

Metilação do DNA e predição de idade para fins forenses



**Matheus de Sousa Ferrari¹, Thássia Mayra Telles Carratto¹,
Silviene Fabiana de Oliveira², Celso Teixeira Mendes Junior¹**

¹Departamento de Química, Laboratório de Pesquisas Forenses e Genômicas, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP

²Departamento de Genética e Morfologia, Laboratório de Genética, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF

Autor para correspondência – ctmendes@ffclrp.usp.br

Palavras-chave: genética forense, epigenética, fenotipagem forense por DNA

Em um cenário de investigações policiais, a possibilidade de extrair o maior número de informações da cena do crime se faz necessária, sendo de extrema importância a existência de metodologias forenses alternativas para a análise das amostras biológicas encontradas, visando a identificação das vítimas e dos agressores. Para isso, a metilação do DNA vem se destacando como um potencial marcador de idade cronológica, permitindo realizar a predição de idade de um indivíduo com alta precisão e acurácia dependendo do método utilizado, das regiões genômicas consideradas e da amostra analisada.

Biologia molecular nas ciências forenses

O DNA é um dos mais importantes objetos de análise das ciências forenses. No âmbito pericial, é comumente utilizado para fins de identificação individual, sendo que a principal análise realizada é a obtenção do perfil genético de um possível criminoso a partir de vestígios biológicos, como sangue, saliva, sêmen, dentre outros, encontrados na cena do crime. Este perfil pode ser comparado com amostras da(s) vítima(s), de suspeitos, de outros vestígios ou, ainda, ser pesquisado na Rede Integrada de Bancos de Perfis Genéticos. O DNA pode ainda ser usado para estimar ancestralidade e para prever características físicas, como cor dos olhos, cabelos e pele. A predição de características físicas, que é a fenotipagem forense, pode potencialmente auxiliar no reconhecimento dos atores de um crime, assim como na busca de pessoas desaparecidas. Outra área que vem ganhando destaque é o uso da epigenética no desenvolvimento de possíveis biomarcadores forenses.

O significado do termo “epigenética” vem sendo discutido na comunidade científica, sendo que uma das definições é: área de estudo das alterações no funcionamento dos genes que ocorrem sem que haja alterações na sequência de bases nitrogenadas do DNA. Dentro dessa área, pesquisas sobre o

que chamamos de mecanismos epigenéticos vêm sendo realizadas, como modificação de histonas, metilação do DNA e RNAs não codificantes, e suas relações com a expressão gênica, a diferenciação e a função celular dentre outros.

Metilação do DNA

No ano de 1948, o bioquímico Rollin Hotchkiss, analisando moléculas de DNA via cromatografia em papel, observou citosinas modificadas que, posteriormente, na década de 1960, foram identificadas como citosinas metiladas. Porém, somente no ano de 1980 foi descrito que a metilação do DNA é um processo que consiste na adição de um grupo metila (CH_3) na posição cinco de uma citosina, convertendo-a em 5-metilcitosina. Hoje, utiliza-se o termo “sítios CpG” para designar sítios no DNA onde uma citosina ocorre vizinha a uma guanina, sendo separadas por um fosfato, do esqueleto açúcar-fosfato do DNA. Essa expressão foi cunhada para descrever um nucleotídeo com citosina vizinho a um nucleotídeo com G na mesma fita do DNA e para distinguir da situação na qual a citosina ocorre emparelhada com uma guanina na fita complementar. Estão distribuídos por todo o genoma sendo classificados conforme a localização. Os sítios CpG mais estudados são aqueles que ocorrem em ilhas CpG, que são definidas como trechos de DNA com tamanho variando de 500 a 1.500 bp, com alta frequência de sítios CpG e com uma razão CG:GC maior que 0,6. Diversos genes têm uma ilha CpG localizada na sua região promotora.

A adição do radical metil é realizada por enzimas DNA metiltransferases (DNMTs) e sua remoção, por enzimas da família TET (*ten-eleven translocations*).

Para avaliar o status de metilação de uma região genômica de interesse, o método mais utilizado consiste no tratamento da amostra investigada com bissulfito de sódio (NaHSO₃), que converte as citosinas não metiladas em uracila, enquanto que as citosinas metiladas permanecem como citosinas.

