

MIECZYŚLAW CHORAŻY

Gen strukturalny – ewolucja pojęcia i dylematy

Pracę tą dedykuję tym wszystkim byłym i obecnym pracownikom Zakładu Biologii Nowotworów, Zakładu Biologii Molekularnej i Zakładu Radiobiologii Doświadczalnej i Klinicznej Centrum Onkologii – Instytutu im. Marii Skłodowskiej-Curie, Oddział w Gliwicach, których rzetelność, żarliwość i oddanie nauce pozwalały zawsze przetrwać trudne lata.

W ostatnim półwieczu termin „gen”, powszechnie używany wśród biologów, rozpowszechnił się także w szerokich kręgach cywilizowanych społeczeństw. „Gen” jest pojęciem teoretycznym, choć współczesna nauka przyjmuje, że ma on postać materialną. Mimo swej stale narastającej złożoności, termin „gen” jest nieodzownym składnikiem komunikacji w różnych kręgach badaczy, wszedł do terminologii medycznej, znalazł miejsce w praktyce ekonomicznej (firmy biotechnologiczne, patenty, agrobiologia), a także w kulturze, socjologii, filozofii i polityce. Termin ten ma wysoki autorytet w dyskusjach i ocenach naukowych, a także często jest używany w potocznych debatach ludzi wykształconych. Stąd wydaje się pożytecznym upowszechnienie podstawowej wiedzy o tym, czym jest „gen”, jak był pojmowany półtora wieku temu i jak jest rozumiany obecnie. Dlatego też badania, które doprowadziły do współczesnego poglądu na gen jako na materialną jednostkę dziedziczności, omawiam nieco szczegółowiej.

Od wieków ludzi ciekawiło pytanie, dlaczego potomstwo zwierząt i roślin zawsze jest podobne do swoich rodziców, choć nie jest to podobieństwo absolutne. Rolnicy starych kultur (Babilon i Egipt), prowadząc hodowlę roślin i zwierząt, byli świadomi, że pewne cechy użyteczne mogą być przenoszone z organizmów rodzicielskich na potomstwo. W Biblii znalazło się ostrzeżenie wskazujące, że ludzie tamtych czasów mieli świadomość dziedziczenia cech takich jak hemofilia. Biblia wskazywała, że po śmierci dwóch synów z powodu wykrwawienia, u trzeciego syna nie wolno dokonywać obrzezania. W ubiegłych wiekach istniała świadomość rodzinnego przenoszenia różnych zaburzeń zdrowotnych (dodatkowe palce – polidaktylia, nowotwór siatkówki oka – retinoblastoma, samoistne krwawienie – hemofilia).

prof. dr hab. n. med. Mieczysław Chorąży, członek rzeczywisty PAN, Zakład Biologii Nowotworów, Centrum Onkologii – Instytut im. Marii Skłodowskiej-Curie, Oddział w Gliwicach, e-mail: chorazy@io.gliwice.pl

Gen jako jednostka dziedziczności – wczesne koncepcje

Przegląd wczesnych poglądów na temat mechanizmu dziedziczenia podaję na podstawie dzieła Michela Morange: *A History of Molecular Biology* (Morange, 1998).

U wszystkich gatunków potomstwo dziedziczy wszystko to, co jest szczególnie u ich rodziców – ta reguła znana była od czasów Arystotelesa, ale dopiero w XVII i XVIII wieku zaczęto formułować pogląd, że połączenie komórki jajowej, produkowanej przez samice zwierząt, z nasieniem (sperma), produkowanym przez osobniki męskie, może dać początek życia zarodka. Wiązało się to z odkryciem komórki jajowej produkowanej przez osobniki żeńskie zwierząt (W. Harvey i A. van Leuwenhoek). W XVII wieku panowała **teoria preformacji** (J. Swammerdam), która głosiła, że plemniki człowieka zawierają w swojej główce uformowanego, pełnego, choć niewidocznego, miniaturowego człowieka. W czasie rozwoju osobniczego (ontogenezy) następowało więc „odkrywanie” formy, a nie progresywne „wypracowywanie” jej. Brzemie dziedziczenia spoczywało na ojcu. W procesie rozwojowym miały powstawać nowe formy preadaptowane już do presji środowiskowej, na którą są skazane następne generacje. Odmienne stanowisko głosiła **teoria inkapsulacji**, wskazująca na żeńską komórkę jajową jako miejsce, gdzie miały bytować miniaturowe organizmy potomne, umieszczone jedne w drugich, jak w rosyjskiej laleczce „matrioszka”, w których większa mieści w sobie mniejszą, a ta jeszcze mniejszą itd. (Moss i Ch. Bonnet). W istocie termin „ewolucja” zawiera w sobie pierwiastek preformistyczny (powstawanie czegoś nowego z czegoś, co już było przed tym). Inni wyrażali pogląd, że komórki reprodukcyjne nie zawierały preformowanych embrionów, lecz cząsteczki, które miały właściwość formowania zarodka.

Według Herberta Spencera (1863) w całym organizmie istnieją jednakowe „fizjologiczne jednostki” odpowiedzialne za dziedziczenie cech. W roku 1868 Charles Darwin zaproponował **teorię pangenezy**, głoszącą, że każda komórka somatyczna organizmu zawiera swobodnie przemieszczające się, niehomogenne, niezależne od siebie, miniaturowe cząsteczki (gemmule), które w narządach rozrodczych tworzą rozwijającego się zarodka. Osobniki rodzicielskie mieszają owe cząsteczki w akcie poczęcia, a jednocześnie, wraz z aktem transmisji cech, przenosi się na potomstwo zmienność, która to właściwość stała się kanwą dla rozwinięcia jego teorii selekcji jako siły ewolucyjnej. Zatem gemmule miały być odpowiedzialne za dziedziczenie cech gatunkowych i jednocześnie byłyby nośnikami zmienności. Środowisko działające na przypadkową, dziedziczną zmienność miało być sitem dokonującym naturalnej selekcji, czyli doboru osobników najlepiej przystosowanych do środowiska, i jednocześnie być siłą napędową ewolucji. Teorię pangenezy Darwina rozwinął Hugo de Vries. Według niego komórkowe „pangeny” mogły mutować i podlegać zmienności co do liczby i rodzaju, dzięki czemu powstawały odmiany i gatunki.

W XIX wieku August Weismann zaproponował **teorię plazmy zarodkowej**. Weismann postulował wyróżnienie „plazmy zarodkowej” jako oddzielnego od ciała (soma)

trwałego komponentu organizmu żywego. Plazma zarodkowa bytowała jedynie w komórkach rozrodczych (komórka jajowa i plemnik), była niezależna od somy i przekazywana była z pokolenia na pokolenie za pomocą komórek rozrodczych. Plazma zarodkowa zawierała hipotetyczne elementy dziedziczne zwane przez Weismanna „idantami”. Idanty składały się ze zdolnych do samopowielania się „idów”, złożonych z kolei z „determinantów” i „bioforów”. „Idanty” były odpowiedzialne za organizację (specyfikację) organizmu i przносиły cechy przodków na potomstwo. W procesie rozwoju „idy” rozdzielały się do różnych części zarodka. Liczba „idów” była redukowana do połowy w czasie formowania się komórek jajowych (oogeneza) i plemników (spermatogeneza). Międzyosobnicza zmienność miała wynikać z różnych kombinacji „idów” w komórkach jajowych.

Doświadczenia G. Mendla w XIX wieku dały impuls do rozwijania badań nad zjawiskami towarzyszącymi procesowi dziedziczenia, a zatem zjawiskami stojącymi u podstaw życia. Mendel obserwował u krzyżówek groszku ogrodowego przenoszenie na potomstwo prostych cech, takich jak wysokość rośliny, kolor kwiatów lub zabarwienie i wygląd ziaren. Na podstawie swoich badań Mendel sformułował koncepcję, że za przenoszenie cech odpowiedzialne są jednostki dziedziczości, które nazwał „determinantami”, „elementami” lub „związkami”. Pierwsze pokolenie roślin (F1) powstałe z krzyżówki roślin o różnych cechach (np. roślina wysoka × roślina niska, kolor kwiatów czerwony × kolor biały) ma cechę jednolitą, dominującą (wysoki wzrost, czerwona barwa kwiatów), ale już w pokoleniu następnym (F2), powstałym z krzyżówek między osobnikami F1, następuje rozszczepienie cech: powstają rośliny o wysokim i niskim wzroście (lub o czerwonych i białych kwiatach). Cechy takie jak niski wzrost i biała barwa kwiatów występują w populacji F2 w 25%. Dokonując następnych krzyżówek roślin z jedną lub dwoma cechami i analizując statystyczny rozkład cech w potomstwie, Mendel doszedł do przekonania, że rozmieszczenie cech w potomstwie nie jest zjawiskiem przypadkowym, lecz podlega pewnym regułom. Za każdą cechę odpowiedzialna jest para elementów (alleli). Na podstawie takich statystycznych analiz Mendel sformułował prawa dziedziczenia i rozwoju. W czasie formowania komórek rozrodczych następuje segregacja cech – pojedyncze jednostki (allele) z pary czynników determinujących poszczególne cechy rozchodzą się i przechodzą do komórek płciowych, a po zapłodnieniu allele łączą się ponownie w pary. Komórki płciowe mają po 50% determinantów cech (alleli) każdego z rodziców. Jednostki te (allele) mogą mieć charakter dominujący, recesywny lub mogą występować w postaci mieszanej. Allel Mendla był stanem, warunkiem lub procesem, który wpływał na kształtowanie właściwości komórkowej albo na ukierunkowanie rozwoju w bardzo swoisty, choć niejasny sposób. Badania Mendla pozostały w zapomnieniu do roku 1900, gdy ponownie „odkrył” je (Hugo de Vries, Erich von Tschermak i Carl Correns).

Wprowadzenie do biologii terminu „gen” (z greckiego *genos* – początek) przypisuje się duńskiemu badaczowi Wilhelmowi Johannsenowi, który w roku 1909 użył go do okreś-

lenia „czegoś” nieokreślonego, neutralnego, co bytuje w komórce (organizmie) i określa pewne aspekty fenotypu. Gen nie miał żadnej formy materialnej i był używany jako termin operacyjny na użytek badań nad dziedzicznością. „Gen” nie został zdefiniowany, lecz zastąpił wszystkie poprzednie nazwy „jednostek”, „elementów”, „czynników”, które miały mieć określoną strukturę lub funkcję w procesie dziedziczenia i rozwoju. Johannsen wskazał też na zasadnicze różnice między genotypem a fenotypem. Termin „genotyp” odnosił się do przenoszonych z pokolenia na pokolenie zbioru par czynników („genów”), z których jeden pochodził od ojca, a drugi od matki. „Fenotyp” określał całkowity zespół cech i ich różnych form występujących u osobników dojrzałych. Fenotyp był czymś realnym, mierzalnym, dostępnym do obserwacji. „Geny” (i genotypy) były czymś niewidocznym, niedostępnym do obserwacji, nie było więc wiedzy o ich wewnętrznych właściwościach.

Do badań nad dziedzicznością Thomas H. Morgan użył muszki owocowej (*Drosophila melanogaster*) ze względu na jej szybki cykl życiowy, możliwość uzyskiwania krzyżówek i łatwe śledzenie cech fenotypowych. Szczególnie interesujące były badania Morgana nad sprzężeniem oraz wymianą fragmentów chromosomów (rekombinacja) w procesie tworzenia komórek płciowych. Geny miały być zorganizowane w formie liniowej. Im bliżej siebie leżały geny w chromosomie, tym rzadziej występowała ich rekombinacja w procesie *crossing-over*. Jednocześnie chromosomy olbrzymie ślinianki muszki owocowej, posiadające morfologiczne znaczniki struktury (jasne i ciemne pasma poprzeczne) i funkcji (pufy), posłużyły Morganowi do badań zależności między zmianami w genach i relacją tych zmian do znaczników. Badania te pozwoliły na sporządzenie fizycznej mapy chromosomów ślinianki i porównanie jej z mapą uzyskaną z badań nad rekombinacją.

Odkrycie chromosomów (W. Flemming i E. Strasburger) pozwoliło na poznanie zjawiska mitozy i mejozy. Terminem mitozy określa się podział komórki somatycznej na dwie identyczne komórki potomne. Mitozę poprzedza duplikacja chromosomów, a następnie rozmieszczenie ich równoważników do komórek potomnych. Mejoza jest zjawiskiem złożonym, mającym miejsce w czasie formowania komórek płciowych. W czasie mejozy następuje zredukowanie liczby chromosomów do połowy w każdej komórce płciowej obu rodziców, tak jak to postulował Mendel dla swoich „czynników” lub „alleli”.

W wyniku badań Mendla i Morgana sformułowano **chromosomalną teorię dziedziczności**. Według tej teorii geny były jednostkami dziedziczności ułożonymi wzdłuż chromosomu, złączonymi ze sobą jak paciorki różańca i mającymi względem siebie stałą pozycję – *locus*. Sporządzone zostały mapy sprzężeń genów, to jest oznaczono względne odległości, w jakich geny rozłożone są względem siebie. Dotychczas nieokreślone pojęcie genu, ujmowane jako stan lub proces, sprowadzono do nadal jeszcze niezdefiniowanego „czynnika”, ale czynnika zlokalizowanego w materialnym chromosomie, czyli

morfolologicznej, dostępnej do obserwacji strukturze komórki. Morgan przewidział, że indywidualny gen może spełniać wiele funkcji i dawać szeroki wachlarz efektów. Dalej domniemywał, że każdy gen może mieć swoiste działanie (efekt) w danym narządzie, ale także swoiste w innym narządzie, a w wyjątkowych przypadkach we wszystkich narządach i w cechach całego organizmu. Te myśli – jak opóźnione echo – wracają po dziesiątkach lat do współczesnej biologii.

Uczeń Morgana, Hermann J. Muller (późniejszy laureat Nagrody Nobla), za pomocą promieni rentgenowskich i związków chemicznych uzyskał różne fenotypy muszki owocowej, przypisując te nowe fenotypy uszkodzeniom (mutacji) genów. Muller uważał geny za „fundament życia” – stabilne struktury, do których są „doczepione” inne elementy organizmu.

Stworzenie podstaw cytogenetyki zawdzięczamy Walterowi Suttonowi. Głosił on, że chromosomy są nośnikami cech dziedzicznych i w 1903 roku sformułował podstawowe kanony dziedziczenia chromosomalnego: wszystkie organizmy w komórkach somatycznych zawierają podwójny komplet chromosomów; komórki rozrodcze (plemnik i komórka jajowa) zawierają tylko pojedynczy komplet chromosomów – ojca lub matki, a zapłodniona komórka jajowa (zygota) ponownie posiada podwójny komplet chromosomów (ojcowski i matczyne). Z zygoty powstaje organizm potomny o podwójnym komplecie chromosomów, podobnie jak u ojca i matki.

Termin „genetyka” jako nauka o dziedziczności został wprowadzony przez Williama Batesona (początek XIX wieku). Bateson sądził, że istnieją „jednostki właściwości” (niekiedy nazywał je również „genami”) występujące w antagonistycznych parach. Te jednostki nazwał „allelomorfami” (*allelomorph* = inna forma). Zapłodnioną komórkę jajową (zygota), powstającą ze złączenia się w parę podobnych allelomorfów – matczynej gamety (komórka jajowa) i ojcowskiej gamety (plemnik), nazywamy homozygotą. Jeśli w gametach rodzicielskich wystąpią allelomorfy niepodobne do siebie, po akcie zapłodnienia powstanie heterozygota. Do dziś pozostał termin „allel” jako alternatywny stan genu strukturalnego.

„Gen” Batesona i „gen” Johannsena nie miały określonej materialnej struktury. Geny były hipotetycznym „czymś” (czynniki, procesy?), co dawało zdefiniowane właściwości (cechy) i co było odpowiedzialne za przekazanie cech rodzicielskich potomstwu.

Na początku XIX wieku nie było dużego zainteresowania fizyczną strukturą genów. Jedynie H. J. Muller podejmował badania nad zmiennością mutacji wywołanych u *D. melanogaster* promieniami rentgena, sądząc (podobnie jak Max Delbrück), że analiza zmienności mutacji w relacji do energii radiacyjnej pozwoli na zdefiniowanie pewnych cech fizycznych (np. wielkość) genu.

Nowa dyscyplina nauki, genetyka, uważana była za naukę użyteczną jedynie w agromonii, zwłaszcza w hodowli roślin i zwierząt, czego wyrazem było powoływanie, zwłaszcza

cza w St. Zjednoczonych, zespołów i instytucji zajmujących się tą dziedziną. Rozwój genetyki populacyjnej był utrudniony przez fakt, że cechy osobnicze ilościowe mają zwykle charakter wielogenowy (i wieloczynnikowy).

Powiązania genetyki z embriologią były luźne i przez długi czas istniał wyraźny rozdział między tymi dyscyplinami. W dociekaniu podstaw ewolucji zrodziła się teoria neodarwinizmu wskazująca na związki procesu ewolucji z genetyką. Teoria ta odrzucała możliwość dziedziczenia cech nabytych (co postulował Lamarck i co dokumentował naiwny eksperyment Weissmana, wykazujący, że mimo obcinania ogonów myszom w kilku pokoleniach nie rodziły się myszy bez ogonów!), podkreślała mutację (genów) jako podstawę istnienia zmienności i wskazywała na eliminację z populacji osobników słabo dopasowanych na rzecz osobników lepiej dostosowanych do środowiska i wykazujących przeto zwiększoną zdolność do reprodukcji. Do sformułowania syntezy neodarwinizmu przyczyniły się poglądy Theodosiusa Dobzhansky'ego, ale hipoteza miała nadal charakter domniemania, gdyż natura genu, zjawisko mutacji, mechanizm zmienności i rozwój osobniczy pozostawały nadal w obszarze abstrakcji.

Uderzającym są słabe związki genetyki z chemią, która pod koniec XIX wieku poczyniła znaczne postępy w poznawaniu procesów przebiegających w żywych organizmach.

Mimo nieokreślonej natury genów przypisywano im wiele niezwykłych właściwości: geny miały mieć moc działania, być siłą sprawczą (przyczynową), mieć zdolność do samoreprodukcji, co nadawało genom atrybut życia. Geny miały mieć siłę animowania i kontroli rozwoju organizmu, a także miały być białkami lub aperiodycznymi kryształami upakowanymi w chromosomach, zawierającymi „zapisy kodowe” (Schrödinger, 1998).

Okres poszukiwań i niepewności

Żywiłowy rozwój biochemii na przełomie XVIII wieku i w pierwszej połowie XIX wieku umożliwił również rozwój genetyki – nauki o dziedziczności.

W 1909 roku lekarz londyński Archibald Garrod przedstawił dowód na dziedziczny charakter metabolicznego zaburzenia zwanego alkaptonurią i jednocześnie doszedł do wniosku, że zaburzenie to polegało na niemożności metabolicznego rozzerwania pierścienia benzenowego tyrozyny. Dowodził, że najprawdopodobniej u chorych występuje wrodzony brak odpowiedniego enzymu, a ponieważ choroba była dziedziczna, więc musiała być związana z defektem genetycznym.

Inni badacze (Boris Ephrussi, Georg Beadle, Fritz von Wettstein i in.) poczynili obserwacje nad dziedziczeniem koloru oczu u owadów (motyl, muszka owocowa) oraz antocjanin u roślin, wskazując na zależności powstające w wyniku reakcji katalizowanych przez enzymy barwnika a dziedzicznością. Były to zatem pierwsze przybliżone opisy funkcji genów i ich związków z funkcją enzymów.

Ważnym postępowaniem wskazującym na ściślejsze relacje między genetyką a biochemią były badania Geорга Beadle i Edwarda Tatum nad pleśnią *Neurospora*. Za pomocą promieniowania indukowali oni u zarodników pleśni mutacje, które powodowały utratę zdolności syntezy stosunkowo prostych związków chemicznych (np. witamin, aminokwasów) niezbędnych dla wzrostu. W toku doświadczeń znaleziono takiego mutantu pleśni, który nie był w stanie syntetyzować jednego z aminokwasów – tryptofanu. Krzyżowanie mutantów i hodowle pleśni na odpowiednich prostych pożywkach umożliwiły badaczom precyzyjne określenie szlaku biosyntezy tryptofanu i wykazanie, że każdy etap syntezy był kontrolowany przez oddzielne enzymy i różne, odpowiadające im geny. Tak powstała hipoteza „jeden gen – jeden enzym”. Gen nosił więc atrybuty „planu” dla budowy białka.

Badania metaboliczne wykazały, że geny kontrolują nie tylko ogólne, biologiczne cechy organizmu, lecz także enzymy mające ściśle określone, mierzalne funkcje. Stąd zrodziła się hipoteza, że same geny są niczym innym jak samosyntetyzującymi i samoreplikacyjnymi enzymami. A ponieważ wiadano już, że enzymy są białkami hipotezy „jeden gen – jeden enzym” używano również naprzemiennie jako „jeden gen – jedno białko”. Hipoteza ta przez wiele lat była główną koncepcją opisującą naturę genów. Podkreślam ten fakt, aby uświadomić sobie, jakie meandry pokonuje myśl badacza, by na podstawie gromadzonej wiedzy usiłować budować obraz świata.

