of the joined information obtained from STRs and SNPs analysis was made. Although it has been some years since the time when some SNPs multiplexes were developed for human identification purposes, as is the case of the 52-plex, there are few studies that report concrete applications of these markers to specific cases. The reported cases were resumed and compared. Some conclusions were reached after the resumed data analysis.

## Results

A review of some literature that presents a number of concrete examples of joining the information from STRs and SNPs is shown in Table 1.

In 47,3% (9 in 19) of the total number of analyzed cases, the PI value obtained only with the SNP markers was similar or higher than the obtained with STRs alone. Although, in all the analyzed cases the additional information obtained after the different SNP sets used by the different laboratories, was very useful and clearly illustrates that the addition of SNPs can help to reach a higher degree of confidence in the interpretation of many challenging relationship tests using extended STR and, in some cases MiniSTR or Y-STR typing. This is particularly important in parentage or identification cases that involve the analysis of DNA extracted from bones and teeth.

Special attention is given to cases that display one or two single-step mutations across generations on which the application of SNP typing, ignoring STR data, increased the PI by at least two magnitude orders. This is maybe due to the fact that the applied SNP typing multiplex, the *SNPforID-52plex*, provides a higher statistical power on European populations than the 15 human identity STRs contained on the employed amplification *kit* [22]. Moreover, it is highly unlikely that any parent-to-child mutation would be spotted since the mutation rate of SNPs is on average five magnitude orders lower than STRs [13]. Additionally, in all these cases, the combined STR/SNP increased PI results lead to a strongly supported hypothesis of paternity. These results were expected, since besides SNPs show extremely low mutation rates, these methodologies accomplish the multiplexing of a great number of markers. Moreover, these markers have been selected to cover all chromosomes ample distributed across the genome in order to ensure finding recombination-born differences even among members of complex inbred pedigrees [10]. Even with the new developed kits like the GlobalFiler kit (AB) this will tend to happen, because their markers were already applied in some of the reported cases [19].

Table 1. Resume of the casework reported in the reviewed literature where it was used the autosomal SNPs as supplementary markers to STR typing.

Reported Cases	Paternity Index		
	18STRs	43 SNPs	Combined
<b>Børsting C. et al. (2012)</b>	<b>18STRs</b>	<b>43 SNPs</b>	<b>Combined</b>
2 alleged fathers			
LR = P(E H1) / P(E H2) ÄH1: AF1 is the father. H2: AF2 is the father.	2.90E+03	2.80E+9	7.80E+12
<b>Philips C. et al. (2012)</b>	<b>21STRs</b>		<b>Combined</b>
LR = P(E H1) / P(E H2) ÄH1: The Uncle is the father. H2: Unrelated is the father.	4.63E+04	*	4.04+E07
LR = P(E H1) / P(E H2) ÄH1: The Deceased is the father. H2: Unrelated is the father.	1.12+E07	*	3.53E+08
Trio with two incompatibilities (CSF1PO + D19S433)	2.03E+05	*	1.89E+10
<b>Schwark T. et al. (2012)</b>	<b>17 STRs</b>	<b>50 SNPs</b>	
P.Trio 1 - no mutations	1.53 E+08	2.14E+07	*
P.Trio 2 - no mutations	7.32E+07	1.41E+05	*
P.Trio 3 - no mutations	5.13E+07	2.35E+06	*
P.Trio 4 - 2-step paternal mutation (D2S1338)	2.22 E+03	1.10 E+06	*
P.Trio 5 - 1-step paternal mutation (FGA)	8.71E+04	1.65E+08	*
P.Trio 6 - two 1-step paternal mutations (vWA+PentaD)	3.22E+01	6.32 E+08	*
<b>Ibarra A. et al. (2013)</b>	<b>STRs + MiniSTRs</b>	<b>52-plex</b>	<b>Combined</b>
Case 1 (U+C+M)	9.98E+02	1.02E+01	1.02E+04
Case 2 (B+C+M)	1.64E+03	9.53E+01	1.56E+04
Case 3 (Bone+C+M)	8.77E+01	1.22E+06	1.07E+08
Case 4 (Bone+C+M)	no results	1.11E+04	1.11E+04
<b>Lindner I. et al. (2014)</b>	<b>23STRs+16Y-STRs</b>	<b>50 SNPs</b>	
Putative father STRs: 2 mismatches Y-STRs: 1 mismatch SNPs: 0 mismatch	7.15E+06	6.32E+08	*
<b>Our complex cases</b>	<b>18 STRs</b>	<b>52-plex</b>	<b>Combined</b>
F1 (1Mut D2S1338)	2.00E+03	1.29E+02	2.84E+05
F2 (2Mut FGA and SE33)	2.96E+03	1.18E+01	3.50E+04
F3 (GF+2U+C+M)	1.19E+03	4.12E+00	4.92E+03
F4 (2S(M1)+C(M2))	2.41E+02	7.31E+00	1.76E+03
F5 (GM+C+M)	1.55E+02	9.43E+02	1.46E+06
F6 (B+C+M)	3.87E+03	1.76E+01	6.81E+04
F7 (1Mut D16S539)	6.13E+03	9.71E+05	5.96E+09

Footnote: GF- Alleged Paternal Grand-father; U- Alleged Paternal Uncle/Aunt; B- Alleged Brother; C- Child; M- Mother; S- Son of the alleged father; GM ó Alleged Paternal Grand-Mother; Mut ó Inconsistency in STR transmission.

\* no available data.

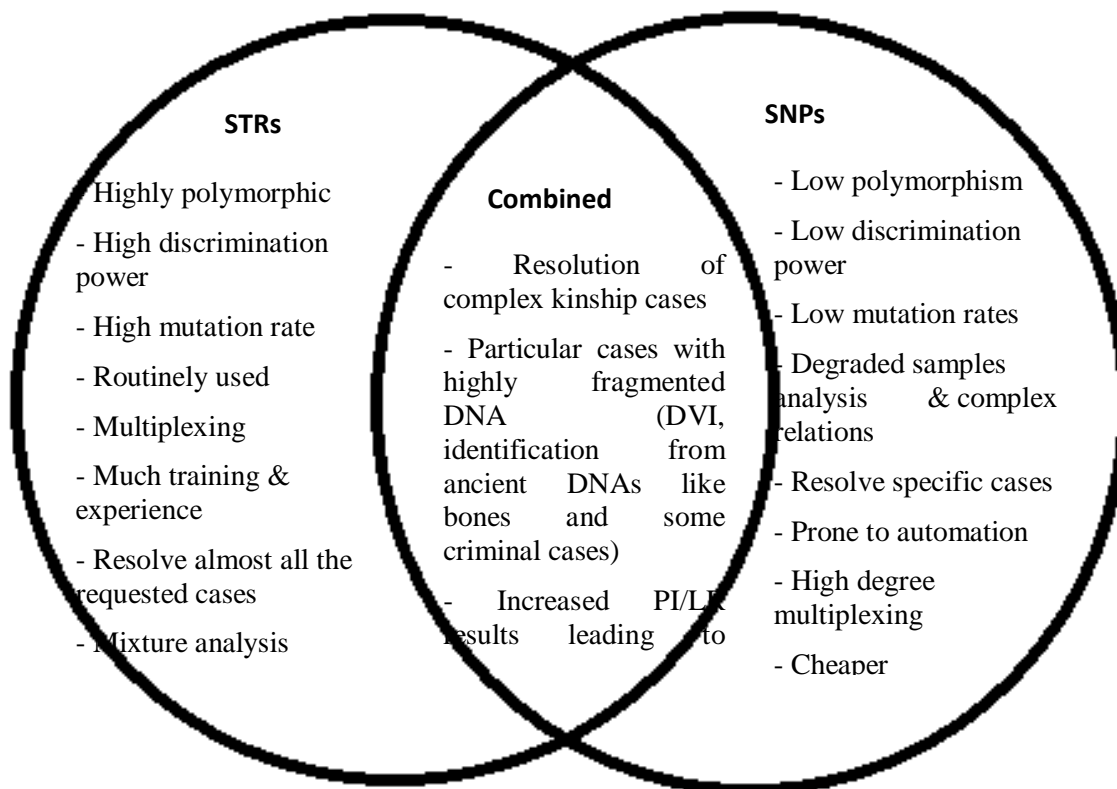


Figure 2. Summary of the features involving the analysis of STRs, SNPs or a combination of both.

---

## Discussion and conclusions

In forensic medicine, namely in forensic genetics is frequent to deal with parentage testing in the presence of close relatives of the true biological father. In this work one case where the brother of the true father was tested as the AF is reviewed. These complex paternity investigations have always represented an increased risk of reporting false inclusions of paternity. In one of the first reported studies by a Brazilian group [28], cases of paternity exclusions were analyzed with a set of 40 autosomal SNPs. In all cases except one, there were found additional inconsistencies in these bi-allelic markers, supporting the previous results. Only in a trio, 2 out of 20 autosomal STR were exclusionary, none of the *SNPforID* panel loci were inconsistent. As it can be observed by the resumed in the reported results (Table 1), many of these complex cases where one or two exclusionary markers are observed, are studied and the respective PI calculations can be incremented by SNP analysis. An experience with the *SNPforID* 52-plex has also revealed two false inclusions of paternity in simulated comparisons [24].

Another type of complex parentage testing are those cases where instead of the father that frequently has already deceased, their relatives are analyzed, as an alleged paternal uncle, a grand-father/mother or a full or half sibs. In these second type of cases, is frequent to reach low PI values after STR are analyzed. As can be seen by the six reported studies by various working groups resumed in Table 1, add the SNPs information represents a good approach to increase the previously obtained PI, as expected.

The last class of cases that often result in low LR or PI are those that involve the analysis of degraded materials which lead to incomplete STR profiles. These are very frequent in Identification of very decomposed or skeletonized human remains. There are also some of these cases reported in the literature, where SNPs are very useful as complement or sole information, depending on the source of the analyzed DNAs and on the degree of their degradation/contamination [9, 11].

Though it can be shown that, as the fact that these markers are less prone to mutation, turns them very useful in parentage testing. In spite of the reduced number of concrete analyzed and reported cases in the literature. This has also been predicted by simulation studies. As a result of this lower mutation rate it can also be seen that the information that results from SNPs alone, sometimes is lower than the obtained from STR analysis [24].

Other advantages of studying SNPs may include the fact that their analysis, in spite of being laborious, it is prone to automation, with the possibility of using pipetting platforms. Also the analysis of autosomal SNPs can be achieved by the application of a *SNaPshot* methodology, followed by capillary electrophoresis (CE) separation of the resulting fluorescent labeled fragments. These analysis methodologies are very easy to implement in most of the existing Forensic Genetic laboratories around the world almost of them equipped with automated sequencers. Furthermore, these strategies are also attractive due to the associated reduced

---

overall cost per analyzed sample.

In conclusion, despite some difficulties that can arise in some cases, especially if the information after the study of SNPs is still not enough, we believe that SNPs are very useful to complement STRs as markers of human identification and genetic relatedness. The number of SNP loci used for testing and the selected SNPs can be improved as more population data becomes available. For instance, some of the markers included in the developed autosomal 52 SNP-plex by the SNPforID Consortium, may not be independent. These loci can also usefully supplement the use of STR loci in forensic casework that sometimes leads to low LR values as summarized in figure 2. Furthermore, other markers have been used by several authors, including lineage markers that may be considered as more prone to the real mutation rate as it is the case of the Y chromosome associated markers [29-31]

Additionally as some authors are becoming to report, with some degree of investment, new approaches that will permit to analyze greater sets of these bi-allelic markers. This new methodology, the next generation sequencing (NGS), with the help of advanced computational devices for the analysis of the produced data, will permit other applications in Forensic Genetics like mitochondrial DNA analysis, that is used since the early 1990s, but that will be more fast and easy to integrate the results, and even in other Medico Legal areas like sudden death among other interesting uses [32,33].

## References

- [1] Kayser M, de Knijff P: Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet.* 2011;12(3):179-192.
- [2] Nelson MR, Marnellos G, Kammerer S, Hoyal CR, Shi MM, Cantor CR, Braun A. Large-Scale Validation of Single Nucleotide Polymorphisms in Gene Regions. *Genome Res.* 2004; 14(8): 1664-1668.
- [3] Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467:1061-1073.
- [4] Qin P, Li Z, Jin W, Lu D, Lou H, Shen J, Jin L, Shi Y, Xu S. A panel of ancestry informative markers to estimate and correct potential effects of population stratification in Han Chinese. *Eur J Hum Genet* 2013; 22:248-253.
- [5] Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* 2007; 80:727-739.
- [6] Shastri BS, SNPs in disease gene mapping, medicinal drug development and evolution. *J Hum Genet.*

---

2007;52(11):871-80. Epub 2007 Oct 11.

[7] Pulker H, Lareu MV, Phillips C, Carracedo A. Finding genes that underlie physical traits of forensic interest

using genetic tools. *Forensic Science International: Genetics* 2007; 1:1006104.

[8] Kidd JR, Friedlaender FR, Speed WC, Pakstis AJ, De La Vega FM, Kidd KK. Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. *Investigative Genetics* 2011; 2:1 (doi:10.1186/2041-2223-2-1).

[9] Børsting C, Mikkelsen M, Morling N: Kinship Analysis with Diallelic SNPs ó Experiences with the SNPforID Multiplex in an ISO17025 Accredited Laboratory. *Transfus Med Hemother* 2012;39:1956201.

[10] C. Phillips, M. Fondevila, M. Garcia-Magarinos, A. Rodriguez, A. Salas, et al.. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Sci. Int. Genet.* 2008; 2:1986204.

[11] Fondevila M, Phillips C, Naveran N, Fernandez L, Salas A, Carracedo A, Laréu MV: Case report: Identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. *Forensic Sci. Int. Genet.* 2008;2:2126218.

[12] Børsting C, Mogensen HS, Morling N: Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples. *Forensic Sci Int Genet.* 2013;7:345-352.

[13] Schneider PM, Beyond STRs: The Role of Diallelic Markers in Forensic Genetics. *Transfus Med Hemother.* 2012 Jun; 39(3): 1766180

[14] The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; 491:56-65.

[15] Choudhury A, Hazelhurst S, Meintjes A, Achinike-Oduaran O, Aron S, Gamiieldien J, Sefid Dashti MJ, Mulder N, Tiffin N, Ramsay M: Population-specific common SNPs reflect demographic histories and highlight regions of genomic plasticity with functional relevance. *BMC Genomics* 2014; 15:437 (doi:10.1186/1471-2164-15-437).

[16] Borsting C, Morling N. Mutations and/or close relatives? Six casework examples where 49 autosomal SNPs were used as supplementary markers. *Forensic Sci. Int. Genet.* 2011; 5:2366241.

[17] Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, et al.. Insertion/deletion polymorphisms: a multiplex assay and forensic applications. *Forensic Sci. Int. Genet. Suppl. Series* 2009; 2:5136515.

