

Przewidywanie cech wyglądu człowieka na podstawie markerów DNA do celów medyczno-sądowych i kryminalistycznych

prof. dr hab. Maria Kaczmarek¹✉,

lic. Wojciech Zysek²

¹Institut Biologii i Ewolucji Człowieka, Wydział Biologii, Uniwersytet im. Adama Mickiewicza w Poznaniu

²Institut Biologii Molekularnej i Biotechnologii, Wydział Biologii, Uniwersytet im. Adama Mickiewicza w Poznaniu

https://doi.org/10.18388/pb.2021_449

✉ autor korespondujący: maria.kaczmarek@amu.edu.pl

Słowa kluczowe: genetyka sądowa, DNA, polimorfizm, cechy wyglądu zewnętrznego

Wykaz skrótów: FDP – kryminalistyczne fenotypowanie DNA (ang. *forensic DNA phenotyping*); STR – krótkie powtórzenia tandemowe (ang. *short tandem repeats*); EVC – cechy wyglądu zewnętrznego (ang. *externally visible characteristics*); GWAS – badanie asocjacyjne całego genomu (ang. *genome-wide association study*); SNP – polimorfizmy pojedynczego nukleotydu (ang. *single nucleotide polymorphism*); MPS – masowe sekwencjonowanie równoległe (ang. *massively parallel sequencing*); NGS – sekwencjonowanie nowej generacji (ang. *next-generation sequencing*); DIP – polimorfizm delecyjno-insercyjny (ang. *deletion-insertion polymorphism*); SBE – wydłużanie łańcucha DNA o pojedynczą zasadę (ang. *single-base extension*); MSE – błąd średniokwadratowy (ang. *mean square error*); mtDNA – DNA mitochondrialne; EWAS – badania asocjacyjne obejmujące cały epigenom (ang. *epigenome-wide association study*); MPB – łysienie typu męskiego (ang. *male pattern of baldness*); AGA – łysienie androgenowe (ang. *androgenic alopecia*)

STRESZCZENIE

Celem niniejszej pracy jest przedstawienie aktualnego stanu wiedzy na temat markerów genetycznych umożliwiających przewidywanie cech fenotypowych człowieka na podstawie próbki DNA wyekstrahowanego z materiału biologicznego pobranego z miejsca zdarzenia w zastosowaniu do celów kryminalistycznych i w naukach sądowych. Taka próbka DNA określana jest jako „biologiczny świadek”, a samą procedurę nazywa się kryminalistycznym fenotypowaniem DNA (FDP). Analityczną część pracy stanowi systematyczny przegląd literatury z lat 2015–2021 dokonany według metodologii PRISMA, przeprowadzony w marcu 2021 roku za pomocą multiwyszukiwarki EBSCO Discovery Service (EDS) dostępnej w bibliotece Uniwersytetu im. Adama Mickiewicza w Poznaniu. Omówiono podstawy molekularne FDP, markery DNA stosowane do przewidywania płci, wieku, pochodzenia biogeograficznego oraz cech fizycznych takich jak: kolor skóry, oczu, włosów, morfologia włosów, morfologia twarzy, obecność piegów, wysokość ciała, masa ciała (otyłość), łysienie typu męskiego i krótkowzroczność. W dyskusji zwrócono uwagę na trudności metodyczne wynikające z poligenicznego modelu dziedziczenia badanych cech, problemy społeczno-etyczne towarzyszące kryminalistycznemu fenotypowaniu DNA oraz nakreślono perspektywy rozwoju tych badań.

WPROWADZENIE

Materiał biologiczny (pochodzenia tkankowego, wydzieliny lub wydaliny) zabezpieczony w trakcie oględzin miejsca zdarzenia może być wykorzystany do wykreślenia profilu genetycznego w oparciu o analizę markerów znajdujących się w różnych miejscach genomu i charakteryzujących się wysokim stopniem różnorodności DNA – polimorfizmami, czyli krótkimi powtórzeniami tandemowymi (ang. *short tandem repeats*, STR) [1]. Prawdopodobieństwo związku z osobą, której ten materiał dotyczy jest prawie stuprocentowe (większe niż 99,99%), wynika bowiem z wysokiej mocy dyskryminacyjnej STR. Oznacza to, że każdy człowiek w populacji światowej, z wyjątkiem bliźniąt jednojajowych, posiada w swoim genomie część indywidualną, która jest niepowtarzalna [2]. Dlatego też dowody pochodzące z analiz DNA są uważane za złoty standard wśród technik kryminalistycznych i jedno z najskuteczniejszych narzędzi w identyfikacji osobniczej [3]. Profil DNA (profil STR) wyizolowany z materiału biologicznego zabezpieczonego do identyfikacji osób zostaje następnie porównany z materiałem pobranym od osoby podejrzanej, co pozwoli potwierdzić lub wykluczyć jej udział w zdarzeniu. Może również posłużyć do identyfikacji osób zaginionych lub niezidentyfikowanych ciał gdy te profile zostaną porównane z profilami osób spokrewnionych (rodzice, rodzeństwo, kuzynostwo itd.). Zasadniczym ograniczeniem tej standardowej kryminalistycznej analizy DNA jest jednak to, że nie pozwala na identyfikację osób, których profile genetyczne nie zostały jeszcze uzyskane przez organy prowadzące dochodzenie i nie figurują w bazie danych [4].

W obliczu takich trudności uwaga badaczy została przekierowana na możliwość wykorzystania dostępnych technik analitycznych do przewidywania cech wyglądu zewnętrznego (ang. *externally visible characteristics*, EVC) na podstawie markerów DNA uzyskanych bezpośrednio z próbki materiału biologicznego pobranej z miejsca zdarzenia lub uzyskanej z nieznanego ciała. Taką próbkę określa się mianem „biologicznego świadka”, a metodykę przewidywania cech fizycznych określa się pochodzącym z języka angielskiego terminem *forensic DNA phenotyping* (w skrócie FDP), które w Polsce funkcjonuje jako kryminalistyczne fenotypowanie DNA [5]. Zasadniczo wyniki FDP potencjalnie mogą dostarczać jeszcze dokładniejszych informacji niż naoczni świadkowie, których wiarygodność jest często poddawana w wątpliwość [6]. W związku z tym oczekuje się, że FDP może być szczególnie przydatne w sprawach dochodzeniowych, w których nie ma potencjalnych podejrzanych i nie ma zgodności między badaną próbką dowodowego DNA a profilami genetycznymi wprowadzonymi do kryminalnych baz danych. Oczekuje się, że FDP będzie również przydatny do identyfikacji osób za-

ginionych, tj. w przypadkach, gdy referencyjny profil DNA jest niedostępny. Należy dodać, że FDP nie zastępuje profilowania STR w procesie dochodzeniowym, lecz służy jako dodatkowe narzędzie pozwalające zawęzić krąg potencjalnych sprawców przez co ułatwia prowadzenie śledztw, szczególnie w przypadkach, w których nie ma żadnego znanego podejrzanego. Wyniki badań FDP mogą stanowić samodzielny dowód w sprawie, najczęściej jednak są jednym z dowodów, w tym dowodów innych niż biologiczne.

Początki badań FDP sięgają 2000 roku i od tego czasu, pomimo wielu trudności, obserwuje się ich rozwój wraz z szerokim wykorzystaniem do celów medyczno-sądowych i kryminalistycznych [7]. Powiązanie badań podstawowych (genetyka molekularna) z praktycznym wykorzystaniem wyników tych badań w kryminalistyce wzbudziło nasze zainteresowanie stanowiąc podwaliny niniejszej pracy. Jej celem jest przedstawienie aktualnego stanu wiedzy na temat markerów genetycznych umożliwiających przewidywanie cech fenotypowych (FDP) na podstawie próbki DNA wyekstrahowanego z materiału biologicznego pobranego z miejsca zdarzenia.

MATERIAŁ I METODY

Niniejsza praca opiera się na systematycznym przeglądzie literatury dokonany według metodologii PRISMA [8]. Wyszukiwanie przeprowadzono w marcu 2021 roku. Do tego celu wykorzystano multiwyszukiwarkę EBSCO Discovery Service (EDS) dostępną w bibliotece Uniwersytetu im. Adama Mickiewicza w Poznaniu. Wybór EDS wynikał z możliwości automatycznego przeszukania wszystkich baz bibliograficznych (16 obszernych naukowych baz danych, w tym MEDLINE i Web of Science) oraz dostęp do pełnego tekstu prac oferowany użytkownikom sieci zasobów UAM. Uzupełniająco przeszukano zasoby bazy Google Scholar. W badaniu uwzględniono empiryczne, oryginalne artykuły naukowe publikowane w recenzowanych czasopiśmie naukowych w okresie 2015-2021. Wykluczono przeglądy literatury by uniknąć powtarzania badań, metaanalizy, studia przypadków, niepublikowane rozprawy doktorskie, komunikaty oraz abstrakty konferencyjne.

Wyszukiwanie odbywało się przez zastosowanie następujących kryteriów włączenia:

1) temat – publikacje powinny zawierać w tytule, streszczeniu lub słowach kluczowych wyrażenia dotyczące „genetic markers”, „forensic DNA phenotyping”, „external visible characteristics”, „technology of DNA profiling”; wyszukiwanie przebiegało w kilku etapach zgodnie ze strategią formowania klas przez łączenie wymienionych wyrażen operatorem „and”;

2) projekt badania – do przeglądu literatury kwalifikowano badania empiryczne;

3) rok wydania – uwzględniono prace opublikowane w okresie od stycznia 2015 do marca 2021;

4) status publikacji – recenzowane, pełnotekstowe publikacje, rekordy z wydawnictw książkowych oraz publikacje zawierające treści powiązane z głównym zagadnieniem pracy;

5) język publikacji – angielski.

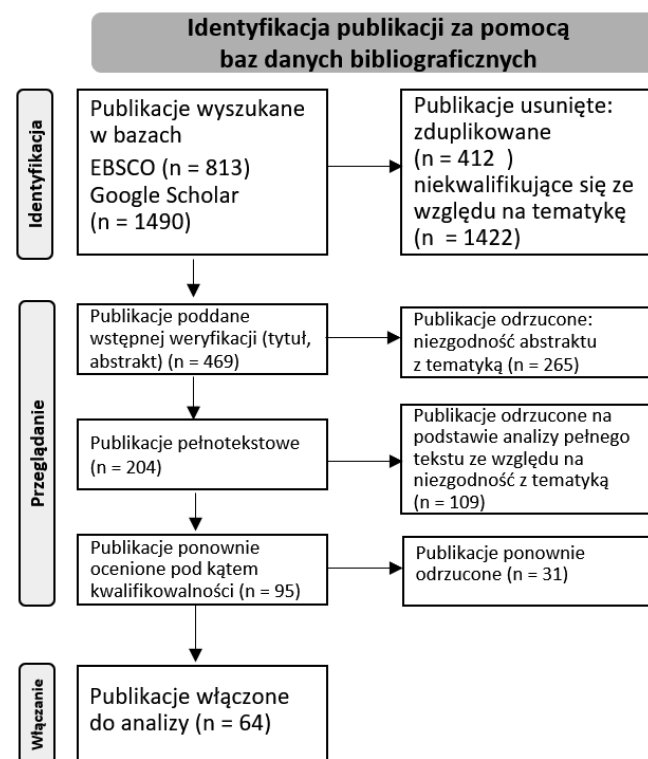
Kolejne etapy selekcji publikacji do niniejszego przeglądu literatury zgodne z metodologią PRISMA ilustruje schemat blokowy na rycinie 1.

Ostatecznie w niniejszym przeglądzie uwzględniono 64 publikacje, z czego najwięcej było opublikowanych w *Forensic Science International: Genetics*.

WYNIKI

PRZEWIDYWANIE FENOTYPU NA PODSTAWIE MARKERÓW DNA: PRZEGLĄD ZAGADNIENI

Przewidywanie cech fizycznych na podstawie znalezionego na miejscu zdarzenia materiału biologicznego takiego jak krople krwi, kosmyki włosów, czy drobne fragmenty ciała, oparte jest na przesłance, że każdy osobnik (z wyjątkiem bliźniąt jednojajowych) posiada jedyny, niepowtarzalny profil genetyczny [2]. FDP, choć zapoczątkowane w 2000 roku, dopiero w ciągu ostatniej dekady obserwuje się wyraźny postęp w wykorzystaniu tych badań jako narzędzia w kryminalistyce i to w zależności od ram prawnych danego kraju. Ma to ścisły związek z opracowanymi w ostatnich latach nowymi technikami analizy całego genomu i wprowadzeniem na szeroką skalę bioinformatycznych metod analizy danych [9]. Chociaż znane były niektóre geny kodujące EVC (najlepiej poznane dla cech pigmentacyjnych) ale dopiero poprzez badania polimorfizmu genów kandydujących, a zwłaszcza badania asocjacyjne całego genomu (ang. *genome-wide association study*, GWAS), które stanowią alternatywę dla badań metodą sprzężeń, pozwoliły na wskazanie genów i polimorfizmów wcześniej niepodejrzanych o związek z daną cechą. W ten sposób uwierzytelniono dotychczasowe i poznano nowe markery DNA leżące



Rycina 1. Proces selekcji artykułów zgodny z metodologią PRISMA

u podstaw EVC. Następnie ustalono moc predykcyjną markerów DNA statystycznie istotnie powiązanych z EVC. To właśnie te markery stanowią podstawę opracowanych narzędzi FDP [10]. Mimo to, nasze obecne zrozumienie genetycznych podstaw EVC jest ciągle niekompletne, być może z wyjątkiem pigmentacji. Problem z EVC polega na tym, że są to cechy wielogenowe, dziedziczone zgodnie z modelem oligogenicznym (udział kilku genów o umiarkowanym efekcie) lub modelem poligenicznym (współdziałanie lub sumowanie działania dużej liczby genów z różnych loci o niewielkim pojedynczym efekcie działania).

Każdy poszczególny gen przyczynia się tylko do niewielkiej części całkowitej wariacji fenotypowej cechy i tylko połączenie dużej liczby czynników genetycznych może wyjaśnić ogólny dziedziczony składnik wariacji [11]. Dodatkowo, niegenetyczne czynniki środowiska, które różnią się zasięgiem w zależności od różnych cech wyglądu, mogą modulować ekspresję i wzajemne oddziaływania między genami. Co więcej, im większy składnik środowiskowy, tym mniej można wyjaśnić za pomocą DNA [11]. Ten złożony, wieloczynnikowy mechanizm wymaga pogłębienia dotychczasowej wiedzy. Dlatego też, kontynuowanie badań nad genetycznymi determinantami EVC, identyfikacja nowych genów i polimorfizmów oraz ocena ich wartości predykcyjnej dla EVC to zadanie dla genetyki sądowej w najbliższej przyszłości.

Markery genetyczne i epigenetyczne stosowane w narzędziach do FDP to polimorfizmy pojedynczego nukleotydu (ang. *single nucleotide polymorphism*, SNP), haplotypy mitochondrialne, markery chromosomu Y, wymienione już krótkie powtórzenia tandemowe STR [12] i markery CpG¹ (wyspy CpG ang. *CpG sites*), których poziomy metylacji są silnie skorelowane z wiekiem [13]. Markery te niosą informacje o cechach fenotypowych człowieka takich jak: cechy fizyczne widoczne zewnętrznie [14], pochodzenie biogeograficzne [15] oraz indywidualny wiek (biologiczny) [16].

Polimorfizmy pojedynczego nukleotydu (SNP) to warianty genomu występujące naturalnie w populacji ludzkiej. Powstają w wyniku mutacji punktowych, które mogą być indelami (InDels - insercja/delecja) lub substytucjami. SNP definiuje się jako mutacja w pojedynczym nukleotydzie DNA w określonym locus w genomie występująca ze znaczną częstotliwością (>1%). SNP może wystąpić w regionie kodującym, ale nie powodować zmiany aminokwasu, może wystąpić w regionie kodującym ze zmianą aminokwasu, może wystąpić w regionie regulatorowym wywołując zmianę ekspresji genów lub może wystąpić w regionie między genami. SNP są najczęstszym typem zmienności w ludzkim genomie, występującym raz na kilkaset par zasad i szacowanym na około 15 milionów powszechnych wariantów DNA [17]. Zdecydowana większość różnic między osobnikami to mutacje punktowe spowodowane SNP. Do tego są one bardziej stabilne niż STR, gdyż przeciętne tempo ich mutacji jest niskie. SNP są najmniejszymi markerami genetycznymi, dlatego są wykorzystywane w analizie najmniejszych fragmentów DNA, co rozwiązuje problem **zdegradowanego** DNA. Główną zaletą SNP jest jednak ¹Wyspy CpG (C - cytozyna, p - reszta kwasu fosforowego, G - guanina) to regiony w genomie o podwyższonej zawartości dinukleotydów 5'-CpG-3' w stosunku do przeciętnej dla całego genomu.

możliwość badania przesiewowego pod kątem obecności polimorfizmu bez elektroforezy żelowej przy użyciu wysokowydajnych metod, takich jak technologia mikromacierzy [18]. Ich największą wadę stanowi to, że nie są tak różnorodne jak STR. Do identyfikacji osoby potrzeba aż 50 SNP, podczas gdy w przypadku STR wystarczy tylko 13 [18].

Krótkie powtórzenia tandemowe (STR) to polimorfizmy generowane przez sekwencję (w tandemie) kopii małych segmentów DNA (od 2 do 6 par zasad). Mogą prezentować kilkanaście alleli, a tysiące takich polimorfizmów zostało już zidentyfikowanych u ludzi; niektóre szacunki wskazują na istnienie około miliona STR rozmieszczonych w ludzkim genomie [18]. Główną zaletą markerów STR jest ich duża różnorodność alleli, co sprawia, że takie markery mają dużą wartość informacyjną. Najbardziej polimorficzne STR mają wysoką moc dyskryminacyjną (prawdopodobieństwo, że dwie losowo wybrane osoby mają różne genotypy) i niskie prawdopodobieństwo dopasowania (prawdopodobieństwo, że dwie losowo wybrane osoby mają identyczne genotypy). Oprócz konieczności posiadania danych referencyjnych, inną wadą jest struktura markerów STR, które składają się z powtórzeń o długości od ~100 do 300 par zasad. Zdarzają się sytuacje, w których materiał biologiczny z miejsca zbrodni jest tak zdegradowany, że uzyskane próbki DNA nie dają możliwości uzyskania wystarczających danych do dokładnej identyfikacji [18].

DNA mitochondrialne (mtDNA) występuje zarówno u kobiet jak i u mężczyzn ale jest dziedziczone wyłącznie w linii matczynej. Jest ono niezwykle użyteczne w identyfikacji silnie zdegradowanego materiału genetycznego. Z racji obecności wielu mitochondriów w komórce, mtDNA ma dużo większą szansę na przetrwanie w niesprzyjających warunkach niż jądrowe DNA. Z tego powodu mtDNA jest wykorzystywane w badaniu historycznych szczątków oraz komórek z małą ilością materiału genetycznego np. kosmków włosów [19]. Oprócz wielu kopii mtDNA w komórce, do jego zalet należą też: możliwość zsekwencjonowania całego genomu mitochondrialnego w jednym przebiegu, akumulacja mutacji w czasie oraz brak rekombinacji przez wiele pokoleń [19].

Chromosom Y, część specyficzna dla mężczyzn, jest szeroko stosowany w badaniach genetycznych i analizie danych sądowych, szczególnie w przypadkach napaści na tle seksualnym, chociaż identyfikacja osób za pomocą analizy DNA chromosomu Y jest ograniczona do osób niespokrewnionych [20].

