

CEZARY ŻEKANOWSKI^{A, 1, 2}, BEATA BORZEMSKA^{B, 1, 2}, EWA BARTNIK^{C, 3}

^A [HTTPS://ORCID.ORG/0000-0001-7332-8494](https://orcid.org/0000-0001-7332-8494)

^B [HTTPS://ORCID.ORG/0000-0002-6987-1993](https://orcid.org/0000-0002-6987-1993)

e-mail: c.zekanowski@imdik.pan.pl

^C [HTTPS://ORCID.ORG/0000-0003-0489-7102](https://orcid.org/0000-0003-0489-7102)

1

Zakład Neurogenetyki i Genomiki Funkcjonalnej,
Instytut Medycyny Doświadczalnej i Klinicznej im. M. Mossakowskiego PAN
Department of Neurogenetics and Functional Genomics,
Mossakowski Medical Research Institute, Polish Academy of Sciences
A. Pawińskiego 5, 02-106 Warszawa, Poland

2

Wydział Wychowania Fizycznego, Akademia Wychowania Fizycznego i Sportu im. Jędrzeja Śniadeckiego w Gdańsku
Faculty of Physical Education, Gdansk University of Physical Education and Sport
K. Górskiego 1, 80-336 Gdańsk, Poland

3

Instytut Genetyki i Biotechnologii, Wydział Biologii, Uniwersytet Warszawski
Institute of Genetics and Biotechnology, Faculty of Biology, University of Warsaw
Pawińskiego 5a, 02-106 Warszawa, Poland

Testy genetyczne wykrywające zdolności sportowe: nadzieje i problemy. Część 1

Genetic tests for athletic ability: hopes and problems. Part 1

https://doi.org/10.36921/kos.2023_2964

Abstrakt

Identyfikacja podłoża genomowego prawidłowych cech organizmu, a do takich należy osiąganie wybitnych wyników sportowych, nie jest zadaniem łatwym. W trzeciej dekadzie XXI wieku jest nawet trudniejszym, niż na początku bieżącego stulecia, ponieważ dużo więcej wiadomo o złożonej architekturze genetycznej cech poligenowych i współdziałaniu wielu wariantów genu w tworzeniu jednego z poziomów organizacji układów żywych. Cechy sprzyjające uprawianiu sportu powstają jako wynik współdziałania genu oraz ponadgenomowych poziomów omicznych: epigenomu, transkryptomu, proteomu i metabolomu oraz świadomej aktywności poszczególnych osób i społeczeństw.

Słowa kluczowe: testy genetyczne, sportowcy, wariant genetyczny, oligogenowość, epigenetyka

Abstract

Identifying the genomic background of normal organismal traits, and one such trait is the achievement of athletic excellence, is not an easy task. In the third decade of the 21st century, it is even more difficult than it was at the beginning of this century, because much more is known about the complex genetic architecture of polygenic traits, about the participation of multiple genomic variants in the formation of one of the levels of organization of living systems. Moreover, athletic traits are formed as a result of the interaction between the genome and supragenomic omic levels: the epigenome, transcriptome, proteome and metabolome, and the conscious activity of individuals and societies.

Keywords: genetic tests, athletes, genetic variant, oligogenicity, epigenetics

FENOTYP TWORZĄ CZYNNIKI GENETYCZNE I ŚRODOWISKOWE

Fizjologia i psychika człowieka są najczęściej rozpatrywane jako wynik współdziałania stałych czynników dziedzicznych („natura”) oraz różnorodnych wpływów środowiska¹. Określenie struktury podwójnej helisy DNA w 1953 r., materialnego podłoża dziedziczności genetycznej, a w cztery dekady później zastosowanie różnych wariantów reakcji łańcuchowej polimerazy (PCR) do powielania i sekwencjonowania genów, znacząco powiększyło wiedzę o pierwszej składowej organizmu ludzkiego: „naturze”. Wydawało się, że łatwo będzie powiązać zmienności konkretnych genów z konkretnymi cechami organizmu, na przykład mutacji sprawczych z fenotypem klinicznym chorób dziedzicznych.

Początkowo dość powszechnie przyjmowano, że ludzki genom liczy 100 000–150 000 genów, czyli dziesięć razy więcej niż u muszki owocowej. Pojawiały się szacunki jeszcze wyższe, np. 350 000 genów². Wielu biologom wydawało się oczywiste, że liczba genów koreluje ze stopniem złożoności organizmu, a zwłaszcza cech psychicznych. W 2000 r. genetycy zebrani na spotkaniu w Cold Spring Harbour Laboratory ogłosili konkurs (z datą rozstrzygnięcia 2003 r.) na podanie najbliższej rzeczywistości liczby genów człowieka³. W międzyczasie, w 2001 r., do publicznej wiadomości podano pierwsze wyniki sekwencjonowania genomu człowieka, w ramach realizacji projektu poznania ludzkiego genomu (HGP, ang. Human Genome Project), co umożliwiło oszacowanie liczby genów kodujących produkty białko-

we na nie więcej niż 30 000 (w następnych latach szacunki zmniejszono do ok. 19 000–22 000)⁴.

Prosty determinizm, łączący konkretny gen z konkretną cechą, okazał się niemożliwy, a ludzki genom okazał się niczym nie wyróżniać na tle innych zwierzęcych genomów. Było to odkrycie na miarę kopernikańskiego i darwinowskiego, wskazujące, że unikatowość gatunku, który jako jedyny zsekwencjonował własny genom, nie jest „zapisana w genach”, lecz wynika z jedynego w historii Ziemi splotu ewolucji biologicznej oraz ewolucji społecznej, kulturowej i technologicznej (których jesteśmy zarówno twórcą, jak produktem) (Corning 2023; Zwart 2015). Jak się wydaje wielopoziomowy, adaptacyjny system homeostatyczny, obejmujący wymiar społeczny, ekonomiczny, kulturowy, polityczny i prawny umożliwia emergentne powstawanie cech jednostek, populacji i społeczności wykraczających poza redukcjonistyczne uwarunkowania biologiczne (Ellis 2023).

HGP zaowocował listą setek milionów wariantów genetycznych, polegających najczęściej na występowaniu w populacji dwu form (alleli, wariantów genowych) w konkretnym miejscu sekwencji nukleotydowej DNA. Część z wariantów jest bardzo rzadka (niekiedy prywatna – ograniczona do jednej rodziny), inne występują w populacji z częstością kilku, kilkunastu lub kilkudziesięciu procent (warianty powszechnie występujące, polimorfizmy).

Prowadzone od końca lat osiemdziesiątych XX w. genetyczne badania asocjacyjne miały na celu wiązanie konkretnych wariantów genów (wybranych najczęściej na podstawie poprzedniej wiedzy

1 A dokładniej genów, organizmu i środowiska. Por.: Richard C. Lewontin – The triple helix: gene, organism, and environment. Harvard University Press 2000; Raymond Noble, Denis Noble – Understanding living systems, Cambridge University Press 2023.

2 Liczbę około 300 000 podaje np. Stephen Jay Gould – Dzieje życia na Ziemi (1993, polskie wydanie 1998).

3 <https://archive.nytimes.com/www.nytimes.com/library/national/science/052300sci-human-genome.html>

4 Bardzo bliskie rzeczywistości oszacowanie liczby genów człowieka (ok. 30 000) pojawiło się w opublikowanej w 1977 roku książce Johna Ecclesa i Karla R. Poppera „The self and its brain” (w rozdziale „Dialog XI” z 29.09.1974). Eccles powołał się na informację ustną od Theodosiusa Dobzhansky’ego. Nieprzypadkowo chyba w rozdziale P1 wspomnianej książki omawiana jest zasada przyczynowości zstępującej (*downward causation*), a więc i udziału ponadgenomowych poziomów organizacji układów biologicznych w tworzeniu złożoności organizmu.

o funkcjach pełnionych przez ich produkty) z konkretnymi cechami organizmu człowieka. Okazało się jednak, że związek cechy z badanym wariantem był najczęściej natury statystycznej, bardzo rzadko przyczynowej. Dopiero dzięki postępowi technologicznemu i zastosowaniu metod wysokoprzepustowego sekwencjonowania nowej generacji (NGS, ang. next generation sequencing), początek XXI w. przyniósł bardziej szczegółową wiedzę o zmienności genów, ale co ważniejsze – rozpoznano zjawiska wzajemnego oddziaływania genów i opisano sieci zależności pomiędzy nimi (a dokładniej między kodowanymi przez nie produktami: białkami lub funkcjonalnymi cząsteczkami RNA). W miarę prowadzenia badań oczywistym stawało się, że każda cecha może być powiązana u danej osoby z wielką liczbą wariantów genomowych, zarówno często występujących w populacji (badanych dotychczas), jak i rzadkich, których poznanie umożliwiły technologie NGS.

Genetyka molekularna początków XXI w. potwierdziła jedno z wielkich osiągnięć biologii początków XX w.: odkrycie, że wiele różnych procesów biologicznych może prowadzić do tego samego fenotypowego efektu, a więc np. identyczny wzrost dwóch osób nie wynika z posiadania identycznych genotypów.

Szczegółowe wyniki uzyskane z sekwencjonowania całych eksomów i genomów (WES/WGS, ang. whole exome/genome sequencing) pacjentów z chorobami dziedzicznymi podważyły klasyczną definicję przyczynowości genetycznej oraz koncepcję zaburzeń ściśle monogenowych (Katsanis 2016; Noble i Noble 2017). Przypuszcza się, że dziedziczenie oligogenowe (poligenowe), gdzie mutacji sprawczej towarzyszy wiele wariantów modyfikujących fenotyp lub współsprawczych, może odpowiadać za znaczną część, zwłaszcza nietypowych, przypadków chorób dziedzicznych. Wielość wariantów tworzących podłoże genetyczne może również wyjaśniać zmienność między- i wewnątrzrodzinną pacjentów z tą samą główną mutacją przyczynową. Jednocześnie niektóre mutacje uznane za przyczynowe, przed pojawieniem się metod NGS, mogą wymagać współistnienia dodatkowych wariantów w innych genach, aby spowodować patologię.

Koncepcja dziedziczenia oligogenowego zmienia nie tylko metodologię identyfikowania mutacji genów sprawczych, ale także wpływa na diagnostykę genetyczną i poradnictwo genetyczne (Niemiec i współaut. 2018). Dlatego też rozróżnienie między możliwymi wariantami współprzyczynowymi lub modyfikującymi, a tysiącami nieistotnych wariantów identyfikowanych w WES/WGS stało się jednym z największych wyzwań stojących przed

genetyką medyczną. Wymaga to zastosowania zaawansowanych metod bioinformatycznych oraz prowadzenia badań podstawowych, identyfikujących funkcjonalne efekty wariantów genowych (a dokładniej: produktów genów niosących wspomniane warianty) na wielu poziomach organizacji układów biologicznych jednocześnie (multiomika). Nie jest to zadanie trywialne.