Após essa etapa, é realizada uma amplificação da região de interesse por uma reação em cadeia da polimerase (PCR). Este procedimento faz com que as citosinas metiladas estejam presentes no produto amplificado como citosinas, enquanto que as citosinas não metiladas, que foram convertidas em uracilas, aparecem no produto amplificado como timinas. Dessa forma, utilizando diferentes métodos de sequenciamento de DNA é possível diferenciar facilmente as citosinas metiladas daquelas não metiladas (Figura 1).

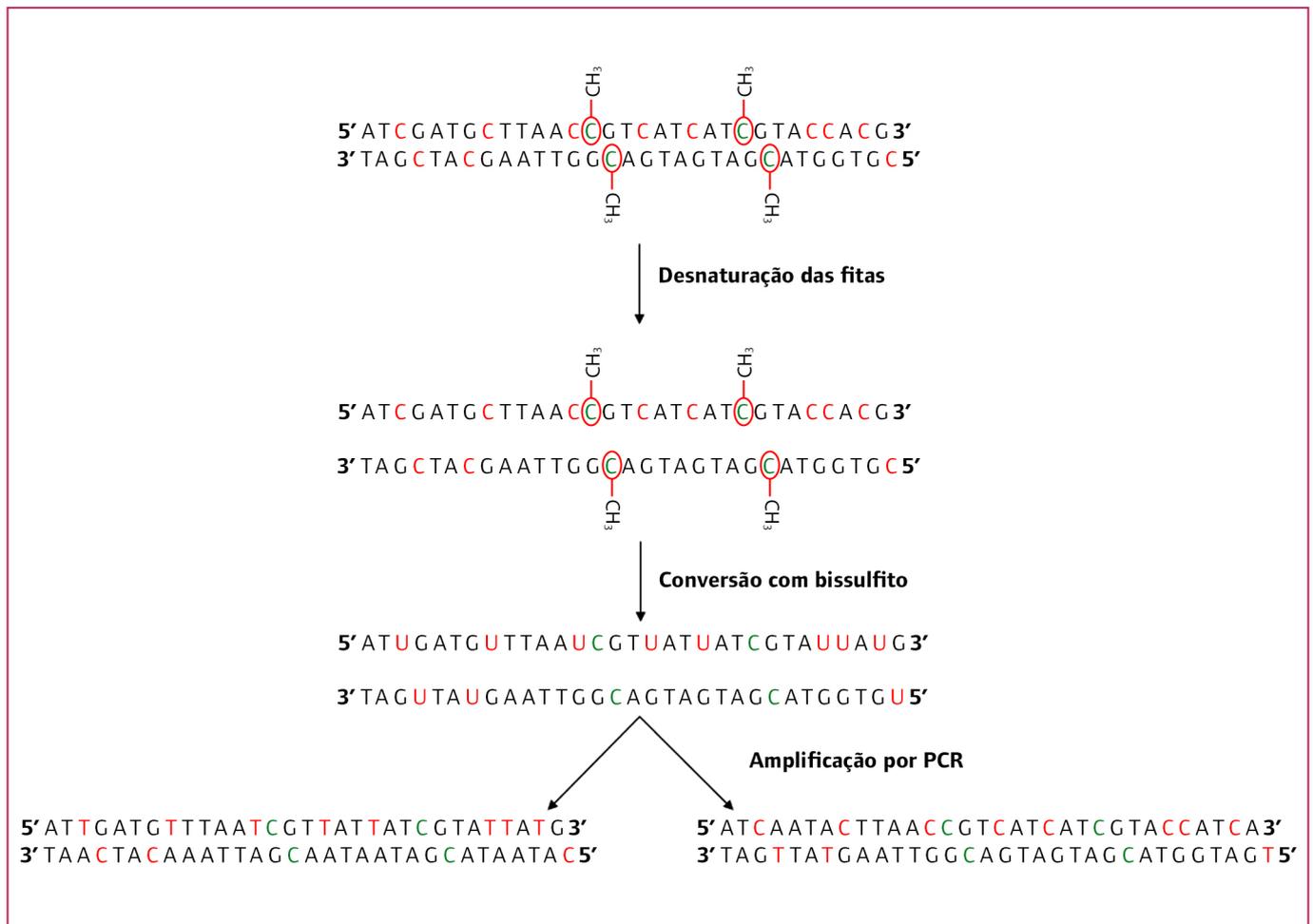


Figura 1. Modelo ilustrativo da conversão com bissulfito de moléculas de citosina não metiladas.