FDP jest ściśle powiązane z dynamicznie rozwijającymi się technologiami, które pozwalają zwiększać prawdopodobieństwo przewidywania cech fizycznych lub pochodzenia biogeograficznego. Skuteczność FDP poprawia rozwój technologiczny w zakresie masowego sekwencjonowania równoległego (ang. *massively parallel sequencing*, MPS). Technologia ta opiera się na nowej generacji sekwencjonowania (ang. *next-generation sequencing*, NGS), która umożliwia łączenie kilku tradycyjnych testów elektroforezy w jeden pojedynczy test. Dzięki temu pojedyncza analiza może być wykorzystana do jednoczesnego wykrywania ogromnej liczby (co najmniej >160 SNP), co stanowi potencjalne narzędzie

do identyfikacji i wnioskowania o pochodzeniu biogeograficznym dawcy próbki [21]. Jednak pomimo większej liczby analizowanych loci moc dyskryminacyjna paneli SNP jest mniejsza niż w STR. Dlatego SNP są raczej uzupełnieniem markerów STR niż ich zamiennikiem. Wydaje się, że pewnym kompromisem między STR a SNP może być wykorzystanie polimorfizmu delecyjno-insercyjnego (ang. *deletion-insertion polymorphism*, DIP) opartego na analizie krótkich amplikonów (jak w przypadku technik stosowanych w analizach STR) do analiz SNP. Większość analiz traktuje SNP jako markery uzupełniające, jednak zdarzają się przypadki gdzie stanowią one główną, a nawet jedyną metodę analizy. Ma to miejsce w sytuacji, gdy analizie zostaje poddane bardzo zdegradowane DNA, bo w tych przypadkach analiza STR często skutkuje niepowodzeniem. SNP może amplifikować nawet na bardzo krótkich amplikonach, które mają mniej niż 100 par zasad, dlatego dają dużo większą szansę na uzyskanie wyników podczas analizy zdegradowanego DNA [22]. Jedną z technik wykorzystywanych w analizie cech fenotypowych jest SNPforID, która pozwala na amplifikację 52 SNP już z minimalnej ilości DNA. Amplifikacja odbywa się na drodze pojedynczej reakcji PCR, a następnie SNP jest analizowane z użyciem dwóch reakcji, podczas których wydłużano łańcuch o pojedynczą zasadę (ang. *single-base extension*, SBE) oraz z użyciem elektroforezy kapilarnej [23]. Metody bazujące na SNP i STR mają różne zalety i wady. Połączenie tych metod zwiększa ich możliwości analizy, szczególnie silnie zdegradowanego DNA, które pochodzi z kości długo zalegających w depozycie.

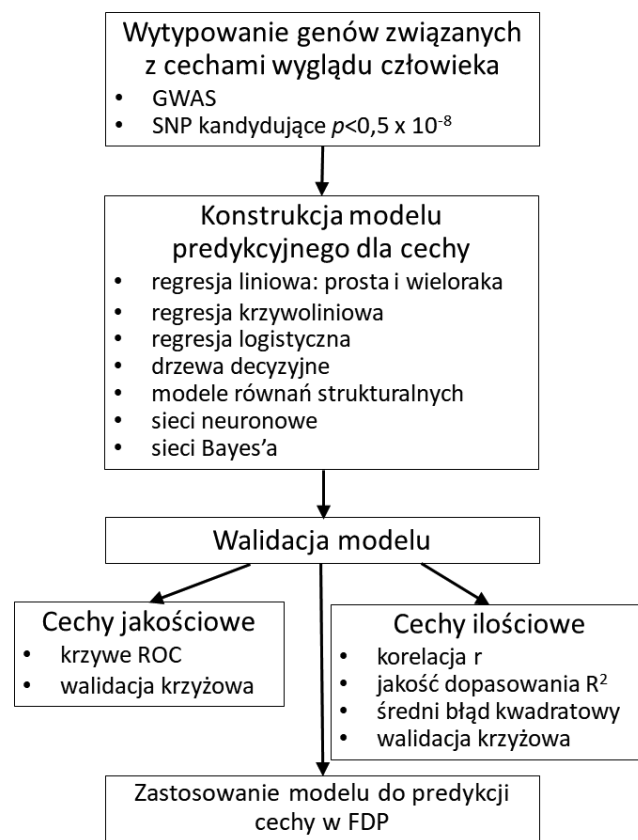
ANALIZA PREDYKCYJNA CECHY FENOTYPOWEJ - OD GENOTYPU DO FENOTYPU

Analiza predykcjna cechy fenotypowej (przy założeniu dziedziczenia wielogenowego - oligogenicznego lub poligenicznego) składa się zazwyczaj z trzech etapów: (1) identyfikacji/wyboru genów, które zgodnie z modelem dziedziczenia danej cechy mogłyby mieć z nią związek, (2) konstrukcji modelu predykcynnego, na podstawie którego można będzie przewidzieć daną cechę fenotypową z danych wejściowych o SNP i (3) walidacji tego modelu (Ryc. 2) [14].

Materiał do analizy stanowią substancje organiczne, określone mianem śladu biologicznego, znajdujące i zabezpieczane na miejscu zdarzenia na przedmiotach należących do ofiary lub na jej ciele. Są to m.in. krew, fragmenty tkanek (kości, zęby, tkanki miękkie), wytwory naskórka (włosy, paznokcie), wydaliny (kał, mocz) lub wydzieliny (ślina, pot, nasienie, wydzielina z pochwy). Plamy krwi, wymaz komórek nabłonka z błony śluzowej jamy ustnej, czy włosy i paznokcie to najczęściej analizowane ślady biologiczne.

Związek między genotypem a fenotypem jest zazwyczaj ustalany w badaniach asocjacyjnych całego genomu (GWAS). Zastosowanie mikromacierzy umożliwia jednoczesne genotypowanie setek tysięcy SNP wszystkich genów w bardzo dużej próbie osób niespokrewnionych ze sobą i posiadających daną cechę oraz podobnej liczbie osób nie posiadających tej cechy (grupa kontrolna). SNP statystycznie istotnie związane z daną cechą ($p < 0,5 \times 10^{-8}$) na poziomie genomu uważane są za markery kandydujące wykorzystywane następnie do budowy modelu. Etap drugi to budowa

modelu predykcynnego, najczęściej w formie liniowej funkcji opisującej zależność między SNP kandydującymi (zmienna objaśniająca - niezależna) a cechą fenotypową (zmienna objaśniania - zależna). Wybór modelu predykcynnego jest arbitralny w ramach klasycznych technik statystycznych takich jak: regresja liniowa (ang. *linear regression*) i logistyczna (ang. *logistic regression*) najczęściej są to układy regresji wielorakiej z wyborem podzbioru zmiennych metodą krokową (ang. *multiple stepwise regression*), regresja nieliniowa (ang. *nonlinear regression*), modele równań strukturalnych (ang. *structural equation modeling*) lub technik maszynowego uczenia i sztucznej inteligencji takich jak: drzewa decyzyjne (ang. *statistics decision tree*, DT), sieci neuronowe (ang. *neural networks*), czy sieci Bayesa (ang. *Bayesian networks*) [14]. Etap trzeci to walidacja modelu. Zmienne objaśniające (SNP) są wprowadzane do sparametryzowanej funkcji modelu celem wygenerowania prognoz dla przewidywania zmiennej objaśnianej (cechy fenotypowej). Oceny jakości modelu dokonuje się za pomocą różnych parametrów. Jeśli model predykcynny dotyczy cech jakościowych (ich zmienność jest opisana kategoriami np. występowanie piegów lub ich brak) wówczas najczęściej używanymi parametrami dokładności przewidywania są obszar AUC (ang. *area under curve*) pod krzywymi ROC (ang. *receiver operator curve*) i traktowanie go jako miarę dobroci i trafności danego modelu. Krzywa charakterystyki operacyjnej odbiornika (ROC) jest wizualną prezentacją działania klasyfikatora binarnego (tak, nie) poprzez wykreślenie współczynnika prawdziwie pozytywnych predykcji względem współczynnika predy-



Rycina 2. Schemat postępowania w predykcji cechy fenotypowej do celów kryminalistycznego fenotypowania DNA

cji fałszywie pozytywnych przy wszystkich możliwych progach. Pole pod krzywą ROC (AUC) to całość krzywej ROC, mieszcząca się w zakresie od 0,5 (przewidywanie losowe) do 1,0 (przewidywanie doskonałe). Wartość AUC może być interpretowana jako prawdopodobieństwo, że model predykcyjny przypisze wyższy wynik losowo wybranemu pozytywnemu wystąpieniu (np. kolor niebieski oczu) niż losowo wybranemu negatywnemu (np. kolor inny niż niebieskie oczy). W ten sposób reprezentuje ogólną ale bardzo skondensowaną miarę dokładności klasyfikatora binarnego. AUC wykorzystano na przykład do przewidywania koloru oczu, włosów, skóry i wieku.

W przypadku cech ilościowych (np. wysokości czy masy ciała, cech o rozkładzie normalnym) dokładność prognoz może być wyrażona współczynnikiem korelacji r lub r^2 lub błędem średniokwadratowym (ang. *mean square error*, MSE). Zarówno r jak i r^2 mieszczą się w zakresie od 0 (przewidywanie losowe) do 1 (przewidywanie doskonałe). r^2 był na przykład używany do przewidywania DNA cech pigmentacyjnych i wieku.

Jeśli wielkość próbki jest ograniczona, można zastosować walidację krzyżową. Polega ona na podzieleniu całego zestawu próbek na komplementarne podzbiory, zbudowaniu modelu predykcji w niektórych podzbiórach i ocenie wydajności modelu w pozostałych zestawach. Zazwyczaj wykonuje się wiele rund walidacji krzyżowej, po czym przyjmowana jest uśredniona dokładność i wartość dla 95 centyla. Krzyżowa walidacja została wykorzystana na przykład do przewidywania koloru włosów.

PRZEWIDYWANIE CECH FENOTYPOWYCH

Przewidywanie płci

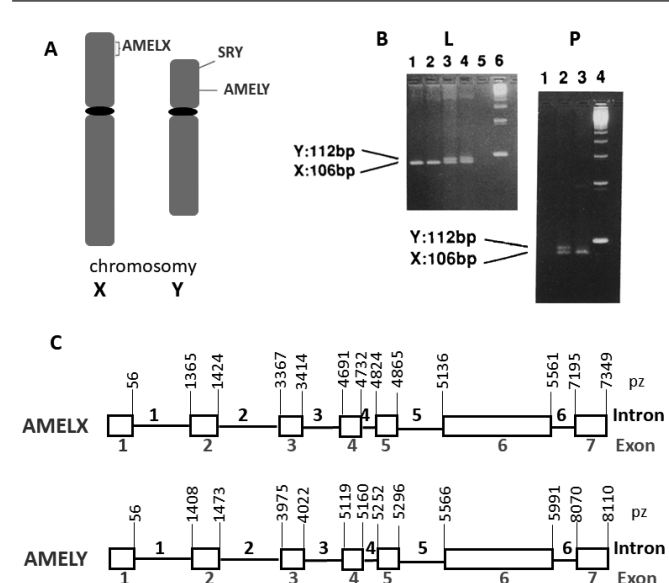
Określenie płci może dostarczyć istotnych informacji w badaniach kryminalistycznych, w szczególności w sprawach dotyczących napaści na tle seksualnym oraz w poszukiwaniach osób zaginionych lub w identyfikacji nieznanymi szczątków [24]. Płeć można określić na podstawie morfologii kości osobników dorosłych (elementy diagnostyczne dla płci można znaleźć na czaszce, miednicy i różnych fragmentach szkieletu; rozwijają się one w okresie pokwitania, dlatego też największym ograniczeniem metod morfologicznych jest niemożność określania płci dzieci) lub jeśli stan zachowania szkieletu uniemożliwia badanie makroskopowe a zabezpieczone zostały ślady biologiczne to na podstawie markerów DNA. Określanie płci molekularnej ma pewną przewagę nad metodami morfometrycznymi ponieważ umożliwia określanie płci ze szczątków płodów i osobników młodocianych, nie wymaga kompletnego materiału kostnego - wystarczy jedna pełna kość lub ząb. Najczęściej analizuje się gen kodujący amelogeninę lub inne markery molekularne np. występujące na chromosomie Y.

Gen kodujący amelogeninę, białko, które odpowiada za rozwój szkliwa, posiada po jednej kopii zlokalizowanej na chromosomie X – *AMELX* (gen amelogeniny) na krótkim ramieniu chromosomu X w regionie p22.1-p22.3 i chromosomie Y – *AMELY* (gen amelogeniny) w pobliżu centromeru chromosomu Y w p11.2. Różnica polega na delecji sześciu par zasad (6 pz) w intronie 1. *AMELX*, względem intronu 1.

AMELY. Można ją wykryć w reakcji PCR; zestaw starterów do PCR celuje w delecję 6 pz, która występuje na chromosomie X, co umożliwia rozróżnienie ampliconów wytworzonych z chromosomów X i Y podczas rozdzielania elektroforetycznego w celu rozdzielania alleli STR. Pojawienie się tylko jednego prążka o długości 106 par zasad będzie świadczyło o tym, że obecny jest tylko wariant *AMELX*, czyli próbka będzie pochodziła od kobiety. W przypadku próbki pochodzącej od mężczyzny pojawią się dwa prążki, jeden o długości 106 par zasad, a drugi o długości 112 par zasad. W próbce będą wtedy dwa warianty genu, kolejno *AMELX* i *AMELY* [24] (Ryc. 3).

Różnice w strukturze i wielkości obu wariantów tego genu stanowią wiarygodny wskaźnik do określenia płci, nawet wówczas gdy dysponujemy małą ilością DNA [25]. Amelogenina, marker predykowania płci, jest włączany do większości dostępnych na rynku zestawów multipleksowych STR i jest obecnie najpowszechniejszym markerem stosowanym w postępowaniach sądowych. Na przykład przez amerykański DNA Index System, NDIS, który jest krajowym poziomem Combined DNA Index System CODIS wykorzystywany przez Federalne Biuro Śledcze (FBI), a na całym świecie przez ponad 90 laboratoriów kryminalistycznych w ponad 50 krajach [26].

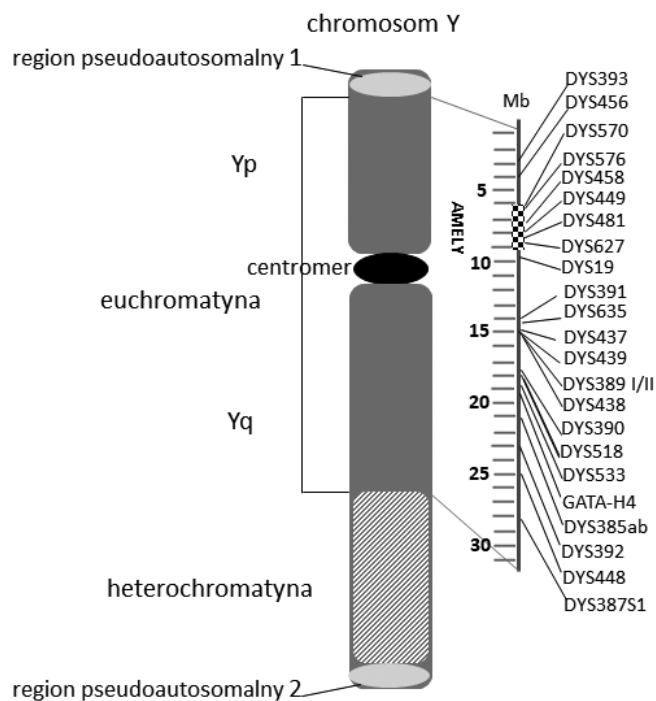
Mimo powszechnego stosowania tej metody nie jest ona w pełni niezawodna. Zdarzają się delecje sześciu par zasad w intronie 1 genu *AMELY*, co sprawia, że ma taką samą długość jak *AMELX*, całkowite delecje genu *AMELY* lub całkowite delecje genu *AMELX*. Takie zaburzenia występują z częstotliwością poniżej 8% [27]. Oprócz amelogeniny badane są alternatywne markery swoiste dla chromosomu X i Y. Są to: *SOX3* (SRY-Box Transcription Factor 3), *STS* (Exon 7)



Rycina 3. Gen kodujący amelogeninę AMEL. A) Lokalizacja genu AMEL na chromosomie X w pozycji p22.1-p22.3 i chromosomie Y w pozycji p11.2. B) Rozdział elektroforezy żelowej genów AMELX i AMELY po PCR. L – Materiał pobrany z mięśni ofiar masowej katastrofy. Pas 1 i 2 zidentyfikowano jako płęć żeńską, 3 i 4 jako płęć męską, 5 ślepa próba, 6 – 123 bp drabinka DNA. P – Materiał pobrany z kości ekshumowanych 74 lata po pochówku. Pas 1 – ślepa próba, 2 i 3, zidentyfikowane odpowiednio jako płęć męską i żeńską, 4 – 123 bp drabinka DNA. (przykłady pochodzą z [80]). C) Struktura ludzkiego genu amelogeniny: AMELX i AMELY

(Steroid Sulfatase), *TSPYL2* (TSPY Like 2) w chromosomie X) oraz *SRY* (Sex Determining Region Y), *STSP1* (Steroid Sulfatase (Microsomal) Pseudogene 1) i *TSPY1* (Testis Specific Protein Y-Linked 1) w chromosomie Y. Zdublikowany locus STR DXYS156 zlokalizowany na długim ramieniu chromosomu X (w pozycji Xq21.31) i krótkim ramieniu chromosomu Y (w pozycji Yp11.2) również pozwala odróżnić płęć męską od żeńskiej. Umożliwia oddzielne szacunki dla pochodzenia matczynego lub ojcowskiego próbki. Jednakże, jak wykazały liczne badania, wiarygodność testów płci jest niska. Wieloallelowe locus tego markera ma allele specyficzne dla ludzi pochodzących z różnych części świata, co znajduje zastosowanie w predykcji pochodzenia biogeograficznego [28]. Wymienione markery stosowane oddzielnie lub najczęściej w połączeniu ze standardowym testem amelogeninowym mogą dostarczać bardziej wiarygodnych informacji o typowaniu płci w dochodzeniu kryminalistycznym i postępowaniu sądowym. Wymagają jednak dalszych badań potwierdzających ich wiarygodność.

Polimorfizmy specyficzne dla chromosomu Y (Ryc. 4) są markerami haploidalnymi, co różni je od markerów diploidalnych autosomalnego DNA. Przekazywane są z ojca na syna, co oznacza, że potomkowie tej samej męskiej linii rodowej mają identyczny haplotyp Y pojawiający się w danej populacji z określoną częstością. Badanie tych markerów w analizie pokrewieństwa ogranicza się do męskich potomków, co obniża ich wartość dowodową w odniesieniu do układów autosomalnych. Istnieją jednak sytuacje prawosądowe, w których analiza Y-STR jest niezastąpiona w rozstrzygnięciu kilkupokoleniowych rodzinnych korelacji oraz w badaniach identyfikacyjnych osób zaginionych, gdy w grę wchodzi ustalanie pokrewieństwa w linii męskiej [28]. Polimorfizmy Y-STR i Y-SNP są już uwzględniane w komercyjnych zestawach kryminalistycznych, na przykład Yfiler®



Rycina 4. Krótkie powtórzenia tandemowe chromosomu Y (Y-STR) stosowane do predykcji płci w kryminalistycznym fenotypowaniu DNA (FDP)

Plus PCR Amplification Kit, szczególnie w sytuacjach gdy polimorfizmy autosomalne nie są wystarczające [29]. W rzeczywistości 27 Y-STR zostało już zwalidowanych, co pozwala na scharakteryzowanie linii ojcowskiej z wysokim stopniem wiarygodności [29]. Połączenie różnych komercyjnych zestawów i dodatkowych multipleksów pozwala obecnie na amplifikację ponad 40 sekwencji Y-STR [30].

W określaniu płci na podstawie chromosomu Y pomocnym okazuje się mtDNA. Ma ono szczególne znaczenie w analizie jądrowego DNA na niskim poziomie, znajdującego np. w zeszkieletowanych szczątkach, włosach bez cebulek czy starych próbach, w których DNA jest bardzo zniszczone. Nierekombinacyjny sposób dziedziczenia chromosomu Y i mitochondrialnego DNA osłabia wagę statystyczną dopasowania między poszczególnymi próbkami, ale sprawia, że metoda jest skuteczna do rekonstrukcji relacji rodzinnych, na przykład w badaniach masowych katastrof [31]. Obecnie kryminalistyczne typowanie chromosomu Y zyskało szeroką akceptację dzięki wprowadzeniu bardzo czułych paneli. Należy jednak zwrócić uwagę na przypadki chimeryzmu, mikrochimeryzmu (najczęstszą postacią jest mikrochimeryzm płodowo-matczyny, w którym komórki płodu przechodzą do organizmu matki, jak w przypadku kobiet w ciąży noszących męski płód), transplacencji lub chorób genetycznych opartych na zmianach np. liczby chromosomów płci jak w zespole Klinefeltera, które mogą dawać fałszywe wyniki identyfikacji płci [32]. Zastosowanie samych sekwencji specyficznych dla chromosomu Y nie jest informatywne, ponieważ brak produktu PCR może wskazywać na płęć żeńską lub nieudaną reakcję. Dlatego takie badania powinno się przeprowadzać w połączeniu z innym fragmentem jako kontrolą pozytywną dla amplifikacji, na przykład mitochondrialnym lub jądrowym DNA, przy czym preferowany jest jądrowy, ponieważ liczba kopii jest podobna do liczby amplifikowanych markerów specyficznych dla chromosomu Y. Omówione wyniki zebrano w Tabeli 1.