Konsekwencje wspomnianej probabilistycznej (w odróżnieniu od przyczynowej lub współprzyczynowej) zależności pomiędzy wariantami (genotypami) a cechami organizmu (np. ryzykiem zachorowania), zostały dostrzeżone przez niektórych genetyków bardzo szybko. Już w roku 2000 wybitny antropolog, ewolucjonista i genetyk Kenneth Monrad Weiss, w recenzji nowego wydania „Mendelian Inheritance in Man” Victora McKusicka, przywołał fragment mówiący o chorobach złożonych, powodowanych nie mutacjami przyczynowymi, ale „zwiększoną częstością specyficznych alleli (wariantów) u osób chorych” (Weiss 2000). Profesor Weiss zauważył, że „takie wymęczone wyjaśnienia oznaczają, iż mamy problem teoretyczny” – m.in. przez brak jasnego związku przyczynowego między wariantami genetycznymi, a cechą (np. chorobą). Konstatacje tego typu, w odniesieniu do wariantów powszechnie występujących, wiązanych z fenotypami prawidłowymi, istotnymi np. w uprawianiu sportu, pojawiają się w literaturze przedmiotu stosunkowo od niedawna. Na przykład Eynon i współpracownicy zwracają uwagę, że obserwowane związki statystyczne pomiędzy wariantami genetycznymi a określonym fenotypem nie są w sposób konieczny wynikiem mechanizmu przyczynowego, i dopiero badania funkcjonalne mogą wyjaśnić molekularne mechanizmy łączące wspomniane warianty i fenotyp (Eynon i współaut. 2017).

ISTNIEJE GENETYCZNE PODŁOŻE ZDOLNOŚCI SPORTOWYCH

Badania zależności osiągania wybitnych wyników sportowych od zmienności konkretnych genów rozpoczęto w początkach lat dziewięćdziesiątych XX wieku. Skomplikowanie problemu okazało się jednak daleko większe niż badanych w tym samym czasie chorób, w różnym stopniu uwarunkowanych genetycznie (Lightfoot i współaut. 2019).

Przede wszystkim wyczynowa aktywność fizyczna to złożone zjawisko, warunkowane wieloma współzależnymi czynnikami (Bouchard i Hoffman 2011). Różnica między sportowym sukcesem a porażką bardzo często liczy się w setnych częściach sekundy i w niewielu centymetrach. Wpływ na zróżnicowane osiągnięcia zawodników, wynoszące nawet

0,1%, mają zarówno czynniki natury wewnętrznej jak i zewnętrznej. Wewnętrznej, czyli podłoże genetyczne, ale także inne czynniki, w części zależne od genotypu: aktywność metaboliczna, fizjologiczna, czy cechy psychologiczne. Czynniki zewnętrzne to np. system treningu, żywienia, a także obiektywne możliwości rozwoju biologicznych zdolności, a zatem cały wymiar socjologiczny, ekonomiczny i polityczny sportu, pozostające najczęściej poza tematem rozważań biologów (Sowa i Wolański 2018; Collins 2013). Nie istnieje zresztą uniwersalny sposób na wytrenowanie wybitnego sportowca, choć wiadomo, że aby odnieść sukces, koniecznych jest wiele lat pracy, zorganizowanego i konsekwentnie prowadzonego treningu. To warunek istotny, nie gwarantujący jednak sukcesu⁵.

Sytuację komplikuje fakt, że niektórzy zawodnicy prezentują wysoki poziom sportowy zanim jeszcze zaczną realizować plan treningowy. Inni z kolei reagują na trening lepiej niż zawodnicy trenujący w tych samych warunkach. Są także zawodnicy łatwiej niż inni utrzymujący wysoką formę. We wszystkich wspomnianych przypadkach można poszukiwać genetycznych przyczyn bądź uwarunkowań.

Ostatnie dwie dekady przyniosły istotny postęp w odkrywaniu genetycznych korelatów aktywności fizycznej oraz osiągania wybitnych wyników sportowych (Posthumus i współaut. 2016). Było to pochodną postępu dokonanego w genetyce medycznej, gdyż genetyka sportowa nie różni się pod względem metod i celów od genetyki badającej podłoże chorób o różnym stopniu odziedziczalności. W rzadkich przypadkach chorób dziedzicznych podłoże tworzą mutacje sprawcze, a osoba posiadająca mutację lub dwie mutacje – jest chora lub niechybnie zachoruje. W przypadku chorób występujących powszechnie (np. cukrzyca typu 2 lub nierodzinnych postaci choroby Alzheimera i Parkinsona) podłoże genetyczne zwiększa lub zmniejsza ryzyko zachorowania. W przypadku chorób środowiskowych, np. infek-

cyjnych – wpływ jest jeszcze mniejszy, modulujący stopień odporności organizmu.

Nawet w tak nietypowych przypadkach, jak posiadanie rzeczywistej mutacji sprawczej, powodującej w sposób konieczny wystąpienie konkretnego fenotypu sprzyjającego aktywności fizycznej istotnej dla uprawiania sportu, nie oznacza automatycznie osiągnięcia sportowego sukcesu. Przykładem może być Eero Mäntyranta, fiński biegacz narciarski, rekordzista świata i medalista olimpijski w latach 1960 i 1964. W 1993 r. okazało się, że jego sukcesy przynajmniej częściowo wynikały z posiadania mutacji w genie kodującym receptor erytropoetyny (*EPOR*, W439X, rs121917830). Mutacja powodowała, że organizm Mäntyrynty produkował 25–50% więcej erytrocytów w porównaniu ze średnią populacyjną. W istocie więc cierpiał na łagodną postać erytrocytozy. Wspomniana mutacja występuje w fińsko-szwedzko-lapońskiej rodzinie i dotąd zidentyfikowano ponad 50 nosicieli wśród około 200 członków rodziny, jednak tylko Mäntyranta był czynnym, wybitnym sportowcem. Nie wystarczy zatem sama mutacja czy doping genowy, powodujące wzrost hematokrytu, by zostać mistrzem olimpijskim. U Mäntyrynty nie zsekwencjonowano całego genomu (czy eksomu), więc nie znamy możliwych wariantów współwystępujących, które mogły przyczynić się do sportowego sukcesu.

Od 1993 r. znaleziono kilkanaście mutacji *EPOR*, jednak ich nosiciele nie byli sportowcami. W 1998 r. u chłopca z populacji angielskiej, cierpiącego na erytrocytozę, znaleziono „mutację Mäntyrynty”. Została ona spowodowana niezależnym zdarzeniem mutacyjnym (mutacja *de novo*). W 2007 r. u pięcioletniej dziewczynki z Hiszpanii, również z fenotypem klinicznym erytrocytozy, znaleziono mutację powodującą ten sam efekt na poziomie białka (W439X), jednak spowodowany zmianą innego nukleotydu kodonu 439. Również w tym przypadku obydwój rodzice nie byli nosicielami mutacji. Wszystkie wymienione warianty są zatem mutacjami prywatnymi, ograniczonymi do jednej rodziny bądź do pojedynczych pacjentów. Z oczywistych przyczyn mutacje te nie figurują w bazach danych podających częstości populacyjne wariantów genomowych, a oznaczanie ich testem genetycznym w populacji ogólnej nie jest racjonalne. Niestety, są firmy oferujące wykrywanie wspomnianej „mutacji Mäntyrynty” jako elementu sukcesu sportowego.

Pomijając tak anegdotyczne przypadki, cechy związane z aktywnością fizyczną i uprawianiem sportu pod względem udziału czynników genetycznych przypominają choroby występujące powszechnie (Bouchard 2012). Podobnie jak wiele innych mierzalnych cech fenotypowych, cechy związane

5 „Często zdobycie jednego czy trzech złotych medali gmatwa obraz sytuacji w danej dyscyplinie sportu. Tak też jest w lekkoatletyce. Nie może to jednak świadczyć o ogólnym poziomie. A właśnie strategia, czyli prawidłowe długofalowe działanie (często od podstaw, a na pewno powiązane z masowością), daje gwarancję rozwoju nie na 2–3 lata lub na kilka następnych imprez. Często się o tym zapomina, żyjąc dniami dzisiejszym i sukcesami dnia dzisiejszego. Ale trzeba też pamiętać, że długofalowe myślenie jest możliwe przy ciągłości ośrodków dyspozycyjnych – kierowniczych, a nie przy jednej kadencji i walce o jej przedłużenie. Jan Mulak, z wywiadu dla miesięcznika *Królowa Sportu* 11.2002 cyt. za: <http://bieganie.pl/?show=1&cat=48&id=2026>

z aktywnością fizyczną wykazują w ludzkich populacjach rozkład Gaussa, cechujący fenotypy wieloczynnikowe, będące wynikiem współdziałania czynników dziedzicznych (wielu genów) oraz czynników środowiskowych (Quijada 2016).

Prowadzone systematycznie od lat siedemdziesiątych XX w. – możliwe bez znajomości konkretnych genów – badania stopnia odziedziczalności różnych cech fizycznych pokazały na przykład, że siła i masa mięśni szkieletowych jest odziedziczalna w 21–92%, różne parametry wydolności tlenowej – w 20–66%, podobne wartości wyznaczono dla innych cech fenotypowych: np. ruchliwości stawów, koordynacji nerwowo-mięśniowej, temperamentu (Posthumus i współaut. 2016; Bouchard i współaut. 1997). Klasyczne metody określania odziedziczalności cech fenotypowych polegają na badaniu rodzin i bliźniąt. W badaniach metodami statystycznymi określa się podobieństwo pod względem konkretnej cechy pomiędzy przedstawicielami kolejnych pokoleń (rodzice-dzieci, dziadkowie-wnuki) oraz między rodzeństwem: wychowywanymi wspólnie lub osobno parami bliźniąt monozygotycznych (MZ) i dzygotycznych (DZ). MZ są pod względem genetycznym identyczne, a różnica między nimi wynika (zasadniczo) z wpływu środowiska. Badając MZ wychowywane oddzielnie można określać precyzyjniej wpływ środowiska. DZ mają, podobnie jak każde rodzeństwo, 50% wspólnych genów, zatem różnica ma podłoże genetyczne i środowiskowe.

Jednak z faktu, że cecha ma podłoże genetyczne, nie wynika, że jest ona w każdych warunkach stała i niezmienna. Określona wartość odziedziczalności (teoretycznie od 0,0 do 1,0) jest prawdziwa dla populacji i warunków środowiskowych, w których dokonano oszacowania (nie można wartości prawdziwych dla jednej populacji przenosić wprost na inne populacje) i przy zastosowaniu konkretnej metody pomiaru. Ponadto, szacując odziedziczalność wskazujemy na zmienność cechy a nie samą cechę. Zatem odziedziczalność równa np. 0,80 nie oznacza, że cecha zależy w 80% od genów a w 20% od środowiska, ale że tak rozkłada się wpływ środowiska i genów na zmienność cechy obserwowanej w danej populacji i w danym momencie. Cecha jest dziedziczna, jeśli geny wpływają na jej pojawienie się. Natomiast stopień, w jakim geny wpływają na zakres zróżnicowania natężenia cechy w populacji w odniesieniu do średniej populacyjnej, to stopień odziedziczalności. Wskaźnik odziedziczalności informuje w jakim stopniu odchylenie od wartości średniej wynika z podłoża genetycznego (Płochiński 1968).

Tak definiując odziedziczalność, zakłada się jednak szereg warunków wstępnych: (1) różne warianty

genetyczne reagują podobnie na te same czynniki środowiska; (2) wpływ środowiska na różne genotypy jest identyczny; (3) różne geny nie oddziałują na siebie. Jest oczywiste, że wiele z cech związanych z dziedziczeniem uzdolnień sportowych nie spełnia wspomnianych warunków (Beunen i Thomis 1999).