Okolo roku 1869 lekarz szwajcarski F. Miescher wyizolował z jąder komórek zawartych w ropiejących ranach substancję, której dał nazwę „nukleina”, nazwaną potem kwasem nukleinowym. Podstawową cegiełką kwasu nukleinowego jest nukleotyd. Każdy nukleotyd zbudowany jest z kolei z jednej z czterech rodzajów zasady azotowej o symbolu: A (adenina), T (tymina), G (guanina), C (cytozyna), jednej cząsteczki cukru (deoksyryboza) i reszty fosforanowej (odpowiedzialnej za kwaśny charakter nukleotydu i kwasu nukleinowego). W tekście symbole A, T, G, C używane są wymiennie dla oznaczenia zarówno zasady, jak i nukleotydu.

Duże znaczenie dla rozwoju wiedzy o genie miał fakt rozróżnienia dwóch rodzajów kwasów nukleinowych: DNA, zawierającego deoksyrybozę jako składnik cukrowy, i RNA, zawierającego rybozę. Stąd pochodzą też akronimy utworzone z ich nazwy w języku angielskim: DNA – *deoxyribonucleic acid* i RNA – *ribonucleic acid*. W RNA zamiast tyminy (T) występuje uracyl (U). Metody rozróżniania obu typów kwasów nukleinowych przy pomocy odmiennego sposobu barwienia pozwoliły na ich lokalizację w komórkach zwierzęcych. RNA występował głównie w cytoplazmie (J. Brachet), a DNA w jądrze, w chromosomach (T. Caspersson).

Przez parę dziesiątków lat cząsteczki DNA nie przypisywano specjalnego znaczenia, zwłaszcza że jeszcze do lat 30. ubiegłego stulecia uważano, że jest to prosta cząsteczka. DNA miał być złożony z czterech nukleotydów powtarzających się monotonicznie wzdłuż cząsteczki. Choć szwedzki badacz T. Caspersson wykazał, że DNA jest składnikiem

chromosomów, sądzono, że stanowi on tylko swego rodzaju podporę dla cząsteczek białkowych odpowiedzialnych za swoistość genetyczną. Wybór kilkudziesięciu najważniejszych prac opisujących badania nad odkryciem DNA, poznaniem jego struktury i właściwości znajduje się w obszernym tomie opracowanym przez J. H. Taylora (Taylor, 1965).

W 1928 roku F. Griffith opisał niezwykle spostrzeżenie. Niewirulentne (nie dające efektu patologicznego) dwoinki zapalenia płuc u myszy zmieszane z wirulentnymi, ale zabitymi dwoinkami, ulegały przekształceniu (transformacji) i stawały się wirulentnymi. Co więcej, cecha wirulencji pojawiała się w bakteriach niewirulentnych nawet wtedy, gdy do ich kolonii dodawano jedynie ekstrakt z zabitych, wirulentnych bakterii.

W 1944 roku Oswald T. Avery, badacz z prestiżowego Instytutu Rockefellera, oraz Colin MacLeod i Maclyn McCarthy, idąc śladem odkrycia Griffitha dostarczyli dowodu, że czynnikiem transformującym, przenoszącym cechę wirulencji był DNA. Odkrycie Avery'ego stało się wielkim wyzwaniem dla powszechnie przyjmowanej hipotezy o białkowej naturze genu. Tymczasem doświadczenia Avery'ego wskazywały jednoznacznie na DNA jako cząsteczkę przenoszącą cechę fenotypową, a więc mającą atrybut genu.

Nieformalna „grupa fagowa” amerykańskich badaczy, działająca w latach 1940-1960, obrała bakteriofagi jako model badawczy zjawisk życia. Trzon grupy stanowili m.in. Max Delbrück, Salvatore Luria, Nikolaj Timofeef-Ressovsky, Alfred Hershey. Delbrück sądził, że bakteriofagi są elementarnymi cząsteczkami biologicznymi, które przez swą zdolność do infekowania bakterii i gwałtownego namnażania się stanowią idealny model do badań nad fundamentalną cechą życia – samoreplikacją. Ponadto Joshua Lederberg (1951) wykazał, że bakteriofag może przenosić cechy dziedziczne z jednego szczepu bakterii na inny (zjawisko transdukcji).

Proces replikacji bakteriofaga ma miejsce w zakażonej bakterii. Ku powszechnemu zdumieniu, Alfred Hershey i Martha Chase (1952) udowodnili jednak, że w procesie zakażenia fag nie wchodzi do komórki bakteryjnej w całości, lecz wstrzykuje jedynie do wnętrza bakterii swój DNA. Białkowa otoczka bakteriofaga pozostaje na powierzchni błony komórkowej bakterii. Konkluzja z doświadczenia Hershey'a i Chase była prosta: DNA fagowy zawiera całą informację potrzebną dla reprodukcji i uformowania dojrzałego bakteriofaga.

Odkrycie zjawiska transformacji i transdukcji oraz poznanie mechanizmu infekcji bakterii bakteriofagiem były bodźcem dla intensywnych badań nad budową DNA i jego rolą w przenoszeniu cech dziedzicznych. Szybko zbliżał się rozkwit nowej wiedzy o genie jako fizycznej cząsteczce.

Gen materialny – DNA

W omawianym okresie genetycy sądzili, że podstawową i wyjątkową właściwością genu jest zdolność do samoreplikacji. Ta cecha genu jest właściwością oddzieloną od funkcji genu, zatem „samoreplikacja” może być badana niezależnie od funkcji. W kon-

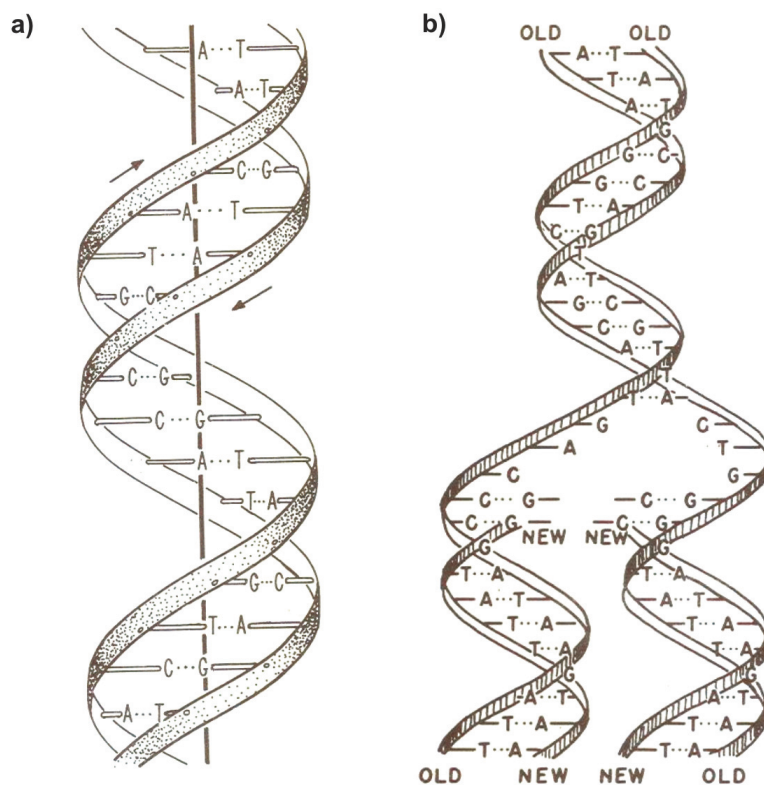
sekwencji badania nad zjawiskiem samoreplikacji DNA nie zajmowały genetyków lecz głównie fizyków oraz chemików i sprowadzały się głównie do konstruowania modeli.

Odkrycie struktury DNA (Watson i Crick, 1953) dokonane w Cavendish Laboratory jest zapewne najbardziej powszechnie znanym odkryciem w biologii. Cytuję tylko pierwszą publikację Watsona i Cricka z „Nature” z 17 kwietnia 1953 roku, zaczynającą się od słów: „Chcemy sugerować strukturę (...) kwasu deoksyrybonukleinowego D.N.A.)”. Następne serie prac tych autorów znajdzie Czytelnik w zbiorze publikacji (Taylor, 1965). Zarówno Crick, jak i Watson nie prowadzili badań laboratoryjnych. Ich zasadniczym źródłem wiedzy były nieformalne dyskusje z Maurice Wilkinsem, rzadziej z Rozalindą Franklin – współpracowniczką Wilkinsa, która również interesowała się badaniem struktury DNA metodą dyfrakcji promieni rentgenowskich. Praca obu badaczy sprowadzała się do konstruowania fizycznego modelu, który uwzględniał fizyczne proporcje, odległości, kąty, wiązania i siły oddziaływania między poszczególnymi atomami składników DNA. Nie bez znaczenia dla konstrukcji modelu struktury DNA były wyniki badań Erwina Chargaffa (Chargaff, 1950) nad DNA i zauważona przez niego reguła, że zasady DNA istnieją w ekwiwalentnych ilościach: stosunek A:T = 1 oraz G:C = 1. Reguła ta nasuwała myśl, że w cząsteczce DNA może zachodzić parowanie zasad, czyli chemiczne wiązanie zasad w pary A z T i G z C. Crick uznawał wkład Chargaffa w budowaniu modelu DNA, czemu wielokrotnie dawał wyraz (Crick, 1988)*.

W modelu struktury cząsteczki DNA Watsona-Cricka (ryc. 1) oba pasma DNA biegną antyrównolegle (mają odmienną polarność) i są skręcone wzdłuż swojej osi (helisa). Oba pasma są utrzymane w rejestrze dzięki wiązaniom chemicznym (wodorowym) między parami zasad A-T i G-C, dopasowującym je wzajemnie do siebie. Dowolny ciąg par nukleotydów w cząsteczce DNA można zatem zapisać jako ciąg sekwencji jednego (np. górnego, por. ryc. 2) pasma: AGCATTCGG.... Każda para nukleotydów A-T lub G-C zajmuje w osi cząsteczki pewną przestrzeń (ok. 3.4 Ångströma) i przeto służy jako miara długości (wielkości) cząsteczki DNA lub jego fragmentu. Mówimy więc, że np. gen ma 1500 par zasad lub 1,5 kilo par zasad (para zasad w języku angielskim – *base pair*; stąd zwykle pochodzą skróty „bp”, tu: 1500 bp lub 1,5 kb).

Dowolną sekwencję nukleotydów możemy zapisać jako ciąg symboli literowych (ryc. 2b) bez zaznaczania cząsteczek cukrowych i reszt fosforanowych, które tworzą jedynie zewnętrzne rusztowanie utrzymujące ciągłość cząsteczki DNA (ryc. 2a). Oś cząsteczki („wnętrze” cząsteczki) stanowią pary A-T i G-C, ułożone jedna nad drugą i powiązane wiązaniami wodorowymi, natomiast reszty fosforanowe i deoksyrybozy są „na zewnątrz”.

* Chargaff odczuwał przez resztę swego życia pewne niedoceniecie swego odkrycia przez przyszłych noblistów i często powtarzał nieco zgryźliwą sentencję (którą parokrotnie bezpośrednio słyszałem na jego wykładach), że „biolog molekularny to jest taki człowiek, który uprawia biochemię bez posiadania licencji chemika”.

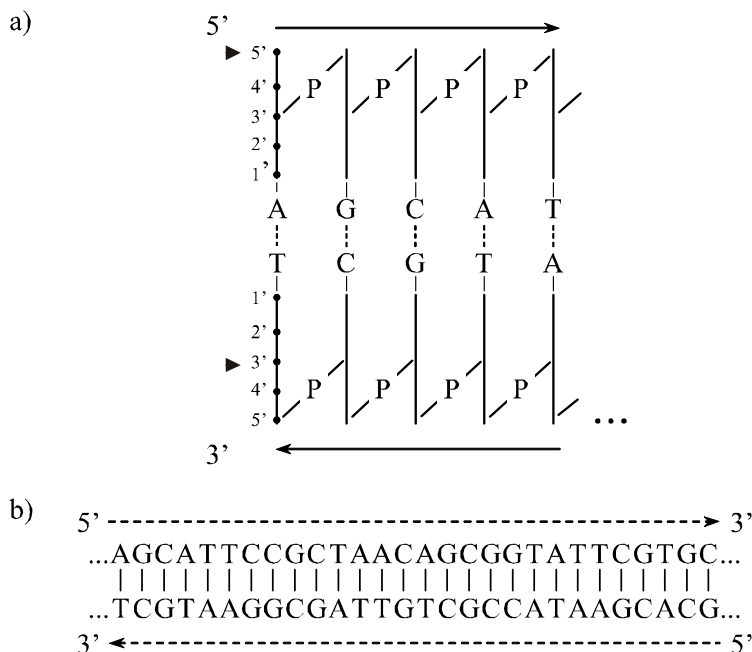


Ryc. 1. Schemat ilustrujący strukturę i replikację DNA. a) struktura DNA – dwa pasma utworzone przez ciąg nukleotydów (nukleotyd to jedna z czterech rodzajów zasad A, C, G, T, reszta cukrowa i reszta fosforanowa) zwinięte wzdłuż osi tworzą helisę. Pasma utrzymane są między sobą przez wiązania chemiczne między zasadami A-T i G-C. Struktura przypomina „spiralne” schody, w których stopnie stanowią płaszczyzny par zasad prostopadłe do osi. Odstęp między kolejnymi stopniami wynosi 3,4 Ångströma. b) schemat replikacji DNA. Na każdym „starym” paśmie DNA odtwarza się nowe pasmo. Po ukończeniu replikacji dwie cząsteczki potomne („nowe”) mają sekwencję nukleotydów identyczną jak w starej cząsteczce.

(Zapożyczono z: J.D. Watson: *The double helix*, Weidenfel and Nicolson, London, 1968)

Reszty fosforanowe nadają charakter kwaśny cząsteczce DNA, stąd w nazwie znajduje się słowo „kwas” (kwas deoksyrybonukleinowy), co niespecjalistom, np. w znanej anegdocie z Łysenką, nasuwa pospolite skojarzenia*.

* Prof. Ilja B. Zbarsky opowiadał mi, że po referacie prof. Biełozierskiego o budowie i znaczeniu kwasu deoksyrybonukleinowego, sceptyczny Łysenko poprosił o próbkę roztworu DNA. Ku zdumieniu asystenta Biełozierskiego, który następnego dnia dostarczył próbkę, Łysenko wylał na dłoń parę kropli roztworu, posmakował je językiem i oznajmił: „Eto na wierno nie kislota” („To na pewno nie kwas”).



Ryc. 2. Schemat zapisu fragmentu dowolnej sekwencji nukleotydów w dwupasmowej cząsteczce DNA. a) Oba pasma są trzymane w komplementarnym „rejestrze” dzięki wiązaniom zasad A-T i G-C. Każda z zasad związana jest z cukrem (deoksyrybozą) o pięciu atomach węgla (1' do 5'), a ten tworzy wiązanie poprzez atom fosforu z następną cząsteczką cukru, między atomami węgla 3' i 5'. W ten sposób powstaje ciągłość pasm DNA wzdłuż osi cząsteczki. Zasada azotowa związana z deoksyrybozą i resztą fosforanową stanowi nukleotyd. Strzałki pokazują polarność cząsteczki DNA od 5' do 3'. Oznakowanie 5' lub 3' bierze się od wolnego wiązania (na końcu cząsteczki DNA) przy 5' lub 3' pozycji atomu węgla deoksyrybozy. b) Uproszczony zapis sekwencji. Literami (A, T, G, C) oznaczone są jedynie zasady azotowe tworzące „kod”. Reszty cukrowe i fosforanowe powtarzają się, zatem nie ma potrzeby uwzględniać ich w zapisie

Żadne z dotychczasowych odkryć w biologii nie miało tak wielkiego wpływu na rozwój nauki, jak odkrycie struktury cząsteczki DNA dokonane przez Watsona i Cricka.

Model Watsona-Cricka sugerował koncepcję semikonserwatywnej replikacji DNA, czyli odtwarzania (syntezy) identycznych względem każdego z komplementarnych pasm kopii (czyli takiego samego ciągu sekwencji zasad), w wyniku czego powstają dwie cząsteczki potomne identyczne z cząsteczką macierzystą (ryc. 1b). Semikonserwatywny model replikacji DNA został potwierdzony w r. 1957 przez Matthew Meselsona i Franklina Stahla (zob. Taylor, 1965). Watson i Crick sądzili, że model cząsteczki DNA zapewnia jedną z podstawowych cech postulowanych dla czynnika dziedziczności: możliwość powielania cząsteczki z zachowaniem budowy i układu sekwencji zasad (nukleotydów) cząsteczki macierzystej.

Struktura cząsteczki DNA złożona z dwóch komplementarnych w stosunku do siebie pasm i semikonserwatywna replikacja sugestywnie, aczkolwiek myląco, ujmowana jest często jako inherentna zdolność DNA „do samoreplikacji”. W istocie utworzenie dwóch cząsteczek potomnych (replikacja) z cząsteczki DNA – od początkowej fazy oddzielania się pasm do ukończenia procesu – prowadzona jest przez złożone systemy enzymatyczne.

W późniejszych latach odkryto, że pasmo DNA służy również jako matryca do syntezy komplementarnej nici RNA. Cząsteczka DNA jest uważana za strukturę stabilną, mimo że DNA ma (wprawdzie ograniczone) zdolności autokatalityczne (DNAzymy) (por. Adamala i Pikuła, 2004). Również niektóre szczególne sekwencje DNA mają zdolność do przemieszczania się (por. niżej), a w komórkach somatycznych zachodzi zjawisko rekombinacji (np. formowanie i rearanżacja genów w procesie odpowiedzi immunologicznej, rearanżacja sekwencji kodujących receptor powierzchniowy limfocytów typu T).

Odkrycie Watsona i Cricka zostało jednoznacznie przyjęte przez biologów i genetyków jako ujawnienie materialnej struktury genu. W euforii towarzyszącej odkryciu badacze ci wypowiedzieli wiele deklaracji cytowanych do dzisiejszego dnia, nadających genowi (DNA) moc sprawczą i plasujących gen (DNA) na centralnej pozycji w zjawisku życia: „Odkryliśmy istotę życia”, „Geny są istotą życia”.

Nawet po upływie dziesiątków lat przekonanie, że geny/DNA są jakąś samoistną siłą sprawczą procesów rozwojowych i samego życia jest przewodnią myślą wypowiedianą przez wielu badaczy. Współczesne definicje genu opierają się na uznaniu go jako realnej, fizycznej jednostki genomowej i formowane są w oparciu o sekwencje DNA. „Genes are made of DNA” stwierdza czołowe dzieło *The Molecular Biology of the Cell* (Alberts i in., 1994).

Kod genetyczny

Identyfikacja DNA jako materialnego nośnika cech dziedzicznych i odkrycie struktury DNA ujawniły następny problem: na czym polega „kod genetyczny” i rola DNA (identyfikowanego już z genem) w syntezie białka, które jest ostatecznym wyrazem funkcji genu. W latach 50. było już ugruntowane przekonanie, że synteza białka w cytoplazmie polega na uformowaniu łańcucha aminokwasów (łańcuch polipeptydowy), który następnie spontanicznie przyjmuje konformację przestrzenną (fałduje się) i staje się cząsteczką funkcjonalną. Podstawowych aminokwasów jest 20. Ustalony też był fakt, że DNA znajduje się w jądrze komórkowym, a synteza białka ma miejsce w cytoplazmie. Pierwsze pytanie brzmiało: jak przekłada się instrukcja zawarta w DNA na ciąg aminokwasów w białku?

Już w pierwszej publikacji w „Nature” (kwiecień, 1953) poświęconej wyłącznie strukturze cząsteczki DNA, Crick i Watson przedstawili krótki komentarz, ekscytując

świat badaczy: „Nie uszedł naszej uwadze fakt, że swoiste parowanie [zasad], które postulujemy, natychmiast nasuwa sugestię co do możliwego kopiowania materiału genetycznego”. To wyważone sformułowanie było podyktowane, jak przyznaje Crick, zapewnieniem sobie pierwszeństwa w kwestii dalszych prac nad kodem genetycznym (Crick, 1988).

George Gamow sugerował, że istnieje bezpośrednia relacja między sekwencjami nukleotydów w DNA a sekwencją aminokwasów białka. Gamow rozważał nawet strukturę „kodu”, która miałaby obejmować trzy zasady (nukleotydy), czyli triplet tworzący kodon, który miał być kodem nakładającym się. Pogląd Gamowa, choć krytykowany przez Cricka, był jednak pewną inspiracją dla niego i nasunął myśl o bezpośrednim związku między nukleotydami DNA i aminokwasami oraz kolinearnością między DNA i łańcuchem białkowym. Jednak molekularny mechanizm rozpoznający taki kod był nadal trudny do wyobrażenia.

Na zasadzie podobnej odległości między nukleotydami w paśmie DNA i wiązaniami peptydowymi aminokwasów w białku Alexander Dounce w roku 1953 sugerował, że DNA może służyć jako matryca dla syntezy RNA, a ten ostatni może być wtórną matrycą dla syntezy białka. Podobną sugestię wysunęli już wcześniej francuscy badacze André Boivin i Roger Vendrely. Oba te fakty są mało znane, cytuję je za Morange (Morange, 1998).