Przewidywanie wieku

Wiek, podobnie jak płęć, można oszacować na podstawie stopnia zaawansowania rozwoju różnych cech morfologicznych kości i zębów, jak np. stadia wyrzynania się zębów mlecznych i stałych oraz stadia kostnienia kości długich i zrastania się nasad z trzonami u dzieci, stopnie zarastania szwów czaszkowych i kostnienia spojenia łonowego. Należy pamiętać, że ocenie podlega wiek biologiczny (zwany też fizjologicznym) wskazujący na stopień zaawansowania procesów rozwojowych prowadzących do osiągnięcia dojrzałości biologicznej (dzieci) lub zmian involucyjnych związanych z procesem starzenia się organizmu (dorośli). Jest on wysoce skorelowany z wiekiem chronologicznym ale ze względu na indywidualny tor rozwojowy obie kategorie wieku mogą się nieznacznie różnić.

W ostatnich dziesięcioleciach zaproponowano różne metody szacowania wieku dawcy próbki biologicznej. Metody te oparte są na pomiarach zmian zachodzących z wiekiem w komórce ciała ludzkiego. Wśród nich można wymienić m.in. modyfikacje epigenetyczne - mechanizm umożliwiający regulację struktury chromatyny oraz jej aktywność

Tabela 1. Markery i geny homologiczne w chromosomach X i Y stosowane do predykcji płci człowieka w kryminalistycznym fenotypowaniu DNA

Gen/Chrom. [Autor]	Zalety	Wady
AMELX Xp22.2 AMELY Yp11.2 [24,25]	- amelogenina dostarcza wyników do identyfikacji płci i służy jako wewnętrzna kontrola PCR - określanie płci testem amelogeninowym możliwe dla zdegradowanego DNA	- brak amplifikacji AMELY, która najczęściej jest spowodowana delecją dużych fragmentów na Yp, skutkuje widocznym brakiem Y-DNA i fałszywą predykcją płci żeńskiej pomimo obecności chromosomu Y (wskaźnik niepowodzeń dla tego testu różni się w zależności od populacji od 0,018% do nawet 8%)
DXYS156 Xp21.31 Yp11.2 [28]	- służy jako wewnętrzna kontrola PCR - jest wieloallelicznym markerem pentanukleotydowym	- brak wyraźnego rozróżnienia między homologami X i Y - przydatny raczej do predykcji pochodzenia etnicznego a nie płci
SOX3 Xp27.1 [28]	- działa jako czynnik transkrypcyjny w kaskadzie genetycznej regulacji różnicowania gonad	- ważny dla prawidłowego rozwoju oocytów i różnicowania jąder męskich ale nieprzydatny do rozróżniania płci
SRY Yp11.31 [24,28,29]	- bezpośredni związek z płcią męską - używany do rozstrzygnięcia niepewnych wyników testu amelogeninowego - mały rozmiar markera - przydatny w przypadku próbek zdegradowanego DNA	- translokacja SRY do chromosomu X może dawać fałszywe wyniki
STS Xp22.31 STSP1 Yq11.221 [25,28,33]	- STSP1 w chromosomie Y służy do wewnętrznej kontroli PCR	- duży rozmiar markera - mniejsza wydajność testu w przypadku próbek zdegradowanego DNA - anomalie u pacjentów z chorobą rybiej łuski sprzężonej z chromosomem X
TSPYL2 Xp11.2 TSPY1 Yp11.2 [21,28]	- wysoka czułość testu - wysoka wydajność w próbkach z zdegradowanym DNA	- brak pozytywnej kontroli wewnętrznej
SNPs [29-31]	- przydatne w analizach mikrośladów i/ lub zdegradowanym DNA - przydatne w predykcji płci osób potencjalnie podejrzanych i nieznanymi	- mają mniejszą moc dyskryminacyjną od STR (50-100 SNP odpowiada 15 STR)

Objaśnienia: AMEL – gen amelogeniny X w chromosomie X, Y w chromosomie Y; DXYS156 – a multi-purpose short tandem repeat locus; SOX3 – SRY-Box Transcription Factor 3; SRY – Sex Determining Region Y; STS (Exon 7) – Steroid Sulfatase; STSP1 – Steroid Sulfatase (Microsomal) Pseudogene 1; TSPYL2 – TSPY Like 2 w chromosomie X; TSPY1 – Testis Specific Protein Y-Linked 1; SNP – single nucleotide polymorphism

transkrypcyjną. Kluczowe dla przebudowy chromatyny są: modyfikacje DNA, warianty i modyfikacje potranslacyjne histonów, długie niekodujące RNA, kompleksy przebudowujące chromatynę i przestrzenne ułożenie chromatyny w jądrze komórkowym. Metylacja DNA jest najlepiej zbadaną formą modyfikacji epigenetycznych pod kątem wykorzystania jej do predykcji wieku w FDP. Z innych metod szacowania wieku należałoby wymienić analizę długości telomerów, delecje w genomie mitochondrialnym, racemizację kwasu asparaginowego (ang. *aspartic acid racemization* AAR), ekspresje mRNA, zaawansowane produkty glikacji (ang. *advanced glycation endproducts*, AGEs), rearanżacje receptorów limfocytów T (ang. *T cell receptor excision circles*, sTRECcs) w krwi obwodowej [34]. Niektóre z tych metod korelują z wiekiem chronologicznym tak silnie, że nazywa się je „zegarem molekularnym” [34]. Nie wszystkie jednak dają wyniki, które można z powodzeniem stosować w FDP do predykcji wieku.

Badania epigenetyczne z wykorzystaniem technologii wykrywania metylacji DNA oraz badania asocjacyjne obejmujące cały epigenom (ang. *epigenome-wide association study* EWAS) zidentyfikowały wiele markerów dinukleotydów CpG (wysp CpG) o zróżnicowanym stopniu metylacji (ang. *differentially methylated regions*, DMR), które silnie korelo-

wały z wiekiem [34,35]. Dokonano selekcji miejsc CpG o różnym poziomie metylacji i opracowano modele predykcji wieku oparte na różnej liczbie miejsc CpG, od 2 do 353, ale najczęściej 3.

Wykazano, że poziom metylacji DNA w tych DMR zmienia się w ciągu życia indywidualnie dla każdego osobnika: wzrasta w dzieciństwie, a następnie zmniejsza się po osiągnięciu dorosłości wraz z postępującym procesem starzenia tak, że zaproponowano metylację DNA określonych loci jako biomarker starzenia, tzw. „zegar epigenetyczny” [36]. Poziom metylacji DNA zależy również od płci, niektórych cech fizycznych (np. otyłość) oraz elementów stylu życia (np. palenie papierosów, stosowanie innych używek, aktywność fizyczna lub jej brak i inne). Poziom metylacji zależy od określonych witamin i minerałów co sugeruje związek metylacji z odżywianiem [36]. Na poziom metylacji DNA wpływa również rodzaj tkanki lub komórki w której ten proces zachodzi [37]. Zmiany te można zmierzyć i wykorzystać do obliczenia szacunkowego wieku osobnika na podstawie próbek biologicznych różnego pochodzenia (może to być krew, ślina, sperma, tkanki zęba: cement, miazga zębowa i zębina) oraz miejsca zdarzenia (nieznane szczątki ludzkie, miejsca zbrodni). Najwięcej badań przeprowadzono z wykorzystaniem krwi.

W 2016 roku Park z zespołem zaproponował model szacowania wieku na podstawie próbki krwi. Trzy CpG (cg16867657, cg04208403 i cg19283806) genów *ELOVL2* (ELOVL Fatty Acid Elongase 2), *ZNF423* (Zinc Finger Protein 423) i *CCDC102B* (Coiled-Coil Domain Containing 102B) pozwoliły oszacować wiek z dokładnością 3,4 lat w próbie testowej [38].

Silva z zespołem przeanalizowali 3 CpG genu *GRIA2* (glutamate ionotropic receptor AMPA type subunit 2) i 6 CpG genu *NPTX2* (Neuronal Pentraxin 2) jako potencjalnych markerów do predykcji wieku z próbek krwi i śliny [39]. Wyniki badań były różne dla krwi i śliny oraz dla *GRIA2* i *NPTX2*. Wiek predykowany z próbki śliny wyjaśniał większą część skorelowanej z wiekiem wariancji wzorców metylacji DNA niż analogiczna predykcja dla próbki krwi (odpowiednio: 80,1% dla *GRIA2* i 65,4% dla *NPTX2* w próbce śliny oraz 55,4% dla *GRIA2* i *NPTX2* w próbce krwi). Ponadto, predykcje wieku z próbki śliny i krwi były słabo skorelowane ($r^2 = 0,313$ dla *GRIA2* i $r^2 = 0,276$ dla *NPTX2*).

Inna grupa badaczy, której przewodził Bekaert, zaproponowała model predykcji wieku oparty na 4 markerach metylacji DNA genów *ASPA* (Aspartoacylase), *PDE4C* (Phosphodiesterase 4C), *ELOVL2* (ELOVL Fatty Acid Elongase 2) i *EDARADD* (EDAR Associated Death Domain) [40]. Model ten opracowano na podstawie analizy 206 próbek krwi zarówno osób zmarłych, jak i żyjących w przedziale wieku od 0 do 91 lat. Wartości metylacji DNA w *ELOVL2* nie wykazały prostej zależności liniowej, dlatego poziomy metylacji *ELOVL2* zostały podniesione do kwadratu (por. tabela 2). Model ten wykazał największą dokładność predykcji przy średnim odchyleniu bezwzględny (MAD) między wiekiem chronologicznym a przewidywanym równym 3,75 lat i skorygowanym R^2 wynoszącym 0,95. Nie zaobserwowano różnicy w dokładności predykcji dla próbek uzyskanych od osób żyjących i zmarłych lub od kobiet i mężczyzn. Te same markery przeanalizowano w próbie liczącej 29 zębów losowo wybranych osób z pierwotnej próby w wieku 19–70 lat. Uzyskane wyniki to MAD = 4,86 lat i skorygowane $R^2 = 0,74$ [40].

Hong z zespołem zaproponował model oparty na sześciu CpG (cg0048195, cg19671120, cg14361627, cg08928145, cg12757011 i cg07547549) genów *SST* (Somatostatin), *CNGA3* (Cyclic Nucleotide Gated Channel Subunit Alpha 3), *KLF14* (Krüppel-like Factor 14), *TSSK6* (Testis Specific Serine Kinase 6), *TBR1* (T-Box Brain Transcription Factor 1) oraz *SLC12A5* (Solute Carrier Family 12 Member 5). Dodatkowo do modelu dołączono CpG specyficzne dla typu komórki cg18384097 genu *PTPN7* (Protein Tyrosine Phosphatase non-receptor type 7) w niezależnym zestawie próbek śliny uzyskanych od 226 osób w wieku od 18 do 65 lat. Metoda ta pozwala szacować wiek z odchyleniem 3,15 roku w stosunku do rzeczywistego wieku chronologicznego na podstawie zaledwie siedmiu znaczników [41].

Alghanim z zespołem [42] po przebadaniu 27 miejsc CpG w trzech loci, *SCGN*, (Secretagoin, EF-Hand Calcium Binding Protein), *DLX5* (Distal-Less Homeobox 5) i *KLF14*, (Kruppel Like Factor 14) opracowali modele predykcyjne

wieku dla próbek krwi i śliny. Do modelu zostały włączone: jedno CpG genu *SCGN* i dwa CpGs genu *KLF14*. Testowane miejsca CpG w *DLX5* nie wykazały korelacji z wiekiem, dlatego też nie zostały włączone do testowanego modelu. Model dla pojedynczego miejsca CpG z próbki śliny wyjaśniał 85% skorelowanej z wiekiem zmienności wzorców metylacji DNA a średnie odchylenie wieku przewidywanego od wieku chronologicznego MAD wyniosło 5,8 lat. Model dla dwóch miejsc CpG wyjaśniał 84% skorelowanej z wiekiem zmienności wzorców metylacji DNA a MAD wyniosła 6,2 lat.

Model predykcyjny wieku dla próbki krwi oparty na jednym CpG w *SCGN* i dwóch CpG w *KLF14* wyjaśniał 71% skorelowanej z wiekiem zmienności wzorców metylacji DNA a dokładność szacowania (MAD) wyniosła 6,6 lat.

Kolejną wydzieliną umożliwiającą szacowanie wieku na podstawie stopnia metylacji DNA jest sperma. Tę metodę opracował Lee wraz z zespołem w 2015 roku. Wykorzystał trzy CpG z genów cg06304190 *TTC7B* (Tetratricopeptide Repeat Domain 7B), cg12837463 i cg06979108 w *NOX4* (NADPH Oxidase 4). Jego metoda cechuje się stosunkowo dużą dokładnością, różnica między wiekiem chronologicznym a uzyskanym z predykcji wyniosła 4,8 lat [43].

Giuliani z zespołem opracowali metody szacowania wieku na podstawie różnych tkanek zęba: cementu, miazgi zębowej i zębiny. Zaproponowane metody wykorzystywały od 5 do 13 CpG pochodzących z genów *ELOVL2* (ELOVL Fatty Acid Elongase 2), *FHL2* (Four And A Half LIM Domains 2) i *PENK* (Proenkephalin). Różnica między wiekiem predykowanym a chronologicznym wyniosła 1,2 roku, gdy DNA pochodziło z cementu i miazgi tego samego zęba, 2,3 lat, gdy pochodziło z miazgi, 2,5 lat w przypadku cementu oraz 7,1 lat w przypadku zębiny [44]. Analiza metylacji DNA w różnych genach przyniosła obiecujące wyniki do wykorzystania w FDP jeśli na miejscu zdarzenia zabezpieczono zęby.

Wyniki analizy czterech genów *ELOVL2* (ELOVL Fatty Acid Elongase 2), *FHL2* (Four And A Half LIM Domains 2), *EDARADD* (EDAR Associated Death Domain) i *PDE4C* (Phosphodiesterase 4C) przeprowadzonej pod kątem poziomu metylacji dinukleotydów CpG w 53 próbkach krwi zdrowych żywych i 51 zmarłych osób z Portugalii wykazały różnice w predykcji wieku w zależności od pochodzenia próbki. Średnie bezwzględne odchylenia (MAD) między wiekiem chronologicznym a przewidywanym wyniosło 5,35 lat, wyjaśniając 94,1% całkowitej zmienności wieku w przypadku próbek od osób żywych natomiast 9,72 lat w próbkach od osób zmarłych [45]. To pewne nowum w FDP albowiem zdecydowana większość badań, w tym omówione wyżej, nie wykazuje różnic w predykcji wieku na podstawie próbek pochodzących od żywych i martwych dawców.

Wiele badań potwierdziło najwyższe wartości predykcyjne dla CpG genów *ELOVL2* i *EDARADD* w próbkach krwi. Zostały one wpisane na listę 44 markerów najsilniej skorelowanych z wiekiem.

Obecnie najbardziej wiarygodne metody przewidywania wieku oparte na markerach CpG z możliwością zastosowania do dowolnej tkanki zawierającej DNA, to modele łączące te markery i nazwane „zegarem epigenetycznym”. W literaturze przedmiotu znane są modele zaproponowane przez Horvatha i współ. [46], Hannuma i współ. [47], Zbieć-Piekarską i współ. [48], czy Montesanto i współ. [49]. Modele te są w stanie dostarczyć dokładnych przewidywań wieku poprzez analizę poziomów metylacji odpowiednio 391 i 71 miejsc CpG. Ważną cechą tych modeli jest to, że zostały opracowane przy użyciu dużych liczebnie próbek obejmujących całe dorosłe życie i różne grupy etniczne; ich predyktory (miejsca CpG) zostały wybrane przy użyciu podobnych modeli regresji; oraz były oparte na technologiach mikromacierzy. Uważa się, że dokładniejsze strategie modelowania statystycznego (algorytmy uczące w uczeniu maszynowym) są w stanie poprawić dokładność modeli matematycznych do przewidywania wieku. A wiek jest o tyle ważny w dochodzeniu kryminalnym, ponieważ przewidywane cechy wyglądu, takie jak siwienie włosów, zmarszczki na twarzy, czy łysienie zmieniają się z wiekiem.

Markery epigenetyczne wydają się być bardzo przydatnym narzędziem do FDP. Ich potwierdzona empirycznie wartość do predykcji wieku wynika z wysokiej swoistości i specyficzności, oraz możliwości szacowania wieku z różnych śladów biologicznych z dużą dokładnością, często mniejszą od 5 lat. Istnieje potrzeba opracowania epigenetycznych markerów metylacji dla różnych cech fenotypowych (EVC), co mogłoby znacząco przyczynić się do

usprawnienia śledztwa. Sumarycznie zagadnienia te ujmuje Tabela 2.

Przewidywanie zewnętrznych cech fizycznych, EVC

Do tej pory następujące EVC zostały ocenione pod kątem FDP: pigmentacja – kolor oczu, włosów i skóry (każda z tych cech rozpatrywana oddzielnie lub w połączeniu, np. kolor oczu i włosów; kolor oczu i skóry; kolor oczu, włosów i skóry), morfologia włosa, morfologia twarzy, obecność piegów, wysokość ciała, masa ciała (otyłość), łysienie typu męskiego i krótkowzroczność. Wśród wymienionych EVC cechy pigmentacyjne, jako najmniej złożone genetycznie w porównaniu z na przykład wysokością ciała, dokładnie przewidywalne poprzez analizę tylko kilku genów, zostały najlepiej przebadane pod kątem możliwości ich wykorzystania w FDP i co zasługuje na uwagę stanowiły pierwszy krok podjęty w kryminalistycznym przewidywaniu EVC [14]. Pozostałe EVC wymagają dalszych pogłębionych badań w celu zwiększenia dokładności ich przewidywania.