ODZIEDZICZALNOŚĆ: GENY, GENOM I ŚRODOWISKO

Jedną z lepiej poznanych od strony genetycznej cech człowieka jest wysokość ciała. Od dawna wiadomo, że odziedziczalność wysokości ciała wynosi około 80–90% (Lettre 2011). W bardzo rzadkich przypadkach mutacje pojedynczych genów znacząco wpływają na wysokość ciała, np. mutacje genu *FGFR3* powodujące achondroplazję, chorobę przejawiającą się m.in. niskim wzrostem. Podobnych genów, których mutacje powodują rzadkie schorzenia genetyczne o objawach obejmujących zaburzenia wzrostu i postawy ciała, jest ponad 100. Natomiast na prawidłową (fizjologiczną) wysokość ciała wpływa kombinacja wariantów wielu genów, każdy modulujący w nieznacznym stopniu ostateczny efekt. Dotychczas poznano ponad 700 wariantów genów, z których część koduje białka istotne np. w tworzeniu chrząstki niezbędnej do wydłużania się kości w trakcie ontogenezy i w twardnieniu tkanki chrzęstnej. Jednak funkcja większości genów związanych z wysokością ciała pozostaje nieznana. Z tego względu nie wiemy, jaki wpływ na obserwowalne cechy fizyczne (fenotyp), w tym przypadku wysokość ciała, ma zdecydowana większość wspomnianych wariantów. Analizując jednak procesy molekularne leżące u podłoża wspomnianych rzadkich chorób genetycznych, powodowanych defektami konkretnych genów, można wnikać w skomplikowane sieci zależności pomiędzy genami (tzn. produktami genów: białkami lub funkcjonalnymi RNA) wpływającymi na prawidłowy wzrost człowieka. Ponieważ wysokość ciała determinowana jest przez bardzo wiele wariantów bardzo wielu genów (jest cechą poligenową), realizowaną przez wiele szlaków biologicznych podatnych w różny sposób i w różnym stopniu na wpływy środowiska, trudno jest dokładnie przewidzieć na podstawie znajomości nawet dużej liczby oznaczonych wariantów, jak wysokie będzie dziecko w dorosłości.

Większość z około 700 wariantów genów nie wpływa na strukturę białek. Są one położone w częściach genów nie kodujących produktów białkowych, ale mogą mieć znaczenie w regulowaniu aktywności niosących je genów. Ponadto warianty te występują często w populacji jako tzw. warianty powszechnie występujące, ang. common variants (częstość wystę-

powania rzadszego allelu, ang. Minor Allele Frequency, MAF, $\geq 5\%$). Poznane dotychczas warianty wyjaśniają jedynie ok. 20% odziedziczalności wysokości ciała (Yengo i współaut. 2022). Natomiast z najnowszych badań wynika, że większy udział w kształtowaniu wysokości ciała mają warianty nieczęste ($0,5\% < \text{MAF} < 5\%$) i rzadkie ($\text{MAF} \leq 0,5\%$) położone w regionach genów kodujących produkty białkowe (Maroul i współaut. 2017). Ponadto istotny wpływ mogą wywierać warianty strukturalne, np. delecje (ubytki części sekwencji) i insercje (dodatkowe odcinki DNA), nie identyfikowane w typowych analizach asocjacyjnych w skali całego genomu (GWAS, ang. genome wide association studies), a także trudniejsze metodologicznie do badania metodami NGS w porównaniu z wariantami jednonukleotydowymi. Warianty strukturalne o wielkości >500 kb mogą stanowić – w zależności od badanej populacji – kilka i więcej procent genomu, więc ich potencjalne znaczenie jest duże (Sherman i współaut. 2019). W tej, nieoznaczonej dotąd frakcji wariantów genomowych, może kryć się pozostałe 60% odziedziczalności tej cechy fenotypowej (Génin 2020).

Pod koniec 2015 r. opublikowano raport z realizacji „Projektu 1000 genomów” (Sudmant i współaut. 2015). Celem projektu było poznanie pełnej sekwencji DNA (genomu) 2 504 osób z 26 populacji z całego świata. W badanej grupie zidentyfikowano ponad 88 000 000 różnych wariantów genetycznych, z czego warianty pojedynczego nukleotydu (SNP, single nucleotide polymorphism) stanowiły ok. 95%. Pozostałe to prawie 4 000 000 krótkich insercji lub delecji sekwencji DNA, ponad 60 000 dużych zmian struktury genomu (duże insercje, delecje lub zmiany położenia sekwencji DNA) oraz warianty długości krótkich sekwencji DNA.

Ze wspomnianych ok. 85 000 000 SNP większość, bo aż 64 000 000 (czyli ponad 75%) stanowiły warianty rzadkie, 12 000 000 to warianty nieczęste i tylko 8 000 000 (mniej niż 10%) to warianty występujące często w populacji, a więc te, które badano dotąd w kontekście sportowym.

Jeśli natomiast rozpatrujemy genom konkretnej osoby (a nie wszystkie genomy osób z konkretnej populacji), to sytuacja przedstawia się nieco inaczej. U konkretnej osoby warianty powszechnie występujące stanowiły zdecydowaną większość. Natomiast tylko od 40 000 do 200 000 wariantów (czyli 1–4% wszystkich) to warianty rzadkie.

Liczba identyfikowanych przez genetyków z całego świata wariantów sekwencji systematycznie rośnie (ponieważ koszty sekwencjonowania całego genomu maleją). Obecnie znanych jest (tzn. zapisanych w bazie dbSNP, w grudniu 2023, <https://www.ncbi.nlm.nih.gov/snp/>) ponad miliard wariantów

zidentyfikowanych u osób z różnych populacji. Zaś najnowszy (sierpień 2023) raport ALFA Project podaje częstości ponad 900 000 000 alleli SNP zidentyfikowanych u ponad 200 000 osób (https://www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/ALFA_20230706150541/). Celem projektu ALFA jest określenie częstości alleli dla ponad miliona osób z różnych populacji.

Oprócz wspomnianych 700 wariantów genowych, na wysokość ciała wpływa szereg czynników zewnętrznych, w tym socjoekonomicznych: dochody rodziców, wykształcenie, praca itp. (Perkins i współaut. 2016). Badania nad imigrantami wskazują, że prawidłowe odżywianie, opieka zdrowotna i inne czynniki cywilizacyjne mogą zdecydowanie zwiększyć wysokość ciała, także w następnych pokoleniach, a więc powodować ujawnianie potencjału, zakrytego przez niesprzyjające okoliczności. Dlatego szacowana odziedziczalność wysokości ciała może być niższa w krajach o niskim dochodzie narodowym.

Inną z przyczyn wpływającą na odziedziczalność cech fenotypowych może być kojarzenie selektywne (ang. assortative mating), czyli wybór partnera zależny od podobieństwa fenotypowego (Sunde i współaut. 2023). Jest to zjawisko powszechne w świecie zwierząt, a u ludzi dowody na istnienie „związków podobieństw” pochodzą dotąd głównie z obserwowanego podobieństwa fenotypowego między małżonkami (np. w wysokości ciała czy inteligencji). Metody molekularne umożliwiają oszacowanie zgodności profili wielu genów między małżonkami, czy poprzez modelowanie zwiększonego podobieństwa profili między krewnymi (Horwitz i współaut. 2023). Badania w skali całego genomu potwierdziły, że podobieństwo wzrostu i inteligencji między małżonkami jest spowodowane wyborem partnera. Zbieżność liczby lat edukacji jest przynajmniej częściowo powodowana podobieństwem cech uwarunkowanych genetycznie i skorelowanych z wykształceniem.

Nie uzyskano na razie podobnych, mocnych dowodów na udział kojarzenia selektywnego w tworzeniu podobieństwa między małżonkami w podatności na choroby. Można przypuszczać, że wnioski dotyczące odziedziczalności wariantów genomu – w tym cech związanych z uprawianiem sportu – mogą być w pewnym zakresie nieobiektywne, co może tłumaczyć część niewyjaśnionej, brakującej odziedziczalności tych cech (Border i współaut. 2022).

Problem „brakującej (ukrytej) odziedziczalności” (ang. missing/hidden heritability) jest jednak bardziej złożony. Pojawiają się na przykład dowody, że w jakiejś części ta frakcja odziedziczalności

może być związana z mikrobiotą jelitową, o której wiadomo, że może wpływać na cechy psychiczne człowieka. Co więcej, np. przeniesienie mikrobioty jelitowej od pacjentów z autyzmem (ASD) do pozbawionych bakterii jelitowych myszy, wywołuje u nich cechy behawioralne autyzmu i wpływa na ekspresję genów, których ortologi u ludzi wiązane są z autyzmem (Sharon i współaut. 2019).

Zmienność materiału genetycznego to zatem tylko jeden z wielu czynników wpływających na osiągnięcie wybitnych wyników w sporcie. Ważną rolę odgrywają także czynniki epigenetyczne, czyli modyfikacje DNA i białek tworzących strukturę chromosomów, dziedziczone od rodziców lub pojawiające się w trakcie życia pod wpływem czynników środowiskowych (np. treningu czy diety). Czynniki środowiskowe poprzez modyfikacje epigenetyczne wpływają na ekspresję genów, zmieniając znaczenie zapisu genetycznego. Precyzyjne określenie zależności pomiędzy genotypem (ang. genotype, G) a środowiskiem (ang. environment, E), wpływających łącznie na predyspozycje sportowe jest więc trudne (Tucker i Collins 2012). Zależności między G i E przedstawia ogólnie przyjęte równanie $P = G \times E$ („P” to fenotyp, konkretna cecha lub ogół cech organizmu, zaś „x” oznacza interakcję obu czynników składowych). O ile genom i epigenom są coraz dokładniej opisywane i poznawane, złożoność szeroko rozumianych elementów środowiskowych jest nazbyt często pomijana w badaniach genetycznych, zwłaszcza dotyczących sportu.

Większość badań nad podłożem genetycznym cech fenotypowych koncentrowała się dotąd na badaniu wpływu pojedynczego czynnika środowiskowego na konkretny przypadek (np. osobę z chorobą genetyczną) lub na grupę osób o określonym fenotypie. Środowisko obejmuje jednak wiele czynników (podobnie jak genom wiele różnych genów), powiązanych ze sobą, tworzących sieci oddziałujących wzajemnie elementów. Rezultaty działania czynników środowiskowych na organizm są przede wszystkim plejotropowe (ten sam czynnik środowiskowy wpływa na wiele cech organizmu). Wiele czynników środowiskowych, choć potencjalnie może dotyczyć każdej osoby, w istocie pojawia się w życiu konkretnej osoby rzadko lub z niewielką intensywnością, co należy uwzględnić w szacowaniu ich wpływu w populacji. Ponadto z tym samym fenotypem mogą być powiązane różne czynniki środowiskowe, jednak w sposób niekonieczny (nie wszystkie biorą udział w tworzeniu fenotypu konkretnej osoby). Czynniki środowiskowe wpływają dodatkowo wzajemnie na siebie, modyfikując (wygaszając, wzmagając) końcowy efekt. Organizm nie reaguje także na czynniki środowiskowe jednakowo w każdym momencie,

czy stadium rozwojowym. Co więcej wpływ na fenotyp zależy od stopnia nasilenia (dawki) czynnika środowiskowego. Wpływ środowiska zależy także od właściwych dla konkretnych osób cech psychologicznych i psychospołecznych. Czynniki te mogą być tak samo ważne jak leżące u podłoża cech fenotypowych wspólne wszystkim ludziom mechanizmy genetyczne i biochemiczne (Richardson 2015).