Tymczasem badania nad syntezą białka w układach bezkomórkowych wskazywały, że dla syntezy niezbędne są m.in. mikrosomy, ekstrakty komórkowe zawierające enzymy i RNA oraz ATP jako źródło energii. Paul Zamecnik i in. wykazali, że w takim systemie znakowany aminokwas był „ładowany” na małą cząsteczkę RNA, a następnie wchodził w łańcuch polipeptydowy. Już w roku 1954 Crick przewidział, że w systemie syntezy białka musi istnieć drobna cząsteczką RNA (*adaptor*), która wchodzi w interakcję z tripletem (kodonem) matrycy (DNA?, RNA?), a drugi region adaptoru wiąże aminokwas i dostarcza go do powstającego polipeptydu. Adaptor miałby pełnić rolę „reduktora”, tłumaczącego trzy sąsiadujące nukleotydy pasma DNA na jeden aminokwas w łańcuchu białka. Musi istnieć, dowodził Crick, przynajmniej 20 takich rodzajów małych cząsteczek RNA – jeden dla każdego aminokwasu. Istotnie, kilka lat później odkryto klasę cząsteczek transportującego (transportującego aminokwas) RNA (tRNA). Okazało się jednak, że rozpoznają one triplety nie DNA, lecz wówczas jeszcze nieznanego informacyjnego RNA – cząsteczki pośredniczącej między matrycą DNA a ośrodkami syntezy białka.

W 1957 roku Crick i Leslie Orgel zaproponowali koncepcję „ramki odczytu” (*reading frame*) jako teoretycznego rozwiązania kodu. Trzy sąsiadujące w paśmie DNA nukleotydy, czyli triplet kodują dany aminokwas, ale gdy ramka odczytu przesunie się o jeden nukleotyd, to wszystkie następujące triplety tracą swój sens kodujący. Crick na posiedzeniu British Society of Experimental Biology przedstawił pogląd, że końcowa konformacja przestrzenna cząsteczki białkowej jest zdeterminowana sekwencją amino-

kwasów i zachodzi spontanicznie, a także sugerował, że swoistość kwasów nukleinowych polega na sekwencji zasad (nukleotydów) i ogłosił wstępną wersję swego słynnego „dogmatu głównego” (*central dogma*), wskazując jedynie, że część „RNA musi być pod kontrolą DNA”. Ostateczna wersja głównego dogmatu, że przepływ informacji biegnie od DNA, przez RNA, do białka, nie mogła być jeszcze sprecyzowana, bo w tym czasie nie została jeszcze odkryta klasa informacyjnego RNA (messenger RNA – mRNA).

„Jest teraz – czytamy w pracy Cricka i in. (Crick i in., 1961) – dużo pośrednich dowodów, które sugerują, że sekwencja aminokwasów wzdłuż łańcucha polipeptydu białka jest określona przez sekwencję zasad wzdłuż pewnej szczególnej części kwasu nukleinowego w materiale genetycznym”.

Informacyjny RNA (mRNA) został odkryty w 1960 r. przez Artura Pardee, François Jacoba i Jaques’a Monoda, którzy na podstawie badań własnych nad indukowalnym enzymem beta-galaktozydazą i jej genem u bakterii postulowali istnienie krótko żyjącego RNA. Taka labilna cząsteczka RNA miałaby pośredniczyć w przenoszeniu informacji między genami (DNA) a mikrosomami jako ośrodkami syntezy białka (mikrosomy określono później jako rybosomy). Do zbliżonych wniosków prowadziły też badania Heinrich Matthaei i Marshalla Nirenberga w National Institutes of Health (NIH), jak też wyniki badań François Grosa.

Tuż przed końcem roku 1961, w „Nature” Crick i in. (zob. Taylor, 1965) podali ogólne cechy uniwersalnego kodu:

- a) grupa trzech zasad [nukleotydów] koduje jeden aminokwas,
- b) kod nie ma charakteru kodu „nakładającego”,
- c) sekwencja zasad jest prawdopodobnie odczytywana od określonego punktu startowego. Nie ma specjalnych „przecinków” wskazujących jak wybierać prawidłowy triplet,
- d) kod jest prawdopodobnie zdegenerowany; to znaczy, że jeden szczególny aminokwas może być kodowany przez kilka tripletów.

W tym czasie ukształtowały się również elementy centralnego dogmatu Cricka, oparte na następujących stwierdzeniach: 1) cała genetyczna (dziedziczna) informacja jest przechowywana w kwasach nukleinowych, 2) struktura podwójnej helisy DNA wyjaśnia, jak informacja jest przechowywana i kopiowana, 3) informacja jest przechowywana jako kod digitalny (literowy), 4) informacja przepływa nieodwracalnie od kwasów nukleinowych do białek. Ostatnie stwierdzenie zostało spopularyzowane jako powiedzenie, które wyszło z pracowni Cricka i Watsona: „DNA robi RNA, RNA robi białko, białko robi nas”.

Mimo że w omawianym okresie zebrano dużo dowodów, że „kod genetyczny” był zapisany w DNA, a cząsteczką pośredniczącą w syntezie białka był „messenger” RNA (mRNA), molekularny mechanizm przeniesienia „kodu” z DNA na białko i natura kodu nie były do końca jasne. Wyniki dziesiątków prac wskazywały na powiązanie DNA z syn-

teżą białka i udział w tym procesie RNA. Oprócz mRNA i tRNA zidentyfikowano także rybosomalny RNA (rRNA), stwierdzono, że wszystkie klasy RNA powstają w jądrze oraz że są one komplementarne do sekwencji DNA. Dowodem, że synteza RNA zależy od DNA, były m.in. obserwacje nad hamowaniem syntezy RNA w różnych komórkach po zablokowaniu matrycy DNA aktynowycyną D. Opracowanie modeli syntezy białka w systemach bezkomórkowych pozwoliło na rozwój koncepcji dotyczących funkcji w tym procesie różnych klas RNA.

Stwierdzono, że w bakteriach zainfekowanych fagiem ma miejsce szybka synteza RNA i że ten RNA bierze następnie udział w syntezie fagowych białek oraz tworzy hybrydową cząsteczkę – fagowy DNA i „nowy” RNA. Wykrycie cząsteczek hybrydowych DNA:RNA, a także oddzielenie dwóch pasm DNA (denaturacja) i ponowne odtworzenie struktury dwupasmowej (renaturacja) wskazywały na ścisłe powiązania strukturalne między tymi cząsteczkami. Jak wspominałem, przegląd tych prac znajdzie Czytelnik w zbiorze pod red. J.H. Taylora (zob. Taylor, 1965).

W 1961 roku, na 5. Światowym Kongresie Biochemicznym w Moskwie, pracujący w National Institutes of Health (NIH) niemiecki biolog Heinrich Matthaei i Marshall Nirenberg przedstawili doniesienie o „złamaniu kodu genetycznego” (Nirenberg i Matthaei, 1961). System syntezy białka opracowany przez tych badaczy zawierał mieszaninę podstawowych aminokwasów (wśród nich, w kolejnych doświadczeniach, jeden aminokwas był znakowany izotopem radioaktywnym), ATP, drobnocząsteczkowy RNA, ekstrakt bakteryjny i syntetyczną matrycę, ale typu RNA – polimer złożony z samych uracyli, U-U-U-U- ... czyli poliU. Inkubacja takiej mieszaniny, a następnie analiza zsyntetyzowanego białka wykazała, że matryca poliU kodowała monotonną cząsteczkę polipeptydu („białka”), złożoną wyłącznie z jednego rodzaju aminokwasu – fenyloalaniny (akronim: Phe). Łańcuch polipeptydowy miał zatem budowę: Phe-Phe-Phe...

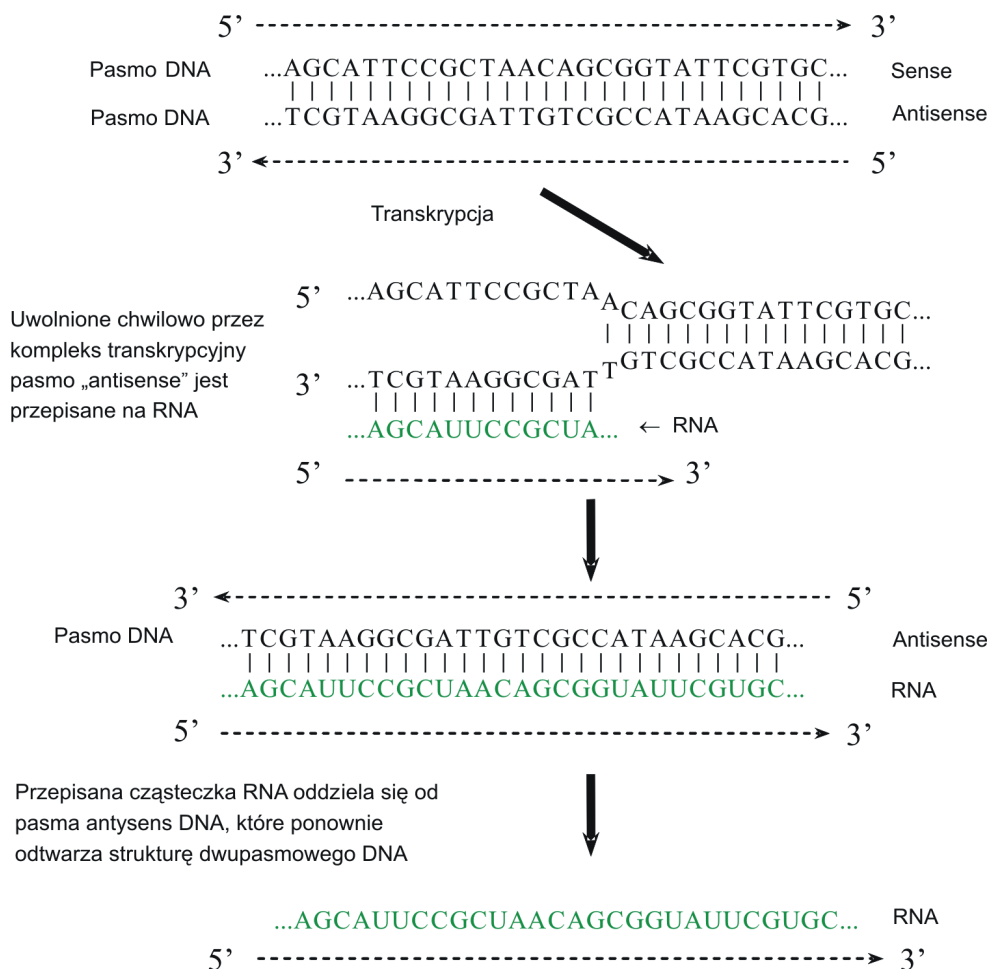
Wkrótce uzyskano syntetyczne polinukleotydy (syntetyczne RNA) o znanej i zaprogramowanej sekwencji nukleotydów (G. Khorana, M. Grunberg-Manago i S. Ochoa), które użyte w systemie Matthaei i Nirenberga pozwoliły w ciągu kilku lat poznać wszystkie kodujące triplety, czyli „złamać” kod genetyczny. Dla przykładu: triplet UUU koduje fenyloalaninę, GCU – alaninę, AGC – serynę, CUG – leucynę itd. Kod genetyczny opisujący triplety (kodony) RNA i odpowiadające im aminokwasy oraz kodony sygnałne „stop” znajdzie Czytelnik w każdym podręczniku biologii.

Gen klasyczny – elementarz

Ówczesna, oparta na hipotezie Cricka, definicja genu głosiła, że gen jest „odcinkiem DNA” kodującym białko (polipeptyd). Dwa główne procesy są niezbędne dla odczytania informacji zakodowanej w genie (DNA) i wyrażenia jej w postaci syntezy swojego białka: transkrypcja i translacja. Proces transkrypcji, czyli przepisanie kodujących sek-

wencji nukleotydowych DNA na liniową, prekursorową cząsteczkę informacyjnego RNA (mRNA, *messenger RNA*), zaczyna się w określonym punkcie sekwencji genu zwanej promotorem. Promotor zwykle poprzedza początek genu (jest „na lewo” od początku albo inaczej: mieści się „w górę” od początku genu). Na sekwencjach promotora usadawia się kompleks transkrypcyjny przesuający się od początku do końca genu i rozrywający chwilowo wiązania w helisie DNA. W wyniku transkrypcji powstaje pojedyncza nić mRNA, która po ukończeniu transkrypcji na końcu genu oddziela się od pasma DNA. Oba pasma DNA ponownie tworzą dwupasmową helisę. Każdy gen może dać początek tysiącom cząsteczek komplementarnego mRNA. „Informacja genetyczna” w postaci sekwencji tripletów nukleotydów (kodonów) zawarta jest na tzw. paśmie sensowym (*sense*) DNA. Informacyjny RNA (mRNA) kopiowany jest z pasma DNA komplementarnego do pasma *sense*, czyli pasma antysensowego (*antisense*). W wyniku transkrypcji pasma antysensowego w cząsteczce mRNA odtworzona zostaje sekwencja pasma kodującego, czyli pasma sensowego (por. ryc. 3), ze znanym nam już wyjątkiem: w miejscu gdzie w paśmie kodującym DNA znajduje się T (tymina) w mRNA pojawia się U (uracyl). Transkrypcja i pojawianie się cząsteczek mRNA w komórce jest wyrazem „aktywacji” genu lub jego „ekspresji”. Przyjęto, że sekwencje kodujące białko będą przedstawiane jako zapis na mRNA, począwszy od lewej strony (koniec 5' cząsteczki mRNA), od kodonu AUG. Kodon AUG jest pierwszym tripletem, od którego zaczyna się synteza białka.

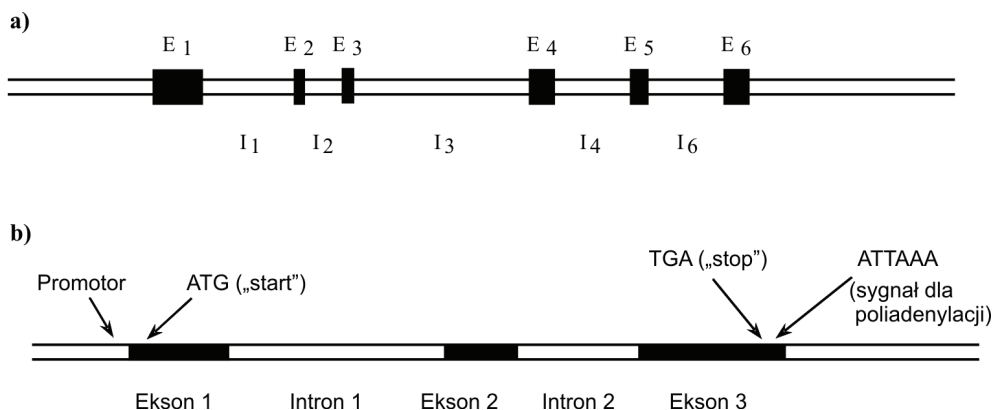
Cząsteczka mRNA, oprócz sekwencji tripletów (kodonów) kodujących poszczególne aminowasy w polipeptydzie, niesie także u obu końców cząsteczki dodatkowe sekwencje (nieulegające translacji) oraz inne specyficzne sekwencje, jak zlokalizowany na początku cząsteczki (5') sygnał (kodon) AUG dla startu odczytywania (translacji) i sygnały „stop” (kodony UAA, UAG i UGA) dla zakończenia odczytu (u końca 3') i inne. Sekwencje te służą jako swoiste miejsca dla lokalizacji białek i ich kompleksów niezbędnych dla procesu odczytu, a także jako sekwencje biorące udział w procesach kontrolnych związanych np. z trwałością cząsteczki mRNA. Odczyt, tzn. translacja informacji zawartej w mRNA na łańcuch białkowy, czyli synteza białka, zachodzi na rybosomach w cytoplazmie komórki z dostępnych tam wolnych aminokwasów i przy użyciu cząsteczek tRNA – „adaptorów” postulowanych przez Cricka i bardzo złożonych systemów „pomocniczych”. Każda cząsteczka mRNA może służyć do syntezy tysięcy cząsteczek białka, które koduje. Jednak żywotność cząsteczki mRNA jest regulowana. Powyższy opis ilustruje „ekspresję” genu i dotyczy w zasadzie genu bakteryjnego, który występuje jako ciągła sekwencja DNA. Sądzono, że w komórkach jądrazastych (eukariotycznych) struktura genu jest podobna do struktury w komórkach bakteryjnych. Ale w końcu lat 70. zaczęły pojawiać się zdumiewające doniesienia, że u niektórych wirusów i w komórkach eukariotycznych organizmów wyższych geny kodujące białko z reguły mają sekwencje kodujące nie w formie ciągłej, lecz są poprzerywane sekwencjami niekodującymi.



Ryc. 3. Schemat transkrypcji. Przed rozpoczęciem transkrypcji pasma DNA zostają chwilowo odsunięte od siebie przez kompleks transkrypcyjny. „Przepisanie” sekwencji z DNA na RNA zachodzi na paśmie antysensowym, komplementarnym do pasma sensowego. Zauważmy, że sekwencja w pre-mRNA oraz polarność są identyczne jak w paśmie sensowym DNA (czyli „genie”), z tym że tyminę (T) zastąpił uracyl (U). Zatem sekwencje genowe możemy odczytywać zarówno na paśmie sensowym DNA, jak i na paśmie RNA począwszy od 5’

Walter Gilbert zaproponował, aby fragmenty kodujące (znajdujące się następnie w „dojrzałym” mRNA) nazywać eksonami (E), zaś przedzielające je sekwencje niekodujące intronami (I). Geny o strukturze intron-ekson-intron-ekson... Gilbert nazwał genami mozaikowymi. Nazywa się je także genami „podzielonymi” – *split genes* (Pierre Chambon i in.). Graficznie strukturę takich genów przedstawiamy jako podwójną linię, a kodujące sekwencje genowe (eksony, E) jako pionowe krótkie linie, o różnej grubości, co ma ilus-

trować ich względną długość (ryc. 4a), a także i względne proporcje między eksonami (E) i intronami (I). Innym, wygodnym przedstawieniem graficznym jest pokazanie eksonów (E) jako odcinków wypełnionych na podwójnej linii DNA (ryc. 4b).



Ryc. 4. Powszechnie używane schematy ilustrujące budowę genu mozaikowego: a) gen o 6 eksonach; pokazano względną długość eksonów (E) i intronów (I); b) gen o trzech eksonach; zaznaczone są podstawowe sekwencje sygnałowe. Opis w tekście. Sygnały „start” i „stop” pokazują miejsca początku i końca przepisywania „informacji” na sekwencję aminokwasów białka

Zatem w komórkach eukariotycznych (jądrzastych) procesy związane z ekspresją genu wyglądają inaczej. Skopiowana z genu wielka cząsteczka zwana pre-mRNA zawierająca zarówno kopie eksonów, jak i intronów podlega procesowi „składania” (*splicing*). Z cząsteczki pre-mRNA zostają usunięte introny, a eksony zostają połączone ze sobą w ciągłą sekwencję stanowiącą teraz ostateczną, dojrzałą (znacznie mniejszą niż pre-mRNA) cząsteczkę informacyjnego RNA (mRNA). Cząsteczka mRNA jest teraz gotowa podjąć funkcję kodowania białka w cytoplazmie komórki.

Liczba eksonów/intronów jest różna (od kilku do kilkudziesięciu) w różnych genach i u różnych gatunków. Największa znana mi z literatury liczba eksonów występuje u muszki owocowej *D. melanogaster* w genie o symbolu *Dsam* i wynosi 115. Eksony te są rozrzucone na wielkim, sięgającym milionów par zasad fragmencie DNA. Ludzki homolog tego genu posiada ponad 100 eksonów podobnie rozsianych na dużym obszarze w znacznych odległościach od siebie.

Introny we wszystkich genach są z reguły znacznie dłuższe od eksonów i ich łączne sekwencje zajmują często ponad 95% sekwencji pierwotnego transkryptu (por. ryc. 4a i 5a). Jeszcze kilkanaście lat temu introny były uważane za „śmieciowego” DNA (*junk* DNA, por. niżej). Tymczasem okazało się, że w samych intronach mogą się znajdować sekwencje kodujące należące do innego genu, a także sekwencje regulatorowe, np. tzw.

sekwencje wzmacniające transkrypcję (enhancery), często leżące w znacznej odległości „w dół” (czyli „na prawo”) od początku genu. Sądzono, że po wycięciu z pre-mRNA introny jako transkrypt „śmieciowy” ulegają kompletnej degradacji jako sekwencje niepotrzebne. Obecnie wiemy, że introny będące głównym składnikiem w pierwotnym transkrypcie (pre-mRNA) wraz z transkryptami intergenowych sekwencji unikatowych niekodujących białka stanowią ważny materiał użyteczny w regulacji wielu procesów biologicznych. Ta porcja sekwencji, która znajduje się w intronach, najlepiej koreluje ze złożonością rozwoju ewolucyjnego (Mattick i Gagen, 2001, także por. niżej).