Cechy pigmentacyjne – kolor skóry, włosów i oczu

Główną cząsteczką odpowiedzialną za cechy pigmentacyjne człowieka jest barwnik zwany melaniną. Melanina jest syntetyzowana w wieloetapowym procesie utleniania i polimeryzacji (melanogeneza) w którym tyrozyna przekształcana jest w dopachinon, a następnie w jedną z form melaniny, feomelaninę lub eumelaninę (Ryc.5). Proces ten odbywa się w melanosomach, specyficznych pęcherzykach o średnicy ok. 500 nm, które są organellami w przeznacz-

Tabela 2. Markery genetyczne stosowane w kryminalistyce do predykcji wieku człowieka

Material [Autor]	Próba	CpG symbol genu	Wynik
Krew [16]	725 próbek 371 K 354 M wiek dawców 18-104 lata 52 pary jednojajowych bliźniąt płci żeńskiej w wieku 42-69 lat	177 CpG genów <i>ELOVL2</i> , <i>ASPA</i> , <i>PDE4C</i> , <i>FHL2</i> , <i>CCDC102B</i> <i>C1orf132</i> chr16:85395429	- markery najsilniej skorelowane z wiekiem: cg21572722 <i>ELOVL2</i> ($r^2 = 0,939$) cg02228185 <i>ASPA</i> ($r^2 = -0,775$) CpG_27.28.29 <i>PDE4C</i> ($r^2 = 0,906$) cg06639320 <i>FHL2</i> ($r^2 = 0,898$) cg19283806 <i>CCDC102B</i> ($r^2 = -0,889$) CR_21_CpG_11 <i>C1orf132</i> ($r^2 = -0,908$) cg07082267 chr16:85395429 ($r^2 = -0,855$) - model predykcyjny wieku wyjaśnia 93,9% skorelowanej z wiekiem zmienności wzorców metylacji DNA - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym: MAD = ±3.07 lat - średni bezwzględny błąd procentowy MAPE = 6,3% średni wskaźnik prawidłowej klasyfikacji = 78,48%
Krew [38]	765 próbek 380 K 385 M (Korea) wiek dawców 11-90 lat	cg16867657 <i>ELOVL2</i> cg04208403 <i>ZNF423</i> cg19283806 <i>CCDC102B</i>	- model predykcyjny wieku wyjaśnia 91,4% skorelowanej z wiekiem zmienności wzorców metylacji DNA - dokładność predykcji: 77,3% dla <60 lat 57,3% dla ≥60 lat - odchylenie wieku przewidywanego od wieku chronologicznego: SE = 6,32 lat; MAD = 3,16 lat - formuła modelu: wiek predykowany $Y = 39.73167 - 0.69994 \times CCDC102B$ (CpG ₁) + 1.19242 × <i>ELOVL2</i> (CpG ₂) - 0.28914 × <i>ZNF423</i> (CpG ₃) gdzie: <i>CCDC102B</i> CpG ₁ = cg19283806; <i>ELOVL2</i> CpG ₂ = cg16867657; <i>ZNF423</i> CpG ₃ = cg04208403 - model CpG genów <i>ELOVL2</i> , <i>ZNF423</i> i <i>CCDC102B</i> pozwala dokładnie predykować wiek z próbki krwi; model przydatny do predykcji wieku populacji azjatyckich

Krew [39]	23 próbki wiek dawców 5-72 lata	CpG <i>GRIA2</i> 6 CpG <i>NPTX2</i>	- w modelu predykcyjnym wieku z próbki krwi, <i>GRIA2</i> i <i>NPTX2</i> wyjaśniają 55,4% skorelowanej z wiekiem zmienności wzorców metylacji DNA
Ślina [39]	44 próbki wiek dawców 5-72 lata		- w modelu predykcyjnym wieku z próbki śliny, <i>GRIA2</i> wyjaśnia 80,1% a <i>NPTX2</i> 65,4% skorelowanej z wiekiem zmienności wzorców metylacji DNA - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym: MAD = 6,9 lat (<i>GRIA2</i>) i MAD = 9,2 lat dla <i>NPTX2</i> słaba korelacja między wiekiem predykowanym z próbki śliny a wiekiem z próbki krwi: $r^2 = 0,313$ dla <i>GRIA2</i> i $r^2 = 0,276$ dla <i>NPTX2</i>
Krew [40]	206 próbek od dawców zmarłych i żyjących wiek dawców 0-91 lat	cg02228185 <i>ASPA</i> cg17861230 <i>PDE4C</i> cg21572722 <i>ELOVL2</i> cg24724428 <i>ELOVL2</i> cg16867657 <i>ELOVL2</i> cg09809672 <i>EDARADD</i>	- model predykcyjny wieku wyjaśnia 95% skorelowanej z wiekiem zmienności wzorców metylacji DNA - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym: MAD = 3,75 lat dla całej próby; MAD = 5,28 lat dla ≥ 60 lat; MAD = 2,32 dla < 20 lat (zwiększa się z wiekiem) MAD = 3,53 lat dla kobiet; MAD = 3,95 lat dla mężczyzn ($p = 0,8$) (nie zależy od płci) - formuła modelu: wiek predykowany $Y = 26,444119 - 0,201902 \times ASPA (CpG_1) - 0,239205 \times EDARADD (CpG_1) + 0,0063745 \times ELOVL2 (CpG_2)^2 + 0,6352654 \times PDE4C (CpG_3)$ gdzie: <i>ASPA</i> CpG ₁ = cg02228185; <i>EDARADD</i> CpG ₁ = cg09809672; <i>ELOVL2</i> CpG ₂ brak ID, and <i>PDE4C</i> CpG ₃ brak ID (por. Table 2: 923) - model CpG genów <i>ASPA</i> , <i>PDE4C</i> , <i>ELOVL2</i> , <i>EDARADD</i> umożliwia bardzo dokładną predykcję wieku z próbek krwi od osób zmarłych i żywych niezależnie od płci badanej osoby
Ślina [41]	226 próbek wiek dawców 18-65 lat zestaw walidacyjny 113 próbek	cg00481951 <i>SST</i> cg19671120 <i>CNGA3</i> cg14361627 <i>KLF14</i> cg08928145 <i>TSSK6</i> cg12757011 <i>TBR1</i> cg07547549 <i>SLC12A5</i> cg18384097 <i>PTPN7</i>	- model predykcyjny wieku wyjaśnia 94,5% skorelowanej z wiekiem zmienności wzorców metylacji DNA - silna korelacja między wiekiem predykowanym a chronologicznym, Spearman's rho = 0.952 - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym MAD = 3,13 lat - model składający się z 7 wybranych CpG pozwala z dużą dokładnością predykować wiek z próbki śliny
Krew [42]	72 próbki wiek dawców 5-72 lat zestaw walidacyjny 40 próbek	27 CpG <i>SCGN</i> , <i>DLX5</i> , <i>KLF141</i>	testowane miejsca CpG w <i>DLX5</i> nie wykazały korelacji z wiekiem - model predykcyjny wieku dla jednego CpG w <i>SCGN</i> i dwóch CpG w <i>KLF14</i> wyjaśnia 71% skorelowanej z wiekiem zmienności wzorców metylacji DNA - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym MAD = 6,6 lat
Ślina [42]	91 próbek wiek dawców 6-73 lat zestaw walidacyjny 52 próbki		- model predykcyjny wieku dla pojedynczego CpG genu <i>SCGN</i> wyjaśnia 85% skorelowanej z wiekiem zmienności wzorców metylacji DNA - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym MAD = 5,8 lat - model predykcyjny wieku dla dwóch CpG genu <i>KLF14</i> wyjaśnia 84% skorelowanej z wiekiem zmienności wzorców metylacji DNA - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym MAD = 6,2 lat - wysoka jakość predykcyjna modeli stanowi o ich przydatności do szacowania wieku, szczególnie, jak wykazano, młodszych osób (około 40 lat), dla których MAD wyniosła 5,1 lat - silna korelacja między wiekiem predykowanym a chronologicznym, Spearman's rho = 0,907 - formuła modelu: predykowany wiek = $46,24 + (-0,519) \times cg06304190 + (-0,178) \times cg12837463 + 0,541 \times cg06979108$ - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym w grupie 20-latków MAD = 2,9 lat; w grupie 50-latków MAD = 7,2 lat błąd średniokwadratowy RMSE w grupie 20-latków = 3,6 lat, w grupie 50-latków = 7,7 lat
Sperma [43]	12 próbek mężczyźni w wieku 24-57 lat	cg06304190 <i>TTC7B</i> cg12837463 <i>NOX4</i> cg06979108 <i>NOX4</i>	- model predykcyjny wieku dla pojedynczego CpG genu <i>TTC7B</i> wyjaśnia 85% skorelowanej z wiekiem zmienności wzorców metylacji DNA - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym MAD = 5,8 lat - model predykcyjny wieku dla dwóch CpG genu <i>NOX4</i> wyjaśnia 84% skorelowanej z wiekiem zmienności wzorców metylacji DNA - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym MAD = 6,2 lat - wysoka jakość predykcyjna modeli stanowi o ich przydatności do szacowania wieku, szczególnie, jak wykazano, młodszych osób (około 40 lat), dla których MAD wyniosła 5,1 lat - silna korelacja między wiekiem predykowanym a chronologicznym, Spearman's rho = 0,907 - formuła modelu: predykowany wiek = $46,24 + (-0,519) \times cg06304190 + (-0,178) \times cg12837463 + 0,541 \times cg06979108$ - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym w grupie 20-latków MAD = 2,9 lat; w grupie 50-latków MAD = 7,2 lat błąd średniokwadratowy RMSE w grupie 20-latków = 3,6 lat, w grupie 50-latków = 7,7 lat

Zęby [44]	21 współczesnych zębów wiek 17-77 lat metylacja z cementu, zębiny i miazgi	cement 17 zębów CpG2.3 CpG5 CpG9 CpG10 CpG11.12.13.14 <i>ELOVL2</i> CpG2 CpG11.12 CpG16.17 CpG18 <i>FHL2</i> CpG11 <i>PENK</i> zębina 21 zębów CpG6.7 CpG 8 CpG 9 <i>ELOVL2</i> CpG16.17 <i>FHL2</i> CpG21.22 <i>PENK</i> miazga zęba 20 zębów CpG8 CpG9 CpG10 CpG11.12.13.14 CpG15.16.17 <i>ELOVL2</i> CpG5 CpG6 CpG11.12 CpG16.17 CpG18 <i>FHL2</i> CpG3 CpG 4 CpG 26 <i>PENK</i>	- dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym wynosi: 1,20 lat (SD = 1,9) jeśli DNA ekstrahowane z cementu i miazgi tego samego zęba 2,25 lat (SD = 2,5) w przypadku DNA z miazgi zębowej 2,45 lat (SD = 3,3) jeśli DNA jest ekstrahowane z cementu 7,07 lat (SD = 7,0) jeśli DNA jest ekstrahowane tylko z zębiny. - wiarygodne narzędzie do przewidywania wieku na podstawie cementu, zębiny i miazgi zębów
Krew [45]	53 próbki żywych dawców 35 K 18 M (Portugalia) wiek 1-95 lat zestaw walidacyjny 18 dawców 10 K 8 M w wieku 1-93 lat 51 próbek martwych dawców 7 K 44 M wiek 24-86 lat	9 CpG <i>ELOVL2</i> 12 CpG <i>FHL2</i> 4 CpG <i>EDARADD</i> 12 CpG <i>PDE4C</i>	- istotna różnica w predykcji wieku na podstawie próbki pochodzącej od żywych dawców w porównaniu z próbką od martwych dawców - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym w próbie żywych dawców MAD = 5,35 lat, R ² = 0,941; w próbie martwych dawców MAD = 9,72 lat - model sugeruje, że w kontekście FDP należy uwzględnić różnice we wzorcu metylacji DNA między próbkami krwi od osób żyjących i zmarłych
Krew [48]	420 próbek wiek dawców 2-75 lat	41 CpG <i>ELOVL2</i> C1orf132 <i>TRIM59</i> <i>KLF14</i> <i>FHL2</i>	- model predykcyjny wieku wyjaśnia 94% skorelowanej z wiekiem zmienności wzorców metylacji DNA - dokładność predykcji: średnia różnica między wiekiem przewidywanym a wiekiem chronologicznym: MAD = 3,9 lat SEE = 4,5 lat - liczba trafnych predykcji ±5 lat 86,7% dla wieku 2-19 lat 50% dla wieku 60-75 lat - formuła modelu: wiek predykowany Y = 3,26847784751817 - 0,355450171437202 × C1orf132 (CpG ₁) + 0,465445549010653 × <i>ELOVL2</i> (CpG ₂) + 0,237081243617191 × <i>FHL2</i> (CpG ₃) + 0,832684435238792 × <i>KLF14</i> (CpG ₄) + 0,306488541137007 × <i>TRIM59</i> (CpG ₇) - opracowany model włączono do internetowego kalkulatora przewidywania wieku http://biovectis.com/forensic1/age-calculator

nych do tego celu komórkach barwnikowych zwanych melanocytami. Melanocyty są rozmieszczone w warstwie podstawnej naskórka na całej powierzchni skóry z wyjątkiem wewnętrznych powierzchni dłoni i stóp, w mieszkach włosowych, w tęczęwce i naczyńcówce oka, uchu wewnętrznym (region przedsionkowy i prążek naczyniowy ślimaka) i oponach mózgowo-rdzeniowych (pajęcznej i miękkiej). Liczba melanocytów na jednostkę powierzchni naskórka jest podobna u wszystkich ludzi i to nie liczba decyduje o zróżnicowaniu międzysobicznym i międzypopulacyjnym w cechach pigmentacyjnych a rozmieszczenie melanosomów w obrębie melanocytów i keratynocytów, wielkość ziaren melaniny oraz skład tego barwnika i po części ekspozycja na promieniowanie ultrafioletowe [50]. Ilościowe dane ilustrujące powyższe stwierdzenia znaleźliśmy w publikacjach spoza wyselekcjonowanych do niniejszego przeglądu (datowane na lata przed 2015 rokiem). Wzór rozmieszczenia melanosomów w keratynocytach skóry osłoniętej od słońca przedstawia się następująco: w przypadku skóry jasnej (u ludzi pochodzenia europejskiego) 88,9% melanosomów występuje w postaci pęczków, natomiast melanosomy w keratynocytach skóry ciemnej (u ludzi pochodzenia afry-

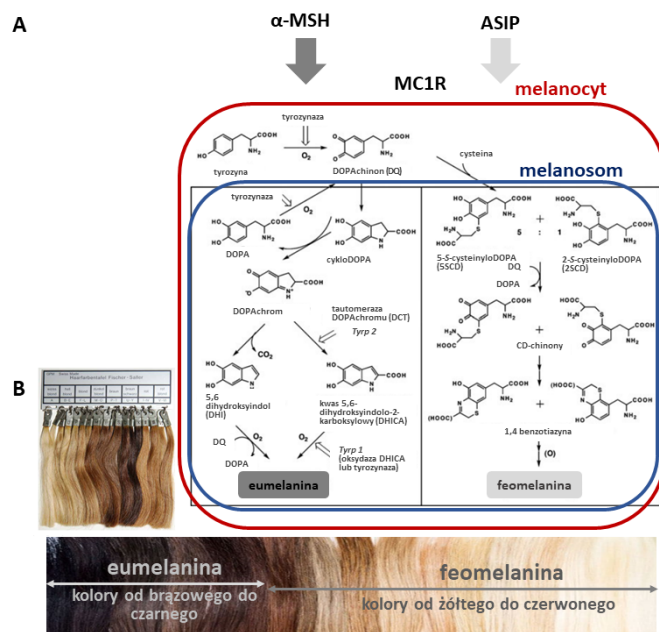
kańskiego) są większe i 84,5% z nich występuje pojedynczo; u ludzi pochodzenia azjatyckiego wzór ten jest kombinacją melanosomów pojedynczych i występujących w wiązkach w proporcji 62,6% do 37,4% [96]. Melanosomy rozmieszczone indywidualnie mają na ogół większe rozmiary niż te w wiązkach. Pomiar wielkości ziaren wykazały, że największe ziarna występują w skórze ciemnej u ludzi pochodzenia afrykańskiego ($1,44 \pm 0,67 \mu\text{m}^2 \times 10^{-2}$), następnie w skórze ludzi pochodzenia azjatyckiego ($1,36 \pm 0,15 \mu\text{m}^2 \times 10^{-2}$), najmniejsze w skórze jasnej ludzi pochodzenia europejskiego o rodowodzie celtyckim ($0,94 \pm 0,48 \mu\text{m}^2 \times 10^{-2}$) [96]. W okolicach ciała eksponowanych na działanie promieniowania słonecznego stężenie melaniny jest blisko dwukrotnie większe od stężenia na obszarach skóry niepoddanej nasłonecznieniu. Na przykład w skórze ludzi pochodzenia afrykańskiego eksponowanej na UVR stężenie melaniny wynosi 40,1 $\mu\text{g}/\text{mg}$ a w skórze niepoddanej UVR 22,0 $\mu\text{g}/\text{mg}$. Podobne proporcje występują u ludzi pochodzenia europejskiego, odpowiednio 19,9 $\mu\text{g}/\text{mg}$ i 12,3 $\mu\text{g}/\text{mg}$ [97].

W badaniach dotyczących składu pigmentowego włosów, które przeprowadzono dla barwy ciemnobrązowej i

barwy czarnej posługując się skalą Fitzpatricka (odpowiednio typ IV i V) wykazano, że głównym barwnikiem w obu przypadkach jest eumelanina. Jej stężenie wynosiło odpowiednio $8,1 \pm 0,5 \mu\text{g}/\text{mg}$ i $19,6 \pm 0,5 \mu\text{g}/\text{mg}$ we włosach ciemnobrązowych i czarnych. Natomiast stężenie feomelaniny wynosiło $2,9 \pm 0,2 \mu\text{g}/\text{mg}$ we włosach ciemnobrązowych i $3,8 \pm 0,0 \mu\text{g}/\text{mg}$ w czarnych. Proporcja feomelaniny do eumelaniny we włosach ciemnobrązowych wynosiła 0,36 i czarnych 0,19, co wskazuje na tendencję, że im ciemniejszy jest kolor włosów, tym mniej feomelaniny znajduje się w składzie pigmentu włosa [98].

Melanogeneza jest inicjowana i regulowana przez liczne kaskady sygnalizacyjne i czynniki transkrypcyjne. Należą do nich m.in. ścieżka sygnałowa Wnt (białko sekrecyjne typu *Wingless* *Wingless-type*), związany z białkiem G receptor endoteliny B (EDNRB *endothelin receptor B*), endotelina 3 (EDN3 *endothelin 3*), receptor kinazy tyrozynowej KIT (*receptor tyrosine kinase*) oraz jego ligand KITL (*KIT-ligand*), czynnik wzrostu komórek macierzystych SCF (*stem cell factor*) a także czynnik transkrypcyjny związany z mikroftalmią (MITF *microphthalmia associated transcription factor*) i inne. Jak wskazują wyniki badań genetycznych, biochemicznych i farmakologicznych kluczową rolę w melanogenezie odgrywa sygnalizacja receptora melanokortyny MC1R (*melanocortin 1 receptor*). Receptor ten znajduje się na powierzchni melanocytów i jest kodowany przez gen MC1R. MC1R wchodzi w skład podrodziny receptorów sprzężonych z białkiem G klasy A GPCR (*G Protein-Coupled Receptor*) która obejmuje pięć rodzajów receptorów MC1R do MC5R. Pośredniczą one w przekazie sygnałów do wnętrza komórki w różnych procesach fizjologicznych lub patologicznych. Aktywacja MC1R przez agonistów α -MSH (*alpha-melanocyte stimulating hormone*) i ACTH (*adrenocorticotropic hormone*) zwiększa syntezę cAMP, co wywołuje szereg reakcji chemicznych wewnątrz melanocytów, które stymulują te komórki do syntezy eumelaniny. Jeśli receptor nie jest aktywowany lub jest zablokowany, melanocyty wytwarzają feomelaninę zamiast eumelaniny. Synteza feomelaniny jest stymulowana przez ASIP (*agouti signaling protein*) białko kodowane przez gen *ASIP* [51].

Eumelanina jest czarnobrązowym nierozpuszczalnym pigmentem, polimerem, pochodną typu indolowego. W skład eumelaniny wchodzi 5,6-dihydroksyindol (DHI) charakteryzujący się czarną barwą i kwas 5,6-dihydroksyindolo-2-karboksylowego (DHICA) o barwie brązowej. Wzajemny stosunek składowych eumelaniny decyduje o ostatecznej intensywności jej barwy (por. Ryc. 5). Feomelanina o żółto-czerwonej barwie posiada w swej strukturze podjednostki benzotiazyny i benzotiazolu oraz 5,6-dihydroksyindolu (por. Ryc. 5). Kolor skóry, włosów i tęczówki oka danej osoby, jak już powiedziano, zależy zarówno od typu melanin (ilościowego stosunku feo- do eumelaniny, który w ludzkim naskórku wynosi 1:3 – około 26% feomelaniny i 74% eumelaniny, niezależnie od stopnia pigmentacji) jak i całkowitej ilości melaniny oraz zawartości, aktywności i liczby melanosomów [52]. Na przykład niebieskie oczy, jasne włosy czy jasna skóra mają mniej melaniny/melanosomów niż oczy brązowe, ciemna skóra czy ciemne włosy [5]. Potwierdziły to wyniki badań uzyskane ilościową i bardziej obiektywną klasyfikacją koloru skóry w oparciu o parametry



Rycina 5. Eumelanina i feomelanina – dwa rodzaje melanin stosowane w kryminalistycznym fenotypowaniu DNA (FDP) do predykcji cech pigmentacyjnych ludzi. A) Schemat melanogenezy i struktura chemiczna eumelaniny i feomelaniny. B) Zakres kolorów, w których bierze udział eumelanina i feomelanina na przykładzie włosów (skala Fisher-Sallera)

try kolorymetryczne układu $L^*a^*b^*$ i określenie indywidualnego kąta typologii niż powszechnie stosowana skala Fitzpatricka [52].

