# Prova Escrita – Concurso de Livre Docência

# Candidato: Prof. Dr. Celso Teixeira Mendes Junior

# Tema sorteado: 03 – Análise de STRs/SNPs

# Título:

# Análise de STRs e SNPs na prática forense:

# contribuição atual e perspectivas futuras

## *Short Tandem Repeats*

Desde o desenvolvimento da técnica de PCR (*Polymerase Chain Reaction*), os exames de identificação humana por DNA se baseiam na análise de marcadores do tipo STRs (*Short Tandem Repeats*), também conhecidos como microssatélites. Desde a descrição dos primeiros STRs para fins forenses, no início da década de 90, vários marcadores potencialmente aplicáveis à prática forense foram identificados. Estes marcadores são caracterizados pela existência de dezenas de alelos, os quais diferem no número de unidades de repetição que os caracterizam. Devido ao elevado número de alelos e à distribuição de frequências alélicas, altos níveis de heterozigose esperada (alcançando até 95%) são encontrados, o que está correlacionado a elevados poder de discriminação e poder de exclusão.

Os STRs são convencionalmente analisados em kits que atualmente incluem todos os 20 marcadores pertencentes ao sistema CODIS desenvolvido pelo FBI. Estes STRs são amplificados em uma única PCR multiplex, sendo que cada primer se encontra marcado com um fluoróforo dentre os quatro ou cinco fluoróforos distintos utilizados pelos kits. A utilização de fluoróforos distintos para marcar STRs cujos fragmentos apresentam tamanhos semelhantes permite a diferenciação de todos os marcadores em um processo de eletroforese capilar. Adicionalmente, estão presentes nestes kits de um a três marcadores voltados para a identificação do sexo do indivíduo que deu origem à amostra, como, por exemplo, o gene da amelogenina, que se encontra presente no cromossomo X e na região não recombinante do cromossomo Y, apresentando uma deleção de 6 pares de base (pb) na versão encontrada no cromossomo X.

A análise de STRs presentes no cromossomo Y também é rotineiramente empregada em exames de paternidade ou em situações em que se deseja estabelecer vínculo entre dois homens conectados a uma mesma linhagem paterna. Abordagens inovadoras permitem a determinação de haplótipos de Y-STRs no plasma materno, o que representa uma abordagem não invasiva para a realização de exames de paternidade (Barra et al., 2015). Entretanto, os Y-STRs presentes nos kits tradicionalmente comercializados pela *Promega* e *Applied* *Biosystems* (*ThermoFisher* *Scientific*) normalmente não permitem diferenciar dois parentes próximos de sexo masculino (pai/filho, ou dois irmãos). Para tanto, Ballantyne et al. (2012) identificou um conjunto de 13 Y-STRs caracterizados por alta taxa de mutação (superior a 10-2 vs. 10-3). Estudo populacional Analisando um conjunto de 604 homens revelou que este conjunto de marcadores apresenta um maior poder de discriminação quando comparado aos 17 Y-STRs que compõem o sistema *Y-fyler* (*ThermoFisher* *Scientific*), sendo sido observado apenas 3 haplótipos compartilhados por 8 homens, enquanto que o Y-fyler apresentou 35 haplótipos compartilhados por 85 homens. Este novo sistema apresenta maior potencial de diferenciar parentes próximos: quase 50% das duplas pais/filhos (comparado com 7,7% pelo *Y-fyler*) e 60% dos irmãos (comparado com 8,0% pelo *Y-fyler*)