Dokładne określenie środowiska człowieka jest jednak zadaniem trudnym. Ludzie żyją w skomplikowanym otoczeniu, którego struktura i skład wynikają z połączenia wpływów cywilizacji, kultury, ekonomii i historii. Ludzie żyją ponadto w różnych środowiskach geograficznych, do których są w różnym stopniu zaadoptowani. W ostatnich 10–15 latach, posługując się narzędziami teorii systemów, wypracowano model środowiska, definiowanego jako całość warunków zewnętrznych, obejmujących złożone czynniki fizyczne, chemiczne i biotyczne, określające formę i sposoby życia organizmów i systemów ekologicznych (enwirom, przez analogię do genomu czy proteomu) (Riggs i współaut. 2018). W badaniach genetycznych w sporcie pojęcie enwiromu nie jest jeszcze wykorzystywane.

Wspomniane równanie należy zatem rozszerzyć, rozpatrując całą złożoność środowiska (ang. polyenvironment score, pE) oddziałującego z wieloma genami (ang. polygenic score, pG) powiązanymi z pojedynczą cechą lub ich zespołem (fenotyp). Jest to ujęcie zasadniczo różne od stosowanego dotychczas, gdy złożoność środowiska zredukowana była w oparciu o przyjęte założenia (hipotezę), podobnie jak złożoność genomu ograniczana bywa do jednego lub kilkunastu genów o znanej funkcji. Tym samym można uzyskać bardziej realistyczne ujęcie kompleksowych i dynamicznych oddziaływań pomiędzy czynnikami środowiskowymi i genetycznymi (Gauderman i współaut. 2017; Padmanabhan i współaut. 2017; Ritz i współaut. 2017). Sytuację komplikuje fakt, że geny nie określają cech fenotypowych wprost, gdyż nie są czynnikami bezpośrednio sprawczymi. Oddziałują one ze środowiskiem poprzez swe produkty, tworzące dynamiczne sieci zależności, których zmiany wpływają modyfikująco na procesy ekspresji informacji zmagazynowanej w genach.

W tym miejscu należy podkreślić, że o ile genów kodujących produkty białkowe jest wg aktualnych szacunków 19 396 (na matrycy których powstaje 89 067 transkryptów), to genów kodujących różne klasy funkcjonalnych RNA (ncRNAs, niekodujące RNA) jest 27 488, a liczba transkryptów samych tylko genów kodujących długie ncRNA jest ponad 58 tysięcy (<https://www.encodegenes.org/human/stats.html>; dostęp 18.09.2023). Skala skomplikowa-

nia wspomnianej sieci zależności, stanowiącej system odbioru i przetwarzania informacji oraz modyfikowania ekspresji genów, jest daleko większa niż przypuszczano jeszcze dwie dekady temu (Hrdlickova i współaut. 2014).

Jedną z przyczyn trudności wyjaśnienia wpływu zmienności genomu na osiąganie wybitnych wyników sportowych może być nie rozpatrywana dotąd kontrola ekspresji genomu przez wyższe poziomy organizacji układów biologicznych (Noble i Noble 2017). Na przykład niedawno Bathgate i współaut. (2018) opublikowali wyniki badań nad parą bliźniąt jednojajowych: bracia przez pierwsze 20 lat życia żyli w bardzo podobnych warunkach środowiskowych, natomiast przez kolejne 30 lat życia cechowały ich całkowicie rozbieżne wzorce ćwiczeń fizycznych. Jeden z braci prowadził regularny trening wytrzymałościowy (udział w biegach na długich dystansach i w triathlonie), natomiast drugi bliźniak nie uprawiał sportu. Okazało się, że różnice między braćmi pod względem wielu wskaźników (różnych rodzajów włókien mięśniowych, wydolności aerobowej organizmu i in.) przekraczają znacząco obserwowane dotąd różnice między parami bliźniąt jednojajowych. Bathgate i współaut. podkreślają, że układ sercowo-naczyniowy i mięśnie szkieletowe wykazują daleko większą plastyczność niż wcześniej sądzono. Osoby o takim samym wyposażeniu genetycznym mogą prezentować zasadniczo różne cechy fenotypowe istotne w uprawianiu sportu, w zależności od stosowanego treningu i stylu życia (Bathgate i współaut. 2018). Wyniki te potwierdzają publikowane już wcześniej dane pokazujące znaczący wpływ treningu na parę bliźniąt jednojajowych uprawiających różne dyscypliny sportu (Keul i współaut. 1981).

Jedną z cech istotnych w uprawianiu sportu i ogólnej aktywności fizycznej jest zdolność maksymalnego pobierania tlenu przez organizm ($\text{VO}_{2\text{max}}$ czyli pułap tlenowy, wydolność aerobowa organizmu). Wartość $\text{VO}_{2\text{max}}$ wyznacza górną granicę obciążeń, przy której organizm może osiągnąć równowagę czynnościową (Górski 2008). Odziedziczalność tej istotnej cechy fizjologicznej człowieka szacuje się na 55–60% (Schutte i współaut. 2016). Mimo wielu analiz molekularnych (prostych badań asocjacyjnych, GWAS czy analizy transkryptomu) nie udało się bezspornie zidentyfikować podłoża odziedziczalności: wariantów genomu różnicujących zawodników o wysokich wartościach $\text{VO}_{2\text{max}}$ czy zdolności do zwiększenia $\text{VO}_{2\text{max}}$ w wyniku treningu (wytrenowalności) od osób nie uprawiających sportu. W 2018 r. Hans Hoppeler zaproponował, że $\text{VO}_{2\text{max}}$ jest cechą emergentną, wyłaniającą się z oddziaływania sieci w obrębie transkrypto-

mu, powstających w rezultacie ekspresji wielu genów, modyfikowanych dodatkowo przez czynniki nadrzędne w stosunku do transkryptomu (epitranskryptom), będące molekularnymi odbiornikami wpływów szeroko pojętego środowiska (Hoppeler 2018). Zrozumienie funkcjonowania cech emergentnych wymaga dodania do analiz genetycznych znajomości ponadgenomowych czynników określających funkcjonowanie transkryptomu i będących molekularnymi odbiornikami wpływów szeroko pojętego środowiska.

BADANIA MOLEKULARNE W SPORCIE

Systematyczne badania molekularne cech istotnych w uprawianiu sportu są stosunkowo młodą dziedziną wiedzy. Pierwszy wariant genowy powiązany z aktywnością wytrzymałościową w 1998 roku (Gayagay i współaut. 1998). Do początków 2023 r. z cechami istotnymi w uprawianiu sportu wiązano co najmniej 251 wariantów genetycznych, zlokalizowanych w genomie jądrowym i mitochondrialnym (Semenova i współaut. 2023). Jednak tylko 128 wariantów potwierdzono w co najmniej dwu badaniach replikacyjnych. Z aktywnością wytrzymałościową powiązano co najmniej 41 wariantów, a 87 z siłą i szybkością.

Jednym z wielu problemów związanych z identyfikowaniem wariantów znaczących czyli wpływających na fenotyp jest fakt, że na podstawie znajomości samej tylko sekwencji DNA zidentyfikowanego po raz pierwszy genu, nie można w prosty sposób przewidzieć struktury i funkcji kodowanego białka (Jisna i Jayaraj 2021). W przypadku gdy sekwencja aminokwasowa białka jest podobna (>30% podobieństwa) do białka o znanej strukturze przestrzennej (tzw. matrycy), można zastosować modelowanie homologiczne. Modelowanie homologiczne jest możliwe, ponieważ w trakcie ewolucji struktura białek jest zachowywana w większym stopniu niż stosunkowo zmienna sekwencja aminokwasowa (ta sama struktura przestrzenna może być przyjmowana przez białka o podobnej, lecz nie identycznej sekwencji aminokwasowej). Podobieństwo struktury przestrzennej świadczyć może najczęściej (choć niekoniecznie) o podobieństwie funkcjonalnym białka. Możliwe jest też modelowanie *ab initio*, w którym model przestrzenny buduje się na podstawie wiedzy o oddziaływaniach międzycząsteczkowych i obliczaniu globalnego minimum energetycznego cząsteczki modelowanego białka, w połączeniu ze znajomością ewolucyjnych zależności między sekwencją aminokwasową, a strukturą przestrzenną. Metoda ta wymaga znaczących mocy obliczeniowych i obarczona jest dużym marginesem wyników błędnych.

Jednak wraz z postępem wiedzy i rozwojem metod informatycznych wykorzystujących głębokie sieci neuronowe (np. algorytmy AlphaFold i RoseTTA-Fold), łączących cechy modelowania homologicznego i wielokrotnego dopasowania sekwencji, przewidywanie trójwymiarowej struktury białek staje się coraz bardziej dokładne. Nadal otrzymuje się jednak tylko struktury przewidywane a nie rzeczywiste, nie zawsze dokładne i użyteczne, co pokazuje porównanie struktur AlphaFold ze strukturami wyznaczonymi eksperymentalnie (David i współaut. 2022). AlphaFold bazuje na danych treningowych (ang. Protein Data Bank, PDB), budując modele analogiczne do znanych już struktur z PDB. Algorytm nie radzi sobie dobrze z wariantami zmutowanymi, które nie mają ewolucyjnie powiązanych sekwencji (Callaway 2022). Co ważniejsze, modelowanie przez analogie nie mówi nic o związkach przyczynowych (podobnie jak nic o przyczynowości nie mówi związek statystyczny pomiędzy wariantami genetycznymi a fenotypem). Otrzymując poprawną strukturę, nadal nie uzyskujemy odpowiedzi dlaczego łańcuch polipeptydowy fałduje w taki, a nie inny sposób (Lowe 2022). Jest to szerszy problem podejścia badawczego „bez hipotez”, a więc odchodzenia od teorii i modeli stworzonych przez człowieka i przyjmowania modeli oraz metod wyznaczanych przez dane (ang. data-driven, big data) i statystykę.

Warianty genów związane z sukcesem sportowym warunkują sprawność wielu układów: mięśniowego, sercowo-naczyniowego, oddechowego, wpływają też na wskaźniki antropometryczne, wydajność metaboliczną czy funkcje mitochondriów. Istotne są także cechy neurologiczne, psychiczne i psychospołeczne zawodników. Na szanse uzyskania wybitnych wyników sportowych wpływa nie tylko koordynacja nerwowo-mięśniowa czy zdolność do szybkiego odbierania i przetwarzania informacji, ale również temperament, umiejętność radzenia sobie ze stresem, pokonywania zmęczenia podczas treningu i rywalizacji, zdolność do współpracy, cechy przywódcze czy chęć poszukiwania nowych wyzwań. Cechy te i ich kombinacje wykazują duże zróżnicowanie w populacji, a ich odziedziczalność zawiera się w granicach 30–50%. Dodatkowo różne dyscypliny sportu wymagają różnych predyspozycji psychicznych (Ben-Zaken i współaut. 2019; Wilson 2019).

Dotąd poznano ok. 20 wariantów genetycznych związanych z procesami molekularnymi tworzącymi podstawę aktywności mózgu sportowca. Większość lokuje się w genach kodujących białka szlaków neurotransmisji, np. elementy systemu dopaminergicznego (receptory: DRD1, DRD2, DRD3, DRD4, DRD5.), serotonergicznego (receptory 5HT1A i 5HT2A, transporter 5HTTLPR), no-

radrenergicznego (SLC6A2), mediatory transmisji (CHRNA3, ADORA2A, EAAT, GABRG2), czynniki wzrostu neuronów (BDNF), enzymy układu syntezy/rozkładu/wychwytu neurotransmiterów (TPH1, MAOA, COMT) czy niektóre hormony. Aktywność szlaków neurotransmisji może stanowić jeden z elementów tworzących profil psychologiczny zawodnika, istotny dla zwycięskiej aktywności sportowej (Ben-Zaken i współaut. 2019).