Kolor skóry, cecha uważana za ewolucyjne przystosowanie do natężenia promieniowania UV w różnych rejonach świata (melanina odgrywa kluczową rolę w fotoprotekcji ze względu na zdolność pochłaniania promieniowania ultrafioletowego) jest funkcją wielkości, liczby i rozmieszczenia melanosomów, a nie gęstości melanocytów. W rzeczywistości liczba melanocytów jest taka sama u wszystkich ludzi niezależnie od ich zróżnicowania morfologicznego wynikającego z pochodzenia biogeograficznego. Jednak melanocyty ciemno zabarwionej skóry mają grubsze, dłuższe, rozgałęzione dendryty.

Kolor skóry jest jednym z najbardziej złożonych fenotypów pigmentacji u człowieka. Fakt ten sprawia, że dokładność przewidywania koloru skóry jest niższa niż pozostałych cech pigmentacyjnych (oczu i włosów). Na predykcję wpływa również pojęcie rasy, różnie definiowane w zależności od kultury czy tradycji [53].

Istotny związek z kolorem skóry stwierdzono dla wariantów genu *MC1R* (Melanocortin 1 Receptor), w tym rs1805006, rs1805007, rs1805008, rs1805009, rs11547464, rs201326[89] i rs3212345. Wykazano, że rs3212345: C>T jest związany z jasną skórą, rudymi włosami i słabą zdolnością do opalania, podczas gdy rs3212346: G>A jest związany z ciemną skórą, czarnymi włosami i silnymi zdolnościami do opalania [54]. W genie *HERC2* (HECT And RLD Domain Containing E3 Ubiquitin Protein Ligase 2) istotnymi loci są rs1636232, rs1133496 i rs2238289, przy czym haplotyp TCT wykazuje związek z jasnym kolorem skóry (a także kolo-

IrisPlex kolor oczu 6 markerów			HIrisPlex kolor oczu i włosów 24 markery			HIrisPlex-S kolor oczu, włosów i skóry 41 markerów		
Gen	SNP	Allel	Gen	SNP	Allel	Gen	SNP	Allel
<i>HERC2</i>	rs12913832	T	<i>HERC2</i>	rs12913832	T	<i>HERC2</i>	rs12913832	T
<i>IRF4</i>	rs12203592	T	<i>IRF4</i>	rs12203592	T	<i>IRF4</i>	rs12203592	T
<i>LOC105370627</i>	rs12896399	T	<i>LOC105370627</i>	rs12896399	T	<i>LOC105370627</i>	rs12896399	T
<i>TYR</i>	rs1393350	T	<i>TYR</i>	rs1393350	T	<i>TYR</i>	rs1393350	T
<i>SLC45A2</i>	rs16891982	C	<i>SLC45A2</i>	rs16891982	C	<i>SLC45A2</i>	rs16891982	C
<i>OCA2</i>	rs1800407	A	<i>OCA2</i>	rs1800407	A	<i>OCA2</i>	rs1800407	A
			<i>TYR</i>	rs1042602	T	<i>TYR</i>	rs1042602	T
			<i>MC1R</i>	rs1110400	C	<i>MC1R</i>	rs1110400	C
			<i>MC1R</i>	rs11547464	A	<i>MC1R</i>	rs11547464	A
			<i>KITLG</i>	rs12821256	G	<i>KITLG</i>	rs12821256	G
			<i>MC1R</i>	rs1805005	T	<i>MC1R</i>	rs1805005	T
			<i>MC1R</i>	rs1805006	A	<i>MC1R</i>	rs1805006	A
			<i>MC1R</i>	rs1805007	T	<i>MC1R</i>	rs1805007	T
			<i>MC1R</i>	rs1805008	T	<i>MC1R</i>	rs1805008	T
			<i>TUBB3</i>	rs1805009	C	<i>TUBB3</i>	rs1805009	C
			<i>MC1R</i>	rs201326893	A	<i>MC1R</i>	rs201326893	A
			<i>MC1R</i>	rs2228479	A	<i>MC1R</i>	rs2228479	A
			<i>PIGU</i>	rs2378249	C	<i>PIGU</i>	rs2378249	C
			<i>SLC24A4</i>	rs2402130	G	<i>SLC24A4</i>	rs2402130	G
			<i>SLC45A2</i>	rs28777	C	<i>SLC45A2</i>	rs28777	C
			<i>MC1R</i>	rs312262906	A	<i>MC1R</i>	rs312262906	A
			<i>LOC105374875</i>	rs4959270	A	<i>LOC105374875</i>	rs4959270	A
			<i>TYRP1</i>	rs683	G	<i>TYRP1</i>	rs683	G
			<i>MC1R</i>	rs885479	T	<i>MC1R</i>	rs885479	T
						<i>ANKRD11</i>	rs3114908	T
						<i>OCA2</i>	rs1800414	C
						<i>BNC2</i>	rs10756819	G
						<i>HERC2</i>	rs2238289	C
						<i>SLC24A4</i>	rs17128291	C
						<i>HERC2</i>	rs6497292	C
						<i>HERC2</i>	rs1129038	G
						<i>HERC2</i>	rs1667394	C
						<i>TYR</i>	rs1126809	A
						<i>OCA2</i>	rs1470608	A
						<i>SLC24A5</i>	rs1426654	G
						<i>ASIP</i>	rs6119471	C
						<i>OCA2</i>	rs1545397	T
						<i>RALY</i>	rs6059655	T
						<i>OCA2</i>	rs12441727	A
						<i>MC1R</i>	rs3212355	A
						<i>DEF8</i>	rs8051733	C

Dane z <https://hirisplex.erasmusmc.nl>

Rycina 6. Narzędzia IrisPlex, H-IrisPlex i H-IrisPlex-S i zawarte w nich markery DNA do predykcji koloru oczu, oczu + włosów i oczu + włosów + skóry (dane pobrane z <https://hirisplex.erasmusmc.nl> 20.05.2022)

rem oczu i włosów), podczas gdy haplotyp CTC był skorelowany z ciemnymi cechami [55]. Podobnie haplotyp CGG wynikający z *OCA2* (*OCA2* Melanosomal Transmembrane Protein gene) rs1800416 i rs1800407 okazał się dobrym predyktorem cech o ciemnym zabarwieniu.

Maroñas i współ. [56] wybrali 10 SNP genów związanych z kolorem skóry u osób pochodzenia europejskiego i pozaeuropejskiego, wśród których są: *SLC45A2* (Solute Carrier Family 45 Member 2) rs16891982 i *SLC24A5* (Solute Carrier Family 24 Member 5) rs1426654 (dwa najważniejsze markery, pierwszy dla pośredniego, drugi dla czarnego i białego koloru skóry z dokładnością przewidywania odpowiednio 77,6% dla pośredniego, 87,6% dla czarnego i 95,7% dla białego koloru), *ASIP* (Agouti Signaling Protein) rs60580017 (drugi ważny marker do klasyfikowania czarnej skóry i czwarty do klasyfikowania pośredniej karnacji), *TYRP1* (Tyrosinase-Related Protein 1) rs1408799 (do różnicowania pośredniej karnacji), *OCA2* (*OCA2* Melanosomal Transmembrane Protein) rs1448484 (ważny wkład w czerni i biel), *SLC45A2* (Solute Carrier Family 45 Member 2) rs13289 (oliwkowy kolor skóry u Europejczyków) oraz *KITLG* (KIT Ligand) rs10777129, *TPCN2* (Two Pore Segment Channel 2) rs3829241 i *SLC24A4* (Solute Carrier Family 24 Member 4) rs2402130.

Kolor oczu i włosów można przewidywać z dużą dokładnością, mając do dyspozycji stosunkowo niewielką liczbę markerów DNA. Kolor oczu, zróżnicowany od niebieskiego, poprzez szary, zielony, żółty, brązowy do czarnego,

zależy od różnic w ilości melaniny u osobnika oraz w ilości melanosomów w zewnętrznej warstwie tęczówki [57,58]. Podobnie międzyosobnicze zróżnicowanie koloru włosów jest spowodowane różnymi proporcjami eumelaniny i feomelaniny w melanosomach cebulki włosa [59].

Jednym z najbardziej rozpowszechnionych narzędzi predykcyjnych do przewidywania cech pigmentacyjnych człowieka jest narzędzie IrisPlex, opracowane w 2011 roku w oparciu o europejską bazę danych przez Walsh i współ. [60] ale jak się okazało odpowiednie również dla osób pochodzenia pozaeuropejskiego. Narzędzie to opiera się na 6 genach: rs12913832 *HERC2* (HECT and RLD Domain Containing E3 Ubiquitin Protein Ligase 2), rs1800407 *OCA2* (*OCA2* Melanosomal Transmembrane Protein Gene), rs12896399 *LOC105370627* (Uncharacterized LOC105370627), rs1393350 *TYR* (Tyrosinase) i rs12203592 *IRF4* (Interferon Regulatory Factor 4) i pozwala przewidywać niebieski i brązowy kolor oczu z dokładnością wynoszącą ponad 90%, jednak tylko w próbach pochodzących z populacji europejskich [60].

Na bazie IrisPlex opracowano narzędzia do jednoczesnego przewidywania koloru oczu i włosów (HIrisPlex) a także skóry, oczu i włosów (HIrisPlex-S) [61]. Narzędzie HIrisPlex składa się z 24 markerów, które tworzą SNP z narzędzia IrisPlex oraz 18 SNP dla barwy włosów (Ryc. 6). Są to: rs28777 *SLC45A2* (Solute Carrier Family 45 Member 2), rs12821256 *KITLG* (KIT Ligand), rs1042602 *TYR* (Tyrosinase), rs2402130 *SLC24A4* (Solute Carrier Family 24 Member 4), rs2378249 *PIGU* (Phosphatidylinositol Glycan

Anchor Biosynthesis Class U), rs683 *TYRP1* (Tyrosinase Related Protein 1), rs11547464, rs885479, rs1805008, rs1805006, rs1805007, rs2228479, rs1110400 *MC1R* (Melanocortin 1 Receptor), rs1805005, rs1805009, Y1520CH, N29insA *MC1R* (Melanocortin 1 Receptor) i rs4959270 *EXOC2* (Exocyst Complex Component 2) (Ryc. 6). Cechuje się dokładnością wynoszącą od 75 do 92%. Jednakże zmiana koloru włosów z wiekiem (włosy ciemnieją w wieku od 6 do 13 lat) obniża dokładność przewidywania blond włosów przez system H-IrisPlex do 69,5%. W przypadku włosów brązowych, rudych i czarnych, dokładność przewidywania wynosi kolejno: 78,5%, 80% i 87,5%. Wskazuje to na konieczność identyfikacji nowych markerów, które pozwoliłyby zwiększyć dokładność tego narzędzia.

Narzędzie H-IrisPlex-S opiera się łącznie na 41 markerach genów, które tworzą markery zestawu IrisPlex i H-IrisPlex i markery wymienione na Ryc. 6 [61]. Rycina 6 przedstawia sumaryczne zestawienie SNP w omówionych narzędziach: IrisPlex, H-IrisPlex i H-IrisPlex-S.

Określenie koloru oczu i włosów może też wskazywać na pochodzenie biogeograficzne. W przypadku oczu badania wskazują, że brązowe oczy liczniej występują na wschodzie Europy, barwy pośrednie w jej centralnej części (Francja, Niemcy), a niebieskie najliczniej występują w krajach skandynawskich (Dania, Szwecja, Norwegia) i na Islandii. Jeśli chodzi o włosy, to największy odsetek brązowowłosych osób jest we Francji i części Europy środkowej, osób blondwłosych w krajach Europy środkowej i krajach skandynawskich, a osób rudowłosych na Islandii oraz wyspach brytyjskich. Takie mapy, które przedstawiają zagęszczenie pewnych fenotypów, mogą stanowić przydatne narzędzie w kryminalistyce [62].

Warto dodać, że wyniki badań z IrisPlex, H-IrisPlex i H-IrisPlex-S zostały połączone w publicznie dostępne narzędzie służące do przewidywania koloru skóry, oczu i włosów. Pod adresem <https://hirisplex.erasmusmc.nl> można wstawić dane genotypu spośród 41 dostępnych markerów i otrzymać prawdopodobieństwo wyglądu spośród trzech dostępnych koloru oczu, czterech kolorów włosów oraz pięciu odcieni skóry.

Niedawno konsorcjum VISAGE (VISible Attributes through GENomics) finansowane przez UE opracowało różne narzędzia FDP do zastosowania w rutynowych analizach kryminalistycznych. Między innymi narzędzie o nazwie VISAGE Basic do przewidywania wyglądu zewnętrznego i pochodzenia biogeograficznego. Panel ten tworzy 41 SNP dla koloru oczu, włosów i skóry z H-IrisPlex-S i 115 markerów do wnioskowania o pochodzeniu biogeograficznym (trzy nakładające się na zestaw z H-IrisPlex-S), łącznie 153 markery [63]. Uzyskane wyniki (wysoka powtarzalność i zgodność w odniesieniu do pojedynczych przypadków) sugerują, że ten test może być użytecznym narzędziem do przewidywania podstawowych cech wyglądu i pochodzenia biogeograficznego w FDP [63].

Dotychczas zidentyfikowano trzy geny zaangażowane w morfologię ludzkich włosów, dwa u Azjatów, *EDAR* (Ectodysplasin A Receptor) i *FGFR2* (Fibroblast Growth Factor

Receptor 2) związane z grubością włosów i jeden u Europejczyków, gen *TCHH* (Trichohyalin) związany z prostymi włosami i wyjaśniający 6% wariacji fenotypowej w badanej populacji [14]. Jednakże, ponieważ grubość włosów u Azjatów jest mało zróżnicowana geny te nie są przydatne do celów FDP [64]. Chociaż badania bliźniąt oszacowały wysoką dziedziczność kręconych włosów u Europejczyków, odpowiedzialne za to geny nie zostały jeszcze zidentyfikowane. Biorąc pod uwagę różnorodność morfologii włosów wśród Europejczyków, przewidywanie DNA dotyczące włosów prostych, falistych lub kręconych byłoby w zasadzie przydatne do celów FDP jeśli w przyszłych badaniach będzie dostępna wystarczająca liczba predykcyjnych markerów DNA [65].

Łysienie typu męskiego

Łysienie typu męskiego (ang. *male pattern of baldness*, MPB), zwane również łysieniem androgenowym (ang. *androgenic alopecia*, AGA), jest formą wypadania włosów u dorosłych mężczyzn. Proces ten objawia się cofającą się linią włosów i/lub wypadaniem włosów na czubku lub przodzie głowy, co prowadzi do zmiany wyglądu fizycznego osoby. Na ten stan wpływają zarówno męskie hormony płciowe (androgeny) jak i predyspozycje genetyczne.

Badania przeprowadzone na bliźniątach wykazały, że odziedziczalność genów odpowiedzialnych za proces łysienia sięga do 80%, zarówno w przypadku wczesnej, jak i późnej utraty włosów. Jednak nowsze badania, oparte na SNP, oszacowały dziedziczność powszechnie występującej odmiany łysienia na ok. 50% [5]. Wśród wielu loci, które są prawdopodobnie zaangażowane w proces łysienia, najistotniejsze są te zlokalizowane w regionie q12 chromosomu X. Obecne są tam geny *AR/EDA2R* (Androgen Receptor/ Ectodysplasin A2 Receptor), które są bezpośrednio powiązane z produkcją receptora androgenowego i receptora ektodysplazyny A2. Inne istotne w tym procesie geny to: *EBF1* (EBF Transcription Factor 1), *TARDBP* (TAR DNA Binding Protein) oraz *HDAC9* (Histone Deacetylase 9), a także region 20p11. Wymienione pięć SNP wykazuje najwyższe wartości asocjacji, które umożliwiają przewidywanie wzorca męskiego łysienia z dokładnością rzędu 76,2%. Można ją zwiększyć do 86,4% po dodaniu kolejnych 15 markerów [5].

Marcińska i współ. [66] potwierdzili rolę 29 SNP w oznaczaniu MBP u Europejczyków w różnym wieku. Wśród nich najsilniejsze powiązanie wykazywały dwa SNP, rs5919324 w pobliżu genów *AR/EDAR2* (Androgen Receptor/ Ectodysplasin A2 Receptor) (chrX) i rs1998076 (chr20), a następnie trzy inne SNP: rs929626 w *EBF1* (EBF Transcription Factor 1), rs12565727 w *TARDBP* (TAR DNA Binding Protein) i rs756853 w *HDAC9* (Histone Deacetylase 9). Na tej podstawie stworzyli oni model predykcyjny MPB złożony z 20 SNP, który wykazywał wyższą swoistość (90%), ale niższą czułość (67,7%) w populacji mężczyzn powyżej 50 roku życia w porównaniu z mężczyznami poniżej 50 roku życia (czułość 87,1% i swoistość 42,4%) oraz lepszą moc predykcyjną dla MBP zapoczątkowanym w młodszym wieku (AUC = 0,761) niż dla MPB w późniejszym wieku (AUC = 0,657) [66].

W 2017 r. Hagenaaers i współ. [67] zbadali warianty genetyczne w kohorcie 52 000 angielskich mężczyzn będących w różnych stadiach łysienia, znajdując ponad 250 genetycznych loci związane z ciężką utratą włosów i opracowali algorytm przewidywania w celu identyfikacji tych, które związane z największym ryzykiem wypadania włosów (AUC = 0,78).