Avanços tecnológicos relacionados à análise de STRs também visam contribuir para combater uma das principais dificuldades enfrentadas pelos órgãos periciais envolvidos com a análise de DNA, que consiste no processamento de um volume muito maior de amostras biológicas em relação ao que era analisado no início deste século. Este aumento no volume de análise decorre da ampla difusão das contribuições e aplicações tecnológicas relacionadas aos exames de DNA. Entretanto, muitas das amostras que apresentam DNA, acabam não sendo amplificadas com sucesso, quer seja pela presença de elementos inibidores, ou então pelo avançado estado de degradação. Visando avaliar precocemente este problema e reduzir a demanda de análises em laboratórios forenses, diversos estudos vêm buscando metodologias alternativas para análise de marcadores genéticos e triagem de amostras. Uma inovação de sucesso consiste no *ParaDNA® Screening System*, que consiste em um instrumento portátil de amostragem de material biológico e processamento do mesmo em cerca de 75 minutos, que pode ser manuseada por investigadores não especializados neste tipo de análise, na própria cena de crime (Dawnay et al., 2014). Este instrumento apresenta placas de análises já preparadas para realização de PCR direta (sem necessidade de extração de DNA) acoplada a detecção por sondas HyBeacon™. Além de detectar a capacidade de amplificação de cinco STRs convencionalmente empregados na prática forense (D3S1358, D8S119, D16S538, D18S1358 e TH01), e apresentar uma quantificação relativa da amostra, o sexo do indivíduo que deu origem à amostra é identificado pela análise do *locus* da Amelogenina. Com esta análise de triagem, o responsável pela análise pode então determinar se a amostra deve ser descartada pela ausência de DNA, deve ser analisada pela metodologia convencional, ou se deve ser analisada por metodologias alternativas mais adequadas para análise de material degradado (miniSTRs, isto é, STRs caracterizados por menores *amplicons*, SNPs ou mtDNA, por exemplo). Deste modo, aumenta-se a eficiência de todo o processo de obtenção de perfis de DNA. Este ensaio foi validado com sucesso, em termos de reprodutibilidade, sensibilidade, acurácia, tolerância a inibidores e performance em uma grande quantidade de amostras forenses simuladas, envolvendo sangue, saliva e DNA de toque (*touch DNA*) (Dawnay et al., 2014). Adicionalmente, a análise dos STRs por meio deste ensaio, em uma amostra de 381 indivíduos do Reino Unido revelou alta concordância de genotipagem quando comparados aos kits tradicionalmente utilizados na prática forense, o que indica que seus resultados compõem um pequeno perfil altamente consistente (Ball et al., 2015).

## *Single Nucleotide Polymorphisms*

SNPs (*Single Nucleotide Polymorphisms*) presentes nas regiões controladoras I e II (também conhecidas como hipervariáveis) da molécula de DNA mitocondrial (mtDNA) também vem sendo analisados em contextos forenses desde o advento da PCR. Embora a análise de mtDNA apresente como principal característica a facilidade de amplificação em material biológico escasso e altamente degradado, devido ao grande número de cópias da molécula de mtDNA dentro de uma mitocôndria e ao grande número de mitocôndrias por células, sua informatividade é bastante limitadas, uma vez que todos os indivíduos conectados por uma mesma linhagem materna apresentam mesmos haplótipos mitocondriais (desconsiderando-se a possibilidade de mutações). Deste modo, ao contrário dos perfis de marcadores autossômicos, os SNPs (e algumas indels – inserções/deleções) se organizam em um número limitado de haplótipos existentes em uma população. Ainda assim, são extremamente úteis em situações nas quais deseja-se estabelecer vínculos maternos (por exemplo, na identificação da família real russa vitimada durante a revolução Bolchevique) ou em desastres de massa caracterizados pela existência de múltiplas vítimas e material biológico altamente degradado (por exemplo, na identificação de vítimas do atentado ao *World Trade Center*)

Apesar da alta informatividade dos STRs, como abordado no item anterior, uma das limitações de tais marcadores esta associada à alta taxa de mutação (cerca de 10-3), o que faz com que inconsistências decorrentes de mutações sejam detectadas em um exame envolvendo um pai biológico e seu filho, ou então na análise de diferentes tecidos de um mesmo indivíduo. Caso em que três inconsistências em um conjunto de 13 STRs analisados em um caso de paternidade legítima foi descrito (Sun et al., 2012). Para piorar, o mesmo mecanismo mutacional observado *in vivo*, que envolve o deslizamento da polimerase, faz com que aproximadamente 15-20% do produto amplificado durante a PCR apresente um menor número de repetições (normalmente uma repetição) do que o alelo tomado como molde (*stutter*), o que complica o processo de detecção de misturas de perfis de DNA. Estas deficiências abrem as portas para o uso de SNPs autossômicos na prática forense.