Najszerzej zakrojone badania podłoża genetycznego cech osobowości dotyczyły cechy neurotyczności, a opublikowana w 2018 r. metaanaliza używanych dotąd wyników (przeanalizowano dane prawie pół miliona osób), doprowadziła do wskazania 136 miejsc w genomie o istotnym statystycznie wpływie na wspomniany fenotyp. W miejscach tych znajduje się 599 genów o potencjalnym związku funkcjonalnym z badaną cechą (Nagel i współaut. 2018). Jest oczywiste, że analogiczne badania dla cech związanych z osiąganiem wybitnych wyników w sporcie nie byłyby możliwe do przeprowadzenia, gdyż nie jest możliwe zgromadzenie tak dużej liczby zawodników najwyższej klasy, zwłaszcza z jednej dyscypliny i tej samej populacji (a są to warunki istotne w prowadzeniu badań genetycznych, o czym poniżej). Niemniej, badania prowadzone nawet na mniejszych grupach zawodników (kilkaset osób) mogą dostarczyć wskazówek, co do mechanizmów biologicznych leżących u podłoża cech psychologicznych istotnych w sporcie, a zwłaszcza podłoża różnic indywidualnych. Są to jednak badania podstawowe, o nikłym lub żadnym bezpośrednim znaczeniu aplikacyjnym.

STATYSTYCZNA METODOLOGIA I WYNIKI BADAŃ ASOCJACYJNYCH

Większość ze wspomnianych markerów genetycznych wyznaczono stosując proste badania asocjacyjne typu „case-control”, polegające na porównaniu metodami statystycznymi częstości występowania konkretnych polimorfizmów w grupie badanej (wybitnych sportowców) i w grupie kontrolnej (osób nigdy nie uprawiających sportu wyczynowego). Można w ten sposób analizować jedynie polimorfizmy powszechnie występujące w populacji. Częstsze występowanie konkretnego wariantu w grupie zawodników może wskazywać, że wpływa on w sposób pozytywny na tworzenie predyspozycji do uzyskiwania znaczących wyników sportowych. Efekt ten jest jednak natury statystycznej, widoczny dopiero w odpowiednio licznych grupach zawodników. Posiadanie przez konkretnego zawodnika wariantu predysponującego nie jest w żadnej mierze paszportem do sukcesu, lecz w najlepszym wypadku – obok wariant-

tów wielu innych genów i czynników środowiskowych – elementem współtworzącym podłoże fizjologiczne sprzyjające aktywności fizycznej, niezbędne dla uprawiania sportu. Wytypowanie wariantu genu w analizach asocjacyjnych może wskazywać na jego udział w sieci zależności kształtujących predyspozycje sportowe, lecz w żadnym wypadku nie powinien być rozpatrywany w kontekście sukcesu sportowego pojedynczego zawodnika.

Hipotetyczny przykład pozwoli lepiej zilustrować zagadnienie. Jeśli występują dwa warianty genu: A i B, możliwości genotypów są trzy: AA, AB i BB. W przypadku, gdy wariant rzadziej występujący (A) ma w populacji częstość 17%, a częściej występujący – 83%, to w populacji częstości genotypów (wyliczone zgodnie z prawem Hardy’ego i Weinberga (H-W), które zakłada istnienie populacji w stanie równowagi genetycznej przez kojarzenie losowe, brak działania doboru naturalnego, brak migracji i zdarzeń mutacyjnych), wynoszą: AA: 2,9%, AB: 28,2%, BB: 68,9%. Zakładając bardzo uproszczoną sytuację, gdy wariant A sprzyja uprawianiu sportu, proporcja osób predysponowanych do sukcesu sportowego w populacji ogólnej byłaby całkiem duża: 3 osoby AA i 28 osób AB na 100. Przy czym najczęściej (choć nie zawsze) wpływ genotypu AA byłby większy niż genotypu AB (efekt dawki). Odchylenia od równowagi H-W wskazują, że konkretna populacja jest niejednolita genetycznie, co może mieć istotne znaczenie w badaniach asocjacyjnych.

Polimorfizmy genetyczne analizowane w badaniach asocjacyjnych to najczęściej jednonukleotydowe warianty genowe powszechnie występujące w populacjach. Zmiana konkretnego nukleotydu w DNA ludzkim ma miejsce mniej więcej 1 raz na 300 nukleotydów, zatem teoretycznie każdy człowiek może mieć takich polimorficznych miejsc 10 milionów (DNA liczy 3×10^9 nukleotydów). Ze względu na łatwość metodologii, analiza SNP była dotąd przedmiotem zainteresowania genetyki sportowej. Pozostałe polimorfizmy polegające na większych zmianach sekwencji DNA np. delecje lub insercje były dotąd badane rzadko w kontekście genetycznego podłoża zdolności sportowych.

Wariant, który występuje częściej, jest formą wcześniejszą ewolucyjnie. Zatem warianty powszechnie występujące powstawały zasadniczo wcześniej niż warianty występujące rzadko. Tym samym warianty powszechnie występujące mogły powstać jako przystosowanie dawnych populacji ludzkich do warunków, które w typowym współczesnym środowisku już nie występują, ale mogą nadal mieć znaczenie przy uprawianiu sportu.

Jak się wydaje w trakcie ewolucji gatunku ludzkiego, w zależności od nacisków selekcyjnych, po-

wstawały w populacjach warianty bardziej sprzyjające albo szybkiej lecz krótkotrwałej aktywności biegowej, albo wymagającemu wytrwałości, powolnemu podążaniu za zwierzyną łowną. Możliwe, że w taki sposób rozprzestrzenił się 40–60 tysięcy lat temu w Eurazji wariant 577X genu kodującego α -aktyninę 3 (Houweling i współaut. 2018). Wariant ten jest bardziej rozpowszechniony u sportowców wytrzymałościowych.

Proste analizy asocjacyjne w grupie zawodników i osób nie uprawiających sportu, obarczone są ponadto pewnym zasadniczym mankamentem: wybór do badania konkretnego genu (wariantu) zależy od aktualnego stanu wiedzy. Analizuje się zatem gen, o którym w świetle zgromadzonej wiedzy można domniemywać, że ma związek funkcjonalny z procesami fizjologicznymi (lub psychicznymi) istotnymi w sporcie. Wybór może być (i jest bardzo często) błędny, a uzyskane wyniki bywają jedynie przypadkowo pozytywne. Im więcej przeprowadza się analiz, tym szansa na generowanie przypadkowych, fałszywie pozytywnych wyników wzrasta. Dysponując wynikami analiz kilkunastu pojedynczych wariantów, można niemal mieć pewność, że niektóre z nich wskazują na istnienie związku, którego w rzeczywistości nie ma. Stąd tak ważne są badania replikacyjne: badanie tego samego wariantu, lecz w innych grupach, w innych populacjach i regionach geograficznych. Wiarygodność wyników zwiększają metody statystyczne, np. różnego rodzaju poprawki na testy wielokrotne, znacznie zaostrzające kryteria istotności statystycznej wariantu.

W ostatnich dwu latach opublikowano pierwsze wyniki analiz prowadzonych z wykorzystaniem GWAS. W badaniu tym określa się związek z cechą fenotypową setek tysięcy polimorfizmów jednocześnie. Bada się wybrane, znacznikowe polimorfizmy SNP (od ok. 300 000 do ponad 2 000 000 wariantów), pokrywające równomiernie cały genom, stosując metody wielkoskalowego sekwencjonowania. Sama jednak zasada badania jest analogiczna, jak w przypadku określania związku konkretnego pojedynczego polimorfizmu: porównanie częstości występowania wariantów znacznikowych w grupie zawodników oraz kontroli. Wynik wskazuje na związek statystyczny a nie przyczynowy.

Każda analiza GWAS to w istocie setki tysięcy testów statystycznych. Standardowy stosowany próg istotności statystycznej w analizie asocjacji pojedynczego wariantu, $p=0,05$ oznacza, że w przypadku 5 testów na 100 mamy do czynienia z wynikiem fałszywie pozytywnym. Łatwo wyobrazić sobie liczbę fałszywych asocjacji w badaniu GWAS. W związku z tym za istotne uznaje się związki o wartości $p < 10^{-8}$. Z tego powodu stosując GWAS trudno wy-

kryć warianty wpływające jedynie w niewielkim stopniu na cechy fenotypowe. Giną one w szumie statystycznym. Jednocześnie warianty o wysokiej istotności statystycznej stanowią niewielką część zmienności genetycznej, nawet w przypadku cech o wysokiej odziedziczalności. W metodologii więc tkwić może tajemnica ukrytej odziedziczalności.

Dodatkowo, jak wiadomo z dotychczasowych badań chorób występujących powszechnie, wiele znaczących statystycznie wariantów zidentyfikowanych poprzez GWAS wywiera jedynie niewielki wpływ na ryzyko zachorowania ($1,1 < OR < 1,4$; ang. OR=odds ratio, iloraz szans, czyli stosunek szansy wystąpienia pewnego zdarzenia (np. genotypu) w jednej grupie do szansy jego wystąpienia w innej grupie). Porównując wartości OR z wpływem czynników środowiskowych, powodujących przykładowo dwukrotny wzrost ryzyka, warianty o wartości OR ok. 1,33 mogą być łatwo „przeoczone”, jeśli warunki środowiskowe nie są dostatecznie ściśle kontrolowane.

Pewnym rozwiązaniem jest zwiększenie mocy statystycznej badania przez powiększenie liczebności grupy badanej (zawodników) i kontrolnej do tysięcy osób. Rozwiązanie jest jednak jednocześnie problemem, gdyż może wpływać na homogenność grup: zgromadzenie dużej liczby zawodników pojedynczej dyscypliny, pochodzących z konkretnej populacji (jak wspomniano związki wariantów z fenotypem mogą być specyficzne populacyjnie), zwłaszcza w przypadku zawodników klasy światowej, jest praktycznie niewykonalne.

Rezultatem GWAS jest zasadniczo wyznaczenie regionu (czy regionów) genomu, w którym mogą znajdować się geny i warianty istotne w uprawianiu sportu. Dlatego badania takie warto prowadzić na dużych grupach, łączących zawodników ze spokrewnionych populacji. Jeśli bowiem powiększanie grup badanych o zawodników i kontrole z tej samej populacji (np. kraju) powodować będzie zawężenie wyznaczanego miejsca w genomie, wskazywało będzie na wariant wpływający silnie na fenotyp. Jeśli przeciwnie, po włączeniu do badania zawodników i kontroli z różnych populacji, utraci się lokalizację pozytywnego sygnału, będzie to oznaczało, że badana cecha warunkowana jest przez wiele wariantów, położonych w różnych miejscach genomu.

Analiza wyników GWAS, ze względu na wielką liczbę danych wymaga skomplikowanych metod obliczeniowych, zaawansowanych metodologii matematycznych, modelowania komputerowego i analiz systemowych, łączących dane z różnych poziomów organizacji komórki i organizmu. Dalsze jednak badania wymagają powrotu do rzeczywistości biologicznej i ustalenia związku fenotypu z polimorfizmami występującymi w wyznaczonym regionie

oraz analiz metodami biochemii, biologii molekularnej czy biofizyki wpływu zidentyfikowanych polimorfizmów na strukturę i funkcje kodowanych w tym rejonie białek, a zwłaszcza na regulowanie procesów ekspresji informacji genetycznej (Gallagher i Chen-Plotkin 2018).