- 
- [18] Phillips C, Fondevila M, Garcia-Magarinos M, Rodriguez A, Salas A, et al.. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Sci. Int. Genet.* 2008; 2:1986204.
- [19] Lindner I, von Wurmb-Schwark N, Meier P, Fimmers R, Büttner Andreas. Usefulness of SNPs as Supplementary Markers in a Paternity Case with 3 Genetic Incompatibilities at Autosomal and Y Chromosomal Loci. *Transfus Med Hemother* 2014;41:1176121 (DOI: 10.1159/000357989).
- [20] Boonyarit H, Mahasirimongkol S, Chavalvechakul N, Aoki M, Amitani H, Hosono N, Kamatani N, Kubo M, Lertrit P. Development of a SNP set for human identification: A set with high powers of discrimination which yields high genetic information from naturally degraded DNA samples in the Thai population. *Forensic Sci Int Genet.* 2014;11:166-173.
- [21] Branicki W, Brudnik U, Kupiec T, Wolańska-Nowak P, Wojas-Pelc A: Determination of phenotype associated SNPs in the MC1R gene. *J Forensic Sci.* 2007 Mar;52(2):349-54.
- [22] Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevilla M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 2006; 27: 171361724.
- [23] Amorim A, Pereira L. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Sci Int.* 2005; 150(1): 17-21.
- [24] Børsting C, Sanchez JJ, Hansen HE, Hansen AJ, Bruun HQ, Morling N. Performance of the SNPforID 52 SNP-plex assay in paternity testing. *Forensic Sci. Int. Genet.* 2008;2:2926300.
- [25] Pinto N, Magalhães M, Conde-Sousa E, Gomes C, Pereira R, Alves C, Gusmão L, Amorim A. Assessing paternities with inconclusive STR results: The suitability of bi-allelic markers. *Forensic Science International: Genetics* 2013;7:16621.
- [26] Schwark T, Meyer P, Harder M, Modrow J-H, von Wurmb-Schwark N. The SNPforID Assay as a Supplementary Method in Kinship and Trace Analysis. *Transfus Med Hemother* 2012;39:1876193.
- [27] Magalhães M, Pinto N, Gomes C, Pereira R, Amorim A, Alves C, Gusmão L: When the alleged father is a close relative of the real father: The utility of insertion/deletion polymorphisms. *Forensic Science International: Genetics Supplement Series, Volume 3, Issue 1, December 2011, Pages e96e10, Progress in Forensic Genetics 14* ô *Proceedings of the 24th International ISFG Congress.*
- [28] Whittle MR, Favaro EC, Sumita DR, Paternity investigation experience with a 40 autosomal SNP panel. *Forensic Science International: Genetics Supplement Series 2* (2009) 1496150.

---

[29] Gomes, C; Magalhaes, M; Alves, C; Amorim, A; Pinto, N; Gusmao, L; Comparative evaluation of alternative batteries of genetic markers to complement autosomal STRs in kinship investigations: autosomal indels vs. X-chromosome STRs. *international journal of legal medicine*, volume: 126, issue: 6, pages: 5, published: 2012

[30] Sánchez-Diz P, Alves C, Carvalho E, Carvalho M, Espinheira R, García O, Pinheiro MF, Pontes L, Porto MJ, Santapa O, Silva C, Sumita D, Valente S, Whittle M, Yurrebaso I, Carracedo A, Amorim A, Gusmão L;GEP-ISFG (The Spanish and Portuguese Working Group of the International Society for Forensic Genetics): Population and segregation data on 17 Y-STRs: results of a GEP-ISFG collaborative study:*Int J Legal Med*. 2008;122(6):529-33.

[31] Carvalho R, Pinheiro MF, Medeiros R. Localization of candidate genes in a region of high frequency of microvariant alleles for prostate cancer susceptibility: the chromosome region Yp11.2 genetic variation: *DNA Cell Biol*. 2010;29(1):3-7

[32] Yang Y, Xie B, Yan J, Application of Next-generation Sequencing Technology in Forensic Science. *Genomics Proteomics Bioinformatics* 12 (2014) 190-197.

[33] Mingkun Li, Anna Schönberg, Michael Schaefer, Roland Schroeder, Ivane Nasidze, Mark Stoneking. Detecting Heteroplasmy from High-Throughput Sequencing of Complete Human Mitochondrial DNA Genomes. *The American Journal of Human Genetics* 2010; 87:237-249.

Tabela 1A . Resultados do Auto 1 para a população do norte de Portugal (n=131)

	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23																							
1	C	T	C	C	T	T					A	A	A	G	C	T	C	G	C	T	C	T	A	G	A	A						A	A			T	T	G	G	A	A	G	G	A	A	
2	C	T	C	C	T		T		A	G		G	G	C	T	C	G	C	T	C	T	G	G	G	A	A	T	C	T	G	T	A	G	T	T	C	T	G		A	A	G	G	G		
3	C	T	C	T	T	T	T	A	G	A	T	A	G	C	T	C	G	C	T	T	T	G	G	G	A	A	T	C	T	G	T	A	G	G	T	T	T	G	G	A	G	G	G	G	G	
4	C	T	C	C	A	A	T	T	A	G	T	T	G	G	C	C	C	C	C	T	C	T	A	A	G	A	A	A	C	T	G	G	A	A	T	T	C	T	G	G	A	A	A	G	A	G
5	C	T	C	C	T	T	A	T	A	G	A	T	G	G	C	T	C	G	C	T	C	T	A	G	G	A	A	T	C	T	G	G	A	G	T	T	C	T	G	G	A	G	G	G	G	G
6	C	T	C	C	T	T	A	T	A	G	A	T	A	G	C	T	C	G	C	T	T	T	A	A	A	A	A	A	C	T	G	T	A	G	T	T	C	T	G	G	A	G	G	G	G	G
7	C	T	C	T	T	T	A	T	A	G	A	T	G	G	C	T	C	G	C	T	T	T	A	A	G	G	A	A	C	T	G	G	A	A	T	T	T	T	G	G	A	G	G	G	G	G
8	C	T	C	C	T		T	T	A	G			G	C	T	C	G			C	T			G		A	T	T	T	G	T	A	G	T	T	T	T			A	G				A	
9	C	T	C	C	A	T	T	T	A	G	A	T	G	G	C	T	C	G	C	T	T	T	G	G	G	G	A	T	T	T	G	T	A	G	T	T	C	T	G	G	A	G	G	G	A	G
10	C	T	C	C	T	T	T	T	A	G	A	T	G	G	C	T	C	C	C	T	C	T		G	G	G	A	A	T	T	G	G	A	A	G	T	C	T	G	G	A	G	G	G	A	G
11	C	T	C	C	T	T	A	T	A	G	A	A	G	G	C	T	C	C	T	T	T	T	A	G	G	A	A	T	C	T	G	G	A	A	G	T	T	T	G	G	G	G	G	G	G	
12	C	T	C	C	T	T	T	T	A	G	A	T	G	G	C	T	C	C	C	T	T	T	A	G	G	A	A	T	T	T	G	T	A	G	G	T	T	T	G	G	A	G	G	G	G	G
13	C	T	C	C	T	T	A	T	A	G	A	T	G	G	C	T	C	C	C	T	C	T	A	G	G	A	A	T	C	T	G	T	A	G	T	T	C	T	G	G	A	G	G	G	G	G
14	C	T	C	T	T	T	A	T	A	G	A	T	G	G	C	T	C	G	C	T	C	T	A	G	G	A	A	A	C	T	G	T	A	G	T	T	C	T	G	G	A	G	G	G	G	G
15																																														
16	C	T	C	C	T	T	T	T	G	G	A	T	G	G	C	T	C	G	C	T	C	T	A	G	G	A	A	T	C	T	G	G	A	G	T	T	C	T	G	G	A	G	G	G	G	G
17	C	T	C	T	A	T	A	T	A	G	A	T	A	G	C	T	C	C	C	T	C	C	G	G	G	A	A	T	T	T	G	G	A	G	T	T	T	T	G	G	A	A	G	G	A	G
18	C	T	C	C	T	T	T	T	G	G	A	T	A	G	T	T	C	C	C	T	C	T	G	G	G	A	A	A	T	T	G	G	A	G	G	T	C	T	G	G	A	A	G	G	A	G
19	C	T	C	C	T	T	T	T	G	G	T	T	G	G	C	T	C	G	C	T	C	T	A	G	G	A	A	T	T	T	G	T	A	G	T	T	T	T	G	G	A	A	G	G	A	A

20	C	T	C	C	T	T	T	T	G	G	T	T	A	G	C	T	G	G	C	T	C	T	A	A	G	A	A	T	C	T	G	T	A	G	T	T	C	T	G	G	A	A	A	G	A	G
21	C	T	C	T	T	T	T	T	G	G	A	T	G	G	C	T	C	G	C	T	C	T	A	G	G	A	A	T	C	T	G	T	A	G	T	T	C	T	G	G	A	G	G	G	G	G
22	C	T	C	T	A	T	A	T	A	G	A	A	A	G	C	T	C	C	C	T	T	T	G	G	G	A	A	T	C	T	G	T	A	A	T	T	T	T	G	G	A	G	G	G	A	G
23	C	T	C	C	T	T	T	T	G	G	T	T	A	A	C	C	C	C	C	T	T	T	G	G	A	A	A	T	C	T	G	T	A	G	T	T	C	T	G	G	A	G	A	G	A	G
24	C	T	C	T	T	T	A	T	G	G	A	T	G	G	C	T	C	C	C	T	T	T	G	G	G	A	A	T	T	T	G	T	A	G	T	T	C	T	G	G	A	G	G	G	G	G
25	C	T	C	C	T	T	T	T	G	G	A	T	G	G	C	T	C	G	C	T	T	T	G	G	G	A	A	T	C	T	G	T	A	G	T	T	C	T	G	G	A	G	G	G	G	G
26	C	T	C	C	T	T	T	T	G	G	T	T	A	G	C	T	C	G	C	T	T	T	G	G	G	A	A	T	T	T	G	T	A	G	G	T	C	T	G	G	A	G	G	G	G	G
27	T	T	C	C	T	T	T	T	G	G	A	T	G	G	C	T	C	C	C	T	T	T	A	G	G	A	A	T	T	T	G	G	A	G	T	T	C	T	G	G	A	G	G	G	G	G
28	C	C	C	T	T	T	A	T	G	G	A	T	G	G	C	T	C	G	C	T	T	T	A	A	G	A	A	T	C	T	G	G	A	A	G	T	C	T	G	G	A	G	G	G	A	G
29	C	T	T	T	A	T	A	T	A	G	A	T	A	G	C	T	C	G	C	T	C	T	A	G	G	A	A	A	C	T	G	T	A	G	G	T	T	T	G	G	A	G	G	G	G	G
30	T	T	C	T	T	T	T	T	A	G	A	T	A	G	C	T	C	C	T	T	C	T	A	A	G	A	A	T	T	T	G	T	A	A	G	T	C	C	G	G	A	G	G	G	G	G
31	C	T	C	T	T	T	A	T	G	G	A	T	G	G	C	T	C	C	C	T	C	T	A	G	G	A	A	A	C	T	G	T	A	G	G	T	C	T	G	G	A	A	G	G	A	G
32	C	T	C	T	A	T	T	T	G	G	A	A	A	A	C	C	C	G	C	T	T	T	A	G	A	A	A	T	C	T	G	T	A	G	T	T	C	T	G	G	A	A	G	G	A	A
33	C	T	C	C	A	T	T	T	G	G	A	T	G	G	C	T	G	G	C	T	T	T	G	G	G	A	A	A	C	T	G	G	A	G	G	T	T	T	G	G	A	A	G	G	A	G
34	C	T	C	C	A	T	A	T	G	G	A	T	A	G	C	T	C	G	C	T	C	T	A	G	G	A	A	T	T	T	G	T	A	A	T	T	C	T	G	G	A	G	G	G	G	G
35	C	T	C	C	A	T	A	T	A	G	A	T	G	G	C	T	C	C	T	T	T	T	A	G	G	G	A	T	C	T	G	T	A	G	T	T	C	T	G	G	A	G	G	G	A	G
36	C	T	C	T	A	A	T	T	A	G	A	T	G	G	C	T	C	C	T	T	C	T	A	G	G	G	A	T	C	T	G	T	A	G	T	T	C	T	G	G	A	A	G	G	A	G
37	C	T	C	C	T	T	A	T	A	G	T	T	A	G	T	T	C	C	C	T	C	T	A	A	G	A	A	T	C	T	G	T	A	G	T	T	C	C	G	G	A	A	G	G	G	G
38	C	C	C	C			A	T	A	G	A	T	G	G	C	T	C	G	C	T	T	T	G	G	G	G	A	A	C	T	G	T	A	G	T	T	C	T	G	G	A	G	G	G	A	G
39	C	C	C	T	A	T	A	T	A	G	A	T	A	G	C	T	C	G	T	T	C	T	A	G	G	A	A	T	T	T	T	A	A	G	T	C	T	G	G	A	G	G	G	A	A	
40	C	T	C	C	T	T	A	T	A	G	A	T		A	C	T		C			T	T	G	G	G		A	A	C	T	G	T	A	G	T	T	C	C	G		A	A			A	

41	T	T	C	C	A	T	T	T	A	G	A	T	G	G	T	T	C	C	T	T	T	A	A	A	A	A	T	T	T	G	T	A	A	T	T	C	C	G	T	A	G	A	G	A	G	
42	C	C	C	T	A	A	A	T	A	G	A	T	A	G	C	T	C	C	C	T	T	T	A	A	A	A	A	T	C	T	G	T	A	G	T	T	C	T	G	G	A	A	G	G	G	G
43	C	T	C	C	A	A	A	T	A	G	A	A	A	G	C	T	C	C	C	C	T	T	G	G	G	A	A	T	T	T	G	T	A	G	G	T	C	C	G	G	A	G	G	G	G	G
44	C	C	C	T	A	T	T	T	A	G	A	A	G	G	C	T	C	C	C	T	T	T	G	G	G	G	A	A	T	T	G	T	A	G	T	T	T	T	G	T	A	G	G	G	G	G
45	C	C	C	T	A	T	T	T	A	G	A	T	G	G	C	T	C	C	T	T	T	T	G	G	G	A	A	T	T	T	G	T	A	A	T	T	C	T	G	G	A	A	G	G	A	G
46	C	C	C	T	T	T	T	T	A	G	A	A	G	G	C	T	C	C	C	T	T	T	G	G	G	G	A	T	T	T	G	T	A	G	T	T	C	T	T	T	A	A	A	G	A	G
47	C	C	C	C	T		T	T	A	A	T	T	G	G	C	T	C	C	C	T	C	T	A	G	G	G	A	T	C	T	G	T	A	A	G	G	C	T	G	T	A	A	A	A	A	G
48	T	T	T	T	A	A	T	T	A	A	A	T	G	G	C	C	C	G	C	T	C	T	A	G	G	A	A	T	C	C	T	T	A	A	G	T	T	T	G	T	A	G	G	G	A	G
49	C	T	C	T	A	A	A	T	A	A	A	T	G	G	C	C	C	C	C	T	C	T	A	A	G	G	A	T	T	T		A	A	T	T	C	T	G	G	A	A	A	G	A	G	
50	C	T	C	C	T	T	T	T	G	G	A	A	A	G	C	C	C	C	C	T	T	T	A	A	G	A	A	T	T	T	G	T	A	A	T	T	C	T	G	T	A	A	A	G	G	G
51	C	C	C	T	A	T	A	T	A	G	A	T	A	G	C	T	C	C	C	T	T	T	G	G	G	A	A	A	T	T	G	G	A	A	T	T	C	T	G	G	A	G	G	G	G	G
52	C	T	C	T	A	T	A	T	A	G	A	A	A	G	C	T	C	C	C	T	T	T	A	G	G	A	A	A	T	T	G	T	A	A	T	T	C	C	G	G	A	G	A	G	A	G
53	T	T	C	T	A		T	T	A	G	A	T	G	G	C	C	C	G	C	T	T	T	A	G	G	A	A	T	T	T	G	T	A	A	G	T	C	C	G	G	G	G	G	G	G	
54	C	C	C	C	A		T	T	A	A	A	T	G	G	T	T	C	C	C	C	C	T	A	A	A	A	A	T	T	T	G	T	A	A	T	T	C	T	G	G	A	G	G	G	G	G
55	C	C	C	C	T	T	A	T	A	A	A	T	G	G	C	T	C	G	C	C	C	C	A	A	G	A	A	A	T	T	G	T	A	A	G	G	C	T	G	G	A	G	G	G	A	G
56	C	C	C	C	T	T	T	T	A	G	A	T	A	G	C	T	C	G	C	T	C	T	A	A	G	G	A	T	T	T	G	T	A	G	G	T	C	T	G	G	A	G	G	G	G	G
57	C	C	C	C	T	T	T	T	A	G	T	T			C	T	C	C	C	T	T	T	A	A	G	A	A	A	T	T	G	T	A	A	T	T	C	T	G	G	A	G	G	G	A	A
58	C	T	C	T	T	T	A	T	G	G	A	T	G	G	C	T	C	G	T	T	T			G	A	A	T	T	T	G	G	A	A	T	T	C	T			A	G	G	G	G	G	
59	C	C	C	C	T	T	A	T	A	G	T	T			C	T	C	C	C	T	C	T	A	A	A	A	A	T	C	T	G	T	A	G	T	T	C	T	G	G			G	G	A	A
60	C	T	C	C	T	T	T	T	A	G	T	T			C	T	C	C	T	T	C	T	A	A	G	G	A	A	C	T	T	T	A	A	T	T	C	T	G	G	A	A	G	G		
61	C	T	C	C	A	T	T	T	A	A	T	T	A	G	C	T	C	C	T	T	C	T	A	A	G	A	A	T	C	T	G	G	A	G	T	T	C	C	G	G	A	G	G	G	G	G