Następująca definicja dobrze opisuje gen „klasyczny”, taki jaki znaliśmy w końcu ubiegłego stulecia: „Gen jest segmentem DNA wciągniętym w produkowanie łańcucha polipeptydowego; gen obejmuje regiony poprzedzające i następujące po regionie kodującym, jak również sekwencje wtrącone (introny), leżące między indywidualnymi segmentami kodującymi (eksonami)” (zob. Lewin, 1994). Gen „klasyczny” miał swoją ściśle zdefiniowaną liniową strukturę, swój początek i koniec. Geny miały być ułożone jeden za drugim jako oddzielne jednostki funkcjonalne, jak paciorki różańca. Obecnie wiemy, że geny mogą nakładać się (*overlapping*) i wówczas z pasma *sense* odczyt zachodzi z przesunięciem ramki odczytu, w wyniku czego na jednym paśmie *sense* mogą być zakodowane dwa polipeptydy. Także pasmo *antisense*, do niedawna uważane za niekodujące, może służyć do kodowania i jest wykorzystane do transkrypcji.

Genom – ile DNA, ile genów?

DNA komórek prokariotycznych (beźjadrzastych, np. bakterie) jest zorganizowany jako jeden chromosom składający się z jednej cząsteczki DNA „rozpuszczonej” w cytoplazmie. Cała taka cząsteczka DNA obejmuje geny ułożone blisko jeden za drugim, każdy z nich ma swój wymiar przestrzenny, dość wyraźne „granice” (początek i koniec) i każdy z nich ma nieprzerwany ciąg sekwencji kodujących.

Natomiast DNA organizmów eukariotycznych (posiadających jądro), zarówno u jednokomórkowych i wielokomórkowych, upakowany jest w chromosomach i zamknięty w jądrze komórkowym. Jeden chromosom zawiera jedną cząsteczkę DNA. Człowiek posiada 22 chromosomy autosomalne oznaczone według malejącej długości od 1 do 22 i dwa różne chromosomy płciowe X i Y (u kobiet XX, u mężczyzn XY). Uporządkowany według wielkości zestaw chromosomów nazywamy kariotypem. W jądrze komórek człowieka poszczególne chromosomy zajmują ściśle określone terytoria. Nawet w okresie cyklu życiowego komórki zwanym interfazą, gdy ma miejsce duża aktywność genów i gdy chromosomy wydają się rozproszone, a DNA chromosomalny jest pozornie rozpuszczony i wymieszany, szkielet każdego indywidualnego chromosomu jest utrzymany, a jego DNA „nie miesza” się z DNA innego chromosomu. DNA każdego chromosomu zajmuje określone terytorium jądra.

W organizmach diploidalnych, takich jak człowiek, każdy chromosom występuje w dwóch kopiach – jedna pochodzi od matki i jedna od ojca. Genomem nazywamy całkowitą „informację genetyczną” czyli DNA, (choć nie są to terminy równoważne) zawartą w komórce i charakterystyczną dla danego organizmu. Wielkość genomu najczęściej podajemy jako długość wyrażoną liczbą par zasad (nukleotydów) i odnosimy zwykle do haploidalnego zestawu chromosomów. „Typowy” genom człowieka, czyli sumaryczna długość cząsteczek DNA wszystkich chromosomów w komórce haploidalnej (zawierającej zestaw jednej kopii chromosomów danego gatunku) ma wielkość ok. 3×10^9 par zasad/nukleotydów (3×10^9 bp), czyli 3 mld par zasad. W przeliczeniu na miarę metryczną DNA każdej komórki człowieka ma długość ok. 1,5 metra! Ta wielkość daje nam przybliżone pojęcie o tym, jak silnie skondensowany i gęsto upakowany musi być DNA w indywidualnym chromosomie i w jądrze komórkowym, które – jeśli przyjmiemy jego kulisty kształt – ma średnicę zaledwie od kilku do kilkunastu μm .

Wielkość genomów różnych gatunków mieści się w szerokich granicach, od ok. 10^6 bp u mykoplazmy i bakterii aż do 10^{11} bp dla kwitnących roślin i zimnokrwistych kręgowców klasy *Amphibia*. Czyli w haploidalnych genomach różnych gatunków ilość DNA waha się do 100 000 razy. Wielkość genomu bakteriofagów jest rzędu 10^5 bp. Jeszcze ok. 20 lat temu spekulowano, że złożoność organizmu będzie ściśle korelować z ilością materiału genetycznego (ilością DNA), a zatem również z liczbą genów i liczbą chromosomów. W miarę gromadzenia danych doświadczalnych i rozwijania metod badawczych okazało się, że sprawa jest bardziej złożona. Nie ma prostej korelacji między biologiczną złożonością gatunków a ilością DNA ani liczbą chromosomów, ani ilością kodującego lub niekodującego DNA.

U bakterii sekwencje DNA kodujące białko (czyli geny) zajmują cały lub prawie cały genom. Ale w organizmach wyższych (ssaki) sekwencje kodujące białko stanowią tylko małą część genomu, np. u człowieka najwyżej ok. 1,5% genomu. Jeśli przyjęliśmy, że liniowa długość DNA genomu człowieka wynosi ok. 1,5 m, to łączna długość sekwencji kodujących (tj. genów) zajmuje zaledwie ok. 2 cm. U niektórych kręgowców klasy *Amphibia*, mających genom wielkich rozmiarów, ta proporcja jest jeszcze mniejsza. W genomie człowieka i innych gatunków zwierząt oraz roślin jest znacznie więcej DNA, niż byśmy mogli oczekiwać z oczywistych różnic w złożoności tych organizmów. Przez wiele lat uznawano niesłusznie (por. niżej), że ten „nadmiarowy” DNA u człowieka i innych organizmów wyższych nie ma znaczenia funkcjonalnego i jest wynikiem procesów ewolucyjnych, a nawet ukuto nazwę dla „nadmiarowego” DNA – ewolucyjne „śmiecie” (*junk* DNA). Organizmy prokariotyczne takich „nadmiarowych” sekwencji DNA nie posiadają.

W latach 90. badaczy intrygowało głównie zagadnienie, jak dużo jest genów u każdego gatunku. Sądzone, że im gatunek jest na wyższym stopniu rozwoju ewolucyjnego, tym więcej będzie posiadał DNA i genów. Jeszcze kilkanaście lat temu szacowano, że

człowiek powinien posiadać przynajmniej 100 000 genów, pojmowanych nadal jako kodujące białko sekwencje DNA mające swój materialny początek i koniec. Szacunki bardzo różniły się w zależności od użytej metody. Analiza sekwencji DNA organizmów posiadających mały genom i poznanie sekwencji aminokwasowej różnych białek pozwalało „uśredniać” wielkość genu i uściślać liczbę genów. Geny kodujące rybosomalną RNA (rRNA), transportującą RNA (tRNA), a także geny kodujące histony występują zwykle w wielu kopiach. Często nazywa się je „genami powtarzalnymi”. Należy je odróżnić od „sekwencji powtarzalnych” lub „powtarzających się” (sekwencje repetytywne). Liczne kopie jednego genu występują czasami w komórkach niektórych nowotworów.

W latach 90. w Stanach Zjednoczonych opracowano i rozpoczęto projekt badawczy zmierzający do ustalenia sekwencji (czyli kolejności występowania w DNA) 3 mld par zasad/nukleotydów DNA całego genomu człowieka (Human Genome Project, HGP). Ustanowienie projektu HGP przez USA (Narodowe Instytuty Zdrowia, NIH oraz Departament Energii, DOE) i międzynarodowej organizacji Human Genome Organization (HUGO) było poprzedzone usilnymi zabiegami naukowców, takich jak Walter Gilbert, James Watson, Charles Cantor, Leroy Hood (cyt. za Lewontin, 1991). Wierzono, że HGP pozwoli ustalić liczbę genów i zrozumieć, na czym ostatecznie polega zjawisko życia, ustalić cały, liniowo zakodowany „plan organizmu” z wszelkimi anatomicznymi i fizjologicznymi cechami i zachowaniem. W. Gilbert obrazowo ujął swój entuzjazm wobec HGP. Sekwencję trzech miliardów par zasad DNA człowieka będzie można zapisać na dysku CD, wyciągnąć taki dysk CD z kieszeni i oznajmić: „Oto jest istota ludzka: to Ja!”.

Projekt wzbudzał wielkie nadzieje. Sądono, że poznanie sekwencji pozwoli na ustalenie liczby genów oraz genów i sekwencji unikatowych dla człowieka, czyli na ustalenie, czym człowiek różni się ewolucyjnie od zwierząt wyższych i najbliższych mu naczelnych (np. szympansa). Oczekiwano dalej, że zostaną poznane defekty i uszkodzenia genów determinujące powstawanie chorób (łącznie z rakiem), co z kolei pozwoli na rozwinięcie metod diagnostycznych i opracowanie nowych skutecznych technik leczniczych, np. „terapia genowa” (Zalewski, 2001). Każdy człowiek będzie mógł mieć swój „paszport genetyczny” dla identyfikacji, prosty i niezbędny np. w sprawach ubezpieczeniowych. Powstały też złudne iluzje, że znając zestaw genów i sekwencji DNA człowieka, będzie można ten gatunek uszlachetniać przez zabiegi inżynierii genetycznej, a nawet zwalczyć negatywne ludzkie zachowania.

W roku 2001 dwa ośrodki badawcze w USA, rządowy (NIH) pod kierunkiem F. Collinsa i prywatny (Celera) J. C. Ventera, ogłosiły równocześnie w tygodnikach „Nature” (Lander i in., 2001) oraz „Science” (Venter i in., 2001) sekwencję ludzkiego DNA i pokazały „mapę genetyczną” człowieka. Był to wielki sukces badawczy, który jednocześnie okazał się wielkim zaskoczeniem dla nauki. Wbrew wcześniejszym oczekiwaniom, liczbę „genów” człowieka oszacowano w przedziale od 25 000 do 30 000. Liczba ta obejmowa-

ła sekwencje posiadające atrybuty genu strukturalnego, kodującego białko, jak również geny kodujące te klasy RNA, które były ostatecznymi produktami (rRNA, tRNA). Okazało się, że trudno jest zdefiniować i rozpoznać indywidualne geny. Rozrzut w szacunkowej liczbie genów zależy od metod przyjętych do identyfikacji opartych na swoistych sekwencjach (sekwencje sygnałne charakterystyczne dla genu, jak np. promotor, sekwencje stop, otwarte ramki odczytu).

Inne grupy badawcze zaczęły wkrótce donosić o ustaleniu sekwencji i szacunkowych liczbach genów u kilku gatunków zwierząt i roślin. Porównanie liczby genów (sekwencji kodujących białko) między gatunkami wyraźnie wskazało, że nie ma także oczekiwanej korelacji między liczbą genów a złożonością organizmów. Obecnie liczbę genów kodujących białko w genomie człowieka szacuje się na ok. 22 700, a nawet zaledwie 20 000 (cyt. wg Pheasant i Mattick, 2007). Człowiek, który umieścił siebie na szczycie drzewa ewolucyjnego, ma porównywalną liczbę genów do myszy (22 500), jeżowca (ok. 23 000) i nicienia *Caenorhabditis elegans* (od 19 000 do 20 000) i tylko o jedną trzecią więcej niż muszka owocowa *Drosophila melanogaster* (ok. 14 000), a znacznie niższą niż jednokomórkowy protista, orzęsek *Tetrahymena thermophila* – 27 000 genów (cyt. wg Pheasant i Mattick, 2007). Te szacunkowe dane wskazują, że w procesie rozwoju osobniczego geny kodujące białko *per se* nie są odpowiedzialne za „programy rozwojowe”. Program (plan) rozwoju jest przypisywany genomowym mechanizmom regulacyjnym lub innym nieznanym jeszcze procesom. Drożdże posiadają ok. 6000 genów, a bakterie między 5000 a 10 000 genów. Sekwencje genomu człowieka i szympansa różnią się zaledwie niewiele ponad 1,0%. Czyżby – jak chcą niektórzy badacze – człowiek i człowieczeństwo, cechy osobowości człowieka, jego wielki ładunek intelektualny tkwił w tej drobnie DNA, stanowiącej różnicę z najbliższym krewniakiem ewolucyjnym?

Pozagenowe sekwencje DNA występują u wielu gatunków i różne opisane niżej sekwencje posiadają często zdumiewające podobieństwo, a poprzez ich porównywanie można prześledzić ich koleje ewolucyjne (Jurka i in., 2007).

Co zawiera ludzki, nadmiarowy DNA uważany do niedawna za ewolucyjne „śmiecie”? Jak wspomnieliśmy, mniej niż 2% ludzkiego DNA zawiera sekwencje kodujące, zatem nadmiarowa część DNA zajmuje aż ok. 98% genomu. Na tę część składają się sekwencje powtarzające się (*repetitive sequences, repetitive DNA*), występujące w licznych kopiach, oraz fragmenty sekwencji unikatowych o nieznannej funkcji, nieposiadające swoich homologów w genomie. Sekwencje powtarzające się o prostej budowie, tandemowe, złożone z kilku (2-6) nukleotydów powtarzających się zwykle kilkadziesiąt razy, np. dinukleotydy – [AT]_n dające trakty ATATATA..., trinukleotydy np. [AAT]_n lub [CGG]_n tworzące odpowiednio ciągi: AATAATAAT... i CGGCGGCGGCGG..., nazywane są mikrosatelitami. Bardziej złożone sekwencje powtarzające się, mające motyw podstawowy od kilkunastu do kilkudziesięciu nukleotydów, definiowane są jako minisatelity. Zajmują

one odcinki DNA o długości nawet kilkunastu tysięcy pz. Niekiedy „inwazja” krótkich (kilka nukleotydów) traktów w rejony kodujące ma miejsce w chorobach degeneracyjnych, głównie typu neurodegeneracyjnego. Do grupy powtarzających się sekwencji rozproszonych zaliczamy także sekwencje typu SINE (*short interspersed nucleotide elements*), z których najlepiej poznana sekwencja *Alu* o długości ok. 300 nukleotydów występuje w ok. 500 tys. kopii w genomie człowieka. Grupa sekwencji LINE (*long interspersed nucleotide elements*), o długości kilku tysięcy nukleotydów, występuje w setkach tysięcy kopii. Znanych jest wiele innych rzadziej występujących elementów.

Wiele z sekwencji powtarzających się ma zdolność przemieszczania (transpozycji) w genomie. Grupy rodzin takich sekwencji nazywamy ogólnie sekwencjami/elementami transpozonalnymi (TE – *transposable elements*) lub wprost transpozonomi. Istnienie elementów transpozonalnych, zwanych również elementami mobilnymi, postulowała w latach 30. Barbara McClintock na podstawie badań nad genetyką kukurydzy. Elementy transpozonalne są to sekwencje o dużej złożoności co do struktury, pochodzenia i różnorodności. Ich funkcja w procesach ewolucji, specjacji (powstawaniu gatunków), w rozwoju osobniczym i rola w genomie współczesnym nie jest do końca poznana. Transpozony odegrały prawdopodobnie główną rolę w procesie ewolucji (Jurka i in., 2007; Kapitonov i in., 2004). W warunkach stresowych transpozony mogą się uaktywniać, i dzięki zdolnościom autokatalitycznym (aktywność własnego enzymu, transpozazy, umożliwiającym wycinanie i insercję) mogą zmieniać miejsce w genomie tego samego osobnika lub gatunku bądź ulegać transpozycji horyzontalnej do genomu innego gatunku. Masowy, horyzontalny współczesny transfer transpozonów opisano u wrotki (Gladyshev i in., 2008). Znane są przykłady chorób powstałych wskutek transpozycji aktywnych TE. Z kodujących sekwencji transpozonów (transpozaza, gen *gag* u retrotranspozonów) ewoluowały niektóre geny kodujące białko u ssaków. Transpozony ingerowały także w powstanie i funkcję sekwencji kodujących microRNA. W „pozagenowym” DNA znajdują się liczne powtarzające się kopie genów nieczynnych (pseudogeny, por. niżej).

Między genem a białkiem

Obraz prostego przekazywania informacji z genu (DNA) przez mRNA do białka i prostej liniowej relacji gen (DNA), struktura białka i jego funkcja bardzo komplikują poznanie mechanizmów zmieniających na poziomie mRNA treść informacji pierwotnej zawartej w DNA (genie). W wyniku działania tych mechanizmów sekwencja aminokwasów w cząsteczce białka (a zatem jej struktura i funkcja) będą różnić się od tej „zapisanej” w genie.

Jedna z hipotez wyjaśniających ewolucję głosiła, że wprowadzenie do procesu ewolucyjnego „nowości” miałyby polegać na duplikacji genu. Jedna kopia genu służyć miała do kodowania „starego” białka, nowa kopia byłaby uwolniona dla ewolucyjnego „ekspe-

rymentowania” i ostatecznie posłużyłaby do wykształcenia nowych funkcji kodujących. Ta hipoteza nie wyjaśnia jednak faktu, że w miarę powstawania organizmów o coraz większej złożoności nie wzrasta proporcjonalnie liczba genów kodujących białko.

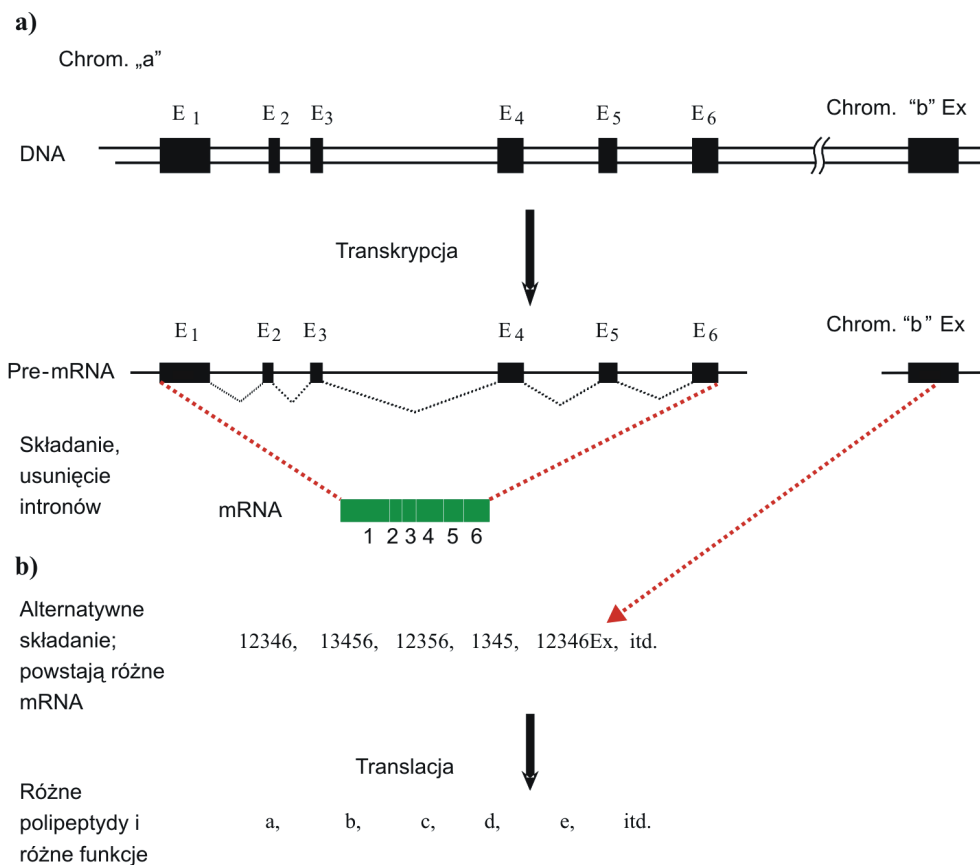
Dogmat: jeden gen – jedno białko – jedna funkcja, uległ obaleniu, gdy okazało się, że liczba rodzajów cząsteczek białek jest przynajmniej o jeden rząd większa niż liczba genów kodujących białko. W dodatku uległ również obaleniu pogląd, że przestrzenna trójwymiarowa konformacja cząsteczki białkowej jest determinowana wyłącznie sekwencją aminokwasów w polipeptydzie, co pierwotnie miał zdeterminować gen kodujący. Przestrzenna forma białka zmienia się w zależności od mikrośrodowiska, np. oddziaływań danej cząsteczki białkowej z innym białkiem, różnymi drobnymi cząsteczkami niebiałkowymi, jonami itd. Przestrzenna forma zależy też od chemicznych potranslacyjnych modyfikacji białek. Takie modyfikacje obejmują fosforylację, metylację, acetylację, sumoylację, ubikwitynację, glutamylację itp. Modyfikacje potranslacyjne zmieniają funkcję białka. Funkcjonalna różnorodność białek określana terminem „pleiotropia” zwiększa złożoność biologiczną białek.

W ostatnich dziesięcioleciach poznano mechanizmy (zapewne nie wszystkie), które wyjaśniają, jak można osiągnąć różnorodność strukturalną i funkcjonalną białek kodowanych przez jeden gen. Najważniejsze z nich zostaną przytoczone niżej.

Alternatywne składanie mRNA

Oprócz możliwości odtwarzania identycznych kopii genu (DNA) głównym zadaniem genu jest programowanie syntezy białka. Istnieje wielka różnorodność białek oraz konieczność ich wybiórczo kontrolowanej syntezy w czasie rozwoju embrionalnego, a potem prowadzenia syntezy w różnych komórkach, tkankach i narządach, a także w odpowiedzi na zmienne bodźce środowiskowe. Jednym z mechanizmów zapewniających generowanie z jednego genu (DNA) różnych polipeptydów o różnych funkcjach jest alternatywne składanie pre-mRNA.