Wysokość ciała

Wysokość ciała jest cechą modelową w genetyce cech ilościowych i mapowaniu genów cech wspólnych człowie-

ka, wykorzystywaną w kryminalistyce do opisu sylwetki sprawcy, ofiary lub nieznanego ciała. Przewidywanie wysokości ciała człowieka dorosłego na podstawie śladów DNA napotyka poważne trudności spowodowane brakiem dokładnych testów. Pierwsze badania FDP wysokości ciała dla tysięcy holenderskich Europejczyków, opublikowane w 2009 roku przez Aulchenko i współ. wykazały, że 54 SNP pozwalają przewidywać wysokość ciała najwyższych osób w próbie (stanowią 5% całego rozkładu zmienności wysokości ciała) z dokładnością AUC 0,65 (czyli tylko nieznacz-

Tabela 3. Markery genetyczne powiązane z cechami wyglądu zewnętrznego stosowane do predykcji tych cech w kryminalistycznym fenotypowaniu DNA

Cecha fenotypowa [Autor]	Gen	Lokalizacja chromosomowa				
Kolor skóry [49,52,53,54,59]	<i>SLC45A2</i>	5p13.2	<i>SLC45A2/MATP</i>	5p13.2	<i>TYRP1</i>	9p23
	<i>BNC2</i>	9p22.3-p22.2	<i>TPCN2</i>	11q13.3	<i>TYR</i>	11q14
	<i>KITLG</i>	12q21.32	<i>SLC24A4</i>	14q32.12	<i>CEP152</i>	15q21.1
	<i>CTSN2</i>	15q21.1	<i>HERC2</i>	15q13.1	<i>MYEF2</i>	15q21.1
	<i>DUT</i>	15q21.1	<i>OCA2</i>	15q12-q13.1	<i>RALY</i>	20q11.22
	<i>SLC24A5</i>	15q21.1	<i>ANKRD11</i>	16q24.3	<i>DEF8</i>	16q24.3
Kolor włosów [49,53,57,59]	<i>MC1R</i>	16q24.3	<i>ASIP/PIGU</i>	20q11.22		
	<i>SLC45A2</i>	5p13.2	<i>SLC45A2/MATP</i>	5p13.2	<i>EXOC2</i>	6p.25.3
	<i>TYRP1</i>	9p23	<i>TYR</i>	11q14.3	<i>KITLG</i>	12q21.32
	<i>SLC24A4</i>	14q32.12	<i>OCA2</i>	15q12-q13.1 2	<i>TPCN</i>	11q13.3
	<i>HERC2</i>	15q13.1	<i>MC1R</i>	16q24.3	<i>SLC45A2/MATP</i>	5p13.3
Kolor oczu [53-56,58-60]	<i>ASIP/PIGU</i>	20q11.22	<i>GPR50</i>	Xq28		
	<i>SLC45A2/MATP</i>	5p13.3	<i>IRF4</i>	6p25-p23	<i>TYRP1</i>	9p23
	<i>TYR</i>	11q14.3	<i>SLC24A4</i>	14q32.12	<i>HERC2</i>	15q13.1
Morfologia włosów [64,65]	<i>OCA2</i>	15q12-q13.1				
	<i>ERRFI1/SLC45A1 TCHH</i>	1p36.23		1q21.3	<i>TCHHL1</i>	1q21
	<i>EDAR</i>	2q13	<i>WNT10A</i>	2q35	<i>SLC45A2/MATP</i>	5p13.3
	<i>LGR4</i>	11p14.1	<i>PRSS53</i>	16p11.2	<i>FRAS1</i>	4q21.21
Morfologia twarzy [76-81]	<i>PTK6</i>	20q13.33				
	<i>EDAR</i>	2q13	<i>LYPLAL</i>	1q41	<i>PRDM16</i>	1p36.32
	<i>PAX3</i>	2q36.1	<i>ALX3</i>	1p13.3	<i>TEX41</i>	2q22.3
	<i>MASP1</i>	3q27.3	<i>RAB7A/ACAD9</i>	3q27.3		
	<i>CACNA2D3</i>	3p21.1-p14.3	<i>DCHS2</i>	4q31.3	<i>HBEGF</i>	5q31.3
	<i>C5orf50</i>	5q35.1	<i>SUPT3H</i>	6p21.1	<i>SUPT3H/RUNX2</i>	6p21.1
	<i>EYA1</i>	8q13.3	<i>MSRA</i>	8p23.1	<i>DKK1</i>	10q21.1
	<i>ABCC9</i>	12p12.1	<i>MIPOL1</i>	14q13.3-q21.1	<i>PAX9</i>	14q13.3
<i>DICER1</i>	14q32.13	<i>GSC</i>	14q32.13	<i>TNFSF12</i>	17p13.1	
Typ męskiego łysienia [66,67]	<i>DHX35</i>	20q11.23-q12		20q12	<i>PAX1</i>	20p11.22
	<i>HDAC8</i>	Xq31				
Piegi [82-84]	<i>TARDBP</i>	1p36.22	<i>EBF1</i>	5q33.3	<i>HDAC9</i>	7p21.1
	<i>AR/EDA2R</i>	Xq12	20 p 11			
Wysoka krótkowzroczność [85,86]	<i>MC1R</i>	16q24.3	<i>IRF4</i>	6p25-p23		
	<i>MYP8</i>	3q26	<i>MYP9</i>	4q12	<i>MYP10</i>	8p23
	<i>MYP7</i>	11p13	<i>MYP6</i>	22q13.33	<i>MYP13</i>	Xq23
Wysokość ciała wybrane markery ¹ [5,70]	<i>MYP1</i>	Xq28				
	<i>DNM3</i>	1q24.3	<i>IHH</i>	2q35	<i>GHSR</i>	3q26.31
	<i>LCORL</i>	4p15.31	<i>HHIP</i>	4q31.21	<i>FBXW11</i>	5q35.1
	<i>NPR3</i>	5p13.3	<i>HMGA1</i>	6p21.31	<i>MICA</i>	6p21.33
	<i>GPR26</i>	10q26.13	<i>SDR16C5</i>	8q12.1	<i>PIF</i>	10q22.3
	<i>EFEMP1</i>	2p16.1	<i>SOCS2</i>	12q22	<i>ACAN</i>	15q26.1
	<i>PML</i>	15q24.1	<i>NOG</i>	17q22	<i>GH1</i>	17q23.3
	<i>STC2</i>	5q35.2	<i>ZFAT</i>	8q24.22	<i>AR</i>	Xq12
	<i>CRISPLD2</i>	16q24.1				
	CODIS [26]	<i>CSF1PO</i>	5q33.3-34	<i>D3S1358</i>	3p21.31	<i>D5S818</i>
<i>D7S820</i>		7q21.11	<i>D8S1179</i>	8q24.13	<i>D13S317</i>	13q31.1
<i>D16S539</i>		16q24.1	<i>D18S51</i>	18q.21.33	<i>D21S11</i>	21q11.2-q21
<i>FGA</i>		4q31.3	<i>TH01</i>	11p15.5	<i>TPOX</i>	2p25.3
<i>vWA</i>		12p.13.31	<i>D1S1656</i>	1q42	<i>D2S441</i>	2p14
<i>D2S1338</i>		2q35	<i>D10S1248</i>	10q26.3	<i>D12S391</i>	12 341 Mb
<i>D19S433</i>		19q12	<i>D22S1045</i>	22q12.3		

¹Wybrano tylko kilkanaście loci z 697 zidentyfikowanych w GWAS, które były przedmiotem analiz w FDP; dokładność predykcji (AUC~65%) tylko nieznacznie przekraczała wartość losową AUC=50%.

nie lepiej niż losowo lub rzut monetą, co odpowiada AUC 0,5 [14].

W 2014 roku międzynarodowe konsorcjum GIANT (Genetics of Anthropomorphic Traits) opublikowało wyniki badań GWAS dotyczące wysokości ciała z udziałem 250 000 osób, w których wykryto 697 SNP istotnie związanych z wysokością ciała. Okazało się jednak, że te polimorfizmy łącznie wyjaśniają zaledwie 15% całkowitej wariancji fenotypowej wysokości ciała w badanej populacji i to w dodatku europejskiego pochodzenia [68]. Nowsze badania z udziałem 711 428 osób z podobnej kohorty określiły 83 warianty kodowania i intronów z częstością małych alleli (MAF) od 0,1% do 4,8% w tej populacji i wielkości efektu do 2,1 cm na ostateczną wysokość ciała osoby dorosłej [69]. Rotwein przeprowadził analizę częstości wariantów genów w genomach 213 158 osób reprezentujących różne populacje z całego świata zebranych przez konsorcjum Genome Aggregation Database (GnomAD) [70]. Okazało się, że większość alleli powiązanych ze zwiększaniem się lub zmniejszaniem ostatecznej (dorosłej) wysokości ciała jest znacznie mniej rozpowszechniona w GnomAD niż stwierdzono wcześniej w innych kohortach (np. w GIANT), przy czym 4 z 5 wariantów podstawienia aminokwasów o największym wpływie na ostateczną (dorosłą) wysokość ciała występuje częściej w populacji europejskiej niż w innych grupach [70]. Jing i współ. podjęli próbę predykcji wysokości ciała opartą na SNP z programu GIANT w odniesieniu do populacji wschodnioazjatyckich. I choć nie uzyskali spektakularnych wyników jeśli chodzi o zakres wyjaśnianej zmienności wysokości ciała przez te SNP to pokrywała się ona z zakresami zmienności wyznaczonymi metodami antropologicznymi (morfologicznymi) [71].

Tymczasem, jak wykazały badania bliźniąt, odziedziczalność wysokości ciała (udział tylko czynnika genetycznego – wariancji addytywnej w całkowitej wariancji fenotypowej) szacuje się na 80%, chociaż ta wartość zmienia się do 20 roku życia i nieznacznie różni się w zależności od płci i pochodzenia etnicznego i geograficzno-kulturowego badanych [72]. Wśród czynników poza genetycznych wpływających na całkowitą wariancję wysokości ciała wskazują się: warunki w życiu prenatalnym (zdrowie matki, przebieg ciąży, cechy łożyska, warunki środowiska wewnątrzmacicznego), wzrastanie po narodzinach, w szczególności w okresie dzieciństwa i adolescencji (neurohormonałna regulacja procesu wzrastania, zdrowie ogólne, odżywianie, warunki bytowe i inne). Wobec tak złożonej determinacji genetycznej i środowiskowej wysokości ciała dokładność jej przewidywania z DNA nie jest wysoka. Podczas gdy początkowe badania osiągnęły dokładność ~65% (AUC 0,65 gdzie 50% oznacza przewidywanie losowe), w najnowszym modelu przewidywania wysokości ciała zaproponowanym przez Liu i współ. w 2019 roku i opartym na 689 SNP nie udało się podnieść tej wartości powyżej 75% (AUC 0,75) [73]. Dalsze badania zatem będą zmierzać do znalezienia polimorfizmów, które pozwolą na przewidywanie wysokości ciała na podstawie DNA z dokładnością wystarczająco wysoką by móc je wstawić do modeli predykcyjnych FDP. Może się okazać, że wiele tysięcy SNP będzie potrzebnych do wyjaśnienia 80% dziedzicznej zmienności wysokości ciała. Dlatego też SNP wymienione w tabeli 3 należy traktować jako

wybrane przykłady, które próbuje się, z różnym skutkiem, włączać do FDP.

Nie udało się również opracować wiarygodnych testów do przewidywania masy ciała. Badania szesnastu SNP mitochondrialnego genu cytochromu b wykazały słaby związek otyłości z dwiema niesynonimicznymi mutacjami (odpowiadającymi podstawieniom nukleotydów, odpowiednio A15043G i C15677A) u dwóch otyłych mężczyzn i dwóch otyłych kobiet z 66 badanych [74]. Tak więc, możliwość zastosowania genu cyt-b do przewidywania otyłości wymaga dalszych badań [75].

Morfologia twarzy

Wygląd (morfologia) twarzy jest najbardziej widoczną cechą fizyczną, której dokładne przewidywanie ma istotne implikacje dla kryminalistycznych badań fenotypu. To struktura zbudowana z wielu elementów morfologicznych takich jak nos, oczy, czoło i inne. Cechy twarzy rozwijają się w ontogenezie twarzoczaszki na który wpływa duża liczba genów (większość jest wciąż nieznana, choć w ciągu ostatnich kilku lat przeprowadzono kilka badań asocjacyjnych całego genomu GWAS i zidentyfikowano wiele loci związanych z cechami twarzy) a także stresory środowiskowe, czy płęć. W FDP dodatkowym czynnikiem jest wiek, który moduluje zmiany zachodzące w wyglądzie twarzy w czasie życia osobniczego. Badanie morfologii twarzy sprowadza się do analizy powiązań markerów DNA z elementami twarzy np. jej wysokością i szerokością, długością nosa, rozstawem oczu czy szerokością ust, opisanymi ilościowo na podstawie pomiarów antropometrycznych lub wyliczonymi w oparciu o te pomiary, bądź dwuwymiarowymi wskaźnikami opisującymi kształt danej struktury [5]. Podjęto również próby wykorzystania trójwymiarowego modelu twarzy (3dDFM) pełniej odzwierciedlającego złożoność jej morfologii. Próby podejmowane przez badaczy począwszy od 2014 roku (Paternoster i współ., Claes i współ.) nie przyniosły oczekiwanych rezultatów w postaci narzędzi umożliwiających przewidywanie cech twarzy z dokładnością wyższą od losowej [5]. Należy dodać, że do opisanych badań włączono markery genetyczne związane z chorobami powodującymi deformacje twarzy, takie jak rozszczep podniebienia czy rozszczep wargi. Niektóre z tych markerów genetycznych powiązane z rozwojem twarzoczaszki, co w konsekwencji doprowadziło do powiązania ich z normalną zmiennością rysów twarzy [5]. Przykładem takiego markera jest gen *PAX3* (*Paired Box 3*). Gen ten koduje czynnik transkrypcyjny, który jest obecny w komórkach grzebienia nerwowego i ma związek z zespołem Waardenburga (grupa odziedziczalnych schorzeń, obejmuje głuchotę i błądź skóry, włosów i koloru oczu a w morfologii twarzy szeroki grzbiet nosa lub poszerzenie podstawy nosa) [5]. Podobnie jak gen *PAX3* (*Paired Box 3*) zidentyfikowano później inne markery genetyczne, takie jak *EDAR* (*Ectodysplasin A Receptor*), *LYPLAL1* (*Lysophospholipase Like 1*), *PRDM16* (*PR/SET Domain 16*), *DKK1* (*Dickkopf WNT Signaling Pathway Inhibitor 1*), *TNFSF12* (*TNF Superfamily Member 12*), *CACNA2D3* (*Calcium Voltage-Gated Channel Auxiliary Subunit Alpha2delta 3*) i *SUPT3H* (*SPT3 Homolog, SAGA And STA-GA Complex Component*) [76]. Jednak tak jak w przypadku markerów stosowanych w przewidywaniu wysokości ciała

tak i te geny zdają się mieć niewielki wkład w ogólną zmienność morfologii twarzy.

Shaffer i współ. [77] wykazali, że *MAFB* (MAF BZIP Transcription Factor B), *PAX9* (Paired Box 9), *MIPOL1* (Mirror-Image Polydactyly 1), *ALX3* (ALX Homeobox 3), *HDAC8* (Histone Deacetylase 8) i *PAX1* (Paired Box 1) odgrywają istotną rolę w rozwoju twarzoczaszki lub w zespołach patologicznych dotyczących twarzy. Li i współ. [78] przetestowali wpływ 125 SNP związanych z kształtem twarzy na rysy twarzy w mieszanej populacji europejsko-azjatyckiej składającej się z 612 osób. Osiem SNP wykazało istotny związek z jedną lub więcej cechami twarzy: rs3827760 *EDAR* (Ectodysplasin A Receptor), rs5781117 *LYPLAL1* (Lysophospholipase Like 1), rs4648379 *PRDM16* (PR/SET Domain 16), rs7559271 *PAX3* (Paired Box 3), rs1194708 *DKK1* (Dickkopf WNT Signaling Pathway Inhibitor 1), rs80067372 *TNFSF12* (TNF Superfamily Member 12), rs56063440 *CACNA2D3* (Calcium Voltage-Gated Channel Auxiliary Subunit Alpha2delta 3) i rs227833 *SUPT3H* (SPT3 Homolog, SAGA and STAGA Complex Component), który jednak wyjaśniał tylko 6,49% zmienności cech twarzy, po uwzględnieniu płci, wieku i BMI. Wszystkie z ośmiu SNP miały inną częstość alleli u Europejczyków i Azjatów, a cztery z nich (rs4648379, rs3827760, rs7559271 i rs1194708) wykazały odwrotną częstość alleli w obu grupach.

W 2019 roku Mbadiwe z zespołem [79] zaproponowali, w oparciu o dane literaturowe, panel złożony z 6816 SNP związanych z ludzką twarzą nazwany FaceSNP. Ten panel jest dostępny na życzenie. Ponadto, zidentyfikowali chromosomy, 10, 17, 1 i 5, których badania wydają się dawać lepsze wyniki w przewidywaniu fenotypu cech twarzy człowieka niż dotychczasowe.

Liu M. i współ. [80] w badaniach populacji chińskiej wykazali istotny związek cech twarzy z 12 SNP, które są łącznie odpowiedzialne za do 3,89% wariacji kontrolowanej na wiek i BMI. Są to: rs17479393 *TEX41* (Testis Expressed 41), rs974448 *PAX3* (Paired Box 3), rs2977562 *RAB7A/ACAD9* (*RAB7A* Member RAS Oncogene Family/Acyl-CoA Dehydrogenase Family Member 9), rs9995821, rs2045323, C5orf50 rs6555969 *DCHS2* (Dachsous Cadherin-Related 2), rs1852985 *SUPT3H/RUNX2* (SPT3 Homolog, SAGA and STAGA Complex Component/ *RUNX* Family Transcription Factor 2), rs11782517 *MSRA* (Methionine Sulfoxide Reductase A), rs10504499 *EYA1* (*EYA* Transcriptional Coactivator and Phosphatase 1), rs2224309 *GSC* (Goosecoid Homeobox), rs7161418 *DICER1* (Dicer 1, Ribonuclease III) i rs2206437 *DHX35* (DEAH-Box Helicase 35).

Najnowsze badanie GWAS White'a i współ. [81] wskazuje na 203 miejsca w genomie związane z wielowymiarową morfologią twarzy u zdrowych osób. Wśród nich 53 znajdują się w regionie, który nie był wcześniej znany. Co więcej, wykazano interakcję między wariantami w różnych loci wpływającymi na zmienność kształtu twarzy. Równocześnie zidentyfikowano zestawy genów, które współdziałają w budowaniu struktur ludzkich twarzy.

Mówiąc o przewidywaniu rysów twarzy, należy dodać, że możliwe jest dość dokładne przewidywanie obecności

piegów. Pieggi to nierównomiernie rozłożone skupiska melaniny w skórze, które mogą pojawiać się na różnych obszarach ciała (twarz, szyja, ramiona, plecy i nogi) głównie u osób o jasnej karnacji i jasnych włosach (dlatego też najczęściej występują u ludzi pochodzenia europejskiego). Ich obecność, zwłaszcza na odsłoniętych obszarach, takich jak twarz, reprezentuje szczególną i łatwo widoczną cechę fenotypową. Obok genetycznych uwarunkowań istotną rolę w powstawaniu piegów odgrywa słońce, które wpływa na przyciemnienie wcześniej jasnych przebarwień, które pojawiają się wówczas na plecach, karku i ramionach, czyli w miejscach najbardziej narażonych na działanie promieni słonecznych.

Cao i współ. [14] analizowali związek między wariantami (SNP: rs33832559, rs2228478, rs2228479, rs885479) genu receptora melanokortyny-1 (*MC1R*) a obecnością piegów u 225 chińskich pacjentów. Badania te nie wykazały statystycznie istotnej różnicy u osób z piegami w porównaniu z grupą kontrolną, chociaż *MC1R* (Melanocortin 1 Receptor) został wskazany jako główna genetyczna determinanta fenotypu piegów.

Zaorska i współ. [82] wykazali silny związek między piegami a innym locus genetycznym, rs12203592 w *IRF4* (Interferon Regulatory Factor 4) u 222 Polaków. Również w tym przypadku nie zaobserwowano wysokiej korelacji z genem *MC1R* (rs1805007).

Na podstawie predyktorów genetycznych wcześniej skorelowanych z pigmentacją u człowieka, Kukla-Bartoszek i współ. [83] opracowali model predykcyjny dla obecności piegów, podzielony na trzy kategorie: brak piegów, średnia liczba piegów i duża liczba piegów, i uzyskali umiarkowaną dokładność (odpowiednio AUC = 0,75, 0,66 i 0,79).

Model zaproponowany przez Hernando i współ. [84] w 2018 roku do przewidywania piegów obejmuje pięć determinantów genetycznych: warianty R genu *MC1R* (Melanocortin 1 Receptor) takie jak: rs1805006, rs11547464, rs1805007, rs1110400, rs1805008 i rs1805009, zdefiniowane jako allele „R” ze względu na ich silny związek z fenotypem rudego koloru włosów w populacji, rs12203592 *IRF4* (Interferon Regulatory Factor 4), warianty r genu *MC1R* (rs1805005, rs2228479 i rs885479, zdefiniowane jako allele „r” ze względu na ich niższy związek z fenotypem rudego koloru włosów), rs4911442 *ASIP* (Agouti Signaling Protein) i rs2153271 *BNC2* (Basonuclin 2). Trafność prognostyczna modelu predykcyjnego sięga 74,13%.

Krótkowzroczność

Wysoka krótkowzroczność (ang. *high myopia*) to stan charakteryzujący się wysoce ujemną wadą refrakcji (>-6 do -8 dioptrii) w kontekście wydłużenia oka (26-26,5 mm). Chociaż jej patogenezą nie jest do końca poznana, aktualna wiedza sugeruje, że wysoka krótkowzroczność jest złożoną cechą, na której występowanie wpływa wiele genów [85].

Metaanaliza autorstwa Tang i współ. [85] ujawniła związek z rs644242 na genie *PAX6* (Paired Box 6), kodującym białko niezbędne do rozwoju oczu u osób pochodzenia

azjatyckiego. Xie i współ. [86] stwierdzili statystycznie istotną korelację w chińskiej populacji Han dla rs8004825 i rs12976445, obydwie zlokalizowane w sekwencjach mikro-RNA (małe sekwencje RNA zaangażowane w regulację ekspresji genów). Jednak wychodząc z tych wstępnych wyników dotyczących korelacji, oba artykuły stwierdzają, że biorąc pod uwagę poligeniczny charakter schorzenia, aby najpierw skorelować, a następnie z dużą dokładnością przewidzieć wysoką krótkowzroczność, należy zbadać więcej markerów genetycznych w większej liczbie próbek.