83	C	T	C	C	A	T	T	T	A	G	A	A	A	G	C	T	C	C	C	T	T	T	A	G	G	G	A	A	T	T	G	T	A	G	T	T	C	T	G	G	A	G	A	G	G	G
84	C	T	C	C	A	A	A	T	A	G	A	T	G	G	C	T	C	C	C	T	T	T	A	A	G	G	A	T	T	T	G	T	A	G	G	T	T	T	G	G	A	A	G	G	G	G
85	C	T	C	T	T	T	A	A	A	G	T	T	A	G	C	C	C	G	C	T	T	T	G	G	G	A	A	A	T	T	G	T	A	G	T	T	T	T	G	T	A	G	G	G	A	G
86	C	T	C	C	A	T	T	T	A	G	A	T	A	G	C	T	C	G	C	T	C	T	A	A	G	A	A	T	T	T	G	T	A	G	T	T	C	T	G	G	A	A	G	G	A	G
87	C	T	C	C	T	T	T	T	A	G	T	T	G	G	C	T	C	G	C	T	C	T	A	A	G	A	A	T	T	T	G	T	A	G	T	T	C	T	G	G	A	A	G	G	A	A
88	C	T	C	T	A	A	A	A	A	G	T	T	A	G	C	T	C	C	C	C	T	T	A	G	A	A	A	T	C	T	G	T	A	A	T	T	C	T	G	G	A	G	G	G	A	A
89	T	T	C	T	A	A	T	T	G	G	T	T	G	G	C	T	C	C	C	T	T	T	A	A	G	A	A	T	T	T	G	T	A	G	T	T	C	C	G	G	A	A	G	G	G	G
90	C	T	C	C	A	T	A	A	A	G	A	T	G	G	C	T	C	G	C	C	C	T	A	A	G	A	A	A	T	T	G	T	A	G	G	T	C	T	G	G	A	G	G	G	G	G
91	C	T	C	C	A	A	A	A	G	G	T	T	A	G	C	C	C	C	C	T	C	T	A	A	A	A	A	T	C	T	G	G	A	G	T	T	T	T	G	G	A	A	G	G	A	A
92	T	T	C	T	A	A	A	T	A	G	A	T	A	G	C	T	C	C	T	T	T	T	A	A	G	A	A	A	T	T	G	T	A	G	T	T	T	T	G	G	A	A	G	G	A	G
93	C	T	C	C	A	A	A	A	A	G	A	T	G	G	C	C	C	C	C	C	T	T	A	G	A	A	A	A	C	T	G	T	A	G	T	T	T	T	G	G	A	A	G	G	A	G
94	C	T	C	C	T	T	A	A	G	G	A	T	A	G	C	C	C	C	C	T	T	T	A	A	G	A	A	T	T	T	G	T	A	G	T	T	T	T	G	G	A	A	G	G	G	G
95	T	T	C	C	A	A	A	A	A	A	A	T	G	G	C	T	C	C	C	C	C	T	A	A	G	G	A	T	T	T	T	T	A	G	G	T	T	T	G	G	A	A	G	G	A	G
96	T	T	C	T	A	T	A	A	A	G	T	T	G	G	T	T	C	C	C	C	C	T	A	G	A	A	A	T	T	T	G	T	A	G	T	T	T	T	G	G	A	G	A	G	A	G
97	T	T	C	C	A	T	A	A	G	G	A	T	G	G	C	T	C	C	T	T	T	T	A	A	G	A	A	T	C	T	G	T	A	G	T	T	C	T	G	G	A	G	G	G	A	G
98	T	T	C	T	A	T	A	A	A	G	T	T	G	G	C	T	C	C	T	T	T	T	A	G	G	A	A	T	T	T	G	T	A	G	T	T	T	T	G	G	A	G	G	G	A	G
99	T	T	C	C	A	A	A	T	A	G	A	A	G	G	C	T	C	C	C	T	C	T	A	A	G	G	A	A	T	T	G	T	G	G	T	T	C	T	T	T	A	A	G	G	G	G
100	T	T	C	C	A	A	A	T	A	G	A	A	A	G	T	T	C	C	C	T	T	T	A	A	G	A	A	T	C	T	G	G	A	G	T	T	C	T	G	T	A	A	G	G	A	G
101	T	T	C	T	A	A	T	T	A	G	T	T	G	G	C	T	C	C	C	T	T	T	A	A	G	A	A	T	T	T	G	T	A	G	T	T	C	T	G	G	A	G	G	G	A	G
102	T	T	C	T	A	A	A	T	A	G	A	T	G	G	C	C	C	C	C	T	C	T	A	G	G	A	A	T	T	T	G	T	A	A	T	T	C	C	G	G	A	A	A	G	A	G
103	T	T	C	T	A	T	A	A	A	G	A	T	G	G	C	T	C	C	C	C	C	T	A	A	G	A	A	T	T	T	G	T	A	G	T	T	C	T	G	G	A	A	G	G	A	A

104	C	T	C	C	A	T	A	A	A	G	T	T	A	G	C	T	C	G	T	T	C	T	A	G	G	A	A	A	C	T	G	T	A	A	T	T	T	T	G	G	A	G	A	A	G	G
105	T	T	C	T	A	T	A	A	G	G	A	A	A	G	C	T	C	G	C	T	C	T	A	G	G	A	A	T	T	T	G	G	A	G	T	T	T	T	G	G	A	A	G	G	A	A
106	C	T	C	T	A	A	A	A	G	G	A	A	G	G	T	T	C	G	C	T	T	T	A	G	A	A	A	A	T	T	G	T	A	A	T	T	C	T	G	T	A	G	G	G	A	A
107	T	T	C	T	A	A	A	A	A	G	A	A	A	G	C	T	C	C	C	T	C	T	A	G	G	A	A	A	T	T	G	T	A	G	T	T	T	T	G	G	A	G	G	G	A	G
108	T	T	C	C	A	A	A	A	G	G	A	T	G	G	T	T	C	C	C	T	C	T	A	G	A	A	A	A	C	T	G	T	A	G	T	T	C	T	G	G	A	A	G	G	G	G
109	C	C	C	C	T	T	T	T	G	G	A	T	G	G	C	C	C	C	C	T	T	T	A	A	G	G	A	T	T	T	G	T	A	G	T	T	C	C	T	T	G	G	G	G	G	G
110	C	C	C	T	T	T	T	T	G	G	A	T	G	G	C	T	C	G	C	T	C	T	A	G	G	A	A	A	C	T	G	T	A	A	T	T	C	T	G	G	A	G	G	G	A	G
111	C	T	C	C	A	A	A	T	A	G	A	T	G	G	C	T	C	C	T	T	T	T	G	G	A	A	A	A	T	T	G	T	A	G	T	T	C	T	G	T	A	G	G	G	G	G
112	C	C	C	C	A	A	A	A	A	G	T	T	G	G	C	T	G	G	C	T	T	T	G	G	A	A	A	A	T	T	G	T	A	G	T	T	C	C	G	G	G	G	A	A	G	G
113	C	C	C	C	A	A	A	A	A	G	A	A	A	A	C	T	C	C	C	T	C	T	A	A	A	A	A	T	T	T	G	T	A	G	T	T	C	C	T	T	G	G	A	A	A	A
114	C	T	T	T	A	A			A	G	T	T	A	A	C	T	G	G	C	C	T	T	G	G	A	A	A	T	T	T	G	G	G	G	T	T	C	C			A	A	A	A	A	A
115	C	T	C	C	T	T	T	T	A	G	T	T	A	G	C	T	C	C	T	T	T	T	A	A	A	A	A	A	T	T	G	G	A	G	T	T	C	C	T	T	A	G	G	G	G	G
116	C	C	C	C	T	T	A	A	A	G	T	T	G	G	C	C	C	C	C	T	T	T	G	G	A	A	A	T	T	T	G	T	A	G	T	T	C	T			A	A	G	G	G	G
117	C	C	C	T	T	T	T	T	A	G	T	T	A	G	C	T	C	C	T	T	T	T	G	G	A	A	A	A	T	T	G	T	A	A	T	T	C	C	T	T	A	G	G	G	G	G
118	C	T	C	C	T	T	A	T	G	G	A	T	G	G	T	T	C	C	C	T	T	T	A	A	A	A	A	T	T	T	G	T	A	G	T	T	C	C	T	T	G	G	G	G	A	G
119	C	C	C	C	T	T	T	T	A	G	T	T	G	G	C	C	C	C	C	T	T	T	G	G	A	A	A	T	C	T	G	T	A	A	T	T	C	T	G	G	A	G	G	G	G	G
120	C	T	C	C	T	T	T	T	G	G	T	T	A	G	C	C	C	C	C	C	T	T	G	G	G	A	A	T	T	T	T	T	A	A	T	T	C	C	G	T	A	G	G	G	G	G
121	C	C	C	C	T	T	A	T	A	G	A	T	A	A	C	T	C	C	C	T	T	T	A	A	A	A	A	T	T	T	G	T	A	A	G	T	T	T	G	T	G	G	G	G	G	G
122	C	T	C	C	A	T	T	T	G	G	T	T	G	G	C	T	C	C	C	C	C	T	G	G	G	G	A	T	C	T	T	T	A	G	T	T	T	T	G	G	A	A	G	G	A	A
123	C	T	C	C	A	A	A	T	A	G	A	A	G	G	C	T	C	C	C	T	C	T	A	G	A	A	A	T	T	T	G	G	A	G	T	T	C	T	G	G	G	G	G	G	G	G
124	C	T	C	C	A	A	A	T	A	G	A	T	G	G	C	C	C	C	C	C	C	T	G	G	A	A	A	T	T	T	G	T	A	A	T	T	T	T	G	G	A	A	A	A	A	A

125	C	T	C	C	T	T	T	T	A	G	T	T	G	G	C	T	C	C	C	T	T	T	A	A	G	A	A	T	T	T	G	T	A	G	T	T	C	T	T	T	A	G	G	G	A	A	
126	C	T	C	T	A	T	A	T	G	G	A	A	G	G	C	C	C	C	T	T	T	T	A	A	G	A	A	T	T	T	G	T	A	G	T	T	C	C	G	G	A	G	G	G	G	G	
127	C	T	C	C	T	T			A	A	A	T	G	G	C	C	C	C	C	T	T	T	A	A	A	A	A	T	C	T	G	T	A	A	T	T	C	T			G	G			G	G	
128																																															
129	C	T	C	C	T	T	T	T	G	G	T	T	G	G	C	C	G	G	C	T	T	T	A	A	G	G	A	A	C	T	G	T	A	A	T	T	T	T	G	G	A	G	G	G	A	G	
130	C	T	C	C	A	T	A	A	A	G	A	T	G	G	C	C	C	G	C	T	T	T	G	G	G	A	A	A	C	T	G	T	A	G	G	T	C	C	G	G	A	G	G	G	A	G	
131	C	C	C	C	T	T	A	T	A	G	T	T	G	G	T	T	C	C	T	T	C	T	G	G	G	G	A	T	T	T	G	T	A	G	T	T	C	C	G	G	A	A	G	G	G	G	

Tabela 1B . Resultados do Auto 2 para a população do norte de Portugal (n=131)

	A24	A25	A26	A27	A28	A29	A30	A32	A33	A34	A35	A36	A37	A38	A39	A40	A41	A42	A43	A44	A45	A46	A48	A49	A50	A51	A52	A53	A54																														
1	C	T	A	G	G	G	A	G	A	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	T	C	C	A	T	G	A	C	T	A	A	G	A	T	T	C	T	G	A	T	T	G	A	G	A	G	A	A	T	A	A	G	A	
2	C	T	A	G	G	T	A	A	G	G	A	G	A	C	C	T	C	T	G	C	G	A	A	T	C	C	C	C	A	T	G	G	C	C	A	A	G	A	C	T	C	T	A	A	T	T	G	G	G	G	G	A	A	A	A	A	A	G	G
3	C	T	A	A	G	G	A	G	G	G	A	G	A	C	C	T	C	T	G	C	G	A	A	T	C	C	C	C	A	T	A	A	C	T	A	A	G	A	C	T	C	T	A	A	C	T	G	G	G	A	G	A	A	A	A	A	G	G	
4	C	T	G	G	G	G	A	A	G	G	A	G	C	C	C	T	C	T	G	C	G	A	A	T	C	T	C	C	A	T	A	A	T	T	A	A	G	A	C	T	C	T	A	A	C	T	G	A	G	A	G	A	A	T	A	A	G	G	
5	C	T	A	G	G	T	A	G	G	G	A	G	A	C	C	T	T	T	G	C	G	G	A	T	C	C	C	C	A	T	G	A	C	T	A	A	G	A	C	T	C	T	A	A	T	T	G	G	G	A	G	A	A	T	A	A	G	G	
6	C	T	A	G	G	T	A	A	G	G	A	G	A	C	C	C	C	T	G	C	G	G	A	T	C	C	C	C	A	T	G	A	C	T	A	A	A	A	C	T	C	T	G	A	T	T	G	A	G	A	G	A	A	A	A	A	G	G	
7	C	T	A	G	T	T	A	G	A	G	A	G	C	C	C	T	T	T	G	C	G	G	A	T	C	T	C	C	A	T	G	A	C	T	G	A	A	A	C	T	C	T	G	A	C	T	G	A	G	A	G	A	A	T	A	A	G	G	
8	C	T	A	A	T	T	A	G	G	G	A	G	A	C	C	T	C	T	G	C	G	A	A	T	C	T	C	C	A	T	G	G	C	C	A	A	G	A	C	T	C	T	A	A	T	T	G	A	G	G	A	A	A	A	A	A	A	A	A
9	C	T	A	G			A	G	A	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	C	C	C	A	T	G	A	C	T	G	A	G	A	C	T	C	T	A	A	C	T	G	A	G	A	G	A	A	A	A	A	A	G	G
10	C	T	A	G					G	G	A	G	A	C	C	T	T	T	C	C	A	A	A	T	C	C	C	C	A	T	A	A	C	T	A	A	G	A	C	T	C	T			C	T			G	A	G	A	A	A	A	A			
11	C	T	G	G	G	T	A	G	A	G	G	G	A	A	C	C	C	T	G	C	G	A	T	T	C	T	C	C	A	T	G	A	C	C	G	G	A	A	C	C	C	C	G	A	T	T	G	A	G	A	G	A	A	T	A	A	A	A	
12	C	T	A	A	G	T	A	A	G	G	A	A	A	C	C	T	T	T	C	C	G	A	A	A	C	T	C	C	T	T	G	A	C	T	G	A	G	A	T	T	C	C	G	G	T	T	A	A	G	G	A	A	A	T	A	A	A	A	
13	C	T	A	G	T	T	A	G	A	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	T	G	C	A	T	G	A	C	T	G	A	G	A	C	T	C	T	G	A	T	T	G	G	G	A	G	A	A	T	A	A	G	G	
14	C	T	A	G	G	T	A	G	G	G	A	G	A	C	C	T	C	T	G	C	G	A	A	T	C	T	G	C	A	T	A	A	C	T	G	A	G	A	C	T	C	T	A	A	C	T	G	G	G	A	G	A	A	T	G	A	G	G	
15	C	T	A	G	T	T	A	A	G	G	A	G	C	C	C	T	C	T	C	C	G	G	A	T	C	T	C	C	A	T	G	G	C	C	A	A	A	A	C	T	C	T	A	A	C	T	G	G	A		A	A	A	A	A	A	A		
16	C	T	A	G	G	T	A	G	A	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	T	G	C	A	T	G	A	C	T	G	A	G	A	C	T	C	T	G	A	T	T	G	G	G	A	G	A	A	T	A	A	G	G	
17	C	T	G	G	G	G	A	A	G	G	A	G	A	C	C	T	T	T	C	C	G	A	T	A	C	T	C	C	A	A	A	A	T	T	A	A	G	A	C	C	C	C	G	A	C	T	G	G	A	A	G	A	A	T	A	A	G	A	
18	C	T	A	G	G	G	A	G	G	G	G	A	C	C	C	T	T	T	C	C	G	A	A	T	C	T	G	C	A	A	G	G	C	C			G	A	C	T	C	T	G	G	C	T	G	A	G	A	G	A	A	A	G	A	G	A	
19	C	T	A	G	G	T	A	G	A	G	A	G	A	C	C	T	T	T	G	C	G	G	A	T	C	T	G	C	A	T	G	A	C	T	G	A	G	A	C	T	C	T	G	A	T	T	G	G	G	A	G	A	A	T	A	A	G	G	
20	C	T	A	G	G	T	A	G	A	G	A	G	A	C	C	T	C	T	G	C	G	A	A	T	C	T	G	C	A	T	G	A	C	T	A	A	G	A	C	T	C	T	G	A	T	T	G	A	G	A	G	A	A	T	A	A	G	G	
21	C	T	G	G	G	T	G	G	G	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	C	G	C	A	T	G	A	C	T	G	A	G	A	C	T	C	T	G	A	C	T	G	A	G	A	G	A	A	T	A	A	G	G	
22	C	T	A	G	G	T	A	G	G	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	T	G	C	A	T	G	A	C	T	G	A	G	A	C	C	C	T	G	A	T	T	G	G	G	A	G	A	A	T	G	A	G	G	
23	C	T	A	G	G	T	A	G	A	G	A	G	A	C	C	T	C	T	G	C	G	A	A	T	C	T	G	C	A	T	A	A	C	T	A	A	G	A	C	T	C	T	G	A	C	T	G	A	G	A	G	A	A	T	A	A	G	G	
24	C	T	G	G	G	T	A	G	G	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	T	G	C	A	T	G	A	C	T	G	A	G	A	C	C	C	T	G	A	C	T	G	A	G	A	G	A	A	T	G	A	G	G	
25	C	T	A	G	G		A	G			A	G			C	T	T	T	G	C			A	T	C	T	G	C			G	A	C	T	G	A			C	T	C	T	A	A	C	T	G	G			G	A	A	T	A	A	G	G	