A metilação tem papel importante em diversas atividades biológicas, dentre elas está a regulação da expressão gênica sendo que a metilação contribui para silenciar um gene e, a desmetilação, para estimular sua expressão. Dessa forma, a metilação atua na diferenciação celular de forma que, ao regular a expressão de genes específicos, pode influenciar na especialização da célula para determinada

função. Por ser fortemente influenciada pelo ambiente e estilo de vida, a metilação média o impacto destes agentes externos em diversos processos fisiológicos como, por exemplo, no envelhecimento. Uma vertente nos estudos de metilação e envelhecimento é a possibilidade de predição de idade biológica a partir da avaliação de algumas regiões do genoma (Figura 2).

Relógios epigenéticos

De acordo com estudos, a idade cronológica tem uma relação direta com os níveis de metilação em sítios CpG específicos, o que vem tornando possível desenvolver estimadores de idade chamados de “relógios epigenéticos”. Um dos maiores desafios em se estabelecer um relógio epigenético está no fato de que os níveis de metilação podem ser fortemente influenciados por inúmeros fatores. Em um dado locus, a metilação pode ocorrer em três condições: 0, 1 ou 2 alelos metilados. Ainda, pode haver a heterogeneidade nos padrões entre as células, o que dificulta as análises e as interpretações

dos dados. De acordo com um estudo realizado em 2020, alterações aleatórias na metilação do DNA de humanos são comuns e muitas dessas têm sido associadas a doenças relacionadas à idade, incluindo diabetes do tipo 2, doenças cardiovasculares, doenças neurodegenerativas e alguns tipos de câncer. Além disso, determinados fatores externos podem influenciar a precisão dos modelos de predição de idade como, por exemplo, consumo de álcool, consumo de cigarros, dieta, prática de exercícios físicos, ou até mesmo exposição solar. Portanto, é de fundamental importância a identificação de regiões genômicas cujo padrão de metilação apresente forte correlação com a idade e que não seja influenciada por tais fatores exógenos (Figura 2).