Przewidywanie pochodzenia biogeograficznego

Część markerów DNA może stanowić źródło informacji o składzie genów pochodzących od przodków u danego osobnika. To pozwoliłoby na poznanie jego pochodzenia biogeograficznego: afrykańskiego, europejskiego, azjatyckiego czy amerykańskiego. W związku z tym użycie markerów informujących o pochodzeniu (ang. *ancestry informative marker* AIM) umożliwia wnioskowanie o tym jakie ktoś ma korzenie, co w kryminalistyce może wspomóc zeznania świadków lub rzucić nowe światło na dowody zebrane z miejsca zbrodni [5]. Informacje o pochodzeniu nie mogą jednak stanowić jedyne kryterium określania wyglądu osobnika. Należy rozumieć różnicę między pochodzeniem biogeograficznym a obaloną już przez antropologów typologiczną koncepcją rasy. Procent genów od przodków u danego osobnika niekoniecznie będzie odzwierciedlał jego wygląd. Widać to szczególnie w próbach pochodzących z populacji mieszanych, gdzie AIM wykazują, że nie ma bezpośredniej korelacji między wyglądem a pochodzeniem biogeograficznym [5].

Obecnie stosowane narzędzia kryminalistyczne umożliwiają dość ograniczone wnioskowanie o pochodzeniu biogeograficznym osobnika w przypadku mieszanych genów z różnych rejonów, zwłaszcza, gdy ich wymieszanie się miało miejsce wiele pokoleń wcześniej. Mimo to udało się wyselekcjonować na tyle dużo pochodzących z różnych rejonów świata markerów informujących o pochodzeniu (głównie SNP), że było możliwe opracowanie i uprawomocnienie kilku kryminalistycznych systemów, które pozwalają na przewidywanie pochodzenia osobnika. Testy te są wiarygodne ze względu na dużą liczbę i różnorodność markerów DNA jakie się w nich stosuje do odróżniania jednej populacji od drugiej. Przykładem są dwa komercyjne testy: test Precision ID Ancestry Panel, który korzysta ze 165 SNP i rozróżnia siedem miejsc pochodzenia biogeograficznego lub test ForenSeq DNA Signature Prep Kit, który korzysta z 56 SNP i rozróżnia cztery miejsca pochodzenia próby. Testy te pozwalają na przypisanie pochodzenia biogeograficznego próby na poziomie kontynentalnym. W przypadku testu Precision ID Ancestry Panel może to być pochodzenie: europejskie, afrykańskie, amerykańskie, subsaharyjskie, wschodnioazjatyckie, zachodnioazjatyckie i oceaniczne. Jeśli chodzi o test ForenSeq DNA Signature Prep Kit, to przypisuje on tylko do czterech miejsc pochodzenia: europejskie, afrykańskie, amerykańskie oraz wschodnioazjatyckie [87].

Warto pamiętać, że jakość przewidywania pochodzenia biogeograficznego nie zależy tylko od stosowanych markerów, ale też od danych genetycznych danej populacji

używanych jako dane referencyjne. Testy mogą skutecznie przewidywać pochodzenie tylko wtedy, gdy dysponujemy dobrą bazą danych, do których może zostać zaklasyfikowana dana próba. W związku z tym raport takich testów powinien zawierać opis danych służących za punkt odniesienia.

DYSKUSJA

Dyskusję wyników należy poprzedzić uwagą na temat niesamowitego tempa w jakim publikowane są prace z zakresu genetyki sądowej (ang. *forensic genetics*). Jak wskazuje wyszukiwarka EBSCO, od marca 2021 do marca 2022 roku przybyło 188 prac, które z przyczyn obiektywnych nie zostały uwzględnione w niniejszym przeglądzie. Należy również pamiętać, że autorzy omawianych prac stosowali różne kategoryzacje cech fenotypowych, co jednak nie wpłynęło na ich ostateczny wynik.

Opracowanie różnych metod przewidywania EVC, takich jak kolor oczu, kolor włosów (w tym siwienie włosów), morfologia włosów (w tym wypadanie włosów), kolor skóry, piegi, morfologia twarzy, wysoka krótkowzroczność, otyłość i wysokość ciała osoby dorosłej, na podstawie markerów DNA wyekstrahowanego ze śladu pozostawionego na miejscu przestępstwa lub nieznanego osoby, stanowi dodatkowe narzędzie dochodzeniowe w kontekście kryminalistycznych technik śledczych. Do tej pory uzyskano obszerną wiedzę na temat cech pigmentacji, tj. oczu, włosów i koloru skóry. Na podstawie dotychczasowej wiedzy o cechach pigmentacyjnych skonstruowano narzędzia predykcyjne do jednoczesnego przewidywania koloru oczu i włosów (HirisPlex) oraz oczu, włosów i skóry (HirisPlex-S). Jednak o ile te narzędzia oferują dużą dokładność przewidywania najbardziej ekstremalnych fenotypów, problem pojawia się w przypadku fenotypów pośrednich a także w odniesieniu np. do włosów blond, które wraz z wiekiem ciemnieją lub w starszym wieku siwieją. Istnieje również prawdopodobieństwo popełnienia błędu związane z pochodzeniem biogeograficznym populacji, do której te narzędzia są stosowane, zwłaszcza w przypadku populacji mieszanych, co sugeruje, że powinny być one dostosowane do różnych regionów geograficznych. Ta uwaga odnosi się również do narzędzia VISAGE BT-A&A (PSeq) dla koloru oczu i włosów, poszerzone o wnioskowanie o pochodzeniu biogeograficznym. Należy pamiętać, że wnioskowanie na podstawie DNA o wyglądzie fizycznym osoby musi być ostrożnie stosowane w kryminalistyce do konkretnych przypadków, ponieważ błędna prognoza może prowadzić dochodzenie na niewłaściwe tory. Potrzebne są dalsze badania mające na celu poszerzenie puli SNP dla każdej cechy fenotypowej i dla poszczególnych osobników pochodzących z różnych regionów geograficznych.

Podczas gdy FDP cech pigmentacyjnych dysponuje wiarygodnymi narzędziami do ich predykcji, FDP innych cech, jak np. wysokości i masy ciała, morfologii twarzy są nadal w fazie intensywnych badań, dalekie od wyjaśnienia drogi od genotypu do fenotypu. Część proponowanych narzędzi wciąż nie oferuje wystarczająco dokładnych wyników i wymaga dalszego dopracowania.

Pomimo potencjalnej użyteczności FDP, to innowacyjne podejście nie uzyskało jeszcze powszechnej akceptacji. Toczy się ożywiona debata na temat zasadności jego stosowania w wymiarze sprawiedliwości w sprawach karnych [88]. Koncentruje się ona przede wszystkim na kwestiach prawnych i społeczno-etycznych [89]. Biorąc pod uwagę, że testy FDP są probabilistyczne, uzyskiwane informacje nie identyfikują pojedynczej lub konkretnej osoby (potencjalnego podejrzanego) ale grupę osób o podobnych widocznych cechach, co prowadzi do obaw prawnych i etycznych, takich jak: niemożliwość wykonania masowego badania całej podejrzananej populacji lub ryzyko wygenerowania uprzedzeń rasowych i stygmatyzacji lub dyskryminację już skierowaną na grupy mniejszościowe w społeczeństwie, szczególnie w odniesieniu do wnioskowania o pochodzeniu biogeograficznym, dotyczącym pochodzenia geograficznego niż specyficznego wyglądu fizycznego osobników błędnie kojarzonym z pojęciem rasy [90].

Ponadto, proces FDP (por. Ryc. 2) obarczony jest szeregiem ograniczeń, wśród których warto wymienić cztery główne: 1) często wielkość próby jest zbyt mała i niereprezentatywna aby uzyskać statystycznie wiarygodne wyniki; 2) brak standardowej metody fenotypowania zarówno na etapie badań genetycznych jak i predykcyjnych, nie ma też jednego predykcyjnego modelu statystycznego, co może powodować problemy z powtarzalnością analiz, oceną ich wiarygodności i dzieleniem się wynikami badań z innymi jednostkami (laboratoria kryminalistyczne, grupy badawcze itp.). Powyższa konstatacja wskazuje na 3) konieczność utworzenia standardowej bazy danych, wobec jej braku, a zawierającej dane na temat cech fenotypowych i ich markerów genetycznych w oparciu o wyniki najnowszych badań; taka baza danych pozwoliłaby dodatkowo na ujednoczenie porównywalnych estymatorów prognoz i parametrów modeli. Kolejnym słabym ogniwem FDP są 4) strategie finansowania badań koncentrujące się na przypadkach chorobowych. Brak wyników badań związanych z normalną zmiennością fenotypu. Te bowiem nie uzyskują wsparcia finansowego argumentowanego względami etycznymi.

Równoległe z rozwojem badań FDP podejmowane są systematyczne próby podwyższania ich jakości, w tym ich standaryzacji. W Europie merytoryczne i etyczno-prawne aspekty FDP nadzoruje Europejska Grupa ds. Profilowania DNA (European DNA Profiling Group EDAP), która zapoczątkowała swoją działalność w 1988 roku w ramach Międzynarodowego Towarzystwa Genetyki Sądowej (International Society for Forensic Genetics ISFG) oraz Europejska Sieć Laboratoriów Kryminalistycznych (European Network of Forensic Science Institutes ENFSI) zawiązana w 1995 roku. W USA i Kanadzie standardy jakości w zakresie FDP wytycza naukowa grupa badawcza SWGDAM (The Scientific Working Group on DNA Analysis Methods) na podstawie rekomendacji dyrektora Federalnego Biura Śledczego (FBI).

Szczegółową informację na temat działalności wymienionych grup badawczych wraz z rekomendowanymi loci do badań FDP, dobrymi praktykami i wieloma innymi doku-

mentami można znaleźć w raporcie z działalności INTERPOLU w latach 2016–2019 [91].

W 2017 roku rozpoczęto realizację finansowanego przez Unię Europejską (UE) programu badań i innowacji, którego celem jest predykcja cech wyglądu zewnętrznego, wieku i pochodzenia biogeograficznego z próbek DNA pobranych z miejsca zdarzenia w ramach utworzonego w tym celu konsorcjum VISAGE (Visible Attributes Through Genomics), które wymieniono we fragmencie niniejszej pracy dotyczącym narzędzia VISAGE Basic. Konsorcjum VISAGE tworzy 13 partnerów z 8 państw członkowskich UE: Holandia, Polska, Hiszpania, Austria, Niemcy, Wielka Brytania, Francja i Szwecja (<http://www.visage-h2020.eu>). Działania tej grupy obejmują badania podstawowe nad markerami epigenetycznymi, tworzenie prototypowych narzędzi opartych na masowym równoległym sekwencjonowaniu, pakietów programów do statystycznej obróbki danych wraz z oprogramowaniem. Szczególna uwaga w tym programie jest poświęcona analizom, które pozwolą na opracowanie sposobów minimalizowania etycznych, prawnych i społecznych zagrożeń (ang. *ethical, legal and societal aspects* ELSA) jakie niesie ze sobą FDP by na żadnym etapie nie dochodziło do naruszenia artykułu 8 Europejskiej Konwencji Praw Człowieka. W 2018 r. został opublikowany 123-stronicowy raport "The regulatory landscape of forensic DNA phenotyping in Europe", który analizuje ramy regulacyjne i prawne dotyczące FDP w 8 państwach członkowskich UE uczestniczących w projekcie i zwraca uwagę na kwestie prawne specyficzne dla danego kraju (http://www.visage-2020.eu/Report_regulatory_landscape_FDP_in_Europe2.pdf). Na tej bazie zostały przygotowane rekomendacje dla lokalnych grup językowych, w tym dla polskich uczestników projektu [92].

Świadomość mocnych i słabych stron FDP generuje potrzebę dalszej dyskusji na temat nie tylko zastosowań i użyteczności FDP, ale także związanych z nimi zagrożeń, a ponadto istnieje potrzeba nowych przepisów, zwłaszcza biorąc pod uwagę, że tylko kilka krajów uchwaliło już szczegółowe przepisy w tej sprawie, podczas gdy w większości innych krajów istnieje próżnia prawna lub techniki te są szczególnie zabronione [93]. Obawy te są potęgowane przez firmy nastawione na zysk, takie jak Parabon NanoLabs, które promują FDP jako technologię zdolną do tworzenia wyglądu twarzy osób z samej próbki DNA, często w oparciu o testy, które są niedopracowane, nieudokumentowane naukowo i niepotwierdzone.

Przykłady państw wykorzystujących FDP w postępowaniu sądowym wskazują, że zmienia się podejście, bo zwiększa się znaczenie tej metody w kryminalistyce. Wśród badaczy i ekspertów rośnie przekonanie, że metoda ta ma duży potencjał i może bardzo pomóc w rozwiązywaniu spraw kryminalnych. Metody fenotypowania będą tym skuteczniejsze, im bardziej będą rozwijane, a dzięki temu wzrośnie ich wiarygodność.

W przyszłości będzie musiało się też zmienić podejście sądów. Aktualnie zdarzają się przypadki dopuszczania dowodów opartych na sekwencjonowaniu DNA, jednak dalej sądy preferują tradycyjne, często niebiologiczne dowody

[89]. Fakt ten stoi w opozycji do dążeń naukowców do opracowywania coraz to lepszych technik molekularnych. Przykładem takiego wyzwania dla sądów może być niedawne odkrycie, że ludzie są nosicielami profili bakteryjnych skóry, które są charakterystyczne dla każdej osoby. Odkrycie to jest potwierdzane przez sekwencjonowanie DNA. Jest to dopiero wstępne badanie, jednak wskazuje możliwość porównania kolonii bakterii z różnych przedmiotów np. klawisze komputera, z koloniami na skórze podejrzanego osoby, co stanowiłoby kolejne ułatwienie w rozwiązywaniu spraw kryminalnych [94,95].

Rozwój technik analizy DNA pozwoli też na bardziej spersonalizowane wyniki, ponieważ kod genetyczny różni się u każdego osobnika. Dzięki rozwojowi NGS będzie można zwiększyć liczbę danych analizowanych w kryminalistyce, a do tego analizować te dane szybciej i taniej. Mimo, że wiele fragmentów DNA może dostarczyć informacji, to wydaje się, że w praktyce śledczej największe zastosowanie będą mieć wymieniane już polimorfizmy: STR, których ogromne ilości są zebrane w bazach danych oraz SNP, mtDNA oraz markery z chromosomu Y, które pozwalają na uzupełnienie wiedzy jaką można uzyskać dzięki analizie STR. Eksperci wskazują, że metody NGS wciąż nie są dość skuteczne, aby używać ich w praktyce, jednak jakość tych metod wzrasta pod kątem dokładności oraz niezawodności. Wymaga to jedynie czasu.

Wraz z rozwojem technik sekwencjonowania rośnie zainteresowanie możliwością wykonywania analiz DNA już na miejscu zdarzenia. Urządzenia do analizy DNA stają się coraz mniejsze i poręczniejsze, a to pozwala na ograniczenie uszkodzeń próbek, zmniejszenie rozprzestrzeniania szkodliwych substancji i mikroorganizmów z miejsca zdarzenia oraz niemal natychmiastowe rozpoczęcie poszukiwań potencjalnego sprawcy.

Wydaje się, że w najbliższej przyszłości zostaną rozwiązane problemy natury etycznej. Fenotypowanie DNA oraz przechowywanie danych zebranych dzięki tym badaniom choć obecnie budzi mieszane opinie, jednak panuje powszechne przekonanie, w tym autorów tej pracy, że kontrowersje etyczne to nieodzowny element rozwoju naukowego, szczególnie dotyczącego biologii człowieka. Niemal wszystkie nowatorskie rozwiązania budziły wątpliwości etyczne, gdy się pojawiały, jednak ich rozwój, rozpowszechnienie i praktyczne pozytywne skutki, pozwoliły na przyjęcie ich przez opinię publiczną jako coś normalnego. Toteż, wydaje się wysoce prawdopodobne, że z biegiem czasu coraz więcej państw będzie dopuszczać stosowanie takich metod w kryminalistyce, a korzystanie z nich oraz z baz danych nie będzie budziło społecznych kontrowersji.

WNIOSKI

Rozwój badań nad materiałem biologicznym pobranym z miejsca zdarzenia, w tym profilowanie DNA, zwiększone możliwości oprogramowania i przechowywania dużych zbiorów danych, znacząco zmieniły przebieg śledztw i procesów sądowych. Fakt, że techniki FDP są systematycznie udoskonalane stwarza nowe perspektywy ich zastosowania w kryminalistyce. W przyszłości duże nadzieje wiąże się z

sekwencjonowaniem nowej generacji NGS. Dotychczasowe badania już pokazały, że stosowanie NGS zwiększa ilość informacji, jakie będzie można otrzymać z próbek znalezionych na miejscu zdarzenia. Metoda ta nie tylko zapewnia większą dokładność i większe możliwości rozróżniania, ale też dostarcza informacji o pochodzeniu lub cechach fenotypowych. Systematycznie zwiększająca się liczba specjalistów, którzy będą korzystać z nowych odkryć i rozwijających się baz danych z pewnością przyczyni się do dalszego postępu w kryminalistycznym fenotypowaniu DNA.

PIŚMIENNICTWO

1. Algee-Hewitt BFB, Michael ED, Kim J, Li JZ, Rosenberg NA (2016) Individual identifiability predicts population identifiability in forensic microsatellite markers. *Current Biology* 26: 935-942
2. Abel HJ, Larson DE, Chiang C, Das I, Kanchi K, Layer RM, Neale BM, Salerno WJ, Reeves C, Buyske S, et al. (2018). Mapping and characterization of structural variation in 17,795 deeply sequenced human genomes. *bioRxiv*. <https://doi.org/10.1101/508515>
3. Biju A, Hambly K, Joshi A (2021) The complexity of forensic science in criminal investigations: is there a gold standard? W: Clayton M, Abbas N (red) Are we there yet? The golden standard of forensic science. *Voices of Forensic Science*, 1(1). University of Toronto Libraries Journal Production Services Institute for the Study of University Pedagogy (ISUP), str. 7-29
4. McKiernan HE, Danielson PB (2016) Molecular diagnostic applications in forensic science. *Molecular diagnostics – third edition*. Academic Press, Cambridge, str. 371-394
5. Marano LA, Fridman C (2019) DNA phenotyping: current application in forensic science. *Research and Reports in Forensic Medical Science* 9: 1-8
6. Pike G, Havard K, Harrison G, Ness H (2021) Eyewitness identification procedures: Do researchers and practitioners share the same goals? *International Journal of Police Science & Management* 23(1): 17-28
7. Walsh S, Kayser M (2017) Predicting human appearance from DNA for forensic investigations, W: Amorin A, Budowle B (red) *Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation*, World Scientific Publishing Europe Ltd., str. 415-449
8. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *British Medical Journal* 372:n71. doi: 10.1136/bmj.n71
9. Hopman R (2020) Opening up forensic DNA phenotyping: the logics of accuracy, commonality and valuing. *New Genetics and Society*, 1-17 doi:10.1080/14636778.2020.1755638
10. Wyner N, Barash M, McNevin D (2020) Forensic autosomal short tandem repeats and their potential association with phenotype. *Frontiers in Genetics* 11: 884 doi: 10.3389/fgene.2020.00884. PMID: 32849844; PMCID: PMC7425049.
11. Boyle EA, Li YI, Pritchard JK (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169: 1177-1186
12. Kowalczyk M, Zawadzka E, Szewczuk D, Gryzińska M, Jakubczak A (2018). Molecular markers used in forensic genetics. *Medicine, Science and the Law*, 58(4): 201-209
13. Bekaert B, Kayadanda A, Zapico SC, Van de Voorde W, Decorte R (2015) Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics*, 10(10): 922-930
14. Kayser M (2015) Forensic DNA phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics*, 18: 33-48
15. Phillips C (2015) Forensic genetic analysis of bio-geographical ancestry. *Forensic Science International: Genetics*, 18: 49-65
16. Freire-Aradas A, Phillips C, Mosquera-Miguel A et al. (2016) Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system. *Forensic Science International: Genetics*, 24: 65-74