26	C	T	A	G	G	T	A	G	G	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	T	G	C	T	T	G	A	C	T	A	A	G	G	C	T	C	T	G	A	C	T	G	G	G	A	G	A	A	T	A	A	G	G	
27	C	T	A	G	G	T	A	G	A	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	T	G	C	A	T	G	A	C	T	G	A	G	A	C	C	C	T	G	A	T	T	G	G	G	A	G	A	A	T	G	A	G	G	
28	T	T	A	A	G	T	A	A	A	G	A	G	A	A	C	C	T	T	G	C	G	A	A	A	C	T	C	C	A	A	G	A	C	T	G	G	G	A	C	T	C	C	G	G	T	T	G	G	A	A	G	A	A	T	A	A	A	A	
29	C	T	A	G	G	T	A	G	A	G	A	G	A	C	C	T	C	T	G	C	G	A	T	T	C	T	C	C	T	T	G	A	T	T	A	A	G	A	C	T	C	T	G	A	C	C	G	G	A	A	G	G	A	T	G	A	A	A	
30	T	T	G	G	G	T	A	G	G	G	A	A	A	A	C	T	T	T	G	C	G	G	A	T	C	T	G	G	T	T	A	A	C	C	A	A	G	A	C	T	C	C	G	A	C	T	G	A	G	G	A	A	A	A	A	A	G	G	
31	C	T	A	G	G	T	A	G	G	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	T	G	C	A	T	G	A	C	T	A	A	G	A	C	T	C	T	G	A	C	T	G	G	G	A	G	A	A	T	G	A	G	A	
32	C	T	A	G	G	T	A	G	G	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	T	G	C	A	T	G	A	C	C	A	A	G	A	C	T	C	T	G	A	C	T	G	A	G	A	G	A	A	T	G	A	G	G	
33	C		A	G	G	T	A	G	G	G	A	G	C	C	C	T	C	T	G	C	G	G	A	T	C	C	G	C	A	T	A	A	C	T	A	A	G	A	C	T	C	T	G	A	C	T	G	A	G	A	G	A	A	T	G	A	G	G	
34	C	T	G	G			A	G			A	G			C	T	C	T	G	C			A	T	C	C	G	C				C	T	G	A			T	T			G	A	T	T	G	G			G	A	A	A	G	A	A	A		
35	C	T	A	G	T	T	A	G	A	G	A	G	A	C	C	T	C	T	G	C	G	A	A	T	C	T	C	C	A	T	G	A	T	T	G	A	G	A	C	T	C	T	G	A	C	T	G	A	G	A	G	A	A	T	G	A	A	A	
36	C	T	A	G	G	T	A	G	G	G	A	A	A	C	C	T	T	T	G	G	G	G	A	T	C	T	C	C	A	T	G	A	C	C	A	A	G	A	C	T	C	T	G	A	T	T	G	G	G	A	G	A	A	A	A	A	G	G	
37	C	T	A	A	T	T	A	A	A	A	G	G	A	C	C	C	T	T	G	C	A	A	T	T	T	C	C	A	T	A	A	C	T	A	A	G	A	T	T	C	T	G	G	T	T			A	A	G	G	A	A	G	A	G	G		
38	T	T	G	G	T	T	A	G	G	G	G	G	C	C	C	T	C	T	G	G	G	A	A	T	C	T	C	C	A	T	G	A	C	T	G	G	G	A	T	T	C	T	G	G	C	T	G	G	A	A	G	G	A	A	G	A	G	A	
39	C	T	G	G	T	T	A	A	G	G	A	G	C	C	C	T	T	T	G	G	A	A	A	T	T	T	C	C	T	T	G	A	C	T	A	A	A	A	C	C	C	T	G	G	C	T	G	A	G	G	G	G	A	A	G	G	G	G	
40	T	T		A	T		A	G	G	G	A	G	C	C	C	T	T	T	G	C			A	T	T	T	G	C	A	T	A		C	C	A	A	A	A	C	T	C	T	A		T	T				G	G	A	T	A	A				
41	C	T	A	G	G	T	A	G	A	G	A	G	A	C	C	T	T	T	G	C	G	A	A	T	C	T	C	C	A	T	G	A	C	T	G	A	G	A	C	C	C	T	G	G	T	T	G	A	G	A	A	A	A	T	G	A	A	A	
42	C	T	G	G	G	T	A	G	G	G	A	G	A	C	C	T	T	T	G	G	G	A	A	T	C	T	C	C	A	T	G	A	C	T	G	A	G	A	C	C	C	T	G	G	C	T	G	G	G	A	G	A	A	A	A	A	A	A	
43	C	T	A	A	G	T	A	G	A	G	A	A	A	A	C	C	C	T	G	G	A	A	A	T	C	T	G	C	T	T	A	A	C	C	G	A	A	A	C	C	T	T	G	G	C	T	G	G	G	A	G	A	A	A	A	A	A	A	
44	C	T	A	G	T	T	A	A	G	G	G	G	A	A	C	C	T	T	T	G	C	A	A	A	T	C	T	G	C	T	T	A	A	C	C	G	A	G	G	C	C	C	C	G	G	C	C	G	G	A	A	A	A	A	T	A	A	A	A
45	C	T	G	G	G	T	A	A	A	G	G	A	C	C	T	T	T	T	G	G	A	A	A	A	C	T	G	C	A	A	A	A	T	T	G	A	A	A	C	T	T	T	G	G	C	T	G	A	A	A	G	G	T	T	A	A	A	A	
46	C	T	G	G	T	T	G	G	A	G	A	A	A	C	C	T	T	T	T	G	G	A	A	A	T	C	C	G	C	A	T	G	A	C	T	A	A	A	A	C	C	C	T	G	G	T	T	G	G	A	A	G	G	A	T	A	A	A	A
47	T	T	A	A	G	G	A	G	G	G	A	G	C	C	C	T	C	T	G	G	G	A	A	A	C	T	C	C	A	T	A	A	C	C	A	A	G	A	T	T	T	T	G	A	T	T	G	G	G	G	A	A	A	T	A	A	G	A	
48	C	T	A	G	G	T	A	A	A	G	A	G	C	C	C	C	T	T	T	G	G	G	A	A	A	C	T	G	C	A	T	G	A	C	C	G	G	G	A	C	T	C	C	A	A	C	C	G	G	G	G	A	A	A	T	G	A	G	A
49	T	T	A	G	T	T	A	G	A	G	A	G	A	A	C	C	C	T	G	G	G	A	A	A	C	C	C	C	A	T	A	A	C	C	G	G	G	A	T	T	C	T	G	A	C	T	G	A	G	A	G	G	T	T	G	A	G	A	
50	C	T	A	G	T	T	A	G	G	G	A	G	C	C	C	T	T	T	T	G	C	A	A	T	T	C	C	C	C	T	T	G	A	T	T	A	A	A	A	C	C	T	T	G	A	T	T	G	G	A	A	G	G	A	T	G	A	G	A
51	C	T	G	G	G	T	A	A	G	G	A	A	C	C	C	T	C	C	G	G	A	A	T	T	C	C	C	C	A	A	G	G	C	C	A	A	G	A	T	T	C	T	G	G	T	T	A	A	G	A	A	A	A	A	A	A	A	G	A
52	C	T	A	A	T	T	A	A	G	G	A	A	A	C	C	C	T	T	T	G	G	G	A	A	T	C	T	C	C	A	T	G	A	C	T	G	A	G	A	C	T	C	T	G	A	T	T	G	G	G	A	G	A	A	T	G	A	G	A

53	C	T	A	G	G	T	A	G	G	G	A	G	A	A	T	T	T	T	G	G	G	A	A	T	T	T	C	C	T	T	G	A	C	C	G	A	G	A	C	T	T	T	G	A	C	T	G	G	G	A	G	A	A	T	A	A	G	A
54	C	T	G	G	G	T	A	A	G	G	A	A	A	C	C	C	T	T	G	G	G	A	A	A	C	T	C	C	A	A	G	A	C	T	A	A	G	A	C	C	C	C	A	A	T	T	A	A	G	A	G	A	A	A	G	A	G	A
55	C	C	G	G	T	T	A	A	A	G	A	G	A	A	C	T	T	T	G	C	A	A	T	T	C	C	G	C	A	T	G	A	C	T	G	G	G	A	C	C	C	C	G	A	C	C	G	G	A	A	G	A	A	T	A	A	G	A
56	T	T	A	G	G	G	A	G	A	A	A	A	A	C	T	T	T	T	G	G	A	A	T	T	C	T	C	C	A	A	G	A	C	T	G	G	G	A	C	T	C	T	G	A	C	T	A	A	A	A	G	A	A	T	A	A	G	A
57	C	T	G	G	G	T	A	A	A	G	A	A	A	A	C	T	C	T	G	G	A	A	A	T	C	C	C	C	A	T	G	A	C	T	G	A	G	A	C	T	C	T	G	G	T	T	G	A	A	A	G	G	A	A	A	A	G	G
58	C	T	A	G	G	G	A	A	A	G	A	A	A	A	C	T	C	T	G	C	G	A	A	T	C	T	C	C	A	T	G	A	C	T	A	A	G	A	C	C	C	T	G	A	T	T	G	G	G	A	G	A	A	T	A	A	G	G
59	C	T	G	G	G	T	A	A	A	G	A	G	C	C	C	C	T	T	G	G	A	A	A	T	C	T	C	C	T	T	A	A	C	C	G	A	G	A	C	T	T	T	G	G	C	T	A	A	G	A	G	A	A	T	G	A	G	A
60	C	T	A	G	G	T	A	A	A	A	A	G	A	A	C	C	T	T	G	C	A	A	A	A	C	T	C	C	T	T	G	A	C	C	A	A	G	A	C	T	C	T	G	G	C	T	G	A	G	A	G	A	A	A	G	A	G	A
61	C	T	A	A	G	T	A	A	G	G	A	G	A	C	T	T	T	T	G	G	A	A	A	A	C	T	G	C	A	T	A	A	T	T	G	G	G	G	C	C	C	T	G	G	T	T	G	A	A	A	G	A	A	A	G	A	G	G
62	T	T	A	A	G	T	A	G	A	G	A	G	C	C	T	T	C	T	G	G	G	G	A	T	C	C	C	C	T	T	G	G	C	C	A	A	A	A	C	C	C	T	G	G	C	T	G	G	G	A	G	G	A	T	G	A	G	
63	T	T	A	G	T	T	A	G	G	G	A	A	A	A			C	T			G	G	T	T	C	C			A	T	G	G	C	C			G	A	C	T	C	C	G	A	T	T	A	A	G	A	G	G	T	T	A	A		
64	T	T	G	G	G	G	A	A	A	G	G	G	A	C	C	T	T	T	G	G	A	A	T	T	C	C	G	G	A	T	G	A	C	C	G	G	G	G	T	T	C	T	A	A	C	T	A	A	G	A	A	A	A	A	A	A	A	A
65	T	T	A	A	T	T	A	A	A	G	A	A	A	A	C	C	C	T	G	C	A	A	T	T	C	C	G	C	A	A	G	A	C	T	G	A	G	A	C	T	C	C	A	A	C	T	A	A	A	A	G	G	T	T	G	G	G	
66	C	T	G	G	T	T	A	G	G	G	A	G	A	C	C	T	T	T	G	C	G	A	T	T	C	C	C	C	A	A	A	A	C	T	A	A	G	A	C	T	C	C	G	G	C	T	G	G	G	G	G	A	T	A	A	G	A	
67	T	T	A	G	T	T	A	A	G	G	G	G	A	A	C	T	T	T	G	G	A	A	T	T	C	T	C	C	A	T	G	A	C	C	A	A	G	A	C	C	T	T	G	G	T	T	G	G	G	A	G	G	A	T	A	A	G	A
68	T	T	G	G	T	T	A	A	G	G	G	G	A	A	C	T	C	T	G	G	G	A	A	T	C	T	G	C	A	T	A	A	T	T	A	A	A	A	T	T	T	T	G	A	T	T	G	G			G	A	T	T	A	A	G	A
69	T	T	A	G	T	T	A	A	A	G	G	G	A	C	C	C	T	T	G	C	A	A	A	A	C	C	C	C	A	T	G	A	C	T	G	A	G	A	C	C	T	T	A	A	T	T	G	G	G	G	G	A	A	A	A	A	A	
70	C	T	G	G	G	G	A	A	A	A	A	G	A	C	C	C	C	T	G	C	A	A	T	T	T	C	C	A	T	G	G	C	T	G	A	G	A	C	C	C	T	A	A	T	T	A	A	A	A	G	G	A	T	G	G	G	A	
71	T	T	A	A	G	T	A	A	G	G	A	G	A	A	C	C	T	T	G	G	A	A	A	A	C	T	C	C	A	T	A	A	C	T	G	A	G	A	C	T	C	C	G	A	T	T	G	G	G	G	G	A	A	A	A	G	A	
72	C	T	G	G	T	T	A	A	A	G	A	G	A	A	C	T	T	T	G	G	A	A	T	T	T	C	C	A	T	A	A	C	C	A	A	G	G	C	T	C	T	G	A	T	T	G	G	G	G	G	A	A	A	G	A	A	A	
73	C	T	A	G	G	T	A	G	G	G	A	G	A	C	C	T	C	T	G	G	G	A	A	T	C	T	C	C	A	T	A	A	C	T	A	A	A	A	C	T	C	T	G	A	T	T	G	G	G	A	G	A	T	T	A	A	G	G
74	C	C	A	G	G	T	A	G	A	G	A	G	A	C	C	C	C	T	G	C	G	A	A	T	T	T	C	C	A	T	G	A	C	C	A	A	A	A	C	C	T	T	G	A	T	T	G	G	G	A	G	A	T	T	A	A	G	G
75	C	T	G	G	T	T	A	A	G	G	A	G	A	C	C	T	C	T	G	C	A	A	A	T	C	C	C	C	A	T	A	A	C	T	A	A	G	A	C	C	C	T	G	A	T	T	G	G	G	A	G	A	A	A	G	A	A	A
76	T	T	A	A	G	T	G	G	G	G	G	G	C	C	C	T	T	T	G	G	G	A	A	T	C	T	C	C	A	T	A	A	C	T	A	A	G	A	C	C	C	C	G	A	C	T	G	G	A	A	G	G	A	T	G	A	G	G
77	T	T	G	G	G	T	A	G	A	G	A	G	A	A	C	T	T	T	G	G	A	A	T	T	C	T	G	C	A	A	G	A	C	T	A	A	G	A	C	T	C	C	G	A	T	T	A	A	G	G	G	A	T	T	A	A	G	A
78	T	T	A	A	G	T	A	A	G	G	A	A	A	C	C	C	T	T	G	G	G	A	A	T	T	T	C	C	A	T	A	A	C	T	G	A	G	A	T	T	C	T	G	A	T	T	G	G	G	A	G	A	A	A	A	A	G	G
79	C	T	A	A	T	T	A	A	G	G	A	G	A	A	C	C	C	T	G	G	A	A	T	T	C	C	C	C	A	T	A	A	C	C	A	A	G	A	C	T	C	T	G	A	C	T	G	G	G	A	G	A	A	A	G	A	A	A