Mais recentemente, com o desenvolvimento da metodologia de mini-sequenciamento, a análise de SNPs (*Single Nucleotide Polymorphisms*) autossômicos e presentes em cromossomos sexuais (X e Y) passou a ser incorporada na prática forense. O mini-sequenciamento consiste em uma variação do método de sequenciamento de Sanger. Resumidamente, um fragmento contendo um SNP de interesse é amplificado por PCR, o produto é purificado por meio de enzimas ExoSAP, e este produto é submetido a uma reação de mini-sequenciamento por meio de um primer único que posiciona sua extremidade 3’OH em local imediatamente adjacente ao SNP que se deseja genotipar. Visto que nesta reação de mini-sequenciamento apenas ddNTPs (cada um dos quatro tipos marcados com um fluoróforo específico), são utilizados, uma única base é incorporada ao *primer* em questão. Indivíduos homozigotos se caracterizam por incorporação de apenas um tipo de base ao primer, enquanto que heterozigotos apresentarão fragmentos que incorporam uma base ou outra. Assim como no caso da metodologia tradicionalmente empregada para análise dos STRs, a detecção é feita por meio de eletroforese capilar em sequenciadores automáticos de DNA. Cerca de 30 a 40 SNPs podem ser analisados simultaneamente em uma mesma reação, por meio da utilização de *primers* com tamanhos distintos que acabam apresentando migrações eletroforéticas distintas. Metodologias alternativas, como o pirossequenciamento a associação de mini-sequenciamento a plataformas de microarranjos de DNA também sejam alternativas viáveis, a metodologia aqui abordada é a mais empregada, visto que os equipamentos necessários são os mesmos utilizados na análise de STRs em laboratórios forenses.

Vários kits de mini-sequenciamento distintos vem sendo desenvolvidos pela comunidade forense e permitiram o estabelecimento do conceito de DNA intelligence, uma vez que os SNPs apresentam potencial para contribuir de forma mais contundente na resolução de crimes, ao permitir não só a identificação, mas também a inferência de ancestralidade e de característica morfológicas. Em uma revisão recente, Mehta et al. (2017) apontou a existência de sete ensaios para a análise de SNPs voltados para a identificação humana (IISNPs), 16 ensaios para detecção de linhagens de cromossomo Y ou mtDNA (LISNPs), cinco ensaios para avaliação de ancestralidade biogeográfica (AISNPs) e outros três kits para predição de fenótipos de pigmentação humana (PISNPs). Dentre eles, detacam-se os ensaios SNP*for*ID *52-plex* para identificação humana, SNP*for*ID *34-plex* para determinação de ancestralidade e *HIrisPlex* para predição de cor de olhos e cabelos.

Embora altamente estáveis e caracterizados por baixa taxa de mutação (cerca de 10-8), o que reduz drasticamente o número de inconsistências na análise entre um pai e seu filho, por exemplo, os SNPs normalmente são bi-alélicos e apresentam menor informatividade (cerca de quatro SNPs são necessários para atingir a informatividade de um único STR), sendo pouco útil para a detecção de misturas de dois ou mais perfis de DNA. Deste modo, iniciativas que visam vasculhar o genoma em busca de SNPs multi-alélicos, na tentativa de selecionar marcadores que apresentem maior poder de discriminação, vêm sendo conduzidas. A analise dos dados gerados pela fase 3 do Projeto *1000 Genomes* revelou a existência de 961 SNPs tetra-alélicos que passaram com sucesso pelos filtros de qualidade do próprio projeto (Phillips et al., 2015). Destes, 160 apresentaram heterozigose esperada superior a 50%. Um conjunto composto pelos 24 marcadores mais informativos, foi capaz de detectar contribuintes secundários em 99,9% das amostras misturadas compostas por dois indivíduos de origem africana, ou em 92,6% das amostras compostas por dois indivíduos do leste da Ásia (Phillips et al., 2015). Representam, portanto uma alternativa a ser desenvolvida para complementar a análise de STRs na detecção de misturas envolvendo elevados níveis de degradação.