Ze względu na wynik GWAS będący dopiero początkiem badań zaawansowanych, trudno dziś jednoznacznie ocenić, jak wiele z wytypowanych wstępnie genów i zidentyfikowanych nowych wariantów zostanie potwierdzonych przez innych badaczy oraz w innych populacjach jako powiązane statystycznie (lub lepiej: przyczynowo) z fenotypem (Brandes i współaut. 2022). Do końca 2022 roku opublikowano wyniki czterech badań grup sportowców z zastosowaniem GWAS. Jak dotąd z 41 markerów zidentyfikowanych z zastosowaniem GWAS tylko trzy zostały potwierdzone w trzech lub więcej niezależnych badaniach replikacyjnych (Ahmetov i współaut. 2016).

W ciągu ostatniej dekady badania typu GWAS, prowadzone w grupach badanych liczących nawet ponad milion osób, doprowadziły do zidentyfikowania tysięcy genów i ich wariantów wpływających na ryzyko zachorowania na złożone choroby (np. cukrzycę, schizofrenię czy chorobę Alzheimera), a także tworzących naturalne podłoże genetyczne cech złożonych (np. wzrostu ciała czy ciśnienia krwi) (Evangelou i współaut. 2018). Identyfikacja tak wielu nowych genów i wariantów może wspierać już nie poligenowy, ale omnigenowy model cech powszechnie występujących. W modelu omnigenowym proponowanym w 2017 roku, ograniczona liczba genów posiada bezpośredni związek z ryzykiem choroby (czy z rozwinięciem się fenotypu prawidłowego) (Boyle i współaut. 2017). Geny (tzn. ich produkty) uczestniczą jednak w sieciach zależności z wielką liczbą innych genów, których warianty wpływają w sposób pośredni i łagodny na badaną cechę. Ta sieciowa, rozmyta przyczynowość, skrywa być może pozostałą, brakującą (ukrytą dotąd) część odziedziczalności złożonych cech fenotypowych, takich jak wysokość ciała czy zdolności sportowe.

Nie jest jednak jasne w jaki sposób i które spośród milionów wariantów w genomie ludzkim wpływają na aktywność sportową. Nie wystarczą same analizy związków wariantów z ogólnie określonym fenotypem (dyscypliną sportu, zakresem osiągnięć), nie wystarczą analizy bioinformatyczne (określające ewentualne znaczenie wariantów genetycznych w kontekście wiedzy, zgromadzonej dotychczas w bazach danych). W większości przypadków, ze względu na koszty, nie przeprowadzono badań mających na celu wskazanie spośród tysięcy wariantów tych, które w sposób przyczynowy wpływają na powstawanie cech organizmu poprzez

konkretne mechanizmy molekularne i fizjologiczne. Powiązanie wariantów z cechami fenotypowymi wymaga zaawansowanych badań podstawowych, polegających na badaniu białek kodowanych przez geny noszące warianty, a także badania fizjologii organizmu sportowców, u których zidentyfikowano wspomniane warianty genetyczne (Young i współaut. 2019). Tego typu badania są podejmowane dopiero w ostatnich latach, a uzyskane wyniki sugerują, że szereg wariantów genetycznych wpływa na cechy organizmu poprzez zmiany ekspresji genów. Zmiany w ekspresji są najczęściej subtelne, zależne od wpływu środowiska oraz różne w poszczególnych tkankach organizmu. Jest bardzo możliwe, że to właśnie warianty wpływające na ekspresję genów są istotne w tworzeniu genetycznego podłoża zdolności sportowych. Konieczne jest także coraz lepsze rozpoznanie, opisanie i uwzględnienie szczególnie opisanych warunków środowiskowych jako zmiennej w analizach statystycznych.

RYZYKO ZACHOROWANIA, SZANSA SUKCESU

Znajomość jednego wariantu niewiele znaczy bez znajomości genetycznego kontekstu, a więc zidentyfikowania innych wariantów, rozsianych po całym genomie. W genomie każdej osoby jest takich polimorficznych miejsc średnio 10 milionów (biorąc pod uwagę tylko zmienność jednego nukleotydu). Jak zauważył prof. Timothy Caulfield: "Istnieje tak wiele elementów przyczyniających się do powstania cechy tak złożonej, jak osiągnięcie wybitnych wyników sportowych, że próba sprowadzenia ich do wyniku

pojedynczego testu genetycznego wydaje się absurdalna" (Caulfield 2011).

Użyteczność testów genetycznych identyfikujących konkretne, pojedyncze warianty genowe jest niewielka, co w odniesieniu do chorób w różnym stopniu uwarunkowanych genetycznie, przeanalizowali Neil A. Holtzman i Therese Marteau (2000). Jednym ze wskaźników użyteczności jest dodatnia wartość predykcyjna (PWP), czyli prawdopodobieństwo, że konkretny fenotyp, np. choroba, rozwinie się u osoby z wynikiem pozytywnym testu. Wartości PWP są funkcją częstości genotypu (np. wariantu genetycznego) w populacji, względnego ryzyka choroby związanego ze specyficznym genotypem oraz populacyjnego ryzyka wystąpienia choroby. W przypadku choroby o pięcioprocentowym ryzyku w populacji, wartość PWP większa niż 50% (co jak się przyjmuje jest jeszcze użyteczne klinicznie) jest tylko wtedy, gdy częstość genotypu (np. konkretnego wariantu SNP) nadającego podatność na wspomnianą chorobę wynosi 1% lub mniej oraz względne ryzyko z nim związane jest ok. dwudziestokrotne (por. Tabela 1).

Gdy częstość genotypu jest wyższa (czyli taka, jak w przypadku wariantów wiązanych dotąd z fenotypem sportowym), nie warto szukać konkretnych genotypów w populacji ogólnej, a jedynie w wybranej wstępnie grupie pacjentów wymagających dalszej diagnostyki różnicowej dla zastosowania odpowiedniego leczenia. Natomiast gdy ryzyko zachorowania jest tylko 1–2 razy ponad ryzyko populacyjne, to częstości polimorfizmów nie mają wielkiego wpływu, a jednocześnie dodatnia wartość predykcyjna nie jest wysoka (poniżej 10%, niezależnie od częstości wariantu).

Tabela 1. Zależność między dodatnią wartością predykcyjną (PWP) oraz ryzykiem atrybucyjnym (RA), a częstościami genotypu nadającego podatność na chorobę i względnym ryzykiem z nim związanym. wg Holtzman i Marteau, 2000.

częstość genotypu (%)	względne ryzyko									
	1,5		2,0		5,0		10,0		20,0	
	PWP	RA	PWP	RA	PWP	RA	PWP	RA	PWP	RA
0,1	7,5	0,05	10,0	0,1	24,9	0,4	49,6	0,9	98,1	1,9
0,5	7,5	0,25	10,0	0,5	24,5	2,0	47,8	4,3	91,3	8,7
1,0	7,5	0,5	9,9	1,0	24,0	3,9	45,9	8,3	84,0	16,0
10,0	7,1	4,8	9,1	9,0	17,9	28,6	26,3	47,4	34,5	65,5
30,0	6,5	13,0	7,7	23,1	11,4	5,6	13,5	73,0	14,9	85,1

$PWP = [R(D) \times 100] / [G(R-1) + 1]$, gdzie: R – względne ryzyko wystąpienia choroby, D – częstość populacyjna choroby (tu: 5% – czyli 0,05), G – częstość genotypu związanego z chorobą; $RA = [G(R-1) \times 100] / [G(R-1) + 1]$; gdzie: R – względne ryzyko, G – częstość genotypu związanego z chorobą

Zatem jedynie w przypadku polimorfizmów występujących rzadko (ok. 1% w populacji), zwiększających w znaczący sposób (dziesięciokrotnie) ryzyko zachorowania opłaca się owe polimorfizmy oznaczać w populacji ogólnej (wartość PPW wynosi ok. 50%). Jak na razie nie są znane warianty genetyczne, związane z fenotypem sportowym, charakteryzujące się niską częstością w populacji i jednocześnie dużym wpływem na fenotyp.

Drugim czynnikiem, który jest uwzględniany przy podejmowaniu decyzji o testowaniu genetycznym jest proporcja przypadków choroby, które mogą być przypisane konkretnemu genotypowi, a nie wpływom środowiska czy genów innych niż badany (tzw. ryzyko atrybucyjne) (Holtzman i Marteau 2000). Ryzyko atrybucyjne jest funkcją względnego ryzyka i częstości genotypu (por. Tabela 1).

Ryzyko atrybucyjne jest niskie w przypadku chorób występujących powszechnie. Inne czynniki – jak szeroko pojęte otoczenie – mogą mieć decydujące znaczenie w rozwoju choroby. Negatywny wynik testu w przypadku osoby zdrowej jest więc nieinformacyjny, jeśli chodzi o przyszły stan jej zdrowia. Tylko w przypadku genotypów występujących z częstością 10–30% w populacji i zwiększających jednocześnie ryzyko choroby 10–20 razy, wartość ryzyka atrybucyjnego jest wystarczająca z klinicznego punktu widzenia. Jeśli pacjent posiada genotypy nadające podatność na chorobę w więcej niż jednym genie, ryzyko atrybucyjne związane z określonym genem jest dużo mniejsze.

W przypadku fenotypów przyczyniających się do rozwoju prawidłowych cech fizjologicznych, np. wspomnianej wysokości ciała, skomplikowanie genetycznego podłoża jest daleko większe. Ponieważ nawet znanych genów i wariantów genowych (polimorfizmów) wiązanych w sposób statystyczny z fenotypem wybitnego sportowca jest dużo więcej niż jeden, genotyp wybitnego sportowca tworzyć może szereg różnych wariantów genowych. Możemy zatem mówić o panelu kilkudziesięciu i więcej polimorfizmów w różnych genach, w różnych miejscach konkretnych genów, nie tylko w sekwencjach kodujących produkty białkowe, ale też w sekwencjach regulatorowych wspomnianych genów oraz w genach kodujących funkcjonalne RNA i ich elementy docelowe, a więc dotyczące elementów organizmu odpowiedzialnych za czas, w którym na matrycy genu powstaje białko czy też ilość w jakiej to białko powstaje chociażby w odpowiedzi na trening. Dodatkowo różne kombinacje genów i ich wariantów mogą być powiązane z różną odpowiedzią na trening czy radzenie sobie ze stresem. Określenie tak skomplikowanych zależności wymagało będzie zakrojonych na szeroką skalę badań podstawo-

wych, uwzględniających rzeczywistość treningową zawodników („rozszerzony fenotyp”).

Jest to jedna z przyczyn niepowodzenia badań replikacyjnych, zmierzających do potwierdzenia pozytywnego wyniku badania asocjacyjnego, wykonanego z zastosowaniem grupy zawodników i kontroli, pochodzących z jednej populacji. Ta sama cecha, którą może być osiągnięcie światowych wyników w biegu na 100 m, może mieć różne podłoże genetyczne, może wynikać z korzystnej budowy kości i ścięgien lub ze sprawnej fizjologicznie struktury mięśni. Jeśli zatem w badaniu asocjacyjnym, szukając związku z markerami genów kodujących elementy strukturalne mięśni, nie rozdzielimy tych dwu podgrup sportowców, nie uzyskamy wyniku pozytywnego. Dodatkowo znaczenie może mieć pochodzenie etniczne, gdyż w różnych populacjach ta sama cecha może być warunkowana przez różne zestawy genów lub wariantów.