Wielka cząsteczka pre-mRNA (zwana też jednostką transkrypcyjną) podlega wspomnianemu wyżej procesowi składania (*splicing*). W procesie tym zostają usunięte długie niekodujące sekwencje intronowe, a eksony zostają złączone w jeden ciąg nukleotydów. W wielu przypadkach składanie przebiega w sposób alternatywny (ryc. 5 i 13), generując z sekwencji jednego genu (jednego pre-mRNA) wiele różnych cząsteczek mRNA kodujących różne polipeptydy (białka). Dla przykładu (ryc. 5a), jeśli w pre-mRNA mamy trakt sześciu eksonów (E) oznaczonych kolejno 1 2 3 4 5 6, to w przypadku alternatywnego składania mogą powstać cząsteczki mRNA o ciągu E jak w pre-mRNA, lecz także, oprócz intronów, mogą zostać usunięte niektóre eksony i utworzone zostaną cząsteczki mRNA o ciągach E (ryc. 5b): 1 2 3 4 6 albo 1 3 4 5 6, albo 1 2 3 5 6, albo 1 3 4 5 itd. Takie cząsteczki RNA będą kodowały różne izofomy polipeptydu o różnych funkcjach.



Ryc. 5. Składanie mRNA. a) Zarówno eksony (E), jak i introny genu (DNA) na chromosomie „a” i niekiedy nawet część genu na chromosomie „b” są przepisane (transkrypcja) na długi transkrypt pierwotny (pre-mRNA). W procesie składania z pre-mRNA zostają wycięte sekwencje intronowe, a sekwencje eksonowe zostają spojone w jedną, „dojrzałą” cząsteczkę mRNA. b) Cyfry pokazują kilka różnych alternatywnych sekwencji eksonów w cząsteczce mRNA po alternatywnym składaniu, gdy oprócz intronów zostają wycięte również niektóre eksony. W pierwszej cząsteczce o ciągu 12346 został wycięty ekson E5, w drugim ciągu 13456 został wycięty ekson E2 itd. W ostatnim przykładzie o ciągu 12346x został wycięty ekson E5, ale dodatkowo został „doklejonny” ekson „Ex” z transkryptu odległego chromosomu „b”. Różne cząsteczki mRNA powstałe w wyniku alternatywnego składania, zostają przetłumaczone (translacja) na różne polipeptydy, które mogą mieć różne funkcje

Zupełnie nieoczekiwanym było odkrycie, że w procesie składania transkryptu (pre-mRNA) genu „a” może być pobrany ekson z transkryptu genu „b”, leżącego w dużej odległości na tym samym chromosomie lub nawet na innym chromosomie. Wówczas może powstać mRNA o ciągu eksonów 1 2 3 4 6, Ex (ryc. 5b). Po translacji aktywnie złożo-

nych mRNA powstają różne izoformy białek. To odkrycie burzy pojęcie genu jako ściśle określonej jednostki strukturalnej, mającej swoje zdefiniowane granice.

Skrajnym przykładem zdumiewających możliwości alternatywnego składania jest transkrypt genu *Dscam*. U człowieka gen *DSCAM* (*Down syndrome cell adhesion molecule*) koduje duże polipeptydy biorące udział w rozwoju układu nerwowego. Neurony w trakcie rozwoju osobniczego dostają etykietę w postaci swoistego spektrum polipeptydów, co pozwala im zająć odpowiednie miejsce w mózgowiu i połączyć się z odpowiednimi komórkami nerwowymi. Polipeptydy *DSCAM* biorą udział w złożonych kompleksach umożliwiających sterowany ruch eksonów komórki nerwowej w rozwoju osobniczym.

Dobrze poznany u muszki owocowej homologiczny gen *Dscam* posiada 115 eksonów. Potencjalnie alternatywne składanie pre-mRNA *Dscam* może generować 38 016 różnych cząsteczek mRNA, a zatem tyleż różnych izoform polipeptydów (Schmuckler i in., 2000). Dotychczas zdefiniowano kilka tysięcy takich izoform. Geny *Dscam* należą do rodziny genów immunoglobulinowych, które także wykazują wielką zdolność rekombinowania. U muszki owocowej polipeptydy *Dscam* stanowią także składową systemu odporności. W komórkach systemu nerwowego i gdzie indziej repertuar alternatywnych polipeptydów *Dscam* zmienia się w czasie i jest elementem pozwalającym odróżnić sąsiadujące komórki (np. w różnych typach fotoreceptora). Transkrypty genów kodujące inne rodziny białek wciągniętych w proces neurogenezy podlegają również alternatywnemu składaniu (Neves i in., 2004). Składanie alternatywne przejawia specyficzność narządową (Xu, Modrek i Lee, 2002), zmienia się w stadiach rozwojowych, a także posiada pewne cechy swoistości gatunkowej. Szacuje się, że proces alternatywnego składania obejmuje 35-59% genów człowieka. Proces regulowanego składania zachodzi w trakcie tak podstawowego zjawiska, jak determinacja płci u muszki owocowej (zob. Alberts i in., 1994). W wyniku składania alternatywnego „informacja pierwotna” genu zostaje przetworzona na „informację wtórną”. Alternatywne składanie pre-mRNA może obejmować także początek i koniec cząsteczki zawierającej sekwencje nieulegające translacji, ale wpływające na wydajność translacji, stabilność cząsteczki mRNA lub jej komórkową lokalizację.

Przypuszcza się, że w trakcie ewolucji zachodził proces generowania lub utraty eksonów. Proces ewolucyjny „eksperymentował” i poddawał próbie nowe funkcje alternatywnie złożonych sekwencji, utrzymując podstawowy, stary zapis. Zjawisko alternatywnego składania jest w odniesieniu do genów człowieka częste i komórka ludzka może generować wiele alternatywnych form białka o potencjalnie różnej strukturze i funkcji. Jednak mamy mało dowodów, że alternatywne składanie daje wiele nowych funkcji enzymatycznych i strukturalnych (Tress i in. 2007).

Niekiedy transkrypcja obejmuje dwa sąsiadujące geny i wówczas powstaje białko chimeryczne oraz jego liczne izoformy (Akiva i in., 2006).

Redagowanie mRNA

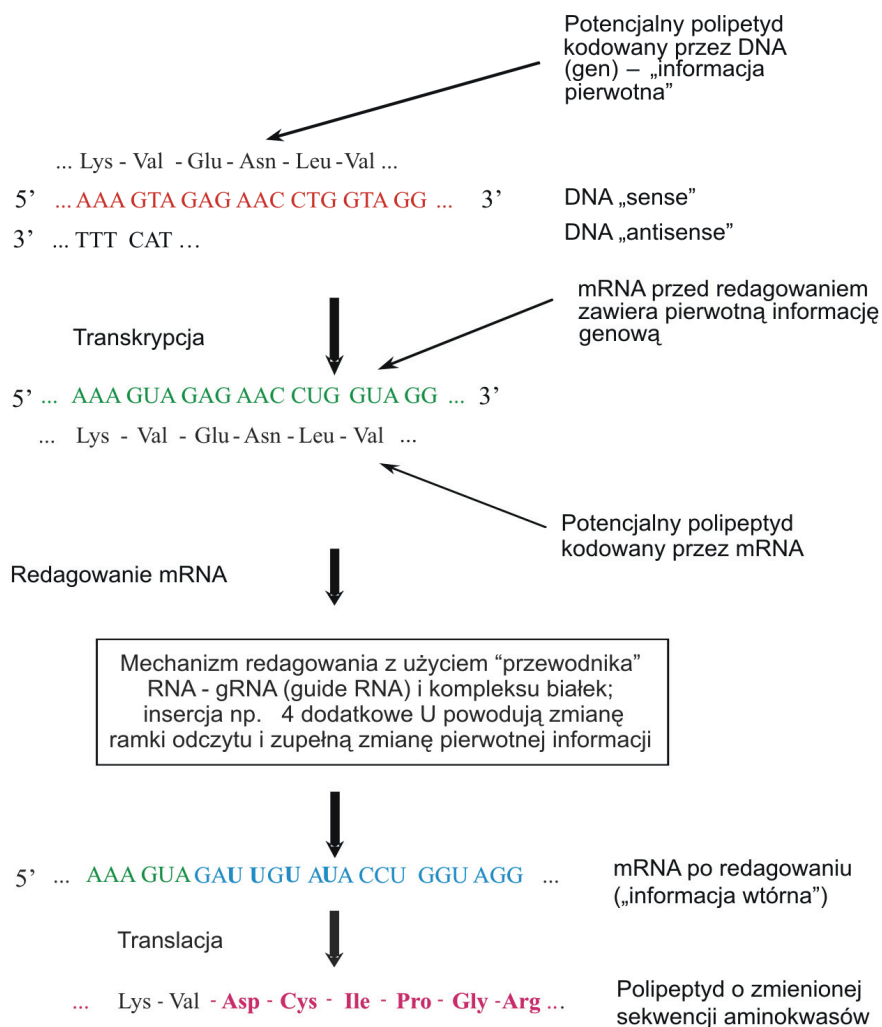
Proces redagowania RNA (*RNA editing*) jest złożony ze skomplikowanych mechanizmów biochemicznych zmieniających na poziomie mRNA treść pierwotnej informacji odczytanej z sekwencji nukleotydowych w DNA (Smith, Gott i Hanson, 1997). Proces ten poznano u bakterii, wirusów, pierwotniaków, grzybów, drożdży, roślin, insektów i ssaków. Proces dotyczy też samego DNA, ale jest tu raczej wyjątkiem (w limfocytach B). Redagowanie RNA obejmuje złożone mechanizmy insercji i/lub usuwania (delecji) cząsteczek urydyny (U) w sekwencji mRNA.

Prosty mechanizm redagowania polega na chemicznym przekształceniu w cząsteczce mRNA w określonych pozycjach kodonów C w U i A w I (I = inozyna, rzadki nukleotyd, gdy znajdzie się w kodonie, odczytywany jest jako G).

Redagowanie mRNA intensywnie zachodzi w mitochondriach *Trypanosoma* i w chloroplastach roślin. Zachodzi również u zwierząt, co umożliwia generowanie narządowych izoform białek. Redagowanie występuje też w syntezie niektórych białek związanych z funkcją kanałów wapniowych oraz neuroprzekaźników w mózgu człowieka. W tych ostatnich przypadkach często zachodzi podstawienie aminokwasu przez redagowanie mRNA z udziałem deaminaz adenozyne zmieniających A w I. Utrata lub deregulacja funkcji deaminaz adenozyne idzie często w parze z zaburzeniami neurologicznymi. Proces zmiany A w I spostrzegano także na granicy intron/ekson, co w efekcie skutkowało tworzeniem nowej formy mRNA.

Konsekwencje redagowania mRNA są głębokie. Dla przykładu, w komórkach wątroby człowieka gen *APOB* daje transkrypt mRNA kodujący białko apolipoproteinę B o długości 4563 aminokwasów. Taki sam mRNA (jako transkrypt genu *APOB*) powstaje w komórkach jelita, ale koduje małą cząsteczkę apolipoproteiny zawierającą jedynie 2152 aminokwasów. Jest to wynik redagowania mRNA w komórkach jelita, polegający na podstawieniu U w miejsce C w środkowej części cząsteczki mRNA. Podstawienie zachodzi w kodonie CAA (kodującym aminokwas glicynę), co przekształca CAA w kodon UAA, który jest sygnałem „stop” dla procesu translacji, w wyniku czego syntetyzuje się jedynie połowa cząsteczki białka. Proces ten zmienił pierwotną treść informacji w genie, co w konsekwencji dało dwie izoformy cząsteczki białkowej, różniące się wielkością i funkcją – apolipoproteina B wątrobowa jest niezbędna dla transportu cholesterolu, podczas gdy apolipoproteina jelitowa bierze udział w absorpcji lipidów.

W mitochondrialnych mRNA u pierwotniaków *Trypanosoma* podczas redagowania zachodzi intensywna insercja i delecja U przy użyciu cząsteczek „RNA-przewodnika” (*guide* RNA, gRNA) (Stuard i Panigrahi, 2002). Jest to wysoce złożony mechanizm, którego zasadę ilustruje ryc. 6. Nie jest jasne, dlaczego, jak się wydaje, intensywne redagowanie zachodzi głównie w mitochondriach.



Ryc. 6. Uproszczony schemat redagowania mRNA. W wyniku redagowania powstaje wtórna informacja w mRNA dająca po translacji polipeptyd odmienny od zakodowanego w genie. Insercję nowych U w czasie redagowania zaznaczono czcionką wytłuszczoną

Redagowanie zachodzi również w rRNA i tRNA. W przypadkach redagowania mRNA proces ten wywiera głęboki efekt na pierwotną strukturę białka, wprowadzając sygnał „stop” kończący translację wcześniej, niż to zakładała informacja pierwotna w genie (DNA), lub wprowadzając zmianę ramki odczytu. Podobnie głęboki wpływ na pierwotną strukturę białka mają zmiany w antykodonie tRNA.

Uniwersalność zjawiska redagowania pre-mRNA zapewnia możliwość generowania wielu nowych izoform białka i przypuszczalnie ma jeszcze wiele innych funkcji biolo-

gicznych. Dla przykładu, dzięki mechanizmom redagowania elementy powtarzające się *Alu* (por. wyżej), występujące w wielkich liczbach w genomie człowieka, podlegają głębokiej przebudowie. W wyniku takich procesów mogą powstawać nowe sekwencje typu eksonów, które po „wyjściu” z sekwencji *Alu* (eksonizacja) zasilają repertuar eksonów w genomie.

Redagowanie ma również miejsce w pri-microRNA przez mechanizm zamiany A na I i jest tkankowo swoiste. Takie redagowanie przenosi swoistość tłumienia ekspresji na inny zestaw genów (Kawakara i in., 2007).

Nie wiemy, co kieruje wysoce swoistym procesem redagowania RNA ani jak ten proces jest regulowany i koordynowany, jak zachodzi w procesie rozwoju i jaka jest jego swoistość. Zjawisko redagowania RNA, podobnie jak składanie mRNA, narusza nasze postrzeganie genu jako autonomicznego odcinka DNA, dziedzicznego i niezmiennego, jako jedyne źródła bezpośredniej instrukcji dla syntezy struktury pierwszorzędowej polipeptydu.

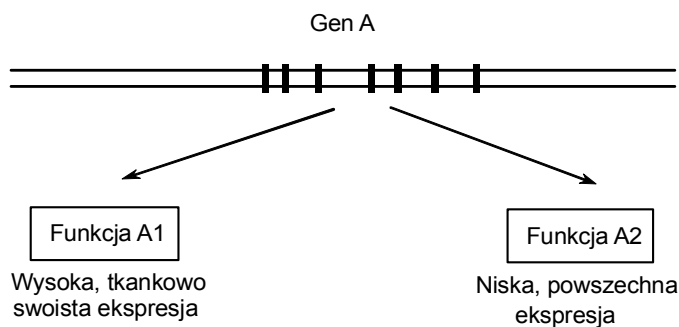
Gene sharing

Termin *gene sharing* (Piatigorsky, 2007) oznacza sytuację, gdy jeden i ten sam gen (DNA) jest używany wspólnie dla syntezy jednego polipeptydu, który w zależności od kontekstu (zapotrzebowania, stężenia polipeptydu w tkance) spełnia zupełnie różne funkcje molekularne. Termin *gene sharing* trudno jest przetłumaczyć na zwięzły termin w języku polskim. Może termin „dzielenie się genem” najbardziej dokładnie oddaje jego sens. W teorii ewolucji panowało przekonanie, że oprócz „korzystnych” mutacji, które wytrzymywały presję selekcyjną, podstawowym mechanizmem wprowadzającym „nowości” ewolucyjne (tj. nowe białka o nowych funkcjach) była duplikacja genów. Zrobienie drugiej kopii genu umożliwiłoby zatrzymanie funkcji „starego” genu, podczas gdy nowa kopia może służyć dla wprowadzenia innowacji w strukturze i funkcji nowego białka. Odkrycie zjawiska „dzielenia się genem” wskazuje na inny możliwy mechanizm innowacyjny: gen może nabywać drugą funkcję bez konieczności duplikacji genu i bez utraty jego funkcji pierwotnej. Co więcej, „dzielenie się genem” i multifunkcjonalność białek istniały prawdopodobnie wcześniej niż zjawisko duplikacji genów. Kilka przykładów zjawiska *gene sharing* przytaczam za Piatigorsky’em (Piatigorsky, 2007).

Pod koniec lat 80. odkryto, że u kaczki, kury i żółwia krystaliny – białka soczewki oka spełniające funkcje optyczne (refrakcja), pełnią w innych narządach funkcje enzymatyczne. Wkrótce okazało się, że delta2 krystalina soczewki oka kury jest białkiem identycznym z liazą argininosukcynyłową, a eta-krystalina soczewki kaczki pełni także funkcję enzymu dehydrogenazy mleczanowej B. U ptaków i gadów są duże różnice międzygatunkowe w budowie krystalin. Krystaliny wywodzą się od drobnocząsteczkowych białek szoku cieplnego i u wielu gatunków zatrzymują tę funkcję nawet w samej

soczewce (jako białka chaperonowe). U kałamarnicy krystalina ma jednocześnie funkcję transferazy-S-glutationowej, u innych mięczaków – funkcję dehydrogenazy aldehydowej.

Białko gelsolina w wysokich stężeniach działa jako regulator szkieletu aktywnego rogówki u ryby danio, ale w niskich stężeniach działa jako czynnik modulujący formowanie się wzoru brzuszno-grzbietowego w czasie rozwoju embrionalnego. Rodopsyna u muszki owocowej posiada funkcję fototransmittera i jednocześnie organizatora cytoszkieletu aktywnego dla morfogenezy fotoreceptora. Białko ksantyno oksydoreduktaza, enzym katabolizmu purynowego, ma również inną funkcję. Używane jest do budowy otoczki wydzielanej kropelki mleka w gruczole mlecznym. Syntaza cytrynowa pełni funkcje metaboliczne, a także funkcje strukturalne w cytoszkielecie.



Ryc. 7. Dzielenie się genem (*gene sharing*). Jeden i ten sam gen koduje jeden polipeptyd, który w zależności od stężenia, miejsca i kontekstu pełni odmienne funkcje molekularne

Te przykładowe białka enzymatyczne mają funkcje białek o właściwościach refrakcyjnych i innych strukturalnych, wówczas gdy ich stężenie jest odpowiednio wysokie, czyli dwie różne właściwości białka wynikają z różnic w ekspresji genu i jego lokalizacji tkankowej/narządowej. Co więcej, ten sam polipeptyd może znajdować się w różnych przedziałach komórki, gdzie przejawia różne funkcje w zależności od interakcji z innymi cząsteczkami białkowymi i lokalnym mikrośrodowiskiem.

Zjawiska te są przykładem zmian struktury i funkcji polipeptydu w zależności od jego stężenia, mikrośrodowiska w obrębie komórki i miejsca syntezy w tkance/narządzie. Należy ostrożnie podchodzić do dogmatu, że w każdym miejscu i w każdym czasie białko będzie zawsze miało tę samą funkcję, jaką wykazuje w danym momencie i danym miejscu, gdyż „funkcja genów i białek zależy od kontekstu” (Piatigorsky, 2007), (ryc. 7). Zjawisko „dzielenia się genem” może być zjawiskiem trwałym, wiecznym lub też może być zdarzeniem przejściowym w procesie ewolucji („testowanie” nowych funkcji białka).

Piatigorsky dyskutuje dość subtelne różnice między zjawiskiem *gene sharing* a pleiotropią. Pierwsze ma miejsce, gdy gen syntetyzuje białko o dwóch oddzielnych funkcjach, drugie zaś oznacza raczej sytuację, gdy gen „kontroluje” różne cechy fenotypo-

we. Tak dla przykładu, liaza argininosukcynylova jest enzymem biorącym udział w syntezie mocznika u ssaków, argininy u ptaków, ale jednocześnie pełni funkcję białka refrakcyjnego w soczewce oka. Natomiast enzym, syntaza tlenku azotu, choć jest również syntetyzowany „na podstawie” informacji jednego genu bierze udział w wielu różnych zjawiskach fenotypowych, jak np. procesy zapalne, neurotransmisja, oddychanie, apoptoza itd.

W żywych systemach polipeptydy (białka) zwykle wykonują liczne funkcje. Enzym GAPDH (dehydrogenaza glicerolaldehydo-3-fosforanowa) jest przykładem białka o wielu funkcjach – jest enzymem glikolitycznym, białkiem powierzchniowym wiążącym się z innymi białkami, posiada rolę w transdukcji sygnału, w soczewce występuje jako krystalina, jest koaktywatorem innych białek itd. Podobnie wielofunkcyjna jest enolaza – inny enzym działający jako białko szoku cieplnego, pełniące rolę w organizacji mikrotubul, które wiąże aktynę i plasminogen, a w dodatku jest również krystaliną.

Zasadnicza nauka, jaka wypływa ze zjawiska „dzielenia się genem” to przypomnienie, że ani gen, ani białko nie są elementami autonomicznymi, działającymi w sposób niezależny, izolowany, ale są elementami dynamicznego systemu (sieci).