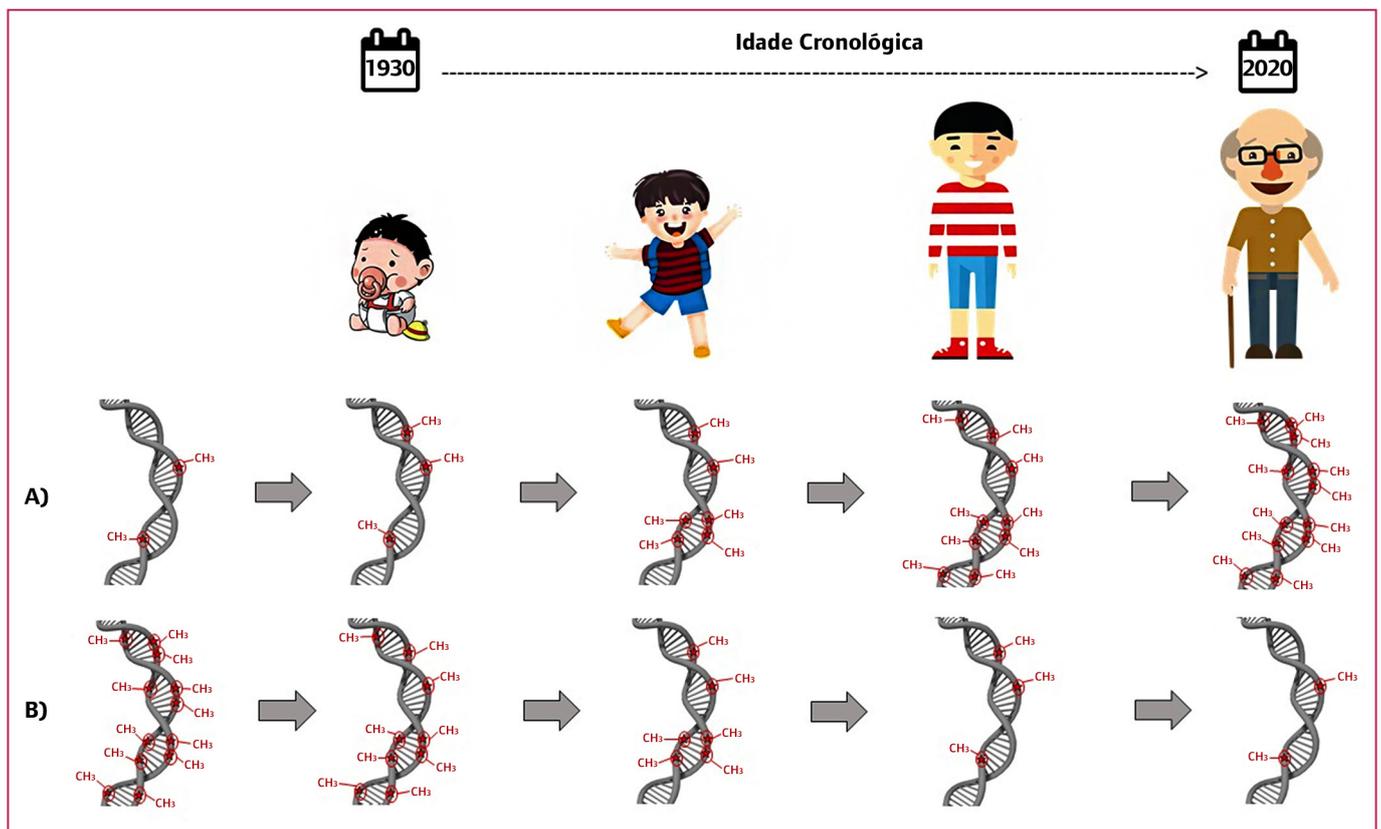


Figura 2.

Alteração do padrão de metilação (representada por estrelas vermelhas) do DNA de acordo com o avanço da idade cronológica. Enquanto existem genes que tendem a ser mais metilados conforme a idade avança (A), como os genes *ELOVL2*, *FHL2* e *KLF14*, outros tendem a se desmetilarem com o passar do tempo (B), como é o caso dos genes *HOXC10* e *HOXC-AS2*.

Segundo esse raciocínio, buscou-se identificar sítios CpG com padrões de metilação fortemente correlacionados com a idade e tentar caracterizar tais “relógios epigenéticos” baseados no nível de metilação de sítios de CpG. Foi constatado que o número de sítios

CpG utilizados influencia a eficácia do relógio, sendo que os resultados mais precisos são obtidos quando um maior número de sítios CpG é utilizado. Além disso, há relógios que são eficazes somente em alguns tecidos e outros, em vários tipos de tecidos.

Variabilidade populacional e a predição da idade a partir da metilação do DNA

A partir do melhor conhecimento acerca dos padrões de metilação em indivíduos de diferentes idades, foi possível desenvolver diferentes modelos para a predição de idade com alta precisão. Os modelos desenvolvidos utilizam diferentes amostras biológicas coletadas de cenas de crime. Nesse contexto é necessária a preocupação com a degradação e contaminação do material coletado que, por estar exposto a insultos ambientais, pode resultar em danos nos sítios CpG, e dificultar a interpretação dos dados.

A amostra biológica mais comumente encontrada em cenas de crimes é o sangue. Por isso, muitos modelos preditivos desenvolvidos no contexto forense são específicos para sangue.

Uma das ferramentas mais conhecidas para predição de idade foi desenvolvida e vem sendo aprimorada por um grupo de pesquisadores da Universidade de Santiago de Compostela. O primeiro modelo foi publicado em 2016, para predição da idade em amostras de sangue utilizando sete sítios CpG, localizados nos genes *ELOVL2*, *ASPA*, *PDE4C*, *FHL2*, *CCDC102B*, *C1orf132* e no locus *chr16:85395429*. O sítio CpG, localizado no gene *ELOVL2* mostrou-se o principal preditor, apresentando um desvio médio de três anos com uma taxa de acerto de aproximadamente 80%. Posteriormente, novos modelos foram criados e a ferramenta passou a incluir modelos preditivos independentes para predição da idade a partir de amostras de sangue, sêmen e saliva. A ferramenta está disponível *online* gratuitamente no site *Snipper app suite 3*.