Podobny wniosek płynie z badań nad zwierzętami modelowymi. Ważną metodą analizy wpływu mutacji na fenotyp organizmu jest tworzenie zwierząt transgenicznych pozbawionych konkretnego genu. Okazuje się, że skutkiem usunięcia tego samego genu u różnych szczepów myszy (to znaczy posiadających różne podłoże genetyczne), może być powstanie myszy o zasadniczo różnych fenotypach (Imenez Silva i współaut. 2017). Istotny jest zatem nie tylko sam gen i jego wariant, nawet wpływający w sposób przyczynowy na fenotyp, ale „kontekst genetyczny”, czyli otoczenie innych genów i wariantów, sekwencji regulacyjnych, regionów chromosomalnych, itp. Kontekstualność informacji genetycznej wynika też ze zjawiska redundancji (nadmiarowości): istotne biologicznie geny czy procesy są dublowane całkowicie lub częściowo przez inne geny i procesy, co pozwala korygować błędy i unikać ewentualnego uszkodzenia istotnych funkcji organizmu. Analogicznie w przypadku wariantów wpływających na aktywność sportową: ich wpływ na organizm zależy zapewne od genetycznego kontekstu, który jest różny u poszczególnych osób.

Skomplikowane podłoże genetyczne może bardzo utrudnić lub uniemożliwić prostą analizę genetyczną predyspozycji do sportu. Z badań przeprowadzonych w populacji hiszpańskiej wynika, że jeden na 1351 Hiszpanów posiada optymalny profil genetyczny siedmiu genów, wiązanych na podstawie ówczesnego (rok 2009) stanu wiedzy z aktywnością sportową (Ruiz i współaut. 2009). Jednak tylko trzech spośród najlepszych hiszpańskich zawodników wytrzymałościowych posiada optymalny profil dla sześciu takich genów. Natomiast wspomniany profil posiada 1 osoba na 600 Hiszpanów z populacji ogólnej. Podobnie w populacji japoń-

skiej nie udało się znaleźć związku pomiędzy panelem 21 polimorfizmów (uznawanych na podstawie danych naukowych za powiązane z aktywnością wytrzymałościową) a statusem wybitnego sportowca dyscyplin wytrzymałościowych. Nie zaobserwowano różnic w tzw. współczynniku TGS (ang. total genetic score, mówiącym jaki procent optymalnych polimorfizmów z panelu ma konkretna osoba) pomiędzy osobami z grupy kontrolnej i grup zawodników o różnym stopniu „wybitności” (Yvert i współaut. 2016).

Ani pojedyncze polimorfizmy, ani „panel” wielu wariantów nie są w sposób przyczynowy związane z osiąganiem wybitnych wyników sportowych. W przeważającej liczbie przypadków nie wiadomo, jaki mechanizm fizjologiczny czy molekularny stoi za wspomnianym związkiem statystycznym.

Wcześniejsze analizy Williamsa i Follanda (2008) przyjmowały za punkt wyjścia, że najwybitniejsi zawodnicy wywodzą się z grupy osób posiadających „idealną” kombinację wariantów (w 2008 roku było to 23 konkretnych polimorfizmów), co oznacza że prawdopodobieństwo bycia mistrzem wynosi 0,0005% dla każdego człowieka (Williams i Folland 2008). Im większa badana populacja, tym większa szansa na znalezienie osoby z wszystkimi „zwykłymi” wariantami. Jednak, jeżeli do profilu genowego będziemy włączali kolejne warianty genów, to tym samym będzie zmniejszało się prawdopodobieństwo znalezienia osoby z profilem optymalnym. Albo też – co bardziej prawdopodobne – w populacji będzie kilka/ kilkanaście/ kilkadziesiąt różnych optymalnych profili, wynikających z kombinacji alleli różnych genów, a zatem taki sam lub podobny efekt fenotypowy (wysoka sprawność fizjologiczna organizmu, związek z osiąganiem wybitnych wyników w sporcie) będzie zależał od różnego wyposażenia genetycznego i co za tym idzie różnego podłoża fizjologicznego. Dotychczas brak jest danych wskazujących, że uda się znaleźć taki profil, złożony z wariantów powszechnie występujących (Rankinen i współaut. 2016).

Identyczne wnioski płyną z badań nad zwierzętami. Gazda i współautorzy (2018) porównując strukturę genetyczną gołębi pocztowych i należących do innych ras, doszli do wniosku, że jest mało prawdopodobne, aby jakkolwiek pojedynczy wariant genetyczny był wystarczający do osiągnięcia lepszych zdolności gołębi do lotu i orientacji. Ponadto, u podłoża genetycznego zdolności gołębi pocztowych (wyścigowych) mogą leżeć różne kombinacje alleli u różnych ptaków. Gazda i współaut. wykryli także znaczną liczbę sygnatur selekcji rozsianych po całym genomie gołębi, co wskazuje na poligenową podstawę adaptacji do osiągnięcia ekstremalnych

wyników lotu (Gazda i współaut. 2018). Jest to – jak podkreślają autorzy – wynik spodziewany, jako że duża liczba elementów fizjologicznych organizmu (układ sercowo-naczyniowy, oddechowy, układ mięśniowo-szkieletowy, odpornościowy i nerwowy) musi współdziałać w celu uzyskania wybitnych wyników: tak gołębi, jak i u sportowców (Ostrander i współaut. 2009).

Warianty powszechnie występujące, identyfikowane w badaniach asocjacyjnych (prostych lub GWAS) wyjaśniają jedynie niewielką część zmienności genetycznej. Większe znaczenie mają zapewne rzadkie warianty, warianty strukturalne (np. delecje dużych fragmentów DNA), ale także zjawiska współwystępowania genów (epistaza), czy złożoność genetyczna populacji powodująca, że wyniki badań asocjacyjnych oddają bardziej genetyczne zróżnicowanie grup badanych, niż związki z fenotypem sportowym.

Rzadkie warianty genetyczne mogą mieć potencjalnie duży wpływ na obserwowalne cechy organizmu. Badając choroby występujące powszechnie czy takie cechy jak poziom lipoprotein o niskiej gęstości (LDL), zidentyfikowano rzadkie warianty o umiarkowanym lub silnym wpływie na fenotyp. Nie można wykluczyć, że również w przypadku zawodników, badania w skali całych genomów doprowadzą do identyfikowania rzadkich wariantów sprzyjających uprawianiu sportu. Jest także możliwe, że fenotyp(y) wybitnego sportowca powiązany będzie (w sposób przyczynowy lub statystyczny) z współwystępowaniem w genomie wielu rzadkich wariantów w wielu genach, tworząc naturalne wzmocnienie (przez analogię do „variant burden”, obciążenia wariantami szkodliwymi, prowadzącymi do choroby). Od stosunkowo niedawna warianty rzadkie zaczynają interesować genetyków zajmujących się sportem, czego dowodem jest powołanie do życia w 2015r. Athlome Consortium (<http://www.athlomeconsortium.org/about/>), którego jednym z celów jest określenie sekwencji genomów reprezentatywnej grupy sportowców z całego świata (1000 zawodników do 2020 r.). Zamierzenia Konsorcjum są jednak szersze, gdyż do zrozumienia znaczenia fenotypowego identyfikowanych wariantów (czyli wyjaśnienia w jaki sposób wpływają na cechy obserwowalne organizmu) niezbędne są zakrojone na międzynarodową skalę badania podstawowe, a następnie działania integrujące dane genetyczne z zaawansowanym profilowaniem zawodników (biochemicznym, anatomicznym, fizjologicznym). Zadanie to wymaga analizy miliardów zmiennych, z wykorzystaniem zaawansowanych metodologii statystycznych i bioinformatycznych.

WNIOSKI

W identyfikacji podłoża genomowego złożonych cech organizmu, do których należy osiągnięcie wybitnych wyników sportowych, największym wyzwaniem jest nie identyfikacja zmian sekwencji DNA (gdyż postęp technologiczny jest niezwykle szybki i łatwość wskazania wariantów jest coraz większa), ale interpretacja wyników, wykazanie że konkretny wariant rzeczywiście powiązany jest z obserwowalnymi cechami organizmu. Określenie związków przyczynowych pomiędzy wariantem a fenotypem, zwłaszcza u konkretnej osoby, jest zadaniem jeszcze bardziej złożonym.

BIBLIOGRAFIA

- Ahmetov I. I., Egorova E. S., Gabdrakhmanova L. J., Fedotovskaya O. N., 2016. *Genes and Athletic Performance: An Update*. *Med Sport Sci* 61, 41–54. <https://doi.org/10.1159/000445240>
- Bathgate K. E., Bagley J. R., Jo E., Talmadge R. J., Tobias I. S. i współaut., 2018. *Muscle health and performance in monozygotic twins with 30 years of discordant exercise habits*. *Eur J Appl Physiol* 118, 2097–2110. <https://doi.org/10.1007/s00421-018-3943-7>
- Ben-Zaken S., Richard V., Tenenbaum G., 2019. *Genetics and the psychology of motor performance*. Routledge/Taylor & Francis Group, New York. <https://doi.org/10.4324/9781315114682>
- Beunen G., Thomis M., 1999. *Genetic determinants of sports participation and daily physical activity*. *Int J Obes Relat Metab Disord* 23 Suppl 3, S55–63. <https://doi.org/10.1038/sj.ijo.0800885>
- Border R., O'Rourke S., de Candia T., Goddard M. E., Visscher P. M. i współaut., 2022. *Assortative mating biases marker-based heritability estimators*. *Nat Commun* 13, 660. <https://doi.org/10.1038/s41467-022-28294-9>
- Bouchard C., 2012. *Genomic predictors of trainability*. *Exp Physiol* 97, 347–352. <https://doi.org/10.1113/expphysiol.2011.058735>
- Bouchard C., Hoffman E. P., 2011. *Genetic and molecular aspects of sport performance*. Vol. XVIII of The Encyclopedia of Sports Medicine. Wiley-Blackwell, Chichester. <https://doi.org/10.1002/9781444327335> Wiley InterScience
- Bouchard C., Malina R. M., Pérusse L., 1997. *Genetics of fitness and physical performance*. Human Kinetics, Champaign, IL.
- Boyle E. A., Li Y. I., Pritchard J. K., 2017. *An expanded view of complex traits: from polygenic to omnigenic*. *Cell* 169, 1177–1186. <https://doi.org/10.1016/j.cell.2017.05.038>
- Brandes N., Weissbrod O., Linal M., 2022. *Open problems in human trait genetics*. *Genome Biol* 23, 131. <https://doi.org/10.1186/s13059-022-02697-9>
- Callaway E., 2022. *What's next for AlphaFold and the AI protein-folding revolution*. *Nature* 604, 234–238. <https://doi.org/10.1038/d41586-022-00997-5>
- Caulfield T., 2011. *Predictive or preposterous? The marketing of DTC genetic testing*. *JCOM* 10, C02. <https://doi.org/10.22323/2.10030302>
- Collins T., 2013. *Sport in capitalist society: a short history*. Routledge.
- Corning P. A., 2023. *Culture–gene co-evolution: Darwin's other theory comes into view*. *Biological Journal of the Linnean Society* 139, 563–569. <https://doi.org/10.1093/biolinnean/blac048>
- David A., Islam S., Tankhilevich E., Sternberg M. J. E., 2022. *The AlphaFold Database of Protein Structures: A Biologist's Guide*. *J Mol Biol* 434, 167336. <https://doi.org/10.1016/j.jmb.2021.167336>
- Ellis G. F. R., 2023. *Efficient, Formal, Material, and Final Causes in Biology and Technology*. *Entropy* 25, 1301. <https://doi.org/10.3390/e25091301>
- Evangelou E., Warren H. R., Mosen-Ansorena D., Mifsud B., Pazoki R. i współaut., 2018. *Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits*. *Nat Genet* 50, 1412–1425. <https://doi.org/10.1038/s41588-018-0205-x>
- Eynon N., Voisin S., Lucia A., Wang G., Pitsiladis Y., 2017. *Preface: genomics and biology of exercise is undergoing a paradigm shift*. *BMC Genomics* 18, 825. <https://doi.org/10.1186/s12864-017-4184-6>
- Gallagher M. D., Chen-Plotkin A. S., 2018. *The Post-GWAS Era: From Association to Function*. *Am J Hum Genet* 102, 717–730. <https://doi.org/10.1016/j.ajhg.2018.04.002>
- Gauderman W. J., Mukherjee B., Aschard H., Hsu L., Lewinger J. P. i współaut., 2017. *Update on the State of the Science for Analytical Methods for Gene-Environment Interactions*. *Am J Epidemiol* 186, 762–770. <https://doi.org/10.1093/aje/kwx228>
- Gayagay G., Yu B., Hambly B., Boston T., Hahn A. i współaut., 1998. *Elite endurance athletes and the ACE I allele—the role of genes in athletic performance*. *Hum Genet* 103, 48–50. <https://doi.org/10.1007/s004390050781>
- Gazda M. A., Andrade P., Afonso S., Dilyte J., Archer J. P. i współaut., 2018. *Signatures of Selection on Standing Genetic Variation Underlie Athletic and Navigational Performance in Racing Pigeons*. *Mol Biol Evol* 35, 1176–1189. <https://doi.org/10.1093/molbev/msy030>