Promotory alternatywne

Promotory są swoistymi sekwencjami DNA, które umiejscawiano zwykle przed początkiem genu strukturalnego. Stanowią one elementy niezbędne dla inicjacji transkrypcji i kontroli ekspresji genów. Okazało się jednak, że ich położenie w stosunku do genu w organizmach eukariotycznych jest bardzo różne, a ponadto komórka używa ich alternatywnie, przez co zwiększa się plastyczność ekspresji genów w organizmie wielokomórkowym. Transkrypcja inicjowana z alternatywnych promotorów różni się poziomem, a transkrypty (a zatem i kodowane przez nie białka) wykazują różne izoformy. Alternatywne promotory wykazują swoistość tkankową i różnią się odpowiedzią na system sygnałów (Ayoubi i Van de Ven, 1996).

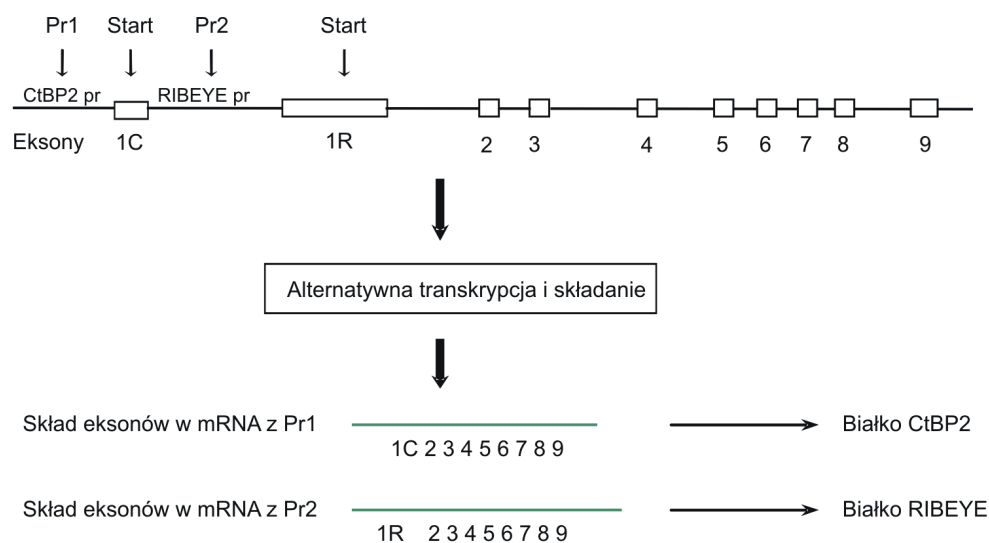
Użycie w procesie transkrypcji alternatywnych promotorów jest jeszcze innym mechanizmem generującym różnorodność mRNA i białka. Transkrypcja DNA (genu) zaczyna się od sekwencji zwanej promotorem, mieszczącej się zwykle u początku genu. Okazuje się, że gen może posiadać kilka miejsc promotorowych.

Przez użycie alternatywnych promotorów uzyskuje się nowe, zróżnicowane i elastyczne narzędzie kontroli ekspresji genów: możliwość generowania białek różniących się u początku cząsteczki, kontrolę poziomu transkrypcji, swoistość tkankową, swoistość związaną ze stadium rozwojowym, możliwość kontroli wydolności translacji i czasu trwania cząsteczek mRNA. Regulacja genów przez mechanizm alternatywnych promotorów jest szeroko rozpowszechniona u organizmów żywych.

U kręgowców wiele genów posiada alternatywne promotory używane w zależności od narządu, tkanki, stanu fizjologicznego komórki, stadium rozwoju, zdolności komórki

do odpowiedzi na szczególne stany mikrośrodowiska i odpowiedzi na warunki metaboliczne. Rozmieszczenie cząsteczek białkowych w komórce, poziom białka i jego funkcja jest także uwarunkowana użyciem odpowiedniego promotora. Wprowadzenie do zagadnienia użycia alternatywnych promotorów znajduje się w przeglądowej pracy (Ayoubi i Van de Ven, 1996).

Użycie alternatywnych promotorów skutkuje głębokimi zmianami fenotypowymi. „Alternatywne miejsca inicjacji transkrypcji mogą spowodować transformację intronów w eksony kodujące białko i wytwarzać białka o różnej strukturze i funkcji” (Piatigorsky, 2007). Przytaczam ten cytat ze świetnej monografii Piatigorsky’ego gdzie podany jest przykład genu kodującego dwa odrębne białka w zależności od użycia alternatywnych promotorów (ryc. 8). Użycie pierwszego promotora genu *CtBP2/RIBEYE* daje białko CtBP2 – posiadające funkcję powszechnego represora transkrypcji. Użycie alternatywnego, drugiego promotora oraz alternatywnego składowania daje białko RIBEYE występujące w synapsach siatkówki, w ślimaku ucha i przysadce. W wyniku użycia alternatywnych promotorów z „informacji” zapisanej w jednym genie powstają dwa różne rodzaje białka o różnej lokalizacji i różnych funkcjach.



Ryc. 8. Alternatywne użycie promotorów Pr1 lub Pr2 oraz alternatywne składowanie daje dwa różne mRNA i dwa białka CtBP2 i RIBEYE o różnej lokalizacji i różnej funkcji. Rycina jest uproszczonym schematem wzorowanym na ryc. 3.1 w monografii Piatigorsky’ego (Piatigorsky, 2007)

Przez użycie alternatywnych promotorów uzyskuje się nowe, zróżnicowane i elastyczne narzędzie kontroli ekspresji genów, możliwość generowania białek różniących się na początku cząsteczki, kontrolę poziomu transkrypcji, swoistość tkankową, swoistość związaną ze stadium rozwojowym, możliwość kontroli wydolności translacji i czasu

trwania cząsteczek mRNA. Regulacja genów przez mechanizm alternatywnych promotorów jest szeroko rozpowszechniona u organizmów żywych.

U kręgowców wiele genów posiada alternatywne promotory używane w zależności od narządu, tkanki, stanu fizjologicznego komórki, stadium rozwoju, zdolności komórki do odpowiedzi na szczególne stany mikrośrodowiska i odpowiedzi na warunki metaboliczne. Rozmieszczenie cząsteczek białkowych w komórce, poziom białka i jego funkcja są także uwarunkowane użyciem odpowiedniego promotora.

Nonsense-mediated mRNA decay

W prawidłowych komórkach translacja mRNA w łańcuch białkowy biegnie od sygnału inicjującego „start” (kodon AUG) do sygnału kończącego „stop” (kodony UAA, UAG, UGA zwane także kodonami *nonsense*). Wskutek mutacji punktowej lub mutacji z przesunięciem ramki odczytu, w mRNA mogą być generowane kodony *nonsense*, będące przedwczesnymi sygnałami terminacji translacji (PTC, *premature termination codons*). Ich obecność uniemożliwia dokończenie syntezy cząsteczki białka, w wyniku czego powstają krótsze („przycięte”) cząsteczki białka. Szacuje się, że tak uszkodzone białka mają związek z jedną trzecią chorób genetycznych, w tym również nowotworowych.

Komórki eukariotyczne posiadają zdolność rozpoznawania i degradowania takich transkryptów mRNA, w których pojawiają się przedwczesne sygnały terminacji translacji (PTC). System degradujący cząsteczki mRNA posiadające sygnały PTC nazywa się systemem NMD (*nonsense-mediated mRNA decay*). System NMD radykalnie redukuje poziom mRNA z przedwczesnymi sygnałami PTC, dzięki czemu synteza uszkodzonych (skróconych) cząsteczek białka spada.

Prawdopodobnie system NMD ma funkcje bardziej rozległe niż tylko rodzaj „kontroli jakości”, możliwe, że reguluje „szum genomowy” i nadzoruje aktywność szerokiego spektrum genów (kinazy białkowe, fosfatazy, czynniki transkrypcyjne, geny biorące udział w metabolizmie aminokwasów), a także bierze udział w mechanizmach składania mRNA. System NMD nadzoruje także niefunkcjonalne transkrypty z retrowirusów i transpozonów, trzymając je w uśpieniu, i jest aktywnym składnikiem regulacyjnym aktywności genowej (Mendel i in., 2004; Alonso, 2005). System NMD jest ewolucyjnie konserwowany i występuje powszechnie u drożdży, robaków, roślin, owadów i u ssaków (Maquat, 2005). Jego rola w rozwoju osobniczym i ewolucji oraz znaczenie w klinice nie jest w pełni poznane (Kuzmiak i Maquat, 2006; Madghalchi i in., 2001).

Inaktywacja genów supresorowych jest związana z dziedziczną formą nowotworów. Większość sporadycznych nowotworów nie wykazuje inaktywacji tych genów. Głęboka mutacja lub utrata jednej kopii genu supresorowego, np. genu *RB* w komórce germinalnej, ma charakter recesywny, a zatem nie wywołuje wyraźnego efektu fenotypowego, czyli nie wiąże się z powstawaniem groźnego nowotworu oka – siatkówczaka, ponieważ

pozostaje druga czynna kopia (drugi allel). Utrata heterozygotyczności, czyli zniesienie funkcji drugiej kopii (drugiego allelu) genu supresorowego (tu genu *RB*) prowadzi do powstania dziedzicznej formy siatkówczaka. To zjawisko znane jest w raku dziedzicznym i odpowiada hipotezie „dwóch trafień” Knudsona. Najczęstszym sposobem nieczynniania allelu w genie *RB* jest nabycie mutacji nonsensowych, które generują kodony przedwczesnej terminacji translacji – PTC (Holbtook i in., 2004). W genie *BRCA1* germinalne mutacje typu PTC (w liczbie ok. 30) zidentyfikowano w 22 kodonach. System NMD degraduje większość mRNA posiadających PTC, ale jednocześnie obserwuje się obecność białek „przyciętych”. Przypuszczalnie pochodzą one z PTC zlokalizowanych blisko kodonu inicjującego, gdyż ta lokalizacja PTC (podobnie jak lokalizacja blisko kodonu terminującego) wymyka się spod nadzoru systemu NMD (Perrin-Vidoz i in., 2002).

NMD jest skomplikowanym systemem i funkcją kompleksu białek kodowanych przez geny z rodziny *Rent1* (*Upf1*), które mają zdolność rozpoznawania mRNA posiadającego sygnał PTC (kodon „stop”), z wyjątkiem lokalizacji w eksonie końcowym lub w pobliżu początku genu. System NMD nie narusza jednak redagowanych cząsteczek mRNA, w których celowo powstał nowy sygnał „stop”, skracający cząsteczkę białka, jak w podanym wyżej przykładzie jelitowej apolipoproteiny B.

Obecność systemu NMD nasuwa konieczność zachowania ostrożności w interpretacji „ekspresji genów” przy użyciu mikromacierzy DNA.

Interferencja RNA

System wyciszania (interferencja RNA, RNAi) jest jednym z elementów kontroli aktywności genów na poziomie DNA lub mRNA przez swoiste cząsteczki RNA zwane ogólnie RNAi (RNA interferencyjny). Zanim wyjaśniono zjawisko interferencji, fenomen ten był znany pod różnymi nazwami historycznymi: potranskrypcyjne wyciszanie genów (PTGS, *postranscriptional gene silencing*), tłumienie genów (*gene quelling*), wyciszanie genów (*gene silencing*) lub kosupresja ekspresji genu.

Zjawisko interferencji zauważono po raz pierwszy u petunii w trakcie prób uzyskania transgenicznej rośliny o ciemnoczerwonym zabarwieniu kwiatów, wprowadzając do rośliny dodatkowe kopie transgeny kodującego enzym odpowiedzialny za powstawanie pigmentu. Jednak zamiast, jak oczekiwano, uzyskania (przez nadekspresję genów i dodatkowe mRNA) kwiatów o ciemniejszym zabarwieniu, zabieg spowodował, że kwiaty były bledsze lub wręcz białe. Fenomen ten określono jako kosupresję genu. U transgenicznej rośliny nastąpiła intensywne degradacja mRNA kodującego enzym odpowiedzialny za syntezę pigmentu wskutek uruchomienia systemu PTGS. Podobne zjawiska obserwowano u roślin w trakcie badań nad opornością na infekcję wirusową. Okazało się, że wprowadzenie do rośliny krótkich, niekodujących sekwencji wirusowego RNA powoduje wzrost oporności na infekcję przez indukcję systemu PTGS.

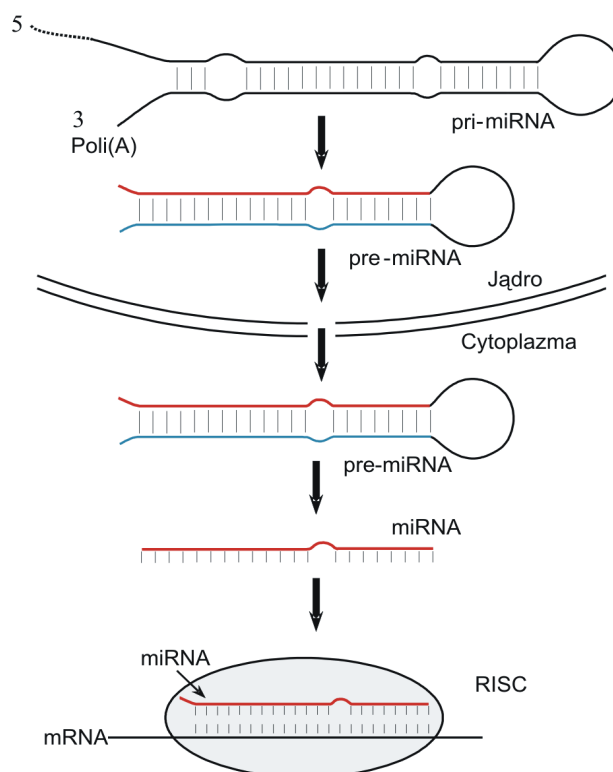
Podstawowy mechanizm zjawiska interferencji opisali późniejsi laureaci Nagrody Nobla – A. Fire i C.C. Mello na modelu nicienia *Caenorhabditis elegans*. (Fire i in., 1998). Aby uzyskać wyciszenie określonego genu (białko mięśni), wstrzykiwano do nicienia mRNA i *antisense* RNA oraz dwupasmowy RNA (dsRNA). Okazało się, że jedynie dsRNA dawał efekt interferencji. Co więcej, dsRNA podany drogą pokarmową do jelita nicienia powodował nie tylko efekt interferencji w różnych organach, lecz także, jak wykazano potem, efekt ten był dziedziczny (najpewniej przez narzucony wzór metylacji DNA).

Doświadczenie to dowodzi, że dsRNA jest pobierany przez komórki nabłonka jelita w stanie nienaruszonym i transportowany między komórkami. W tym złożonym procesie bierze udział zarówno system porów transbłonowych, jak i system TNT (*tunnelling nanotubules*) tworzący tunelowe połączenia między komórkami przez nanotubule w komórkach zwierzęcych i kanaliki (*plasmodesmata*) tworzące połączenie międzykomórkowe u roślin, a także eksocytotyczne pęcherzyki sekrecyjne (*exosomes*). Jest wysoce prawdopodobne, że takie mikrostruktury stanowią wewnątrzkomórkowy układ komunikacyjny dla cząsteczek regulatorowych, takich jak np. generowane przez system RNAi. Być może wkrótce narodzi się molekularna „genetyka narządowa”.

Fenomen interferencji RNA ma znaczenie uniwersalne. System RNAi jest jednym z podstawowych wewnątrzkomórkowych (i międzykomórkowych?) systemów regulacji funkcji genów działających w świecie roślin, owadów i ssaków, a indukujące cząsteczki dsRNA są endogennym produktem.

W dużym uproszczeniu, system interferencji działa w następujący sposób. Końcowe aktywne cząsteczki systemu (szlaku) RNAi należą do klas drobnych cząsteczek RNA o długości 21-23 nukleotydów, miRNA (*micro RNA*) oraz siRNA (*small interfering RNA*). Macierzysta cząsteczka RNA, z której powstaje miRNA, jest długim transkryptem, oznaczanym jako pri-miRNA (*primary miRNA*), posiadającym strukturę typu *stem-loop* (pętla z łydą, szpilka do włosów) o długości ok. 70 nukleotydów. Cząsteczka pri-miRNA ma zatem częściową strukturę dwupasmową. Ta struktura zostaje wycięta w jądrze komórkowym przez kompleks enzymatyczny i jako dwupasmowy pre-miRNA (*precursor miRNA*) zostaje przeniesiona do cytoplazmy, gdzie jest poddana dalszej obróbce enzymatycznej, w wyniku czego powstaje aktywna jednopasmowa, dojrzała cząsteczka końcowa – miRNA (ryc. 9).

Cząsteczki miRNA wchodzą następnie w kompleks interferencyjny (wyciszający) RISC (*RNA – induced silencing complex*) z odpowiednimi komplementarnymi mRNA (zwykle na końcu 3'), degradując je lub uniemożliwiając ich translację. Jeśli w obrębie 21-23 sekwencji pre-miRNA znajdzie się jednonukleotydowy polimorfizm (*SNP, single nucleotide polymorphism*), to okazuje się, że taka zmiana ma wpływ na ostateczną aktywność miRNA (Jazdzewski i in., 2009).

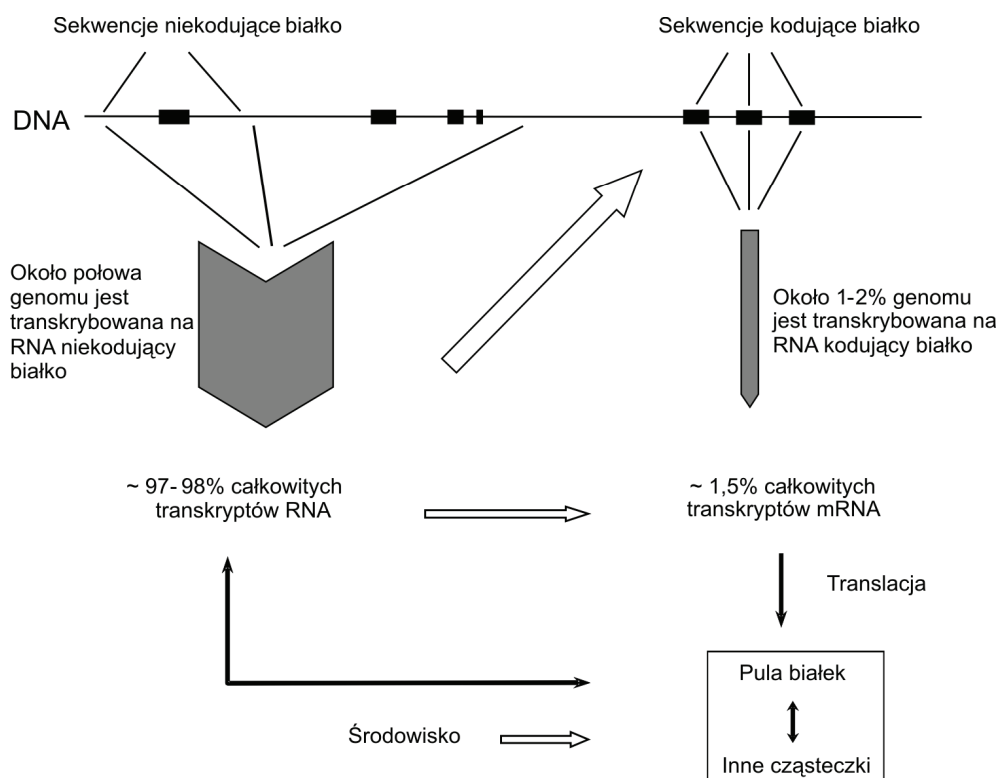


Ryc. 9. Schemat powstawania cząsteczki miRNA i proces wyciszania mRNA

Terminologia drobnych cząsteczek regulatorowych RNA jest nieco zawiła. Termin miRNA odnosi się głównie do cząsteczek endogennych, dających niekompletne parowanie zasad z mRNA w procesie interferencji, przeto miRNA jest mniej swoisty i wywiera efekt wyciszający na różne mRNA. Cząsteczka siRNA ma sekwencje precyzyjnie komplementarne do określonego mRNA i dzięki temu powoduje w końcowym efekcie rozpad powstającego mRNA, czyli wycisza tylko określony gen. Inna klasa drobnych cząsteczek, piRNA, chroni komórki linii germinalnych przez inwazyjnym działaniem transpozonów.

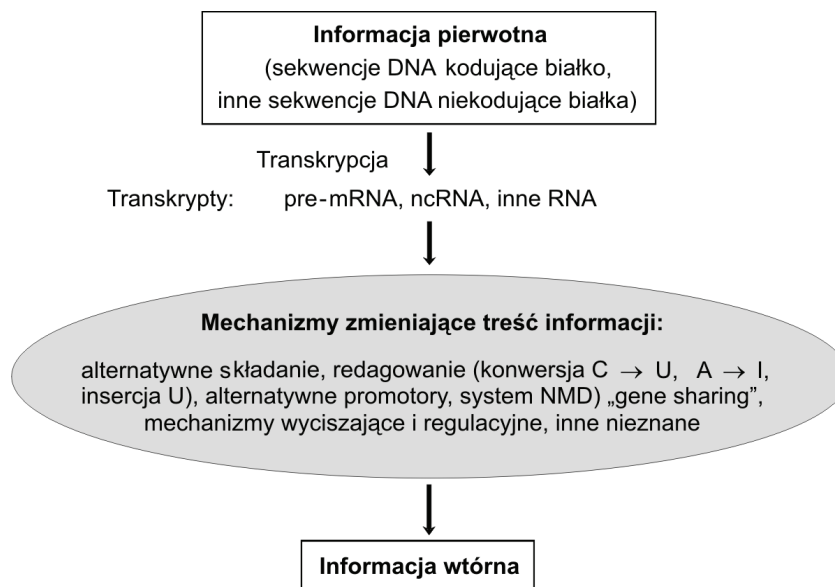
Już parę dziesiątków lat temu pojawiły się prace sugerujące, że proces transkrypcji nie obejmuje wyłącznie genów kodujących białko oraz rRNA i tRNA, lecz także sekwencje międzygenowe i introny są przepisywane na niekodujący RNA (ncRNA, *non-protein coding* RNA). W istocie, w samej populacji cząsteczek pre-mRNA transkrypty z intronów stanowią ok. 95%. Wykazano następnie, że transkrypcji nie podlega jedynie pasmo *sense* DNA, lecz transkrypcja biegnie również z pasma *nonsense* DNA. Szacuje się, że 98% wszystkich transkryptów genomu ludzkiego stanowią ncRNA. Przegląd wcześniej-

szych prac dotyczących ncRNA znajdzie Czytelnik w wielu publikacjach, np. (Mattick i Gagen, 2001; Szymański i Barciszewski, 2002; Szymański i in., 2003).