No ano de 2013, Steve Horvath desenvolveu o primeiro relógio epigenético

“multitecido”, denominado de relógio epigenético de Horvath (*Horvath's epigenetic clock*). Sabendo que cada tecido apresenta padrões distintos de metilação, Horvath e seu grupo analisaram 7.844 amostras para avaliar os padrões e os níveis de metilação de mais de 450 mil sítios CpGs em 51 tipos de tecidos e células. Com isso, desenvolveram uma ferramenta capaz de prever a idade a partir dos níveis de metilação de 353 sítios CpG. Esta ferramenta ficou conhecida pela sua capacidade de predição com alta precisão (erro médio de cerca de quatro anos), independente do tipo de tecido analisado. O modelo preditivo de Horvath está disponível para uso de forma *online* e gratuita no site *DNA Methylation Age Calculator*.

Em 2019, utilizando sequenciamento após conversão por bissulfato, foram identificados e analisados sete sítios CpG nos genes *ASPA*, *EDARADD*, *CCDC102B*, *ZNF423*, *ITGA2B*, *KLF14* e *FHL2* em amostras de sangue de chineses. Os genes que apresentaram maior correlação com a idade foram o *CCDC102B*, o *KLF14* e o *FHL2*. Com esse estudo foi possível identificar que genes fortemente correlacionados com a idade em populações europeias, tais como *ELOVL2* e *TRIM59*, podem não apresentar a mesma correlação em outras populações. Em 2020, amostras de sangue *post-mortem* provenientes de indivíduos estadunidenses foram analisadas. Após o sequenciamento foram identificados cinco sítios CpG nos genes *ELOVL2*, *C1orf132*, *TRIM59*, *KLF14* e *FHL2*, previamente associados à idade. Os genes que apresentaram maior correlação com a idade foram o *TRIM59*, seguido do *ELOVL2*. O modelo preditivo proposto alcançou uma taxa de acerto de 73,1%, associado a uma variação na idade predita em relação à idade cronológica de cerca de 10,57 anos, devido à dificuldade do modelo em prever idades de indivíduos mais velhos, o que não é desejável.

Posteriormente, em 2021, foi realizado um experimento em duas fases, utilizando amostras japonesas. Após a conversão com bissulfato, foi realizado NGS para

determinar os níveis de metilação. Na primeira fase, sete sítios CpG nos genes *ASPA*, *CCDC102B*, *C1orf132*, *ELOVL2*, *ITGA2B*, *KLF14* e *PDE4C* foram analisadas em amostras de sangue *post-mortem* de adultos. Metilação no gene *ELOVL2* mostrou-se novamente altamente correlacionada com a idade. Na segunda fase, foram utilizadas amostras de sangue de indivíduos com idades variando de dois meses até 91 anos. Quatro sítios CpG nos genes *ASPA*, *ELOVL2*, *ITGA2B* e *PDE4C* foram analisados. A partir dos modelos criados foi possível obter um desvio médio em torno de seis anos e uma acurácia de aproximadamente 94%. Não foram observadas diferenças significativas nas idades estimadas para o sexo masculino e para o sexo feminino. Ainda no ano de 2021, foram analisadas amostras de sangue de indivíduos chineses de Singapura, Malásia e Índia, utilizando quatro sítios CpG nos genes *ELOVL2*, *TRIM59*, *KLF14* e *FHL2*. O melhor modelo proposto apresentou um desvio médio de 3,7 anos e taxa de acerto de 75,5%.

De acordo com os estudos realizados com amostras de sangue em indivíduos de diferentes populações, é possível observar que determinados genes são constantemente apontados como correlacionados à idade, tendo o gene *ELOVL2* um papel de destaque na maior parte das vezes. De modo geral, os desvios são baixos e as taxas de acertos dentro do intervalo estabelecido pelo desvio médio são elevadas, o que evidencia o potencial de tais ferramentas para a predição de idade. Entretanto, é importante ressaltar que as pesquisas são ainda recentes, e também evidenciam a possibilidade de se aumentar a acurácia dos resultados de modo a se aproximar a um patamar ideal para que contribua de maneira mais eficiente nas investigações forenses.