80	T	T	A	G	G	G	A	G	A	G	G	G	A	A	C	C	C	T	G	G	G	G	A	T	T	T	C	C	A	A	A	A	C	T	G	A	G	A	C	C	C	T	G	A	T	T	G	G	G	G	A	A	A	G	A	G	G		
81	T	T	A	G	T	T	G	G	A	G	A	G	A	A	C	C	C	T	C	C	G	A	A	A	C	T	G	C	A	A	A	A	C	T	G	G	G	A	C	T	C	C	G	A	T	T	A	A	G	A	G	A	A	A	G	A	G	A	
82	T	T	A	A	T	T	A	A	G	G	A	G	A	C	C	T	T	T	G	C	A	A	A	T	T	T	C	C	A	T	A	A	T	T	A	A	G	A	C	T	C	T	A	A	T	T	G	G	A	A	A	A	A	A	A	A	A	A	A
83	T	T	G	G	G	T	A	G	A	G	A	G	A	C	C	C	T	T	G	G	G	A	A	T	T	T	C	C	A	T	A	A	C	T	A	A	G	A	C	C	C	T	G	A	T	T	G	G	A	A	G	A	A	T	G	A	G	G	
84	T	T	A	A	G	G	A	G	A	A	G	A	C	C	C	C	T	G	G	G	A	T	T	C	T	C	C	T	T	G	A	C	C	G	G	A	A	C	T	C	C	G	G	C	T	A	A	A	A	G	A	T	T	A	A	G	G		
85	C	T	A	G	T	T	A	G	G	G	A	G	A	C	C	T	T	T	G	G	G	A	A	T	C	T	C	C	A	A	A	A	C	T	G	A	G	A	C	T	C	T	G	A	T	T	A	A	G	G	G	G	A	A	A	A	G	A	
86	C	T	A	G	G	T	A	A	G	G	A	G	A	C	C	T	C	T	G	G	G	A	A	T	C	T	C	C	A	T	G	A	C	T	G	A	G	A	C	T	C	T	G	G	T	T	G	G	G	A	G	A	A	T	G	A	G	A	
87	C	T	A	G	G	T	A	A	G	G	A	G	A	C	C	T	C	T	G	G	G	A	A	T	C	T	C	C	A	T	A	A	C	T	A	A	G	A	C	T	C	T	G	G	T	T	G	G	G	A	G	A	A	T	G	A	G	A	
88	C	C	G	G	T	T	A	A	A	A	A	G	A	C	C	C	C	T	G	C	G	G	A	T	C	T	C	C	A	T	G	A	T	T	G	A	G	A	T	T	T	T	A	A	C	C	G	G	G	A	G	A	A	A	G	A	G	G	
89	C	T	A	A	G	T	A	G	A	G	G	G	A	C	T	T	T	T	G	C	G	A	A	T	T	T	C	C	A	T	A	A	C	T	A	A	A	A	C	T	C	T	G	A	T	T	G	G	A	A	G	A	A	T	A	A	G	G	
90	C	T	G	G	T	T	G	G	G	G	A	G	A	A	C	C	C	T	G	C	A	A	A	T	C	C	C	C	A	T	A	A	C	T	G	G	G	A	C	T	C	T	G	A	C	T	G	G	G	A	G	A	A	T	A	A	G	A	
91	T	T	G	G	G	T	A	G	A	G	G	G	A	A	T	T	T	T	G	G	G	A	T	T	C	C	C	C	A	A	G	A	C	T	A	A	A	A	C	C	C	C	G	A	C	T	G	G	A	A	G	A	A	T	G	A	G	A	
92	C	C	G	G	G	T	A	G	A	G	A	A	A	C	C	C	C	T	G	C	G	A	A	T	C	T	C	C	A	T	A	A	C	T	A	A	G	A	C	T	C	T	G	A	T	T	G	G	G	A	G	A	A	T	G	A	G	A	
93	C	C	G	G	T	T	A	A	G	G	A	G	A	C	C	C	C	T	G	C	G	A	A	T	C	T	C	C	T	T	G	A	C	T	A	A	G	A	T	T	C	T	G	A	C	T	G	G	G	A	G	A	T	T	G	A	A	A	
94																																																											
95	C	C	A	A	T	T	A	G	G	G	A	A	A	C	C	T	C	C	G	C	A	A	A	T	C	C	C	C	A	T	A	A	C	C	A	A	G	A	C	C	C	C	G	A	T	T	A	A	A	A	G	G	A	T	G	A	A	A	
96	T	T	G	G	T	T	A	A	A	G	G	G	C	C	C	T	T	T	G	C	A	A	T	T	C	C	C	C	A	T	G	A	C	T	A	A	A	A	C	T	C	T	G	A	T	T	G	G	A	A	G	G	A	T	G	A	G	A	
97	T	T	A	A	G	T	G	G	A	G	A	A	C	C	C	C	T	T	G	C	A	A	A	T	T	T	C	C	A	A	G	G	C	T	A	A	G	A	C	C	C	C	G	A	T	T	G	G	G	G	G	A	T	A	A	A	A		
98	C	T	A	A	T	T	A	G	A	G	A	G	A	A	C	C	T	T	G	C	G	A	A	T	C	T	C	C	A	T	G	A	C	T	G	G	G	A	C	T	C	C	G	G	T	T	G	G	G	A	G	G	A	T	A	A	G	A	
99	C	C	A	G	G	G	A	G	G	G	A	G	A	C	C	C	T	T	G	C	G	G	A	T	C	C	C	C	A	T	A	A	C	T	A	A	A	A	C	C	C	T	G	G	C	T	G	G	A	A	A	A	A	A	A	G	A	G	G
100	C	T	G	G	G	T	A	G	G	G	A	A	A	A	C	T	T	T	G	G	A	A	A	T	T	T	C	C	A	T	A	A	T	T	A	A	A	A	T	T	C	T	G	A	C	T	G	A	G	G	A	A	A	A	G	A	G	A	
101	C	T	A	G	G	T	A	G	G	G	A	A	A	C	C	C	T	T	G	C	G	A	A	T	T	T	C	C	A	T	G	A	C	T	A	A	G	A	C	T	C	T	G	G	T	T	G	G	G	A	G	A	A	T	A	A	G	A	
102	C	T	G	G	G	G	A	G	G	G	A	A	A	C	C	T	C	T	G	G	A	A	T	T	T	T	C	C	A	T	G	A	C	T	G	A	A	A	C	T	C	T	G	A	T	T	G	G	G	A	A	A	A	A	A	A	A	A	
103	T	T	A	A	G	G	G	G	G	A	G	A	A	C	T	T	T	C	C	G	G	A	T	T	T	C	C	A	T	G	A	C	T	G	A	G	A	C	C	T	T	A	A	T	T	G	G	A	A	A	A	A	A	T	A	A	G	G	
104	C	T	A	A	G	G	A	G	G	G	A	A	C	C	T	T	T	T	G	G	G	A	T	T	C	T	C	C	A	A	G	A	T	T	G	A	A	A	T	T	C	C	G	A	C	C	A	A	G	G	G	G	T	T	G	A	G	A	
105	C	T	A	A	G	G	A	A	G	G	A	G	A	C	C	C	T	T	G	G	G	A	A	T	T	T	C	C	T	T	A	A	C	T	G	A	A	A	C	C	C	T	G	G	C	T	G	G	G	A	G	A	A	A	A	A	G	G	
106	C	T	G	G	T	T	G	G	A	G	A	G	C	C	C	T	T	T	G	G	G	G	A	A	C	C	G	C	T	T	G	A	T	T	G	A	A	A	C	C	C	T	A	A	T	T	A	A	A	A	A	A	A	A	T	A	A	G	G

107	C	T	A	A	T	T	G	G	G	G	A	G	A	C	C	C	C	T	G	C	G	A	A	T	T	T	C	C	A	T	A	A	T	T	G	G	G	A	C	T	C	T	G	A	C	T	G	A	A	A	A	A	A	A	A	A	A	A	G	A
108	C	C	A	A	G	T	G	G	G	G	A	A	A	A	C	C	T	T	C	C	A	A	T	T	C	T	C	C	A	T	G	A	C	T	A	A	A	A	C	C	C	T	G	A	C	T	A	A	A	A	G	A	T	T	A	A	A	A		
109	C	T	A	A	G	G	A	G	G	G	G	G	A	C	C	C	T	T	G	C	A	A	A	T	C	T	C	C	A	T	A	A	C	T	A	A	G	A	C	T	C	T	G	A	C	C	G	G	G	G	G	A	A	T	G	A	A	A		
110	C	T	A	G	T	T	A	A	G	G	A	A	A	A	C	C	T	T	G	C	A	A	A	T	T	T	G	C	A	T	G	A	C	T	G	A	G	A	C	C	C	T	G	A	C	T	G	A	A	A	A	A	T	T	G	A	A	A		
111	T	T	A	A	G	T	A	G	G	G	G	G	A	C	C	T	C	T	G	G	A	A	A	T	T	T	C	C	A	T	A	A	C	T	A	A	G	A	T	T	C	C	A	A	T	T	G	G	G	A	G	G	A	T	G	A	A	A		
112	C	T	A	A	G	T	A	A	A	G	A	A	C	C	C	T	C	T	C	C	G	A	A	T	C	T	G	C	T	T	G	A	C	C	A	A	G	A	C	C	C	C	G	A	T	T	A	A	A	A	G	A	A	T	A	A	A	A		
113	C	T	G	G	G	T	A	A	A	G	G	G	C	C	C	C	C	C	C	C	G	G	T	T	C	T	G	C	A	T	G	A	C	T	A	A	G	A	C	T	T	T	G	A	C	T	A	A	G	G	G	G	A	A	A	A	A	A		
114	C	T	A	G	G	T	A	G	A	G	A	G	A	C	C	C	C	T	G	G	A	A	A	T	C	C	C	C	A	T	G	G	C	C	A	A	G	A	T	T	T	T	A	A	T	T	A	A	A	A	A	A	A	T	A	A	A	A		
115	C	T	A	G	T	T	A	A	G	G	A	G	A	C	C	T	C	T	G	G	A	A	A	T	T	T	C	C	A	T	G	A	C	T	A	A	G	A	C	C	C	T	G	A	C	C	G	G	G	A	G	G	A	A	A	A	A	A		
116	C	T	A	A	G	T	A	A	A	G	A	A	C	C	C	C	C	C	G	G	G	A	T	T	C	T	G	C	A	T	A	A	C	C	A	A	G	A	C	C	C	C	G	A	C	T	G	G	G	A	A	A	A	T	A	A	G	A		
117	T	T	A	G	G	T	A	G	G	G	A	G	A	A	C	C	T	T	G	G	G	A	A	T	C	T	C	C	A	T	A	A	C	T	G	A	G	A	T	T	C	C	G	A	C	T	A	A	A	A	G	G	A	T	A	A	G	A		
118	C	C	G	G	T	T	A	A	G	G	A	G	A	A	C	C	C	T	G	C	A	A	T	T	C	C	C	C	A	A	G	G	C	T	A	A	G	A	T	T	C	C	A	A	C	T	G	G	G	A	G	A	A	T	A	A	A	A		
119	C	C	G	G	T	T	G	G	G	G	G	G	A	C	C	T	C	T	G	G	G	A	T	T	C	C	C	C	A	A	A	A	C	C	A	A	A	A	A	T	T	C	T	A	A	C	T	G	G	G	A	A	A	A	A	A	A	G	A	
120	C	C	A	G	T	T	A	G	G	G	A	G	A	C	C	T	C	T	G	C	A	A	A	T	T	T	G	C	T	T	G	G	C	C	A	A	A	A	C	C	T	T	A	A	C	T	G	A	A	A	G	A	A	A	A	A	G	A		
121	C	T	A	G	T	T	A	G	A	G	A	A	A	C	C	T	C	T	G	C	G	A	A	T	T	T	C	C	A	T	G	A	C	T	G	A	G	A	C	T	C	T	G	G	T	T	G	G	G	A	A	A	A	T	G	A	A	A		
122	C	T	G	G	T	T	A	G	A	G	A	A	A	C	C	C	T	T	G	C	A	A	T	T	C	C	C	C	A	A	G	A	C	T	G	A	A	A	T	T	C	C	A	A	T	T	G	A	G	A	A	A	A	A	A	A	A	G	A	
123	T	T	A	G	G	G	A	G	G	G	A	G	A	C	C	C	C	T	G	G	G	A	A	T	C	T	C	C	A	A	A	A	C	T	G	A	G	A	C	C	C	C	G	A	C	C	G	G	G	G	G	A	A	A	A	A	A	A		
124	T	T	A	A	G	G	A	A	A	G	A	G	A	C	C	T	T	T	C	C	A	A	A	T	C	C	C	C	A	A	G	A	C	C	A	A	G	A	T	T	C	C	A	A	T	T	G	G	A	A	G	G	A	A	G	A	G	A		
125	T	T	A	G	G	T	A	A	G	G	A	G	C	C	C	C	T	T	G	G	G	A	A	T	C	T	G	C	A	A	A	A	C	T	G	A	A	A	C	T	C	T	G	A	C	C	G	A	G	A	G	A	A	A	G	A	A	A		
126	C	C	A	G	T	T	A	A	G	G	A	G	A	C	C	T	C	T	G	G	G	A	A	T	C	T	C	C	A	T	A	A	C	T	G	A	G	A	C	T	C	T	G	A	C	T	G	G	A	A	G	A	A	A	A	A	G	A		
127	T	T	A	G	T	T	A	A	A	A	A	G	A	C	C	C	T	T	G	G	A	A	A	T	C	C	G	C	T	T	G	A	T	T	A	A	G	A	T	T	C	T	A	A	T	T	A	A	A	A	G	G	A	T	A	A	G	G		
128	C	T	G	G	T	T	A	A	A	G	G	A	C	C	T	C	T	G	G	G	A	A	A	T	C	C	C	C	A	A	G	G	C	T	A	A	G	A	T	T	C	C	A	A	C	T	A	A	G	A	A	A	T	T	A	A	A	A		
129	T	T	A	G	T	T	A	G	A	G	A	G	A	C	C	T	T	T	G	G	G	A	A	T	C	C	C	C	A	T	A	A	T	T	A	A	A	A	A	T	T	C	T	G	A	T	T	G	G	G	A	G	G	A	T	G	A	A	A	
130	C	C	A	A	G	T	A	A	G	G	G	G	C	C	C	C	C	T	G	G	A	A	A	A	C	T	G	C	T	T	G	A	C	T	G	A	G	A	C	T	C	C	G	A	C	C	G	G	G	A	G	G	A	T	G	A	A	A		
131	C	T	A	A	G	G	A	G	G	G	G	A	C	C	C	T	T	G	G	A	A	A	A	C	T	C	C	A	A	A	A	C	T	A	A	G	A	C	C	C	T	G	A	C	T	G	G	G	G	G	G	A	T	A	A	A	A			