Ryc. 10. Schemat ilustrujący relatywne ilości transkryptów z obszarów międzygenowych i intronów oraz z sekwencji kodujących białko. Pula cząsteczek białkowych w komórce wraz z innymi cząsteczkami stanowi system dynamiczny, oddziałujący na procesy transkrypcyjne oraz działający w kontekście środowiska

Rycina 10 ilustruje szacunkowe proporcje transkryptów kodujących (mRNA, tRNA, rRNA) oraz niekodujących (ncRNA) u człowieka. Ilość transkryptów ncRNA absolutnie dominuje w ogólnej puli transkryptów i fakt ten nie może być przypadkowy, wynikający np. z „rozrzutności” natury, a ncRNA nie może stanowić „szumu” transkrypcyjnego. Wynika z tego, że znaczna część genomu może mieć znaczenie funkcjonalne. Odkrycie tak wielkiego udziału sekwencji niekodujących podlegających transkrypcji oraz identyfikacja ncRNA podważyło centralny dogmat, nakazało zrewidować tradycyjny pogląd, że w organizmach żywych funkcja regulacyjna aktywności genów jest przypisana wyłącznie białkom oraz zmusiło, aby postrzegać „śmieciowy” DNA jako ważny element wprzęgnięty w procesy regulacji genowej (Mattick, 2001; Mattick, 2003; Erdmann i in., 2001).



Ryc. 11. Schemat ilustrujący przetwarzanie „informacji pierwotnej” na „informację wtórną”. Mechanizmy przetwarzania informacji są słabo poznane. Owal ilustruje „szarą strefę”, gdzie zachodzi przetwarzanie informacji przez dynamiczny system biologiczny, aktywnie wykorzystujący „zapis” w sekwencjach DNA stosownie do kontekstu i aktualnych potrzeb komórki

Transkrypty ncRNA stanowią istotną część systemu regulatorowego i kontrolującego różnorodne funkcje, takie jak architektura chromosomów, transkrypcja genów kodujących, obrót mRNA („żywność”, rozpad, stabilizacja), alternatywne składanie i redagowanie mRNA, oraz biorą udział w modulacji czasowej molekularnych procesów rozwoju osobniczego itp. W puli ncRNA powstają także cząsteczki systemu interferencji RNA. Między genem a białkiem operuje niezwykle złożony i mało jeszcze poznany system regulujący istotne funkcje komórki. Wydaje się, że jednym z głównych zadań tego systemu jest przetwarzanie „informacji pierwotnej” genu na „informację wtórną”. Mechanizmy przetwarzania informacji nie są ani inicjowane, ani sterowane przez gen, lecz dyktowane są dynamicznym systemem biologicznym komórki (ryc. 11).

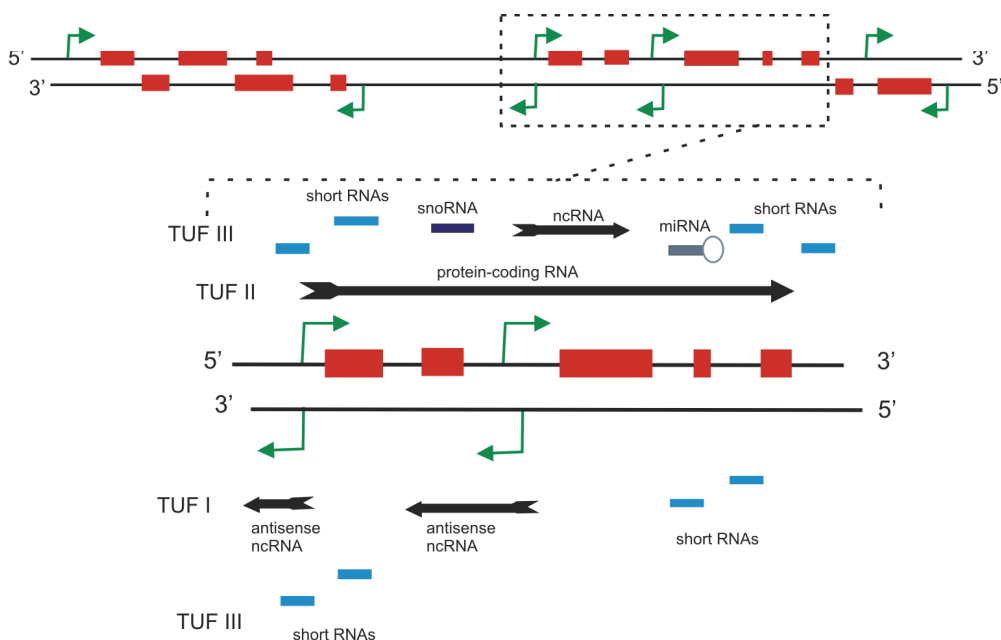
Co dał projekt ENCODE?

W 2003 roku National Human Genome Research Institute, nowa placówka naukowa Narodowych Instytutów Zdrowia (NIH) w USA, uruchomił międzynarodowy projekt badawczy pod nazwą ENCODE (*Encyclopedia of DNA Elements*). Konsorcjum badawcze zgromadziło naukowców różnych specjalności i miało do dyspozycji praktycznie wszystkie współczesne metody i techniki badawcze oraz aparaturę. Projekt ten zmierza do pełnego poznania struktury i funkcji sekwencji ludzkiego genomu, które dotychczas były

uznawane za pozostałości ewolucji, niemające znaczenia biologicznego. Początkowa faza projektu zmierzała do poznania elementów funkcjonalnych tylko w określonej porcji genomu. Pilotażowy projekt programu ENCODE objął 44 wybrane regiony genomu człowieka. W sumie stanowiło to zaledwie około 1% genomu. Badania te dały olbrzymią liczbę nowych informacji i wiedzę, która zmusza nas do rewizji „klasycznej” koncepcji genu strukturalnego oraz rewizji poglądów na funkcję DNA i relacje między genotypem a fenotypem.

Najważniejszym wynikiem projektu było odkrycie, że prawie cały genom jest przepisywany na pierwotne transkrypty, często nachodzące na siebie (*overlapping*) oraz zawierające zarówno sekwencje kodujące białko, jak i sekwencje niekodujące (czyli spisywane z obu pasm DNA – *sense* i *antisense*). Już wcześniej wiadano, że transkrypty niekodującego stanowią w komórce eukariotycznej przeważającą część całkowitych transkryptów (por. Mattick, 2003), jednak dopiero w ramach projektu ENCODE poznano bliżej strukturę takich transkryptów, ich relację do DNA i niektóre ich właściwości. Rozwinięcie projektu ENCODE przyniesie zapewne nowe odkrycia. Dotychczas znane transkrypty nie kodujące białka obejmują rybosomalny RNA (rRNA), transportujący RNA (tRNA), małe jądrowe RNA (*snRNA*, *small nuclear RNA*), microRNA, siRNA, krótkie sekwencje RNA wchodzące w skład kompleksów RNazy P i inne. Wstępne badania w ramach projektu ENCODE ujawniły wielką kolekcję nieznaną uprzednio transkryptów posiadających nikłą lub żadną możliwość kodowania białka, a w zasadzie transkryptów o nieznannej funkcji (TUF, *transcript of unknown function*). W okolicy „genu klasycznego” i regionach międzygenowych zidentyfikowano wiele nowych transkrypcyjnych miejsc startowych (TSS, *transcriptiom start site*).

Gingeras (Gingeras, 2007) wyróżnia trzy potencjalne kategorie transkryptów TUF (ryc. 12). Kategoria pierwsza TUF obejmuje transkrypty niekodujące RNA (ncRNA), spisywane z pasma DNA *sense*, a zatem komplementarne do transkryptów pre-mRNA przez co mogą one tworzyć, przynajmniej z niektórymi eksonami i intronami pre-mRNA lub mRNA, komplementarne dwupasmowe struktury. Takie transkrypty mogą pochodzić z regionów odległych od danego genu i mogą być tylko częściowo komplementarne do transkryptów *sense* (ryc. 12, TUF I). Transkrypty te występują w dużej ilości; szacuje się, że nawet 70% genów kodujących białko może wytwarzać transkrypty niekodujące białka – ncRNA (*non-protein coding RNA*), z których większość należy do transkryptów TUF. Kategoria druga odpowiada izoformom znanych transkryptów kodującym białko, czyli pre-mRNA (ryc. 12, TUF II). Okazuje się, że w przypadku 90% genów objętych projektem występują nowe izoformy poprzednio nienotowanych transkryptów kodujących. Odkryto nowe miejsca startu TSS zlokalizowane daleko „na lewo” (nawet w odległości ponad 100 kb) od końca 5’ genu. W takich transkryptach u końca 5’ odkryto całkiem nowe sekwencje o cechach eksonów (nieulegające jednak translacji), co oznacza, że pierwszy intron może być znacznie dłuższy, niż to poprzednio oznaczano na strukturze genu „klasycznego”.



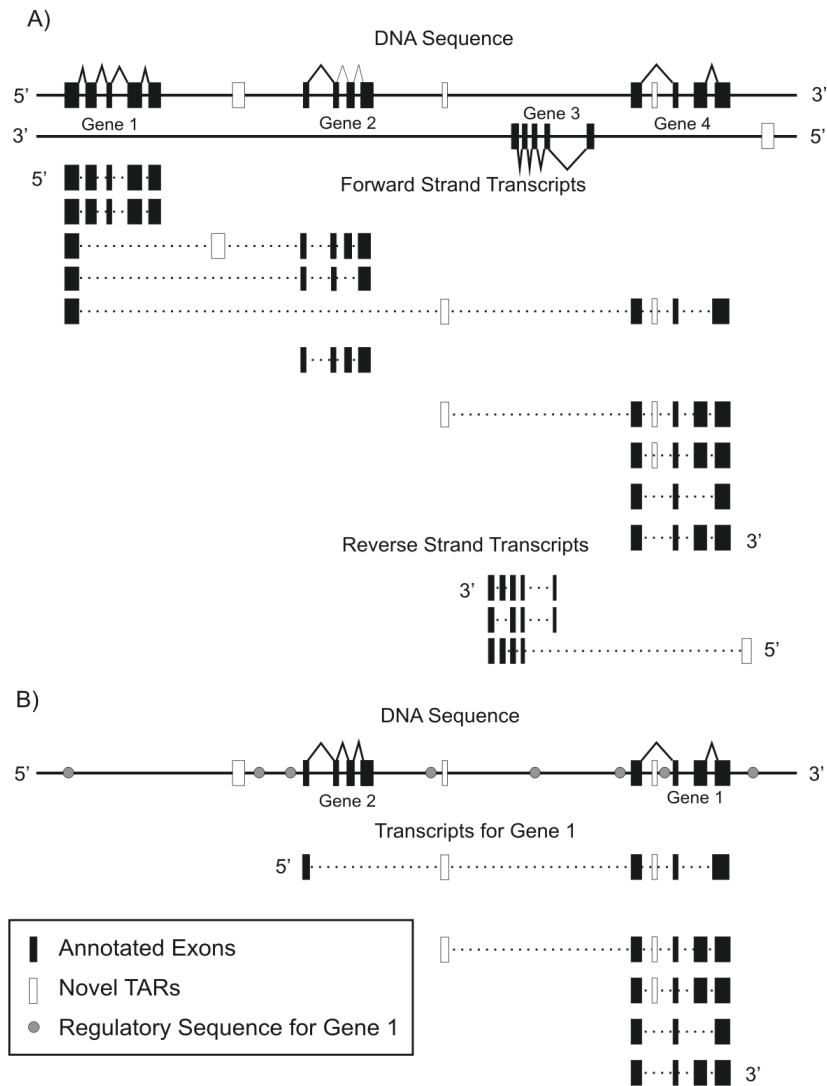
Ryc. 12. Schemat ilustrujący złożoność transkrypcyjną genu. Hipotetyczna grupa genowa (górną część ryciny) jest powiększona dla pokazania poszczególnych transkryptów (część dolna ryciny). Eksony oznaczone są kolorem czerwonym. TSS – strzałki zielone. Transkrypty ujęte jako trzy klasy TUF oznaczone są kolorem czarnym i niebieskim. Skala długości eksonów, intronów i regionów międzygenowych nie jest zachowana. W rzeczywistości długość intronów i regionów międzygenowych jest wielokrotnie większa, niż to pokazuje rycina. (Na dolnej części ryciny nałożone są na siebie zarówno pasmo DNA, jak i homologiczny transkrypt, co wraz z kierunkami strzałek TSS wprowadza pewne utrudnienie co do orientacji w polarności. Należy zauważyć, że np. transkrypt *protein-coding* RNA, ma sekwencje identyczne jak pasmo *sens* DNA biegnące od 5' do 3', ale ten transkrypt (strzałka czarna pokazuje polarność od 5' do 3') został spisany z dolnego pasma DNA o kierunku 3' do 5'. Odwrotnie, transkrypty *antisense* ncRNA są identyczne z sekwencją dolnego pasma DNA, ale zostały spisane z górnego pasma DNA *sense*). Według Gingeras (2007), *Genome Res.* 17, 628-690; copyright: Cold Spring Harbor Laboratory Press. Reprodukowano za zgodą z dnia 25.03.2009

Do tej kategorii TUF należą również transkrypty z pseudogenów, które mimo że utraciły zdolności kodujące, mogą być nadal transkrybowane, a które w genomie człowieka występują w liczbie nie mniejszej, niż geny kodujące białko (ok. 20 000). W kilku przypadkach stwierdzono, że transkrypty z pseudogenów mogą brać udział w procesie składania i tworzyć formy chimeryczne z transkryptami genów kodujących (Zheng i Gerstein, 2007). Do trzeciej kategorii TUF (Ryc. 12, TUF III) należą transkrypty pokrywające się z sekwencjami intronów, ncRNA, snoRNA, miRNA, albo pochodzące z regionów międzygenowych. Niektóre z nich są poliadenylowane.

Wiedza o funkcji niektórych klas niekodujących transkryptów stale się pogłębia. Z transkryptów TUF, głównie z intronów, pochodzą drobnocząsteczkowe, małe jąderkowe RNA (snoRNA, *small nucleolar* RNA), o długości 60-300 nukleotydów, które biorą udział w modyfikacjach rRNA i jednocześnie stanowią wraz z białkami komponentę kreującą kontekst (mikrośrodowisko) dla funkcji innych czynników białkowych. Do transkryptów TUF zaliczamy interferencyjny RNA (miRNA i siRNA – regulatory ekspresji genów o licznych funkcjach; por. rozdz. Interferencja RNA). Zidentyfikowano także nowe klasy krótkich cząsteczek RNA o nieznannej funkcji, powstających na początku i końcu genu. Interesującą cechą transkryptów TUF jest ich zdolność do tworzenia stabilnych przestrzennych struktur o nieznannej roli. W sumie TUF są cząsteczkami o częściowo tylko poznanych funkcjach biologicznych i stanowią nowy poziom w sieci regulacyjnej (Gingeras, 2007).

Odkryte TSS nie są równomiernie rozproszone po genomie, lecz zwykle występują w dużych hierarchicznie zorganizowanych zgrupowaniach (*clusters*), w których rozpoznawane są następne zgrupowania (*clusters in clusters*). W takiej organizacji TSS trudno jest rozpoznać, gdzie znajduje się „rdzeń” promotora i gdzie są jego granice (Frithl i in., 2008). Nie wiemy też, jak komórka wybiera TSS na poziomie chromatyny i na poziomie sekwencji DNA i jak generowane są regiony transkrypcyjnie aktywne. W komórce DNA upakowane jest wraz z histonami i innymi białkami w chromatynie. Wiadomo, że modyfikacje histonów korelują z aktywnością transkrypcyjną, a wzór zmodyfikowanych histonów tworzy swoisty „kod histonowy”, współdziałający ze złożoną siecią regulacyjną genomu (Koch i in., 2007). Gerstein w świetnej pracy przeglądowej (Gerstein i in., 2007) zwrócił uwagę na złożoność strukturalną samego genu kodującego, mnogość izoform transkryptów kodujących wynikającą z alternatywnego składania pre-mRNA oraz zidentyfikowanie nowych miejsc aktywnych transkrypcyjnie (TAR, *transcriptionally active region*) i nowych sekwencji regulatorowych. Ten rodzaj złożoności biologicznej ilustruje ryc. 12. Sekwencje genowe (kodujące) znajdują się na obu pasmach DNA. Nazwy pasmo *sense* i *antisense* stają się pojęciem elastycznym, a nawet wprowadzającym pewne zamieszanie terminologiczne. Stąd proponowane są nowe nazwy: pasmo *forward* (dawniej *sense*) i pasmo *reverse* (dawnie *antisense*). Nowe obszary TAR pojawiają się w miejscach do niedawna uważanych za miejsca niefunkcjonalne. Sekwencje regulacyjne i TSS są liczne, rozproszone.

Niekiedy alternatywny TSS używa promotora zupełnie innego genu znajdującego się przed genem aktualnie transkrybowanym. Takie dwa geny wykorzystują ten sam promotor, a transkrypt spina oba *loci* genowe. Jednak takie izoformy transkryptu kodują jedno białko (por. ryc. 13). Sekwencje u końca 5' transkryptu nie ulegają translacji i mogą mieć jedynie znaczenie regulatorowe w stosunku do transkryptu (np. wpływają na jego trwałość).



Ryc. 13. Biologiczna złożoność genu i jego transkryptów. A) – Typowy region DNA zajęty przez geny. Eksony – prostokąty zaczernione. Nowo odkryte regiony aktywne transkrypcyjne TAR – prostokąty puste. Poniżej grupy transkryptów z trzech genów i alternatywne składanie. Introny w informacyjnym pre-mRNA (linie przerywane) zostają usunięte w czasie składania. Najdłuższy transkrypt objął trzy loci genowe (gen 1, gen 2 i gen 4) używając TSS daleko „na lewo” („w górę”) od genu 1. B) – Przykład lokalizacji różnych sekwencji regulatorowych (siwe kółka) zidentyfikowanych jako swoiste dla genu 1. Gen 1 używa sekwencji regulatorowych poprzedzających gen 2, co może prowadzić do błędnej indentyfikacji genu docelowego. Wg Gerstein i in. (2007), *Genome Res.* 17, 669-681; copyright: Cold Spring Harbor Laboratory Press.

Reprodukowano za zgodą z dnia 25.03.2009)

Miejsca regulatorowe (np. sekwencje promotorowe, sekwencje wzmacniające, „enhancer”) danego genu niekoniecznie znajdują się przed genem (co zwykle antycypujemy, że coś co reguluje „liniowy gen” musi być „na początku”). Często są one ulokowane w eksonach, intronach, a także daleko „w górze” i „w dole” genu w pobliżu innego genu. Już uprzednio poczyniono takie sporadyczne spostrzeżenia.

Klasyczny model genu zakładał, że elementy regulacyjne nie są transkrybowane, a zatem nie są one częścią genu. Tymczasem odkryto, że niektóre elementy regulatorowe ulegają transkrypcji, nawet jeśli leżą z dala od początku genu.