Além do sangue, outros fluídos destacam-se por serem frequentemente encontrados em cenas de crimes. Amostras de sêmen são comumente encontradas em casos de violência sexual. Por ser um fluido biológico, pode ocorrer grandes perdas de amostra até que ocorra a coleta, além da facilidade

de contaminação. Um estudo conduzido em 2015 com amostras de sêmen de coreanos selecionou, a partir do resultado de NGS, dois sítios CpG nos genes *TTC7B* e *NOX4*, além do sítio CpG *cg12837463*, para predição da idade. O gene que melhor apresentou correlação com a idade foi o *TTC7B*, com um desvio médio de 4,7 anos e uma taxa de acerto de 90%. Outro fluido biológico bastante encontrado é a saliva que, por se encontrar na região bucal, recebe uma maior proteção em relação aos agentes externos e se torna menos passível de contaminação. Uma pesquisa realizada em 2017, utilizando amostras de indivíduos coreanos, selecionou para investigação sete sítios CpG nos genes *SST*, *CNGA3*, *KLF14*, *TSSK6*, *TBR1*, *SLC12A5* e *PTPN7*. O modelo apresentou uma alta precisão, com um desvio médio de 3,1 anos e uma taxa de acerto de 94,5%.

Outro tipo de amostra que é facilmente encontrada em cenas de crime são os ossos. Devido à resistência, eles conseguem manter o DNA preservado, e por isso podem ser considerados valiosos para se estudar marcadores de predição de idade, principalmente em ossos fragmentados que não se mostram úteis para se estimar a idade por métodos antropométricos. Em 2019, foram analisados cinco sítios CpG, localizados nos genes *ELOVL2*, *FHL2*, *KLF14*, *C1orf132/MIR29BC* e *TRIM59*, em esqueletos de identidades desconhecidas. Os pesquisadores observaram que, quando os indivíduos possuem 30 ou mais anos de idade, os marcadores predizem uma idade consideravelmente menor do que a idade estimada por métodos antropométricos. Comparando com dados da literatura baseados em sangue, saliva e *swabs* bucais obtidos com os mesmos marcadores, os resultados das análises em ossos apresentaram uma grande variação no nível absoluto de metilação de DNA em três sítios CpG (*ELOVL2*, *FHL2* e *TRIM59*), produzindo um grande erro na predição da idade. A partir dessas primeiras análises, foi realizada uma busca para encontrar novos marcadores que tivessem boa correlação com a idade em ossos, o que culminou em 14 pos-

síveis marcadores CpG adjacentes ou presentes nos genes *C5orf66*, *ARHGAP20*, *TCF21*, *SMPD3*, *BCL11A*, *ANKH*, *TMEM51*, *MIR764-2/CDH1*, *EPHA6*, *LCMT1*, *SPIDR/KIAA0146*, *C14orf72/LINC00239*, *MYO18A* e *LINC00968*. Entretanto, não foram propostos modelos preditivos baseados neste conjunto.

Além dos ossos, os dentes são as estruturas que mais resistem a mortes violentas, como em desastres de massa, por exemplo, por se situarem em regiões extremamente protegidas, além de terem o suporte da mandíbula. Embora a arcada dentária possa ser empregada diretamente para a determinação de idade, os métodos odontológicos apresentam baixa acurácia para estimativas em adultos. Assim, os métodos moleculares tornam-se importantes alternativas para casos como esses. Um estudo de 2021 utilizou amostras de molares para analisar os níveis de metilação, a partir de ensaios por sequenciamento, em 46 sítios CpG localizadas nos genes *ELOVL2*, *KLF14*, *SCGN*, *NPTX2* e *FHL2*. Valores significativos de correlação entre metilação e idade foram obtidos para apenas sete desses sítios CpG localizados nos genes *ELOVL2*, *KLF14* e *NPTX2*. Foi construído um modelo preditivo utilizando a maior parte dos sítios CpG nos genes *ELOVL2*, *NPTX2*, *KLF14*, *SCGN* e alguns CpGs individuais do gene *FHL2*, levando a um erro médio absoluto de apenas 1,55 anos entre a idade estimada e a idade cronológica e uma incrível taxa de acerto de 98%, ainda mais se considerado o menor erro absoluto já descrito.