Tabela 2A . Resultados do Auto 1 para a população de imigrantes de leste residentes em Portugal (n=85)

Amostra	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23																							
U001	C	C	C	C	T	T	A	T	G	A	A	T	G	A	C	T	C	C	C	C	T	T	A	A	G	T	T	T	G	T	A	A	G	G	G	G										
U002	C	C	C	T	T	T	T	T	G	A	A	A	G	G	C	T	G	C	T	T	C	T	A	A	G	A	T	T	T	T	G	G	A	A	G	A	G	A								
U002.1	C	T	C	C	T	T	T	T	G	A	A	T	G	G	C	C	G	C	C	T	T	T	G	A	A	A	T	T	T	T	G	G	A	A	T	T	C	T	G	G	A	A	G	A	G	A
U003	C	T	C	C	T	T	T	T	G	A	T	T	G	A	C	C	C	C	C	T	C	T	G	A	A	A	T	T	T	T	G	G	A	A	T	T	C	C	G	G	G	A	G	A	G	G
U004	C	C	C	T	A	A	T	T	A	A	T	T	G	A	C	T	C	C	C	C	C	T	A	A	G	G	A	T	T	T	G	T	G	A	G	T	T	T	G	G	G	G	G	G	A	
U005	C	T	C	C	A	A	A	T	G	A	T	T	G	G	C	C	C	C	C	T	C	T	A	A	G	A	A	T	T	T	G	T	A	A	T	T	C	T	T	T	G	G	G	G	G	A
U006	C	C	C	T	A	T	T	T	G	A	T	T	G	G	C	T	G	C	C	C	T	T	A	A	G	G	A	T	T	T	G	G	G	A	T	T	C	T	G	G	G	G	G	G	G	
U007	T	T	C	T	T	T	A	T	G	A	A	T	G	G	T	T	C	C	T	T	T	T	A	A	G	G	A	A	T	T	G	T	G	A	T	T	C	T	G	T	A	A	G	A	G	A
U008	C	T	C	T	T	T	T	T	G	A	T	T	G	G	C	T	C	C	C	C	T	T	A	A	G	A	A	T	T	T	G	G	A	A	G	G	C	T	G	G	G	G	G	G	A	
U009	C	C	C	C	A	T	A	T	G	A	A	T	G	G	C	C	G	C	T	T	T	T	G	A	G	G	T	T	T	T	G	G	G	A	T	T	T	T	G	G	A	A	G	G	G	A
U009.1	C	C	C	C	T	T	A		G	A	A	T	G	A	C	C	G	C	C	T	T	T	G	G	G	G	A	T	T	T	G	G	A	A	T	T	T	T	G	G	A	A	G	G	G	G
U009.2	C	C	C	C	T	T	A	T	G	A	A	T	G	A	C	C	C	C	C	T	T	T	G	G	G	G	A	T	T	T	G	T	G	A	T	T	T	T	G	G	A	A	G	G	G	G
U010	C	T	C	C	T	T	T	T	A	A	A	T	G	A	C	C	C	C	C	C	T	T	G	A	G	G	T	T	C	T	G	G	A	A	T	T	T	T	G	G	G	A	G	A	G	A
U011	C	T	C	C	A	T	T	T	G	A	A	A	G	G	C	T	C	C	C	T	T	T	G	A	G	A	A	T	T	T	G	G	G	A	G	T	T	T	G	G	A	A	G	G	G	A
U012	C	T	C	T	A	T	A	T	G	G	T	T	G	G	C	C	G	C	C	T	T	T	G	A	G	G	A	T	T	T	G	G	A	A	G	T	C	T	G	G	G	G	G	G	G	
U012.1	C	T	C	C	T	T	A	T	G	A	T	T	G	G	C	C	G	C	C	T	C	T	G	A	G	G	A	T	C	T	G	T	A	A	G	T	C	T	G	G	A	A	G	G	G	G
U013	C	C	C	T	A	T	A	T	G	G	A	A	G	G	C	C	C	C	C	T	T	T	A	A	G	A	A	A	T	T	G	T	G	A	G	T	T	T	G	G	G	A	G	A	G	A
U014	T	T	C	C	T	T	A	T	G	A	T	T	G	G	C	C	G	C	C	T	C	T	G	A	G	A	A	T	C	T	G	T	G	A	G	T	C	T	G	G	G	A	G	A	G	G
U015	T	T	C	C	T	T	A	T	G	A	T	T	G	G	C	C	C	C	C	T	C	C	A	A	G	A	A	T	C	T	G	T	A	A	G	T	C	T	G	G	G	A	G	G	G	A

U014.1	C	T	C	T	T	T			G	A	A	T	G	G	C	T	C	C	T	T	C	T	A	A	G	A	A	T	T	T	G		G	A	T	T	C	T	G	T	A	A	G	G	G	A
U016	C	T	C	T	A	T	A	T	G	A	A	T	G	G	C	T	C	C	C	T	C	T	A	A	A	A	A	A	C	T	G	T	A	A	G	T	T	T	G	G	G	A	G	G	G	G
U017	C	T	T	T	A	A	T	T	G	A	T	T	G	G	T	T	C	C	C	T	T	T	G	A	G	A	A	T	T	T	T	A	A	G	T	C	T	G	G	A	A	A	A	G	A	
U018	C	T	T	T	A	T	T	T	G	A	T	T	G	G	T	T	C	C	C	T	T	T	G	A	A	A	T	T	T	T	G	T	A	A	T	T	C	T	G	G	A	A	G	A	G	A
U019	C	C	C	C	A	A	A	T	G	G	A	T	G	G	C	T	C	C	C	C	C	T	A	A	A	A	A	T	T	T	G	T	A	A	T	T	C	T	G	T	A	A	G	G	G	A
U020	C	T	C	C	A	T	A	A	G	A	A	T	A	A	C	C	G	C	C	T	C	T	G	A	G	A	A	A	T	T	T	T	G	G	G	T	C	T	G	G	G	A	G	G	G	A
U021.1	C	C	C	T	T	T	A	T	G	G	T	T	G	G	C	T	C	C	C	T	T	T	G	A	G	A	T	T	C	T	G	T	G	G	G	T	T	T	G	G	A	A	G	A	A	A
U021	C	C	C	T	A	T	A	T	G	A	T	T	G	A	C	T	C	C	C	C	C	T	A	A	G	G	T	T	C	T	G	T	G	G	G	T	C	T	G	G	A	A	A	A	G	A
U022	C	C	C	T	A	T	A	T	G	G	A	T	G	G	C	T	C	C	T	T	T	T	G	A	G	A	A	T	T	T	G	T	G	A	G	G	T	T	G	G	A	A	G	G	G	A
U023	C	C	C	C	A	A	T	T	A	A	T	T	G	A	C	T	C	C	C	T	T	T	A	A	G	G	A	A	T	T	G	G	G	A	G	G	T	T	G	G	G	A	G	A	G	A
U024	C	C	C	C	T	T	A	A	G	G	T	T	G	A	C	T	C	C	C	T	T	T	A	A	G	G						G	A			T	T	G	G	G	A	G	A	G	A	
U025	C	T	C	T	T	T	A	T	G	A	T	T	G	A	T	T	C	C	T	T	C	T	A	A	G	G	T	T	T	T	T	A	A	T	T	C	T	G	G	A	A	G	G	G	A	
U026	C	T	C	C	T	T	T	T	G	G	T	T	G	G	C	T	G	C	T	T	C	T	A	A	G	A	A	A	T	T	G	T	G	A	T	T	C	T	G	G	G	A	G	A	G	A
U027	C	T	C	C	T	T	A	T	G	A	A	T	G	A	C	T	G	C	C	T	C	T	G	A	G	G	A	T	T	T	G	G	G	G	G	T	C	T	G	G	G	A	A	A	A	A
U028	T	T	C	T	A	A	A	T	G	A	A	T	G	G	C	T	C	C	C	T	T	T	G	A	A	A	A	A	T	T	G	G	G	A	G	T	C	T	G	T	G	A	G	A	G	G
U029	T	T	C	C	A	T	A	A	A	A	A	T	G	G	C	T	C	C	T	T	T	T	G	A	G	A	A	A	C	T	T	T	A	A	G	T	T	T	T	T	G	A	G	G	G	A
U030	C	C	C	C	T	T	A	T	G	A	A	T	G	G	C	T	G	C	C	T	C	T	G	A	G	A	A	T	C	T	G	T	G	A	T	T	C	T	G	G	G	A	G	G	G	A
U031	C	T	C	C	A	T	T	T	G	A	T	T	G	A	C	T	C	C	C	T	T	T	A	A	G	G	A	A	T	T	G	G	G	A	G	G	T	T	G	T	G	A	G	A	G	A
U032	C	C	C	T	T	T	A	T	G	A	A	T	G	G	C	T	C	C	C	T	T	T	A	A	G	G	A	T	T	T	T	A	A	T	T	C	T	G	G	G	A	G	G	G	G	
U025.1	C	T	C	T	T	T	T	T	G	A	T	T	G	A	T	T	C	C	T	T	T	T	A	A	G	G	A	T	T	T	G	T	G	A	T	T	T	T	G	G	A	A	G	G	G	A
U033	C	T	C	C	A	A	A	A	G	A	T	T	A	A	C	T	C	C	C	T	T	T	A	A	G	A	A	A	C	T	T	T	G	A	T	T	C	T	G	G	A	A	G	G	G	G

U034	C	T	C	C	T	T	A	A	A	A	A	T	A	A	C	T	G	C	C	T	T	T	G	A	G	A	A	T	T	T	T	G	A	G	T	T	T	G	G	G	G	G	G	G	G	
U035	T	T	C	C	A	T	T	T	G	G	A	T	G	G	C	C	C	C	C	T	T	T	G	G	G	A	A	T	C	C	G	G	A	A	T	T	T	T	G	G	A	A	G	A	G	G
U036	T	T	C	C	A	T	A	T	G	G	A	T	G	A	T	T	C	C	C	C	T	G	A	G	A	A	T	C	T	G	G	A	A	T	T	C	T	G	G	G	A	A	A	G	G	
U037	T	T	C	T	A	T	A	T	G	A	A	T	G	G	C	T	C	C	C	C	T	T	A	A	G	A	A	T	T	T	G	T	A	A	G	T	T	T	G	T	G	A	G	G	G	A
U038	T	T	C	C	A	T	A	T	A	A	A	T	G	A	T	T	C	C	T	T	C	T	A	A	A	A	A	T	C	T	G	T	A	A	G	T	T	T	T	T	G	A	G	G	G	A
U039	T	T	C	C	A	T	T	T	G	G	T	T	G	G	T	T	C	C	C	T	T	T	A	A	G	A	T	T	C	T	G	T	A	A	T	T	C	T	G	G	G	A	G	G	G	A
U040	C	T	C	C	A	A	T	T	A	A	T	T	G	A	C	T	C	C	T	T	C	T	A	A	A	A	A	A	T	T	G	T	G	A	G	G	T	T	G	G	G	A	A	A	G	A
U041	T	T	T	T	A	T	T	T	G	G	A	T	G	G	C	T	C	C	T	T	C	T	A	A	G	A	A	A	T	T	G	G	A	A	T	T	C	T	G	T	A	A	G	G	G	G
U042	T	T	C	C	T	T	A	T	G	G	T	T	G	G	C	T	G	C	C	T	T	T	A	A	G	G	T	T	T	T	G	T	A	A	T	T	C	T	G	G	A	A	G	G	A	A
U043	T	T	C	C	T	T	A	A	G	A	A	T	G	G	C	C	C	C	C	C	T	T	G	A	G	A	A	A	C	T	T	T	G	A	T	T	C	T	G	T	A	A	G	G	G	A
U044	T	T	C	T	A	T	T	T	A	A	A	A	G	A	C	T	G	C	C	T	C	T	G	A	G	A	T	T	C	T	G	T	A	A	T	T	T	T	G	G	G	A	G	A	G	G
U045	C	T	C	T	A	T	T	T	G	A	T	T	G	A	C	T	C	C	C	T	T	T	A	A	G	A	A	A	T	T	G	G	G	A	G	T	T	T	G	G	G	A	G	G	G	A
U046	T	T	C	T	T	T	T	T	G	A	A	T	G	A	T	T	G	C	C	T	T	T	A	A	A	A	A	T	T	T	G	G	G	A	T	T	T	T	T	T	G	G	G	G	G	G
U047	T	T	C	C	A	T	T	T	A	A	A	T	G	G	C	C	C	C	C	C	T	A	A	G	A	A	T	T	T	G	G	A	A	T	T	T	T	G	G	A	A	G	G	G	G	
U048*	T	T	C	C	A	T	T	T	G	A	A	T	G	G	C	C	C	C	C	C	T	A	A	G	A	A	T	T	T	G	G	A	A	T	T	T	T	G	G	A	A	G	A	G	G	
U049*	C	T	C	T	A	T	T	T	G	G	A	T	G	G	C	T	C	C	C	T	C	T	G	A	G	G	A	T	T	T	G	T	G	A	T	T	C	T	G	G	G	A	G	G	G	G
U050*	C	T	C	T	T	T	A	T	G	A	A	T	G	G	C	T	C	C	C	T	T	T	G	A	G	A	A	T	C	T	G	T	G	A	T	T	C	T	G	G	G	A	G	G	G	A
U051*	C	T	C	C	T	T	A	T	G	A	A	T	G	G	C	T	C	C	C	T	C	T	G	G	G	A	A	T	T	T	G	T	G	A	T	T	T	T	G	G	A	A	G	G	G	A
U052	C	T	C	C	T	T	T	T	G	A	A	T	G	G	C	T	C	C	C	T	C	T	G	G	G	A	A	T	C	T	G	T	G	A	T	T	T	T	T	T	G	A	G	G	G	A
U053	C	T	C	C	T	T	T	T	G	A	A	T	G	G	C	T	G	G			C	T	G	G	G	G	T	T	C	T	G	T	G	A	T	T	T			G	G	G	A	A	A	
U054	C	T	C	C	T	T	T	T	G	A	T	T	G	G	T	T	C	C	C	T	T	T	A	A	G	A	A	T	T	T	G	T	G	A	T	T	T	T	G	G	A	A	A	A	A	A