## A revolução proporcionada pelo sequenciamento de nova geração

O desenvolvimento das diferentes tecnologias de sequenciamento massivo em paralelo, também denominadas como sequenciamento de nova geração (NGS), resultou na possibilidade de análise de fragmentos (*reads*) cada vez maiores. Atualmente, algumas metodologias consideradas de segunda geração permitem o sequenciamento de *reads* superiores a 500 pb, enquanto que as metodologias de terceira geração resultam no sequenciamento ininterrupto de fragmentos superiores a 10 kb. Apesar de que as tecnologias de terceira geração ainda não se mostrem adequadas a protocolos forenses, quer seja pela necessidade de maior quantidade de DNA molde para a análise, quanto pelas maiores taxas de erros de incorporação de nucleotídeos, as tecnológicas de sequenciamento *illumina* e *Ion* *Torrent* (*ThermoFisher* *Scientifc*) vêm revolucionando a prática forense.

A substituição da metodologia de eletroforese capilar por NGS para análise de STRs permite a análise dos mesmos marcadores. Entretanto, as primeiras iniciativas de análise revelaram que muitos dos marcadores apresentam um nível adicional de variabilidade não detectável pela metodologia tradicional de eletroforese capilar, uma vez que vários alelos acabam diferindo em termos de sequência, mas não em termos de tamanho. Este nível adicional de variabilidade é causado (i) pela presença de SNPs, tanto nas regiões que flanqueiam as repetições, quanto nas regiões repetitivas em si, e (ii) pela existência de diferentes padrões de repetições, que antes não eram diferenciados por eletroforese capilar. Gettings et al. (2015), utilizando dados coletados junto ao *GenBank*, *dbSNP* e Projeto *1000 Genomes*, avaliou a variação genômica ao redor de STRs e comparou os achados com as escadas alélicas presentes nos kits atuais de análise por eletroforese capilar. Observou enorme diversidade previamente desconhecida, o que sugere a necessidade de esforços por parte da comunidade forense para atualização da nomenclatura de STRs. Em estudo subsequente, Gettings et al. (2016) avaliaram o impacto deste nível adicional de variabilidade em 22 *loci* autossômicos, o que levou à identificação de que seis marcadores convencionalmente utilizados na prática forense apresentam o dobro de alelos em relação ao que se conhecia anteriormente.

No que se refere aos SNPs, abordagens recentes permitem incorporar em uma a tecnologia de NGS permitiu o desenvolvimento de microhaplótipos de até 200 pb que incluem três ou mais SNPs. Dois conjuntos de microhaplótipos revelaram-se informativos nas mais diferentes populações humanas, apresentando elevada heterozigose e baixo desequilíbrio de ligação entre si, o que os tornam adequados para as mais diversas aplicações forenses, com potencial de superar os STRs para a análise e interpretação de misturas, uma vez que a amplificação destes microhaplótipos não resultam na produção de *stutter* (Kidd et al., 2014; Kidd et al., 2017). Adicionalmente, considerando-se que a metodologia de NGS permite a inclusão de milhares de alvos distintos em uma mesma reação de sequenciamento, ensaios incorporando os diferentes ensaios normalmente analisados pela metodologia de mini-sequenciamento SNaPshot abordados anteriormente vem sendo desenvolvidos. Neste contexto, Daniel et al. (2015) desenvolveu um ensaio próprio envolvendo a plataforma *Ion Torrent™ PGM System* (*ThermoFisher Scientific*) que incorporou a análise de mais de 136 SNPs pertencentes a cinco kits diferentes, e avaliou o sucesso de análise em amostras caracterizadas por diferentes concentrações de DNA. Os autores observaram elevada concordância com a metodologia SNaPshot (97%), sendo que as duas abordagens comparadas apresentaram níveis de erros equivalentes quando as discrepâncias foram resolvidas por sequenciamento de Sanger.