Por fim, amostras de cabelo também podem ser utilizadas para identificação de vítimas ou agressores e são de fácil coleta, uma vez que os fios desprendem-se facilmente mas, ao mesmo tempo, podem ser contaminados com facilidade, dependendo do ambiente. Embora os estudos sejam escassos, um estudo na Alemanha demonstrou, por NGS, níveis de metilação em sítios CpG nos genes *ELOVL2*, *KLF14*, *RPA2*, *TRIM59* e *ZYG11A*, que apresentaram alta correlação com a idade.

Direções futuras

Os padrões de metilação no DNA, na forma de preditores de idade para a composição de perfis fenotípicos em casos criminais, vêm se mostrando como promissores aliados às ciências forenses. Os estudos realizados realçam como sítios CpG específicos podem se mostrar bons preditores de idade para diferentes tipos de tecido e com boa acurácia na predição. Apesar de se ter um erro muito pequeno entre a idade real e a prevista das amostras, observa-se em alguns casos uma variação maior decorrente de influências externas. Por se tratar de uma metodologia recente, os modelos ainda não permitem lidar adequadamente com tais influências externas, o que requer estudos adicionais para aprimorá-los. Para obtenção de modelos mais consistentes, é imprescindível que a comunidade científica continue os esforços para identificar os sítios CpG com melhor precisão, buscando encontrar aqueles que sejam menos influenciados pelo meio externo, ou mesmo identificando os padrões gerados por essas interferências para incorporação em modelos preditivos.

Enquanto estudos desta natureza são praticamente inexistentes no Brasil, existem na Europa muitos grupos de pesquisa que estudam a predição da idade a partir do DNA. Em 2017, várias instituições de oito países europeus uniram-se no consórcio VISAGE (Atributos visíveis a partir da genômica, do inglês, *Visible Attributes Through Genomics*), cujo objetivo central é o desenvolvimento de painéis forenses para predição de ancestralidade, características físicas e idade. Para determinação da idade, dois tipos de ferramenta estão sendo avaliados, a básica e a avançada, sendo a primeira para predição de idade em amostras de sangue a partir de marcadores presentes em cinco genes, e a segunda para predição em amostras de células somáticas (sangue, células bucais e ossos) ou de sêmen, a partir de oito e treze genes, respectivamente. Tais ferramentas ainda estão sendo validadas sendo que, de maneira geral, os estudos

preliminares apresentam erros médios em torno de três a cinco anos.

Os estudos demonstraram que a metilação de alguns genes manteve-se eficaz na predição de idade mesmo em populações diferentes. Assim, é possível observar que os métodos desenvolvidos podem ser utilizados em qualquer população com uma certa acurácia. Porém, é essencial que estudos desta natureza também sejam conduzidos no Brasil, um país de dimensões continentais caracterizado por uma grande variedade populacional, de ambientes e estilos de vida.

Com o aprimoramento dos modelos preditivos de idade, abre-se a possibilidade de uma nova contribuição da genética para a elucidação de crimes. Ao se predizer a idade de uma vítima ou de um agressor a partir de um vestígio biológico, novas pistas são geradas, o que pode ser essencial para trazer um novo foco às investigações, aumentando, portanto, as chances de se solucionar casos criminais complexos e aparentemente insolucionáveis, tão frequentes no Brasil.

Para saber mais

Calculadora *online* do modelo preditivo do relógio de Horvath: <https://dnamage.genetics.ucla.edu>

Calculadora *online* do modelo preditivo desenvolvido por pesquisadores da Universidade de Santiago de Compostela: http://mathgene.usc.es/snipper/age_models.php

HORVATH, S.; RAJ, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, v. 19, p. 371-384, 2018.

KOOP, B. E. et al. Epigenetic clocks may come out of rhythm—implications for the estimation of chronological age in forensic casework. *International Journal of Legal Medicine*, v. 134, p. 2215–2228, 2020.

SIMPSON, D. J.; CHANDRA T. Epigenetic age prediction. *Aging Cell*, v. 20, p. e13452.

ZHENG, S. C. et al. Epigenetic drift, epigenetic clocks and cancer risk. *Epigenomics*, v. 8, p. 705-719, 2016