- Génin E., 2020. *Missing heritability of complex diseases: case solved?* Hum Genet 139, 103–113. <https://doi.org/10.1007/s00439-019-02034-4>
- Górski J., 2008. *Fizjologiczne podstawy wysiłku fizycznego*. PZWL, Warszawa
- Holtzman N.A., Marteau T.M., 2000. *Will genetics revolutionize medicine?* N Engl J Med 343, 141–144. <https://doi.org/10.1056/NEJM200007133430213>
- Hoppeler H., 2018. *Deciphering $\dot{V}O_{2,max}$: limits of the genetic approach*. J Exp Biol 221, jeb164327. <https://doi.org/10.1242/jeb.164327>
- Horwitz T.B., Balbona J.V., Paulich K.N., Keller M.C., 2023. *Evidence of correlations between human partners based on systematic reviews and meta-analyses of 22 traits and UK Biobank analysis of 133 traits*. Nat Hum Behav 7, 1568–1583. <https://doi.org/10.1038/s41562-023-01672-z>
- Houweling P.J., Papadimitriou I.D., Seto J. T., Pérez L.M., Coso J.D. i współaut., 2018. *Is evolutionary loss our gain? The role of ACTN3 p.Arg577Ter (R577X) genotype in athletic performance, ageing, and disease*. Hum Mutat 39, 1774–1787. <https://doi.org/10.1002/humu.23663>
- Hrdlickova B., de Almeida R.C., Borek Z., Withoff S., 2014. *Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease*. Biochim Biophys Acta 1842, 1910–1922. <https://doi.org/10.1016/j.bbadis.2014.03.011>
- Imenez Silva P.H., Melo D., de Mendonça P.O.R., 2017. *Insights from Systems Biology in Physiological Studies: Learning from Context*. Cell Physiol Biochem 42, 939–951. <https://doi.org/10.1159/000478648>
- Jisna V.A., Jayaraj P.B., 2021. *Protein Structure Prediction: Conventional and Deep Learning Perspectives*. Protein J 40, 522–544. <https://doi.org/10.1007/s10930-021-10003-y>
- Katsanis N., 2016. *The continuum of causality in human genetic disorders*. Genome Biol 17. <https://doi.org/10.1186/s13059-016-1107-9>
- Keul J., Dickhuth H.H., Simon G., Lehmann M., 1981. *Effect of static and dynamic exercise on heart volume, contractility, and left ventricular dimensions*. Circ Res 48, 162–170.
- Lettre G., 2011. *Recent progress in the study of the genetics of height*. Hum Genet 129, 465–472. <https://doi.org/10.1007/s00439-011-0969-x>
- Lightfoot T.J., Hubal M.J., Roth S.M., 2019. *Routledge Handbook of Sport and Exercise Systems Genetics*. Routledge, London. <https://doi.org/10.4324/9781315146287>
- Lowe D., 2022. *The Law of Conservation of Data*. Chemistry World, January 11. <https://www.chemistryworld.com/opinion/the-law-of-conservation-of-data/4014927.article>
- Marouli E., Graff M., Medina-Gomez C., Lo K.S., Wood A.R. i współaut., 2017. *Rare and low-frequency coding variants alter human adult height*. Nature 542, 186–190. <https://doi.org/10.1038/nature21039>
- Nagel M., Jansen P.R., Stringer S., Watanabe K., de Leeuw C.A. i współaut., 2018. *Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways*. Nat Genet 50, 920–927. <https://doi.org/10.1038/s41588-018-0151-7>
- Niemiec E., Vears D.F., Borry P., Howard H.C., 2018. *Readability of informed consent forms for whole-exome and whole-genome sequencing*. J Community Genet 9, 143–151. <https://doi.org/10.1007/s12687-017-0324-6>
- Noble R., Noble D., 2017. *Was the Watchmaker Blind? Or Was She One-Eyed?* Biology (Basel) 6, 47. <https://doi.org/10.3390/biology6040047>
- Ostrander E.A., Huson H.J., Ostrander G.K., 2009. *Genetics of athletic performance*. Annu Rev Genomics Hum Genet 10, 407–429. <https://doi.org/10.1146/annurev-genom-082908-150058>
- Padmanabhan JL, Shah JL, Tandon N, Keshavan MS. *The „polyenviromic risk score”: Aggregating environmental risk factors predicts conversion to psychosis in familial high-risk subjects*. Schizophr Res. 2017 Mar;181:17-22. <https://doi.org/10.1016/j.schres.2016.10.014>
- Perkins J.M., Subramanian S.V., Davey Smith G., Özaltin E., 2016. *Adult height, nutrition, and population health*. Nutr Rev 74, 149–165. <https://doi.org/10.1093/nutrit/nuv105>
- Płochiński N., 1968. *Odziedziczalność*. Państwowe Wydawnictwo Rolnicze i Leśne, Warszawa.
- Posthumus M., Caine D.J., Collins M., Hills A.P., Noakes T. i współaut., 2016. *Genetics and sports, 2nd, revised and extended edition. Medicine and sport science*. Karger, Basel Freiburg.
- Quijada M.R., 2016. *Is the Successful Athlete Born or Made? A Review of the Literature*. Apunts: Educació Física i Esports 123, 7–12.
- Rankinen T., Fuku N., Wolfarth B., Wang G., Sarzynski M.A. i współaut., 2016. *No Evidence of a Common DNA Variant Profile Specific to World Class Endurance Athletes*. PLoS One 11, e0147330. <https://doi.org/10.1371/journal.pone.0147330>
- Richardson S.S., 2015. *Postgenomics: Perspectives on Biology after the Genome*. Duke University Press, Durham; London.
- Riggs D.W., Yeager R.A., Bhatnagar A., 2018. *Defining the human envirome: an omics approach for assessing the environmental risk of cardiovascu-*

- lar disease. *Circ Res* 122, 1259–1275. <https://doi.org/10.1161/CIRCRESAHA.117.311230>
- Ritz B.R., Chatterjee N., Garcia-Closas M., Gauderman W.J., Pierce B.L. i współaut., 2017. *Lessons Learned From Past Gene-Environment Interaction Successes*. *Am J Epidemiol* 186, 778–786. <https://doi.org/10.1093/aje/kwx230>
- Ruiz J.R., Gómez-Gallego F., Santiago C., González-Freire M., Verde Z. i współaut., 2009. *Is there an optimum endurance polygenic profile?* *J Physiol* 587, 1527–1534. <https://doi.org/10.1113/jphysiol.2008.166645>
- Schutte N.M., Nederend I., Hudziak J.J., Bartels M., de Geus E.J.C., 2016. *Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption*. *Physiol Genomics* 48, 210–219. <https://doi.org/10.1152/physiolgenomics.00117.2015>
- Semenova E.A., Hall E.C.R., Ahmetov I.I., 2023. *Genes and Athletic Performance: The 2023 Update*. *Genes (Basel)* 14, 1235. <https://doi.org/10.3390/genes14061235>
- Sharon G., Cruz N.J., Kang D.-W., Gandal M.J., Wang B. i współaut., 2019. *Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice*. *Cell* 177, 1600–1618.e17. <https://doi.org/10.1016/j.cell.2019.05.004>
- Sherman R.M., Forman J., Antonescu V., Puiu D., Daya M. i współaut., 2019. *Assembly of a pan-genome from deep sequencing of 910 humans of African descent*. *Nat Genet* 51, 30–35. <https://doi.org/10.1038/s41588-018-0273-y>
- Sowa J., Wolański K., 2018. *Sport nie istnieje. Igrzyska w społeczeństwie spektaklu*. WAB, Warszawa.
- Sudmant, P., Rausch, T., Gardner, E. Handsaker, R.E., Abyzov A. i współaut., 2015. *An integrated map of structural variation in 2,504 human genomes*. *Nature* 526, 75–81. <https://doi.org/10.1038/nature15394>
- Sunde H.F., Eftedal N.H., Cheesman R., Corfield E.C., Klepepesto T.H. i współaut., 2023. *Genetic similarity between relatives provides evidence on the presence and history of assortative mating*. <https://doi.org/10.1101/2023.06.27.546663>
- Tucker R., Collins M., 2012. *What makes champions? A review of the relative contribution of genes and training to sporting success*. *Br J Sports Med* 46, 555–561. <https://doi.org/10.1136/bjsports-2011-090548>
- Weiss K.M., 2000. *Consulting the oracle: reverence, but criticism*. *Genet Epidemiol* 19, 468–472.
- Williams A.G., Folland J.P., 2008. *Similarity of polygenic profiles limits the potential for elite human physical performance*. *J Physiol* 586, 113–121. <https://doi.org/10.1113/jphysiol.2007.141887>
- Wilson K., 2019. *Personality and physical activity*. [W:] *APA Handbook of Sport and Exercise Psychology, Volume 2: Exercise Psychology, Vol. 2, APA Handbooks in Psychology Series*. American Psychological Association, Washington, 219–239. <https://doi.org/10.1037/0000124-012>
- Yengo L., Vedantam S., Marouli E., Sidorenko J., Bartell E., Sakaue S. i współaut., 2022. *A saturated map of common genetic variants associated with human height*. *Nature* 610, 704–712. <https://doi.org/10.1038/s41586-022-05275-y>
- Young A.I., Benonisdottir S., Przeworski M., Kong A., 2019. *Deconstructing the sources of genotype-phenotype associations in humans*. *Science* 365, 1396–1400. <https://doi.org/10.1126/science.aax3710>
- Yvert T., Miyamoto-Mikami E., Murakami H., Miyachi M., Kawahara T. i współaut., 2016. *Lack of replication of associations between multiple genetic polymorphisms and endurance athlete status in Japanese population*. *Physiol Rep* 4, e13003. <https://doi.org/10.14814/phy2.13003>
- Zwart H., 2015. *Human Genome Project: history and assessment*. [W:] *International Encyclopedia of Social & Behavioral Sciences*. 2nd ed. Elsevier, Oxford, 311–317.