17. The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* 526(7571): 68
18. Pontes L, de Sousa JC, Medeiros R (2017) SNPs and STRs in forensic medicine. A strategy for kinship evaluation. *Archiwum Medycyny Sądowej i Kryminologii* 67(3): 226-240
19. Holt CL, Stephens KM, Walichiewicz P, Fleming KD, Forouzmand E, Wu SF (2021) Human mitochondrial DNA control region and mtGenome: design and forensic validation of NGS multiplexes, sequencing and analytical software. *Genes* 12: 599
20. Jobling MA, Tyler-Smith C (2017) Human Y-chromosome variation in the genome sequencing era. *Nature Review Genetics* 18(8): 485-497
21. Pereira V, Mogensen HS, Børsting C et al (2017) Evaluation of the precision ID ancestry panel for crime case work: a SNP typing assay developed for typing of 165 ancestral informative markers. *Forensic Science International: Genetics* 28: 138-145
22. Hanssen EN, Lyle R, Egeland T, Gill P (2017) Degradation in forensic trace DNA samples explored by massively parallel sequencing. *Forensic Science International: Genetics* 27: 160-166
23. do Valle Silva G, de Souza FDN, Marcorin L, Pereira ALE, Carratto TMT, Debortoli G, de Oliveira MLG, de Aguiar Fracasso NC, de Andrade ES, Donadi EA, Norton HL, Parra EJ, Simoes AL, Castelli EC, Mendes-Junior CT (2019) Applicability of the SNPforID 52-plex panel for human identification and ancestry evaluation in a Brazilian population sample by Next-Generation Sequencing. *Forensic Science International: Genetics* 40: 201-209
24. Masuyama K, Shoji H, Nakanishi H, Inokuchi S, Adachi N (2017) Sex determination from fragmented and degenerated DNA by amplified product-length polymorphism bidirectional SNP analysis of amelogenin and SRY genes. *Plos One*, 12(1), p. e0169348
25. Salih MH (2021) Molecular markers for human sex determination in forensic genetics analysis. *International Journal for Research in Applied Sciences and Biotechnology* 8(6): 25-30
26. Hares DR (2015) Selection and implementation of expanded CODIS core loci in the United States. *Forensic Science International: Genetics*, 17: 33-34
27. Elmrgbni S, Kaddura M (2019) Human identification by amelogenin test in Libyans. *American Journal of Biomedical Science & Research*, 3(6): 543-549
28. Dash HR, Neha R, Das S (2020) Alternatives to amelogenin markers for sex determination in humans and their forensic relevance. *Molecular Biology Reports* 47: 2347-2360
29. Gopinath S, Zhong C, Nguyen V, Ge J, Lagacé RE, et al (2016) Developmental validation of the Yfiler® Plus PCR Amplification Kit: an enhanced YSTR multiplex for casework and database applications. *Forensic Science International: Genetics* 24: 164-175
30. Calafell F, Larmuseau MHD (2017) The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. *Human Genetics* 136(5): 559-573
31. Roewer L (2019) Y-chromosome short tandem repeats in forensics – sexing, profiling, and matching male DNA. *WIREs Forensic Science* 1: e1336
32. Tallman SD, Kincer CD, (2021) Centering transgender individuals in forensic anthropology and expanding binary sex estimation in casework and research. *Forensic Anthropology* 5(2): 161-180
33. Hand JL, Runke CK, Hodge JC (2015) The phenotype spectrum of X-linked ichthyosis identified by chromosomal microarray. *Journal of the American Academy of Dermatology* 72(4): 617-627
34. Adserias-Garriga J, Thomas C, Ubelaker DH et al (2018) When forensic odontology met biochemistry: multidisciplinary approach in forensic human identification. *Archive of Oral Biology* 87: 7-14
35. Vidaki A, Kayser M (2018) Recent progress, methods and perspectives in forensic epigenetics. *Forensic Science International: Genetics* 37: 180-195
36. Fu-Hui X, Hao-Tian W, Qing-Peng K (2019) Dynamic DNA methylation during aging: a “prophet” of age-related outcomes. *Frontiers in Genetics* 10: 107. doi: 10.3389/fgene.2019.00107
37. Jung S, Shin K, Lee H (2017) DNA methylation-based age prediction from various tissues and body fluids. *BMB Reports* 50(11): 546-553
38. Park JL, Kim JH, Seo E, Bae DH, Kim SY, Lee HC et al (2016) Identification and evaluation of age-correlated DNA methylation markers for forensic use. *Forensic Science International: Genetics* 23: 64-70
39. Soares Bispo Santos Silva D, Antunes J, Balamurugan K, Duncan G, Sampaio Alho C, McCord B (2015) Evaluation of DNA methylation markers and their potential to predict human aging. *Electrophoresis* 36(15): 1775-1780
40. Bekaert B, Kamalandua A, Zapico SC, Van de Voorde W, Decorte R (2015) Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics* 10(10): 922-930
41. Hong SR, Jung SE, Lee EH, Shin KJ, Yang WI, Lee HY (2017) DNA methylation-based age prediction from saliva: high age predictability by combination of 7 CpG markers. *Forensic Science International: Genetics* 29: 118-125
42. Alghanim H, Antunes J, Silva DSBS, Alho CS, Balamurugan K, McCord B (2017) Detection and evaluation of DNA methylation markers found at SCGN and KLF14 loci to estimate human age. *Forensic Science International: Genetics* 31: 81-88. doi: 10.1016/j.fsigen.2017.07.011. Epub 2017 Aug 7. PMID: 28854399.
43. Lee JW, Choung CM, Jung JY, Lee HY, Lim SK (2018) A validation study of DNA methylation-based age prediction using semen in forensic casework samples. *Legal Medicine (Tokyo)* 31: 74-77
44. Giuliani C, Cilli E, Bacalini MG, Pirazzini C, Sazzini M, Gruppioni G et al (2016) Inferring chronological age from DNA methylation patterns of human teeth. *American Journal of Physical Anthropology* 159(4): 585-595
45. Correia Dias H, Cunha E, Corte Real F, Manco L (2020) Age prediction in living: Forensic epigenetic age estimation based on blood samples. *Legal Medicine (Tokyo)* 47: 101763. doi: 10.1016/j.legalmed.2020.101763
46. Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, et al. (2018) Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany N Y)* 10(7): 1758-1775
47. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, et al. (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell* 49(2): 359-367
48. Zbieć-Piekarska R, Spólnicka M, Kupiec T, Parys-Proszek A, Makowska Z, Paleczka A, et al. (2015) Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Science International: Genetics* 7: 173-179
49. Montesanto A, D'Aquila P, Lagani V, Papparazzo E, Geracitano S, Formentini L, Giacomoni R, Cardelli M, Provinciali M, Bellizzi D, Passarino G (2020) A new robust epigenetic model for forensic age prediction. *Journal of Forensic Sciences* 65: 1424-1431
50. Markiewicz E, Idowu OC (2020) Melanogenic difference consideration in ethnic skin type: a balance approach between skin brightening applications and beneficial sun exposure. *Clinical, Cosmetic and Investigational Dermatology* 13: 215-232
51. William J. Pavan WJ, Sturm RA (2019) The genetics of human skin and hair pigmentation. *Annual Review of Genomics and Human Genetics* 20(1): 41-72
52. Del Bino S, Ito S, Sok J, Nakanishi Y, Bastien P, Wakamatsu K, Bernerd F (2015) Chemical analysis of constitutive pigmentation of human epidermis reveals constant eumelanin to pheomelanin ratio. *Pigment Cell & Melanoma Research* 28(6): 707-717
53. Queirós F (2019) The visibilities and invisibilities of race entangled with forensic DNA phenotyping technology. *Journal of Forensic and Legal Medicine* 68(101858): 1-7
54. Sarkar A, Nandineni MR (2018) Association of common genetic variants with human skin color variation in Indian populations. *American Journal of Human Biology* 30: e23068
55. Andrade ES, Fracasso NCA, Strazza Júnior PS, Simões AL, Mendes-Junior CT (2017) Associations of OCA2 - HERC2 SNPs and haplotypes with human pigmentation characteristics in the Brazilian population. *Legal Medicine* 24: 78-83

56. Maroñas O, Söchtig J, Ruiz Y, Phillips C, Carracedo Á, Lareu MV (2015). The genetics of skin, hair, and eye color variation and its relevance to forensic pigmentation predictive tests. *Forensic Science Review* 27(1): 13-40
57. Bulbul O, Zorlu T, Filoglu G (2020) Prediction of human eye colour using highly informative phenotype SNPs (PISNPs). *Australian Journal of Forensic Sciences* 52: 27-37
58. Andersen JD, Pietroni C, Johansen P, Andersen MM, Pereira V, Børsting C, *et al.* (2016) Importance of nonsynonymous OCA 2 variants in human eye color prediction. *Molecular Genetics & Genomic Medicine* 4: 420-430
59. Söchtig J, Phillips C, Maroñas O, Gómez-Tato A, Cruz R, Alvarez-Dios J, *et al.* (2015) Exploration of SNP variants affecting hair colour prediction in Europeans. *International Journal of Legal Medicine* 129: 963-975
60. Walsh S, Lindenbergh A, Zuniga SB, Sijen T, de Knijff P, Kayser M, *et al.* (2011) Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics* 5: 464-471
61. Chaitanya, K. Breslin, S. Zuñiga, L. Wirken, E. Pospiech, M. Kukla-Bartoszek, *et al.* (2018) The HirisPlex-S system for eye, hair and skin colour prediction from DNA: introduction and forensic developmental validation, *Forensic Science International: Genetics* 35: 123-135
62. Pośpiech E, Karłowska-Pik J, Ziemkiewicz B, Kukla M, Skowron M, Wojas-Pelc A, Branicki W (2016) Further evidence for population specific differences in the effect of DNA markers and gender on eye colour prediction in forensics *International Journal of Legal Medicine* 130(4): 923-934
63. Palencia-Madrid L, Xavier C, de la Puente M, Hohoff C, Phillips C, Kayser M, Parson W (2020) Evaluation of the VISAGE basic tool for appearance and ancestry prediction using ForenSeq[®]chemistry on the MiSeq FGx[®] system. *Genes (Basel)* 11(6): 708. doi: 10.3390/genes11060708
64. Ho YYW, Mina-Vargas A, Zhu G, Brims M, McNevin D, Montgomery GW, *et al.* (2020) Comparison of Genome-Wide Association Scans for quantitative and observational measures of human hair curvature. *Twin Research and Human Genetics* 23: 271-277
65. Pośpiech E, Chen Y, Kukla-Bartoszek M, Breslin K, Aliferi A, Andersen JD, *et al.* (2018) Towards broadening Forensic DNA Phenotyping beyond pigmentation: Improving the prediction of head hair shape from DNA. *Forensic Science International: Genetics* 37: 241-251
66. Marcińska M, Pośpiech E, Abidi S, Andersen JD, van den Berge M, Carracedo Á, *et al.* (2015) Evaluation of DNA variants associated with androgenetic alopecia and their potential to predict male pattern baldness. *PLoS ONE* 10: e0127852.
67. Hagenaaers SP, Hill WD, Harris SE, Ritchie SJ, Davies G, Liewald DC, *et al.* (2017) Genetic prediction of male pattern baldness. *PLoS Genetics* 13: e1006594
68. Guo MH, Hirschhorn JN, Dauber A (2018) Insights and implications of genome wide association studies of height. *Journal of Clinical and Endocrinological Metabolism* 103(9): 3155-3168
69. Marouli E, Graff M, Medina-Gomez C, *et al.*; EPIC-InterAct Consortium; CHD Exome+ Consortium; ExomeBP Consortium; T2D-Genes Consortium; GoT2D Genes Consortium; Global Lipids Genetics Consortium; ReproGen Consortium; MAGIC Investigators (2017) Rare and low-frequency coding variants alter human adult height. *Nature* 542(7640): 186-190
70. Rotwein P (2017) Revisiting the Population Genetics of Human Height, *Journal of the Endocrine Society*, 4(4): bvaa025
71. Jing X, Sun Y, Zhao W, Gao X, Ma M, Liu F, Li C (2019) Predicting adult height from DNA variants in a European-Asian admixed population. *Int J Legal Med.* 2019 Nov;133(6):1667-1679. doi: 10.1007/s00414-019-02039-8. Epub 2019 Apr 12. Retraction in: *Int J Legal Med.* 2021 Nov;135(6):2151. PMID: 30976986.
72. Jelenkovic, A., Sund, R., Hur, Y.M. *et al.* (2016) Genetic and environmental influences on height from infancy to early adulthood: An individual-based pooled analysis of 45 twin cohorts. *Scientific Reports* 6, 28496. <https://doi.org/10.1038/srep28496>
73. Liu F, Zhong K, Jing X, Uitterlinden AG, Hendriks AEJ, Drop SLS, *et al.* (2019) Update on the predictability of tall stature from DNA markers in Europeans. *Forensic Science International: Genetics* 42: 8-13
74. Mohamed Amer SA, Allah AlHothali BR, Alotaibi MH, Tubaigy SM (2018) Forensic Evidence for cytochrome b gene SNPs in obese and non obese Saudi Arabians. *American Journal of Biochemistry and Biotechnology* 14: 61-66
75. Cho S, Lee EH, Kim H, Lee JM, So MH, Ahn JJ, Lee HY (2021) Validation of BMI genetic risk score and DNA methylation in a Korean population. *International Journal of Legal Medicine* 35(4): 1201-1212
76. Li Y, Zhao W, Li D, Tao X, Xiong Z, Liu J, *et al.* (2019) EDAR, LYPLAL1, PRDM16, PAX3, DKK1, TNFSF12, CACNA2D3, and SUPT3H gene variants influence facial morphology in an Eurasian population. *Human Genetics* 138: 681-689
77. Shaffer JR, Orlova E, Lee MK, Leslie EJ, Raffensperger Z, Heike CL, *et al.* (2016) Genome-Wide Association Study reveals multiple loci influencing normal human facial morphology. *PLoS Genetics* 12: e1006149
78. Li L, Wang Q, Wu S, Li Z, Jiang Y, Luo X, *et al.* (2019) What makes your "eyes" look different? *Forensic Science International: Genetics Supplement Series* 7: 105-106
79. Mbadiwe S, Dawson J, Adjero D (2019) FaceSNPs: Identifying face-related SNPs from the human genome. W: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (s. 279-282). San Diego, CA, USA
80. Liu M, Li Y, Yang YF, Yan YW, Liu F, Li CX, Zeng FM, Zhao WT (2020) Human facial shape related SNP analysis in Han Chinese populations. *Yi Chuan* 42(7): 680-690
81. White JD, Indencleef K, Naqvi S, Eller RJ, Hoskens H, Roosenboom J, *et al.* (2021) Insights into the genetic architecture of the human face. *Nature Genetics* 53: 45-53
82. Zaorska K, Zawierucha P, Nowicki M (2019) Prediction of skin color, tanning and freckling from DNA in Polish population: linear regression, random forest and neural network approaches. *Human Genetics* 38: 635-647
83. Kukla-Bartoszek M, Pośpiech E, Woźniak A, Boroń M, Karłowska-Pik J, Teisseyre P, Zubańska M, Bronikowska A, Grzybowski T, Płoski R, Spólnicka M, Branicki W (2019) DNA-based predictive models for the presence of freckles. *Forensic Science International: Genetics.* 42: 252-259
84. Hernando B, Ibañez MV, Deserio-Cuesta JA, Soria-Navarro R, Vilar-Sastre I, Martinez-Cadenas C (2018) Genetic determinants of freckle occurrence in the Spanish population: Towards ephelides prediction from human DNA samples. *Forensic Science International: Genetics* 33: 38-47
85. Tang SM, Rong SS, Young AL, Tam PO, Pang CP, Chen LJ (2014) PAX6 gene associated with high myopia: a meta-analysis. *Optometry and Vision Science* 91: 419-429
86. Xie M, Song F, Li J, Ma H, Wu J, Hou Y (2017) Characteristics of SNPs related with high myopia traits in Chinese Han population. *Forensic Science International: Genetics Supplement Series* 6: e35-e36
87. Schneider PM, Prainsack B, Kayser M (2019) The use of forensic DNA phenotyping in predicting appearance and biogeographic ancestry. *Deutsches Ärzteblatt International* 51-52(51-52): 873-880
88. Samuel G, Prainsack B (2019) Forensic DNA phenotyping in Europe: views "on the ground" from those who have a professional stake in the technology, *New Genetics and Society*, 38(2): 119-141
89. Tomaszewski T (2019) Kryminalistyczne fenotypowanie DNA - wybrane problemy prawne. *Problemy kryminalistyki* 303(1): 5-12
90. Granja R, Machado H (2020) Forensic DNA phenotyping and its politics of legitimization and contestation: Views of forensic geneticists in Europe. *Social Studies of Science* 30: 306312720945033
91. Butler JM, Willis S (2020) Interpol review of forensic biology and forensic DNA typing 2016-2019. *Forensic Science International: Synergy* 2: 352-367
92. Rębała K, Branicki W, Pawłowski R, *et al.* (2020) Recommendations of the Polish Speaking Working Group of the International Society for Forensic Genetics on forensic Y chromosome typing. *Archiwum Me-*

dycyny Sądowej i Kryminologii/Archives of Forensic Medicine and Criminology 70(1): 1-18

93. Wienroth M (2018) Governing anticipatory technology practices. Forensic DNA phenotyping and the forensic genetics community in Europe. *New Genetics and Society* 37 (2): 137-152
94. Salzman AP, Arora N, Russo G, Kreutzer S, Snipen L, Haas C (2021) Assessing time dependent changes in microbial composition of biological crime scene traces using microbial RNA markers. *Forensic Science International: Genetics* 53:102537. doi: 10.1016/j.fsigen.2021.102537
95. Pamela T, D'Angiolella G, Brun P, Castagliuolo I, Gino S, Caenazzo L (2020) Skin microbiome analysis for forensic human identification: what do we know so far? *Microorganisms* 8(6): 873
96. Thong HY, Jee SH, Sun CC, Boissy RE (2003) The patterns of melanosome distribution in keratinocytes of human skin as one determining factor of skin colour. *British Journal of Dermatology* 149: 498-505
97. Alaluf S, Atkins D, Barret N, Blount M, Carter N, Heath A (2002) Ethnic variation in melanin content and composition in photoexposed and photoprotected human skin. *Pigment Cell Research* 15: 112-118
98. Uyen LDP, Wquyen DH, Kim E (2008) Mechanism of skin pigmentation. *Biotechnology and Bioprocess Engineering* 13: 383-395

ŹRÓDŁA ELEKTRONICZNE:

CODIS pobrano z <https://www.fbi.gov/services/laboratory/biometric-analysis/codis#CODIS-and%20NDIS%20Fact%20Sheet> 20.05.2022

Nazewnictwo genów pobrano z HGNC - HUGO Gene Nomenclature Committee. The resource for approved human gene nomenclature <https://www.genenames.org/oraz> Online Mendelian Inheritance in Man An Online Catalog of Human Genes and Genetic Disorders database: <https://www.omim.org> 20.05.2022

Nazewnictwo STR pobrano z <https://strbase.nist.gov> 20.05.2022

DNA-based prediction of human phenotypic traits for the medico-legal and forensic purposes

Maria Kaczmarek^{1✉}, Wojciech Zysek BSc²

¹Institute of Human Biology and Evolution, Faculty of Biology, Adam Mickiewicz University

²Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University

✉corresponding author: maria.kaczmarek@amu.edu.pl

Key words: forensic genetics, DNA, polymorphism, externally visible traits

ABSTRACT

This study aims to present the current state of knowledge on the DNA-based prediction of human externally visible characteristics of an unknown person based on the crime scene biological material left behind. This DNA sample is referred to as a “biological witness” and the procedure itself is called forensic DNA phenotyping (FDP). The analytic part of this work is based on scholarly articles published between 2015 and 2021. The electronic search of relevant references was conducted according to the PRISMA methodology in March 2021 at EBSCO Discovery Service (EDS) at the Adam Mickiewicz University library and Google Scholar. The molecular basis of FDP, DNA markers used to predict sex, age, biogeographic origin and externally visible traits such as pigmentation (skin, eye and hair colour), hair morphology, facial morphology, presence of freckles, body height, body weight (obesity), male pattern baldness and myopia were described. Furthermore, methodological difficulties resulting from the polygenic inheritance of the studied traits, as well as social and ethical problems accompanying forensic DNA phenotyping were discussed. Finally, key themes for future research related to forensic DNA phenotyping were outlined.