U055	C	T	C	C	T	T	A	T	G	A	A	T	G	A	C	T	G	G	C	T	T	T	G	G	G	A	A	T	T	T	G	T	G	A	T	T	C	C	G	G	A	A	G	A	A	A
U056	C	T	C	C	A	A	T	T	G	G	A	T	G	G	C	T	C	C	C	T	C	T	G	G	G	G	A	A	T	T	G	G	G	A	T	T	T	T	G	T	A	A	G	G	G	G
U057	C	C	C	C	T	T	T	T	G	A	A	T	G	G	C	C	C	C	C	C	T	G	G	G	A	A	A	C	C	G	T	G	A	T	T	C	C	G	T	A	A	G	G	G	A	
U058	T	T	C	C	T	T	A	T	A	A	T	T	G	G	C	C	C	C	C	T	T	T	A	A	G	A	A	T	T	T	G	G	A	A	G	T	C	T	G	G	G	G	G	A	A	
U059	C	C	T	T	T	T	A	T	G	A	A	T	G	G	T	T	C	C	C	T	C	T	A	A	G	A	A	A	T	T	G	T	G	A	G	T	C	T	G	T	G	A	G	G	G	A
U060	C	T	C	C	A	T	A	T	G	A	A	T	G	A	T	T	C	C	C	C	T	A	A	G	G	A	T	C	T	G	T	G	G	T	T	T	T	G	T	G	A	G	G	G	G	
U061	C	C	C	C	T	T	A	T	G	A	A	T	G	A	C	C	C	C	T	T	T	T	A	A	G	A	T	T	C	C	G	T	A	A	T	T	T	T	G	G	A	A	G	G	A	A
U062	T	T	C	C	A	T	A	A	G	A	A	T	G	A	T	T	C	C	C	C	T	T	A	A	G	A	A	T	T	T	G	G	A	A	G	T	T	T	G	T	A	A	G	G	G	A
U063	C	T	T	T	A	T	T	T	A	A	A	T	G	G	C	C	G	C	C	C	T	T	G	A	G	A	T	T	C	T	G	T	A	A	G	T	C	T	G	T	A	A	G	G	G	A
U064	T	T	C	C	A	T	T	T	G	A	T	T	G	A	C	T	C	C	C	T	T	T	A	A	G	G	A	T	T	T	G	T	G	A	G	T	C	T	T	T	A	A	G	G	G	A
U065	C	C	T	T	T	T	A	T	G	A	A	T	G	A	T	T	C	C	C	T	C	C	A	A	G	A	T	T	C	T	G	T	A	A	T	T	C	T	G	G	A	A	G	G	G	A
U066	C	T	C	C	T	T	A	T	G	G	T	T	G	G	T	T	C	C	C	T	T	T	A	A	G	A	A	T	T	T	G	T	A	A	G	T	T	T	G	G	G	A	G	G	G	A
U067	C	T	C	C	T	T	A	T	G	A	T	T	G	G	C	T	C	C	T	T	C	T	A	A	G	A	A	A	C	T	T	T	A	A	G	G	C	C	G	G	A	A	G	G	G	G
U068	T	T	C	C	A	T	A	T	A	A	T	T	G	G	C	T	C	C	T	T	C	T	G	G	G	G	A	T	T	T	G	G	A	A	G	T	T	T	G	G	A	A	A	A	G	A
R001	T	T	C	T	T	T	T	T	G	A	A	T	G	G	C	T	G	C	C	C	C	T	G	G	G	A	A	T	T	T	G	T	G	G	G	T	C	T	G	T	G	A	G	A	G	A
R001.1	C	T	C	C	T	T	T	T	G	A	T	T	G	G	C	T	G	C	C	T	C	T	G	A	G	G	A	T	T	T	G	T	G	A	T	T	C	T	G	T	A	A	G	G	G	A
R002	C	T	C	C	T	T	T	T	G	A	T	T	G	G	T	T	G	C	C	T	C	T	G	G	G	A	A	A	T	T	G	T	A	A	T	T	C	T	G	G	A	A	A	A	G	G
R002.1	C	T	C	C	T	T	T	T	G	A	T	T	G	A	T	T	G	C	T	T	C	T	G	G	G	A	A	A	T	T	G	T	G	A	T	T	C	C			A	A	G		G	A
R003	C	T	C	T	A	T	A	T	G	A	A	T	G	A	C	T	G	C	C	T	C	T	G	A	G	G	A	T	T	T	G	T	A	A	G	T	C	T	G	G	G	A	G	G	G	A
R004	C	T	C	T	A	T	T	T	G	A	T	T	G	G	C	C	C	C	C	T	C	T	G	A	A	A	T	T	C	C	T	T	A	A	T	T	C	T	T	T	A	A	A	A	G	A
R005	C	C	C	C	T	T	T	T	A	A	A	A	G	A	C	T	C	C	C	T	T	T	G	A	G	A	A	T	T	T	G	T	G	A	G	T	C	C	G	G	A	A	G	A	G	A



U014	C	T	G	A	T	T	G	A	G	A	G	A	A	A	A	C	T	C	T	G	C	G	A	A	A	C	T	C	C	A	T	A	A	C	T	A	A	G	G	C	C	C	C	G	A	T	T	A	A	A	A	G	A	A	T	A	A	G	A
U015	T	T	G	A	T	T	G	A	G	A	A	A	A	A	A	T	T	C	T	G	G	G	A	A	T	C	T	G	G	A	A	A	A	C	T	G	A	G	A	C	C	C	C	A	A	C	T	A	A	A	A	G	G	A	A	A	A	G	A
U014.1	C	T	G	G	T	T	G	A	G	G	A	A	C	A	C	T	T	T	G	C	A	A	T	C	C	C	C	A	T	G	A	T	T	A	A	G	A	C	T	C	T	A	A	C	C	G				G	A	A	A	G	A	G	A		
U016	T	T	G	G	G	T	G	A	G	G	G	A	C	A	C	C	T	T	G	C	G	A	A	T	C	T	G	G	A	T	A	A	C	T	A	A	G	G	C	T	C	T	G	G	C	C	G	A	A	A	G	A	A	A	A	A	A	A	
U017	C	T	G	G	T	T	A	A	G	A	G	A	C	C	C	C	T	T	G	C	A	A	T	T	C	T	C	C	A	T	A	A	C	C	G	A	A	A	C	T	C	C	G	G	C	C	G	G	G	A	A	A	A	A	A	A	A	G	A
U018	C	T	G	G	T	T	A	A	G	A	G	A	C	C	C	C	T	T	G	C	A	A	T	T	T	T	C	C	A	T	A	A	C	T	A	A	G	A	C	C	C	T	G	G	C	C	G	A	A	A	A	A	A	A	A	A	G	G	
U019	T	T	G	G	G	G	A	A	G	A	G	A	C	A	C	C	T	T	G	C	A	A	T	T	T	T	C	C	A	T	G	A	C	T	G	G	G	A	C	C	C	C	G	A	C	T	A	A	G	A	A	A	A	A	A	A	G	A	
U020	C	C	G	G	G	T	G	A	A	A	A	A	A	A	C	C	C	T	C	C	G	G	T	T	C	T	C	C	A	T	G	A	C	T	A	A	G	A	C	C	C	T	A	A	C	C	G	A	G	G	G	G	A	T	A	A	G	G	
U021.1	C	T	G	G	G	T	A	A	G	A	G	A	A	A	C	T	T	T	G	C	G	A	A	T	T	T	C	C	A	T	A	A	C	C	A	A	A	A	C	T	C	C	A		T	T	A	A	G	A	G	A	A	A	G	A	G	A	
U021	C	T	G	G	T	T	A	A	G	A	A	A	C	A	C	T	C	T	G	C	G	A	A	T	C	T	C	C	A	T	G	A	C	C	A	A	G	A	C	T	C	C	G	G	T	T	G	A	G	A	A	A	A	A	A	A	G	G	
U022	T	T	G	G	G	T	G	A	G	A	G	A	A	A	C	T	T	T	G	C	G	A	T	T	C	T	G	C	A	T	G	A	C	C	G	A	G	A	T	T	C	C	G	G	T	T	A	A	G	G	G	A	A	T	G	A	G	G	
U023	T	T	G	G	G	G	G	A	G	G	G	G	C	A	C	T	T	T	G	G	G	A	A	A	C	T	C	C	A	T	G	A	C	T	A	A	G	A	C	C	C	T	G	G	C	T	A	A	G	A	G	G	T	T	A	A	G	G	
U024	C	C	A	A	G	G	G	A	G	A	A	A	C	A	C	T	T	T	C	C	G	G	A	T	T	T	C	C	A	T	G	A	C	T	G	A	G	A	C	C	C	T	G	A	T	T	G	A	G	A	G	G	T	T	A	A	G	G	
U025	C	C	G	G	T	T	A	A	G	G	G	A	C	C	T	T	C	T	G	C	A	A	A	T	C	T	G	C	A	T	A	A	C	T	G	A	A	A	C	C	C	C	G	A	C	C	G	G	G	A	G	A	A	T	A	A	G	G	
U026	C	C	A	A	G	G	G	A	G	G	G	G	C	C	C	C	T	T	G	C	A	A	T	T	T	T	C	C	A	T	G	G	C	C	A	A	A	A	C	T	C	T	G	G	C	C	A	A	G	G	G	A	A	A	G	A	G	G	
U027	C	T	G	G	G	G	G	A	G	A	G	A	C	C	C	C	T	T	G	C	A	A	T	T	T	T	C	C	A	T	G	G	C	C	A	A	A	A	C	T	C	C	A	A	C	C	A	A	G	G	G	A	A	A	A	A	G	G	
U028	C	T	G	A	G	T	A	A	G	A	A	A	C	A	C	C	C	C	G	C	G	A	T	T	C	T	C	C	A	T	A	A	C	T	G	A	A	A	C	T	C	C	G	A	C	T	G	A	A	A	A	A	T	T	A	A	G	A	
U029	C	C	G	A	T	T	G	A	G	G	A	A	C	A	C	C	T	T	G	C	G	A	A	T	C	T	C	C	A	T	G	A	C	C	G	A	G	G	C	T	C	C	G	A	C	T	G	A	G	A	G	G	A	A	A	A	G	A	
U030	T	T	G	A	G	T	G	A	A	A	A	A	C	A	C	T	T	T	G	G	G	A	T	T	T	T	G	C	A	T	G	A	C	T	A	A	A	A	C	T	C	C	G	A	T	T	A	A	G	A	G	A	A	T	A	A	G	G	
U031	T	T	G	G	G	G	G	A	G	G	G	G	C	C	C	T	T	T	G	G	G	A	A	A	C	T	G	C	A	T	G	A	C	T	A	A	G	A	C	T	C	T	G	A	T	T	A	A	G	A	G	G	T	T	A	A	A	A	
U032	T	T	G	G	T	T	G	A	G	G	A	A	C	C	C	T	T	T	G	C	A	A	A	A	C	T	C	C	T	T	G	A	C	C	A	A	A	A	C	T	C	C	G	A	T	T	G	G	G	A	G	A	A	A	G	A	G	G	
U025.1	T	T	G	G	G	T	G	A	G	G	G	A	C	C	C	T	C	T	C	C	A	A	A	T	C	T	G	C	A	T	A	A	C	C	G	A	G	A	C	C	C	C	A	A	C	C	G	G	G	A	G	G	T	T	A	A	G	G	
U033	C	C	G	A	G	T	G	A	G	A	G	A	A	A	C	C	T	T	G	G	G	A	A	T	C	T	G	C	T	T	G	A	C	T	G	A	G	A	C	C	C	T	G	A	C	C	G	G	G	A	G	G	T	T	G	A	G	A	
U034	C	C	A	A	T	T	G	A	G	A	A	A	A	A	C	C	T	T	G	G	G	A	A	T	C	C	G	C	A	T	G	G	C	T	G	A	G	A	C	C	T	T	G	A	C	C	G	G	G	A	G	A	T	T	A	A	G	G	
U035	C	T	A	A	G	G	G	A	G	G	G	A	A	A	C	T	C	T	G	G	G	A	A	A	T	T	G	C	T	T	A	A	C	T	G	A	G	A	C	C	C	C	G	G	T	T	A	A	G	G	G	G	A	A	A	A	A	A	
U036	C	T	G	A	G	T	G	A	G	A	G	A	C	A	C	T	T	T	G	C	G	A	T	T	T	T	C	C	T	T	A	A	C	T	A	A	A	A	C	T	C	C	G	A	C	T	G	A	A	A	G	G	A	T	A	A	G	A	

U037	C	T	A	A	G	T	G	A	G	G	A	A	A	A	C	T	T	T	G	C	G	G	A	T	T	T	C	C	A	A	G	A	T	T	A	A	G	G	C	C	C	C	G	A	T	T	G	G	G	A	G	A	A	A	A	A	G	G	
U038	C	C	A	A	G	G	A	A	G	G	G	A	C	A	C	T	T	T	G	C	A	A	A	T	C	T	C	C	A	T	G	A	C	C	A	A	A	A	C	T	C	C	G	A	C	C	A	A	G	A	G	G	A	A	A	A	A	G	A
U039	C	C	G	G	T	T	G	G	G	A	A	A	C	A	C	T	T	T	G	G	A	A	T	T	C	C	C	C	T	T	G	A	C	T	G	A	G	A	C	T	T	T	G	G	C	T	G	A	G	G	G	A	A	T	G	A	A	A	
U040	C	T	A	A	G	T	G	A	G	A	G	A	A	A	C	T	T	T	C	C	G	A	A	T	C	C	C	C	A	T	G	A	C	T	A	A	G	A	C	C	T	T	G	A	T	T	A	A	A	A	A	A	A	T	A	A	G	G	
U041	C	T	G	G	G	G	A	A	G	G	G	A	A	A	C	T	T	T	G	G	G	A	A	T	C	T	C	C	A	T	G	A	C	T	G	A	A	A	T	T	C	C	G	G	C	T	G	G	A	A	G	A	A	T	A	A	G	A	
U042	C	C	G	G	T	T	G	A	G	A	G	A	C	C	C	C	T	T	G	C	G	A	A	A	C	C	C	C	T	T	A	A	C	C	A	A	A	A	C	T	C	T	G	G	C	T	G	G	G	A	G	A	A	A	A	A	G	A	
U043	T	T	A	A	G	T	G	A	G	G	A	A	C	C	C	C	T	T	C	C	A	A	T	T	C	T	C	C	T	T	A	A	C	T	A	A	A	A	C	T	C	T	G	A	C	C	A	A	G	G	G	A	A	T	G	A	G	A	
U044	C	T	G	G	T	T	G	A	G	A	A	A	A	A	T	T	T	T	G	C	G	A	T	T	T	T	C	C	T	T	G	A	C	T	G	A	G	A	C	T	T	T	G	A	C	T	A	A	A	A	G	A	A	A	A	A	A	A	
U045	C	T	A	A	G	T	A	A	G	G	A	A	C	C	T	T	T	T	C	C	G	G	A	T	C	T	C	C	A	A	G	A	T	T	A	A	G	A	C	T	C	C	G	G	C	T	G	A	G	A	G	A	A	T	G	A	G	A	
U046	C	C	G	G	T	T	G	A	G	A	A	A	C	A	C	C	C	T	G	C	G	A	A	A	C	T	C	C	A	T	G	G	C	T	G	A	G	A	C	C	C	C	G	G	T	T	G	G	G	A	G	A	A	A	G	A	G	G	
U047	C	T	G	G	G	T	G	A	G	A	G	A	C	C	C	T	T	T	G	C	A	A	A	T	C	T	C	C	A	T	A	A	C	C	A	A	G	A	C	C	C	C	G	A	T	T	G	G	G	A	G	G	A	A	A	A	G	A	
U048*	C	T	G	A	G	T	G	G	G	A	G	A	C	C	C	T	T	T	G	C	A	A	A	T	C	T	C	C	A	T	A	A	C	C	A	A	G	A	C	C	C	C	G	A	T	T	G	G	G	A	G	G	A	A	A	A	G	A	
U049*	T	T	G	A	G	T	A	A	G	G	G	A	C	C	C	T	T	T	G	C	G	G	A	T	C	C	G	C	A	T	A	A	T	T	G	A	A	A	C	C	C	T	G	A	C	T	G	A	A	A	G	A	A	T	A	A	G	A	
U050*	C	T	G	A	T	T	G	A	G	A	G	A	C	A	C	T	C	T	G	C	G	A	A	T	C	C	C	C	A	T	G	A	C	T	G	A	G	A	C	T	C	T	G	A	T	T	G	G	G	A	G	A	A	T	G	A	G	A	
U051*	C	T	G	A	T	T	A	A	G	G	G	A	C	A	C	T	T	T	G	C	G	A	A	T	C	C	C	C	A	T	G	A	C	T	G	A	G	A	C	C	C	T	G	A	T	T	G	G	G	A	G	A	A	A	A	A	G	A	
U052	T	T	G	G	G	T	G	A	G	G	G	A	C	A	C	C	C	T	G	G	G	G	A	T	C	C	C	C	A	T	G	A	C	C	G	A	G	A	T	T	C	C	G	A	C	C	G	A	A	A	G	A	A	A	A	A	G	G	
U053	C	C	A	A	T	T	A	A	G	A	G	A	A	C	C	T	C	T	G	C	G	A	A	A	C	T	C	C	A	T	A	A	C	C	A	A	A	A	C	T	C	T	G	G	T	T	A	A	G	G	G	A	A	A	A	A	A	A	
U054	T	T	G	G	T	T	A	A	G	G	G	A	C	A	C	T	T	T	G	C	G	G	A	T	C	T	G	C	A	T	A	A	C	T	G	A	A	A	C	T	C	C	G	G	T	T	G	G	A	A	G	A	A	T	A	A	G	A	
U055	T	T	G	A	T	T	G	A	G	G	G	G	C	A	C	T	C	T	G	C	G	A	A	T	T	T	G	C	T	T	G	G	C	C	A	A	G	A	C	T	C	T	G	A	C	T	G	G	A	A	A	A	A	T	A	A	G	A	
U056	T	T	A	A	G	T	A	A	G	A	A	A	C	A	C	C	T	T	G	G	A	A	A	T	C	C	G	C	T	T	A	A	C	C	A	A	G	G	T	T	C	C	G	A	C	C	A	A	G	G	A	A	A	A	G	A	G	G	
U057	C	C	G	G	G	T	A	A	G	A	G	A	C	A	C	T	T	T	G	C	G	A	T	T	C	C	C	C	A	T	A	A	C	T	A	A	G	A	T	T	C	T	G	A	T	T	A	A	G	A	A	A	A	T	G	A	G	A	
U058	C	T	A	A	T	T	G	A	G	G	G	A	A	A	C	T	T	T	C	C	A	A	A	A	T	T	C	C	A	T	A	A	C	T	G	A	A	A	T	T	C	C	A	A	C	T	G	G	A	A	G	A	A	T	G	A	A	A	
U059	C	T	G	G	T	T	G	A	G	G	G	A	C	A	C	T	T	T	G	G	A	A	T	T	T	T	C	C	A	T	G	A	C	T	A	A	G	A	C	C	C	C	G	A	T	T	A	A	G	A	A	A	A	T	G	A	A	A	
U060	T	T	G	G	T	T	G	A	G	G	A	A	A	A	C	T	T	T	G	C	G	G	A	T	C	C	C	C	A	T	A	A	C	C	G	A	G	A	C	T	T	T	G	A	C	C	G	A	G	G	G	A	A	A	A	A	G	G	
U061	C	T	G	A	G	G	A	A	G	G	A	A	C	C	C	C	T	T	G	G	G	A	A	T	C	T	G	C	T	T	A	A	C	C	A	A	G	A	T	T	T	T	G	A	C	T	G	A	G	G	A	A	A	A	A	A	G	A	
U062	T	T	A	A	G	T	G	G	G	G	G	C	C	C	C	C	T	G	C	A	A	T	T	C	T	C	C	A	T	G	G	C	T	G	G	G	A	T	T	C	T	G	A	T	T	G	G	G	A	G	A	A	T	G	A	G	A		