De olho em uma fatia de um mercado crescente, a *ThermoFisher Scientific* e *illumina* vêm desenvolvendo soluções específicas envolvendo sequenciamento de nova geração. A *ThermoFisher Scientific* desenvolveu quatro ensaios distintos voltados para a plataforma *Ion Torrent* (Pereira et al., 2017): *Precision ID Identity Panel* (124 SNPs), *Precision ID Ancestry Panel* (165 SNPs), *Precision ID mtDNA Whole Genome Panel* e *Precision ID mtDNA Control Region Panel*. A *illumina*, por outro lado, desenvolveu o ensaio *ForenSeq* (Guo et al., 2017), que inclui 27 STRs autossômicos, 24 Y-STRs, 7 X-STRs, 94 SNPs para identificação, 56 SNPs para ancestralidade e 22 SNPs para inferência de fenótipos em um ensaio único, além de uma plataforma voltada para NGS de amostras forenses, o *MiSeq FGxTM Forensic Genomics System*.

Em resumo, a tecnologia de Sequenciamento de nova geração revolucionou a aplicação da análise de polimorfismos de DNA na prática forense, abrindo caminho para outras aplicações futuras até pouco tempo não imaginadas, como a identificação de tecidos por RNA-Seq, ou até mesmo a inferência de idade em amostras desconhecidas por *Methyl-Seq.*

## Bibliografia

Ball G et al. (2015). Concordance study between the ParaDNA® intelligence test, a rapid profiling assay, and a conventional STR typing kit (AmpFlSTR® SGM Plus®). Forensic Sci Int Genet 16:48-51.

Ballantyne KN et al. (2012). A new future of forensic Y-chromosome analysis: rapidly mutation Y-STRs for differentiating male relatives and paternal lineages. Forensic Sci Int Genet 6:208-218.

Barra GB et al. (2015). Fetal male lineage determined by analysis of Y-chromosome STR haplotypes in maternal plasma. Forensic Sci Int Genet 15:105-110.

Daniel R et al. (2015). A SNaPshot of next generation sequencing for forensic SNP analysis. Forensic Sci Int Genet 14:50-60.

Dawnay N et al. (2014). Developmental validation of the ParaDNA® screening system: a presumptive test for the detection of DNA on forensic evidence items. Forensic Sci Int Genet 11:73-79.

Gettings KB et al. (2015). STR allele sequence variation: current knowledge and future issues. Forensic Sci Int Genet 18:118-130.

Gettings KB et al. (2016). Sequence variation of 22 autosomal STR loci detected by next generation sequencing. Forensic Sci Int Genet 21:15-21.

Guo F et al. (2017). Massively parallel sequencing of forensic STRs and SNPs using the illumine ForenSeq DNA Signature Prep Kit on the MiSeq FGxTM Forensic Genomics System. Forensic Sci Int Genet 31:135-148.

Kidd KK et al. (2014). Current sequence technologies makes microhaplotypes a powerful new type of genetic markers for forensics. Forensic Sci Int Genet 12:215-224.

Kidd KK et al. (2017). Evaluating 130 microhaplotypes across a global set of 83 populations. Forensic Sci Int Genet 29:29-37.

Mehta B et al. (2017). Forensically relevant SNaPshot® assays for human DNA SNP analysis. A review. Int J Legal Med 131:21-37.

Pereira V et al. (2017). Evaluation of the Precision ID Ancestry Panel for crime casework: a SNP typing assay developed for typing 165 ancestral informative markers. Forensic Sci Int Genet 28:138-145.

Phillips C et al. (2015). Tetra-allelic SNPs: informative forensic markers compiled from public whole-genome sequence data. Forensic Sci Int Genet 19:100-106.

Sun HY et al. (2012). A paternity case with mutations of three CODIS core STR loci. Forensic Sci Int Genet 6:e61-62.