U063	C	T	G	G	G	T	G	A	G	G	G	A	C	C	C	C	T	G	C	G	G	A	T	C	T	C	C	A	T	G	A	C	C	G	G	A	A	C	C	C	C	G	A	C	T	G	A	A	A	G	A	A	T	A	A	G	G		
U064	C	T	G	G	G	T	A	A	G	G	G	A	C	C	C	T	C	C	C	C	G	G	T	T	C	C	G	C	T	T	A	A	T	T	G	G	G	A	T	T	C	C	G	G	C	T	G	G	A	A	G	G	A	A	G	A	G	G	
U065	C	T	G	G	T	T	A	A	G	A	G	A	C	C	C	C	T	T	G	C	A	A	T	T	C	T	G	C	A	T	A	A	T	T	A	A	G	G	C	T	C	C	G	G	T	T	A	A	G	G	G	G	A	T	A	A	G	A	
U066	T	T	G	A	T	T	G	A	G	G	A	A	C	A	C	C	T	T	G	C	G	A	T	T	C	C	G	C	T	T	G	A	C	T	A	A	G	A	C	T	C	T	G	A	C	T	G	G	A	A	G	A	A	A	A	G	A		
U067	T	T	G	G	G	T	A	A	G	A	G	A	C	A	C	C	T	T	G	C	G	G	A	A	C	C	C	C	A	T	G	G	T	T	A	A	G	G	C	T	C	C	G	A	T	T	G	G	G	A	G	A	A	A	A	A	G	G	
U068	T	T	G	G	G	T	G	G	A	A	G	A	C	A	C	C	C	C	G	C	A	A	T	T	T	T	C	C	T	T	A	A	C	C	A	A	G	A	C	C	C	T	G	G	T	T	A	A	G	A	G	A	A	T	A	A	A	A	
R001	C	T	G	A	T	T	G	A	G	A	A	A	C	C	C	C	T	T	G	C	A	A	T	T	C	T	G	C	A	T	A	A	C	T	G	A	G	A	C	C	T	T	A	A	C	C	A	A	A	A	A	A	A	A	T	A	A	G	A
R001.1	C	T	A	A	T	T	G	A	G	A	G	A	C	A	C	T	T	T	G	C	A	A	A	T	T	T	C	C	A	T	G	A	C	T	A	A	G	A	C	T	C	T	A	A	T	T	A	A	A	A	A	A	A	T	G	A	G	A	
R002	C	T	G	G	G	T	G	A	A	A	G	A	C	A	C	T	T	T	G	G	A	A	T	T	C	T	C	C	A	A	A	A	C	C	A	A	A	A	T	T	T	T	G	G	C	T	A	A	G	A	G	A	T	T	A	A	A	A	
R002.1	C	T	G	G	G	T	G	A	G	A	G	A	C	A	T	T	T	T	G	G	A	A	A	T	C	C	C	C	A	A	A	A	C	C	G	A	A	A	C	T	T	T	G	A	T	T	A	A	G	G	G	G	A	T	A	A	A	A	
R003	C	T	G	A			A	A	G	A	G	A	C	A	C	T	C	T	G	C	G	A	A	T	C	T	C	C	A	T	G	A	C	T	A	A	G	A	C	T	C	T	G	A	C	T	G	G	G	A	G	A	A	T	A	A	G	A	
R004	C	T	A	A	G	T	G	A	G	A	G	G	C	C	C	T	T	G	C	A	A	T	T	C	T	C	C	A	A	G	G	C	C	G	A	G	A	C	C	C	T	G	G	T	T	G	G	G	A	G	A	T	T	G	A	G	A		
R005	C	T	A	A	T	T	G	A	G	G	G	A	C	A	C	C	T	T	G	C	A	A	T	T	T	T	C	C	A	T	G	A	C	C	A	A	G	A	C	T	C	T	G	A	C	C	G	A	A	A	A	A	A	A	G	A	G	G	
R005.1	C	C	A	A	G	T	G	A	G	G	G	A	C	C	C	C	T	T	G	C	A	A	T	T	T	T	C	C	A	T	G	A	C	C	A	A	G	A	C	T	C	T	G	A	T	T	G	G	G	A	A	A	A	T	A	A	G	G	
R005.2	C	C	G	A	G	G	A	A	G	A	G	A	C	C	C	C	T	T	G	C	A	A	T	T	C	T	C	C	A	T	G	A	C	C	G	A	G	A	C	C	G	A	G	A	T	T	G	G	G	A	A	A	A	T	G	A	G	G	
K001	C	T	A	A	T	T	A	A	G	G	A	A	C	C	C	T	T	T	G	C	G	A	A	A	C	T	C	C	A	T	A	A	C	T	A	A	A	A	C	T	C	T	A	A	T	T	G	G	A	A	G	A	A	A	A	A	A	G	A

## Referências bibliográficas

---

- [1] Vogel, F. & Motulski, A.G.. **Human Genetics: Problems and approaches**, 229-239 (Springer, Berlin, 1996).
- [2] Brookes A J. **The essence of SNPs**. *Gene* **234** (1999) 177. 186.
- [3] Sobrino B, Brión M, Carracedo A, **SNPs in forensic genetics: a review on SNP typing methodologies**. *Forensic Sci Int.* **154**(2005) 181-194.
- [4] Phillips C, Fondevila M, García-Magariños M, Rodriguez A, Salas A, Carracedo A, Laréu M V, **Resolving relationship testes that show ambiguous STR results using autosomal SNPs as supplementary markers**. *Forensic Science International: Genetics* **2** (2008) 198-204.
- [5] Pinheiro MF, **Identificação Genética: passado, presente e futuro**. *In* Revista do Ministério Público, Ano 30 (2009), Número 118.
- [6] Gill P, **An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes**. *Int. J. Legal. Med.* **114** (2001) 204. 210.
- [7] Amorim A, Pereira L, **Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs**. *Forensic Sci Int.* **150** (2005) 17. 21.
- [8] Sanchez J, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, Harrison C D, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider P M, Carracedo A, Morling N, **A multiplex assay with 52 single nucleotide polymorphisms for human identification**. *Electrophoresis* **27** (2006) 1713-1724.
- [9] Pakstis AJ *et al.* **SNPs for a universal individual identification panel**. *Hum. Genet.* **127** (2010) 315-324.

- 
- [10] Kidd, kk *et al.* **Developing a SNP panel for forensic identification of individuals.** *Forensic Sci Int.* **164** (2006) 20. 32.
- [11] Børsting C, Sanchez J J, Hansen H E, Hansen A J, Bruun H Q, Morling N, **Performance of the SNPfor ID 52-plex assay in paternity testing.** *Forensic Science International: Genetics* **2** (2008) 292-300.
- [12] Vallone PM, Decker AE, Butler JM, **Allele frequencies for 70 autosomal SNP loci with U.S. Caucasian, African-American, and Hispanic samples.** *Forensic Sci Int.* **149** (2005) 279. 286.
- [13] Phillips C, Fang R, Ballard D, Fondevila M, Harrison C, Hyland F, Musgrave- Brown E, Proff C, Ramos-Luis E, Sobrino B, Carracedo A, Furtado M R, Syndercombe-Court D, Schneider P M, **Evaluation of the Genplex typing system and a 49plex forensic marker panel.** *Forensic Science International: Genetics* **1** (2007)180-185.
- [14] Stangegaard M, Tomas C, Hansen A J, Frank-Hansen R, Børsting C, Morling N, **Typing of 48 autosomal SNPs and amelogenin with Genplex SNP genotyping system in forensic genetics.** *JALA October* 2008.
- [15] Tomas C, Stangegaard M, Børsting C, Hansen AJ, Morling N, **Biomek-3000 and Genplex SNP Genotyping in forensic genetics.** *Journal of Laboratory Automation* (2008) **13**: 297.
- [16] Jobling MA, **Y-chromosomal haplotype diversity in forensic analysis.***Forensic Sci Int.* **118**(2001): 158-162.
- [17] Just RS, Irwin JA, O'Callaghan JE, Saunier JL, Coble MD, Vallone PM, Butler JM, Barritt SM, Parsons TJ. **Toward increased utility of mtDNA in forensic identifications.** *Forensic Sci Int.* **146 Suppl** (2004) 147-149.

---

[18] Grimes EA, Noake PJ, Dixon L, Urquhart A, **Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype.** *Forensic Sci. Int.* **122** (2001) 124-129.

[19] Sobrino B, Brión M, Carracedo A, **SNPs in forensic genetics: a review on SNP typing methodologies.** *Forensic Sci Int.* **154** (2005) 181-194.

[20] Kidd KK, Kidd JR, Speed WC, Fang R, Furtado MR, Hyland FCL, Pakstis AJ, **Expanding data and resources for forensic use of SNPs in individual identification.** *Forensic Sci Int. Genet.* **6** (2009): 646-652.

[21] Lareu M, Puente J, Sobrino B, Quintans B, Brion M, Carracedo A, **The use of the LightCycler for the detection of Y chromosome SNPs.** *Forensic Sci. Int.* **118** (2001) 163. 168.

[22] Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvanen AC, **A system for specific, high throughput genotyping by allele-specific primer extension on microarrays.** *Genome Res.* **10** (2000) 1031. 1042.

[23] Landergren U, Kaiser R, Sanders J, Hood L, **A ligase-mediated gene detection technique.** *Science* **241** (1988) 1077. 1080.

[24] Barany F, **The ligase chain reaction in a PCR world.** *PCR Methods Appl.* **1** (1991) 5. 16.

[25] Grossman PD, Bloch W, Brinson E, Chang CC, Eggerding FA, Fung S, Iovannisci DM, Woo S, Winn-Deen ES, Iovannisci DA, **High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation.** *Nucleic Acids Res.* **22** (1994) 4527. 4534.

[26] Chen X, Livak KJ, Kwok PY, **A homogeneous, ligase mediated DNA diagnostic test.** *Genome Res.* **8** (1998) 549. 556.

---

[27] Sanchez J, Børsting C, Hallenberg C, Buchard A, Hernandez A, Morling N, **Multiplex PCR and minisequencing of SNPs: a model with 35 Y chromosome SNPs.** *Forensic Sci. Int.* 137 (2003) 74. 84.

[28] Quintans B, Alvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo A, **Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing.** *Forensic Sci. Int.* 140 (2004) 251. 257.

[29] Brion M, Sobrino B, Blanco-Verea A, Lareu MV, Carracedo A, **Hierarchical analysis of 30 Y-chromosome SNPs in European populations.** *Int. J. Legal Med.* 119 (1) (2005) 10. 15.

[30] Sobrino B, Lareu M, Brion M, Carracedo A, **SNP genotyping with single base extension-tag microarrays.** in: *Prog. Forensic Genet.*, tenth ed., Elsevier, 2004, pp. 331. 333.

[31] Mengel-Jorgensen J, Sanchez JJ, Børsting C, Kirpekar F, Morling N, **Multiplex Y chromosome SNP genotyping using MALDI-TOF mass spectrometry.** in: *Prog. Forensic Genet.*, tenth ed. Elsevier, 2004, pp. 15. 17.

[32] Børsting C, Tomas C, Morling N, **SNP typing of the reference materials SRM 2391b 1E10, K562,XY1, XX74, and 007 with the SNPforID multiplex.** *Forensic Sci Int: Genetics* 5 (2011) e81. e82.

[33] Pontes ML, Pinheiro MF, **Autosomal SNPs study of a population sample from North of Portugal and a sample of immigrants from the Eastern Europe living in Portugal.** *Legal Medicine* 16 (2014) 118. 120.

[34] Excoffier L, Laval G, Schneider S, **Arlequin (versión 3.5): a new series of programs to perform population genetics analyses under Linux and Windows.** *Mol Ecol Resour* 2010;10:564 . 7.

---

[35] Brenner C, Morris J (1990), In. **Proceedings of the International Symposium on Human Identification**. 1989, Promega Corporation.

[36] Evett, IW, Weir BS (1998), **Interpreting DNA Evidence**. Sinauer: Sunderland, MA.

[37] SNPforID browser; 2013 [accessed 25.07.13].

[38] Drobnic K, Børsting C, Rockenbauer E, Tomas C, Morling N. **Typing of 49 183 autosomal SNPs by SNaPshot™ in the Slovenian population**. *Forensic Sci Int* 184 Genet 2010;4:e125. 7.

[39] Vallone PM, Decker AE, Butler JM, **Allele frequencies for 70 autosomal SNP loci with U.S. Caucasian, African-American, and Hispanic samples**, *Forensic Sci. Int.* 149 (2005) 279. 286).

[40] Phillips C, Salas A, Sánchez J, Fondevila M, Gómez-Tato A, Álvarez-Dios J, et al. **187 Inferring ancestral origin using a single multiplex assay of ancestry- 188 informative marker SNPs**. *Forensic Sci Int Genet* 2007;1:273. 80.

[41] Børsting C, Mogensen HS, Morling N (2013) **Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples**. *Forensic Sci Int Genet* 7:345-352.

[42] Gill P, **An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes**. *Int. J. Legal. Med.* 114 (2001) 204. 210.

[43] The STRbase website lists a wide range of observed STR mutation rates at: <http://www.cstl.nist.gov/biotech/strbase/mutation.htm> based on data from the most extensive recent AABB survey, typing trios with standard STRs, outlined

---

at:[http://www.aabb.org/Documents/Accreditation/Parentage\\_Testing\\_Accreditation\\_Program/ptannrpt03.pdf](http://www.aabb.org/Documents/Accreditation/Parentage_Testing_Accreditation_Program/ptannrpt03.pdf)

[44] Børsting C, Sanchez JJ, Birk AH, Bruun HQ, Hallenberg C, Hansen AJ, Hansen HE, Simonsen BT, Morling N, **Comparison of paternity indices based on typing 15 STRs, 7 VNTRs and 52 SNPs in 50 Danish mother-child-father trios.** *Prog. Forensic Genet.* 11 (2006) 436. 438.

[45] Butler J M, Coble M D, Vallone P M (2007), **STRs vs. SNPs: thoughts on the future of forensic DNA testing.** *Forensic Sci Med Pathol* (2007) 3:200. 205.