Współczesne definicje genu molekularnego

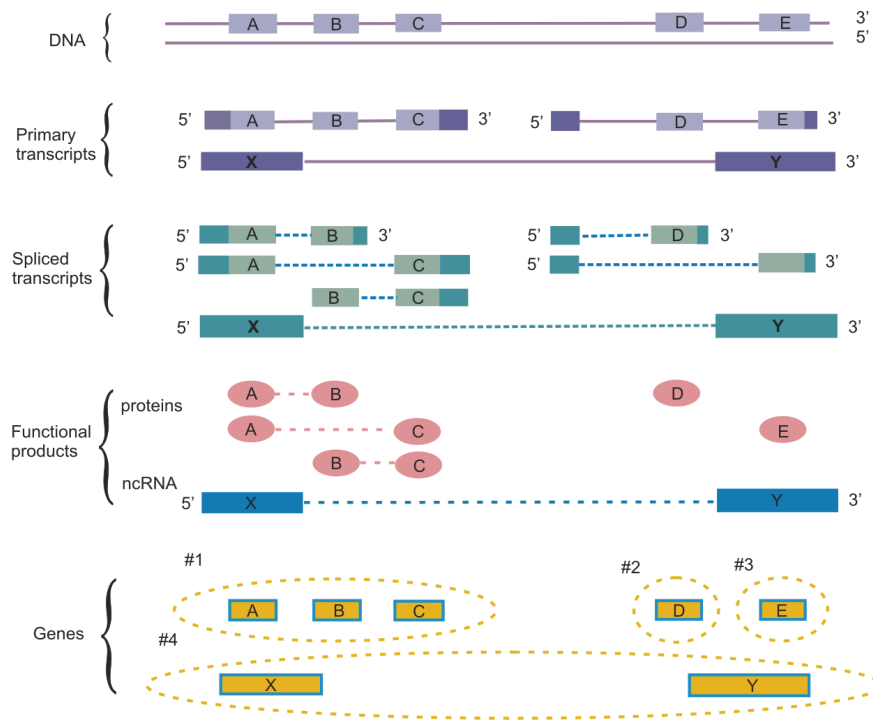
Termin „gen” powstał jako termin teoretyczny. Takie teoretyczne terminy mają zastosowanie w rozmaitych obszarach badań, a ich znaczenie zmienia się w czasie i w zależności od kontekstu. Nie jest zatem niespodzianką, że definicja genu zmieniała się od czasów Mendla i Morgana do czasów współczesnych. Po odkryciu roli DNA w procesie transformacji i poznaniu struktury DNA definicja genu została immanentnie związana DNA i transkryptami. Współczesne definicje genu opierają się na uznaniu genu jako realnej, autonomicznej, fizycznej jednostki genomicznej i formowane są w oparciu o sekwencje nukleotydów DNA, wykorzystywanych w syntezie polipeptydów, lub RNA. „Gen” miałby posiadać zatem swój wymiar przestrzenny, swój początek i koniec, swoją strukturę molekularną.

Definicje genu rozwijały się w oparciu o przekonanie, że DNA ma zdolność do „samopowielania się” oraz na bazie dogmatu: „DNA robi RNA, RNA robi białko, białko robi nas”. Obie te tezy są niemożliwe do przyjęcia w swej czystej postaci.

Pierwotna definicja genu „klasycznego” głosiła, że gen jest „odcinkiem DNA kodującym białko (polipeptyd)”. Pojęcie genu zostało wbudowane w „genocentryczny paradygmat biologii” (Łukow i Żekanowski, 2005).

Klasyczna definicja molekularna była rozbudowywana w miarę jak poznawano strukturę i funkcję sekwencji DNA innych niż sekwencje kodujące białko. Najpierw do definicji dodano sekwencje kodujące te klasy RNA, które są „ostatecznymi produktami genu” (np. rybosomalny RNA, transportujący RNA), potem uzupełniono definicję jeszcze o sekwencje regulatorowe, wprowadzając koncepcję „jednostki transkrypcji”. Następnie okazało się, że i ta definicja nie wyczerpuje wszystkich atrybutów biologicznych przypisywanych pojęciu „gen”.

B. Lewin w dziele z 1994 r. (Lewin, 1994) podaje następującą definicję: „Gen jest segmentem DNA wciągniętym w produkowanie łańcucha polipeptydowego; gen obejmuje regiony poprzedzające i następujące po regionie kodującym, jak również sekwencje wtrącone (introny), leżące między indywidualnymi segmentami kodującymi (eksunami)”.



Ryc. 14. Ilustracja definicji genu wg Gersteina. Region genomu produkuje trzy pierwotne transkrypty (por. Primary transcripts). Po alternatywnym składaniu dwa z nich produkują pięć rodzajów cząsteczek białkowych (por. Functional products): izoformy A-A, A-C i B-C oraz białka D i C, a trzeci produkuje niekodujący RNA (ncRNA). Produkty białkowe są zakodowane przez trzy klastry sekwencji DNA (por. DNA): pierwszy jest trój-segmentowy (A, B, C) i odpowiada genowi nr.1 (por. Genes), drugi (D) odpowiada genowi nr 2, trzeci (E) odpowiada genowi nr 3. Na tym przykładzie jeden transkrypt niekodujący (ncRNA) kodowany jest przez sekwencję X, Y; został on spisany z genu nr. 4. Zatem podany przykładowo region ma wg Gersteina być zajęty przez 4 geny rezydujące na obu pasmach DNA. Wg Gerstein i in. (2007), *Genome Res.* 17, 669-681; copyright: Cold Spring Harbor Laboratory Press. Reprodukowano za zgodą z dnia 25.03.2009

W dziele *Molecular Biology of the Cell* z 1994 r. (Alberts i in., 1994) słownik podaje następującą definicję: „Gen – region DNA, który kontroluje dyskretne cechy dziedziczne, zwykle odpowiadający pojedynczemu białku lub RNA. Definicja zawiera całą funkcjonalną jednostkę, obejmującą kodujące sekwencje DNA, niekodujące sekwencje DNA oraz introny”.

Przykładem bardzo rozbudowanej definicji genu jest ogłoszona w latach 90. propozycja M. Singer i P. Berga: „Gen jest kombinacją segmentów DNA, które łącznie tworzą jednostkę mogącą ulec ekspresji, prowadzącej do utworzenia jednego lub więcej swoistych funkcjonalnych produktów genu, którymi mogą być cząsteczki RNA lub polipeptydów. Segmenty genu zawierają: 1) transkrybowany region (jednostka transkryp-

cji), który obejmuje sekwencje kodujące, sekwencje wtrącone, każde sekwencje wiążące 5' i sekwencje typu „trailer” na końcu 3', które flankują końce sekwencji kodujących oraz każde sekwencje regulatorowe zawarte w jednostce transkrypcji, 2) sekwencje regulatorowe, które flankują jednostkę transkrypcji i które są niezbędne dla swoistej ekspresji”.

H.M. Wein podaje następującą definicję: „Gen – segment DNA, który bierze udział w formowaniu fenotypu lub funkcji. W sytuacji braku obserwowalnej funkcji gen może być opisany przez sekwencję, transkrypcję lub homologie” (Wain i in., 2002). Takie rozszerzanie pojęcia genu stwarza dylemat klasyfikacji „genów” na różne, dobrze zdefiniowane klasy i problem ich liczebności w genomie poszczególnych gatunków.

Prawie każda definicja „genu” nadaje mu siłę sprawczą i pewien rodzaj autonomii, co z miejsca plasuje gen (DNA) na szczycie wszelkich zjawisk biologicznych, a zwłaszcza zjawisk związanych z dziedzicznością, rozwojem osobniczym i ewolucją.

„W tym momencie – pisze M. Gerstein – nie jest jasne, co należy zrobić. W skrajnym przypadku moglibyśmy zadeklarować, że koncepcja genu jest martwa i zaproponować coś zupełnie nowego, co pasuje do wszystkich danych” (Gerstein i in., 2007). A Jean Gayon dodaje: „Prawdopodobnie nie ma nadziei, aby stworzyć ogólną, zunifikowaną (ujednoliconą) koncepcję genu. W najlepszym razie taka koncepcja byłaby nieokreśloną, alternatywną listą wielu możliwych struktur i procesów” (Gayon, 2007).

Gerstein (Gerstein i in., 2007) mimo krytycznego stanowiska pokusił się jednak podać następną, własną definicję genu. Podaję tę definicję w oryginalnym brzmieniu: „A gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products”. Definicja Gersteina jest nadzwyczaj „pojemna”. Jak pokazuje ryc. 14, tą definicją objęto praktycznie oba pasma DNA i wszelkie sekwencje kodujące i niekodujące, a także sekwencje nakładające się. Powstaje tu dużo pytań wychodzących poza ramy tego artykułu.

Czym jest gen?

Wstępne wyniki projektu ENCODE pokazały następne pokłady wielkiej złożoności struktury, funkcji genomu i jego regulacji. Ta złożoność była postulowana intuicyjnie na podstawie wielu obserwacji i odkryć. Uznaje się ją powszechnie, chociaż jak pisze Gerstein (Gerstein i in., 2007), niektóre aspekty złożoności genowej były marginalizowane i „zamiatano je pod dywan”, aby nie naruszyć konstrukcji genocentrycznego obrazu obowiązującego w biologii.

Prosta definicja genu staje się problematyczna. Granice „klasycznego” genu zacieśniają się, stają się płynne. „Gen” rozciąga się również na regiony intergenowe, gdzie odkryto nowe miejsca regulatorowe, gdzie (jak również w intronach genów kodujących) znajdują się RNA niekodujące białka (ncRNA).

Nowe odkrycia i fakty poznane w ostatnich latach burzą naszą wiedzę o genach kodujących jako sekwencjach mających swoją stałą, fizyczną architekturę i granice, a także naruszają nasze dotychczas uporządkowane pojęcia o organizacji regionów DNA, zawierających geny kodujące białko. Te odkrycia komplikują koncepcję genu jako autonomicznej struktury molekularnej. Zapewne będą one wywierać głęboki wpływ na nauki medyczne, agrobiologię, zastosowania „inżynierii genetycznej” i, ogólnie, na rozumienie oraz kształtowanie filozofii przyrody.

W rozdziale „Między genem a białkiem” opisałem niektóre procesy molekularne, zmieniające pierwotną instrukcję dla syntezy białka zawartą w „genie” (DNA). Alternatywne składanie pre-mRNA, alternatywne promotory, zjawisko „dzielenia się genem”, redagowanie RNA, system NMD, interferencja RNA wskazują, że nie ma kolinearności między genem a białkiem, a także prostej zależności między genotypem a fenotypem. Przypisanie terminu „gen” do określonej sekwencji nukleotydów w DNA okazuje się być problematyczne. Informacja genetyczna realizuje się poprzez wielorakie interaktywne warstwy, poprzez dynamiczny komórkowy system złożony z samego DNA, białek i innych molekuł. Cząsteczka DNA jest ważnym, ale raczej biernym, choć stosunkowo stabilnym, elementem tego systemu. Przy pomocy tego systemu komórka realizuje programy zapewniające jej integralność i przystosowanie się do środowiska.

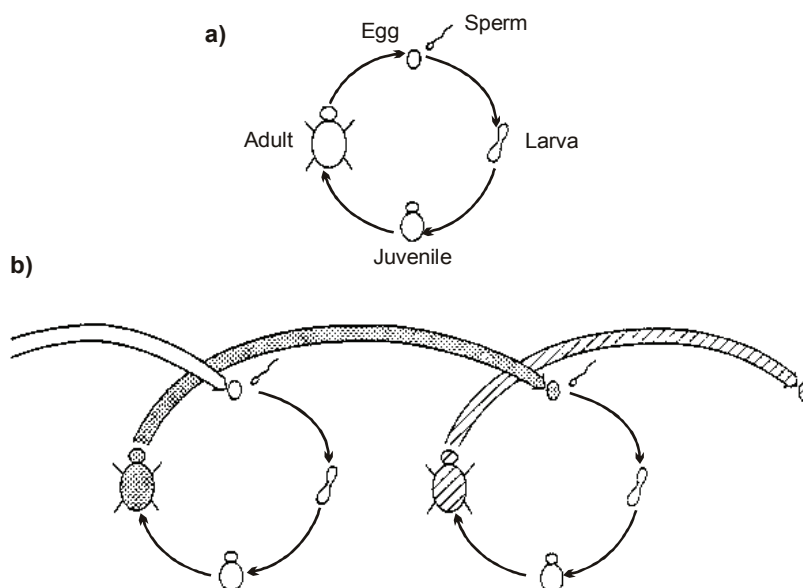
Żywa materia może istnieć tylko w formie komórki. To żywa komórka (narząd) określa, które geny mają być aktywowane, a które wyciszone, określa zapotrzebowanie na rozmaite funkcje białkowe w zależności od stanu fizjologicznego. Co więcej, to żywa komórka zmienia interpretację kodu. „Decyzja” uruchomienia mechanizmów przeróbki „pierwotnej informacji w DNA” na „informację wtórną” wychodzi z kontekstu i aktualnego stanu systemu operującego w komórce, a nie jest inicjowana przez DNA. „Ten sam gen ma wielorakie znaczenie, co jest przedmiotem interpretacji maszyneryjnej komórki w zależności od jej stanu” (Wang, 2005).

Zjawiska epigenetyczne, jak modyfikacja (metylacja) nukleotydów w DNA bez zmiany ich sekwencji, i specyficzne modyfikacje chromatyny (np. acetylacja, metylacja), generujące „pamięć” komórkową, również biorą udział w procesie dziedziczenia (Wierzbicki, 2004).

Międzypokoleniowa ciągłość linii germinalnej bazuje na ciągłości materiału genetycznego (DNA), ale również „na ciągłości fenotypu z całym jego ładunkiem historycznym”, o czym często się zapomina (West-Eberhard, 2003). Dziedziczna ciągłość fenotypu spinająca rodziców z potomstwem (*inherited bridging phenotype*), a jednocześnie poddająca się wpływowi środowiska jest zapewniona przez komórkę jajową matki (ryc. 15).

Cytoplazma komórki jajowej wnosi do rozwoju osobniczego system mitochondrialny generujący energię, wzorce wszystkich złożonych organelli komórkowych, systemów błoniastych, struktur szkieletowych, system syntezy białek, zestaw swoistych, informa-

cyjnych RNA dla uruchomienia najwcześniejszych etapów rozwoju zarodka, system reprogramowania „pamięci komórkowej” i wiele innych, a wreszcie podstawowe, samokontrolne i samoorganizujące się, bardzo słabo poznane systemy biologiczne. Komórka jajowa podlega jednocześnie wpływowi środowiska, które wywiera wpływ na jej stan. „System biologiczny” jest plastyczną koncepcją wymagającą jeszcze definicji i odpowiednich narzędzi do badań analitycznych (por. Keller, 2005). System biologiczny obejmuje złożone procesy komórkowe. System nie tkwi ani wyłącznie w genach, ani w białku, ale odnosi się do oddziaływań, regulacji i komunikacji między różnymi częściami komórkowymi. System ma charakter dynamiczny, ale cechuje się stabilnym atraktorem. Być może gen będzie pojmowany także w kategoriach procesu (Naumann-Held, 2001).



Ryc. 15. Ciągłość fenotypu. a) cykl życiowy zwierzęcia zwykle jest opisywany jako czasowy cykl form fenotypowych. Ciągłość międzygeneracyjna rozumiana jest jako funkcja dziedziczonych genów. b) osobnicze (indywidualne) cykle życiowe są połączone dla wykazania międzypokoleniowej ciągłości fenotypowej (duże strzałki). Komórka jajowa wnosi przed zapłodnieniem dziedziczenie struktur komórkowych i cały komórkowy system biologiczny. Według West-Eberhard, *Developmental plasticity and evolution*, Fig. 5.2, str. 92, Oxford University Press, Oxford-New York, 2003. Reprodukowano za zgodą Oxford University Press

Komórka jajowa matki wydaje się mieć jednocześnie funkcję „strażnika gatunku”. Doświadczenia nad klonowaniem pokazują, że jądro komórki somatycznej dawcy, pochodzące od osobnika gatunkowo obcego i noszące przecież „pełną informację” w swoim genomie, wprowadzone do enukleowanej (pozbawionej jądra) komórki jajowej nie ma

szans na zainicjowanie i rozwój klonu dawcy jądra mimo zawartego tam całego „planu rozwoju”. Sklonowanie dinozaura, nawet w przypadku gdyby uzyskano jego cały genomiczny DNA, jest mrzonką, gdyż brak jest żywej komórki jajowej dinozaura.

Cecha fenotypowa nie jest zwykle związana z jednym genem. Cecha fenotypowa często realizuje się przez zjawisko emergencji („wyłania się”), w którym biorą udział zarówno geny, białka, jak i inne cząsteczki. Złożoność (complexity) zjawiska życia – od biologicznej różnorodności po świadomość człowieka – nie daje się wyjaśnić jako zwykła suma działalności genów i innych cząsteczek. Zjawisko życia i porządek życia mogą być objaśnione jako fenomen emergencji („wyłaniania się”). Badanie komponentów życia: genów, innych makromolekuł, jonów itp. w izolacji, „samych w sobie” nie rozwiązuje zjawiska złożoności. Z takiej analizy nie można przewidzieć, jaki samorzutnie organizujący się układ powstanie. Ale to jest materia do rozważań w innym artykule.

Ewolucja, powstawanie „nowości” fenotypowych i pojawianie się gatunków nie polegają jedynie i wyłącznie (jak często przyjmuje się) na mutacji genów, warunkujących przetrwanie takich organizmów, które są najlepiej przystosowane do środowiska. Akomodacja fenotypowa (efekt Baldwina) do zmiennego środowiska i „asymilacja genetyczna” (West-Eberhard, 2003) mogą wpływać na kierunki ewolucji genetycznej. Podobnie mutacje nie są jedynymi i wyłącznymi czynnikami sprawczymi chorób dziedzicznie uwarunkowanych.

Złożoność „genu strukturalnego”, generowanie wtórnej informacji przez skomplikowane procesy komórkowe, złożoność procesów dziedziczenia i rozwoju osobniczego, pozagenowe dziedziczenia fenotypu, interakcja komórki ze środowiskiem i hierarchiczna organizacja organizmu człowieka narzucają konieczność krytycznego spojrzenia na genocentryczny paradygmat biologii (Keller, 2002; Łukow i Żekanowski, 2005). Pojmowanie „genu” na podstawie redukcjonistycznego podejścia i przekonanie o wyłącznie genetycznej determinacji zjawisk życiowych wymagają nowych przemyśleń.

Proste, liniowe przenoszenie informacji i „twardy”, genocentryczny determinizm nie objaśniają zjawisk życiowych (por. Paszewski, 2005; Jura i in., 2006). Nie oznacza to, że pewne elementy zjawisk życiowych (jak np. ogólny kierunek przepływu „informacji” od DNA do białek) tracą swoje znaczenie. Chociaż mogą czekać nas nowe odkrycia, np. odkrycie prionów wskazuje, że cząsteczki białka prionu mogą między sobą przekazywać „informację” o strukturze przestrzennej.

W dyskusji nad dziedzicznym przekazem nierzadko pomija się dziedziczenie kulturowe, które nie daje się precyzyjnie oddzielić od dziedziczenia genetycznego. Często również złożone funkcje przypisuje się bezpośrednio genom. Dotyczy to zwłaszcza wyższych czynności intelektualnych, talentów, cech charakteru, stanów emocjonalnych, zachowań itd. będących wytworem centralnego układu nerwowego, procesów wychowawczych i doświadczeń w trakcie rozwoju osobniczego. Człowiek jest organizmem o zło-

zonej, hierarchicznej architekturze. Nie można całej jego złożoności biologicznej zredukować do genu. Tym bardziej że ponownie powstaje pytanie: czym jest gen?

Podziękowania. Składam serdeczne podziękowanie Pani dr n. biol. Dorocie Butkiewicz za wykonanie korekty językowej, a Pani Beacie Bęben za wykonanie niektórych ilustracji i pomoc w przygotowaniu tekstu.

Literatura

- Adamala K. i Pikula S. (2004): *Hipotetyczna rola autokatalitycznych właściwości kwasów nukleinowych w procesie biogenezy*. „Kosmos” 53, 123-131.
- Akiva P., Toporik A., Edelheit S. et al. (2006): *Transcription-mediated gene fusion in the human genome*. „Genome Res.” 16, 30-36.
- Alberts A., Bray D., Lewis J. et al. *Molecular Biology of the Cell*. Garland Publishing Inc., New York & London, 1994.
- Alonso C. (2005): *Nonsense-mediated RNA decay: a molecular system micromanaging individual gene activities and suppressing genomic noise*. „BioEssays” 27, 463-466.
- Ayoubi T.A.Y., Van de Ven W. J.M. (1996): *Regulation of gene expression by alternative promoters*. „FASEB J.” 10, 453-460.
- Chargaff E. (1950): *Chemical specificity of nucleic acids and mechanism of their enzymatic degradation*. „Experiencia” VI/6, 201-209.
- Crick F.H.C., Barnett L., Brenner S., Watts-Tobin R.J. (1961): *General nature of the genetic code for proteins*. „Nature” 192, 1227-1232.
- Crick, F. *What Mad Pursuit – A Personal View of Scientific Discovery*, Basis Books, 1988
- Fire A., Xsu S., Montgomery M. K. et al. (1998): *Potent specific genetic interference by double-stranded RNA in *Caenorhabditis elegans**. *Nature* 391, 806-811.
- Erdmann A., Barciszewska M. Z., Hochberg A. et al. (2001): *Regulatory RNAs*. *Cell. Mol. Life Sci.* 58, 960-977.
- Frithl M. C., Valen E., Krogh A. et al. (2008): *A code for transcription initiation in mammalian genomes*. „Genome Res.” 18, 1-12.
- Gayon J., *The concept of the gene in contemporary biology: continuity or dissolution?* [w:] Fagot-Largeaut A., Rahman S., Toress J.M. (red.). *The Influence of Genetics on Contemporary Thinking*. Springer, Dordrecht, The Netherlands, 2007, str. 81-95
- Gerstein M.B., Bruce C., Rozowsky J. S. et al. (2007): *What is gene? post-ENCODE? History and updated definition*. „Genome Res.” 17, 669-681.
- Gingeras T.R. (2007): *Origin of phenotypes: Genes and transcripts*. *Genome Res.* 17, 682-690
- Gladyshev E., Meselson M., Arkhipova I.R. (2008): *Massive horizontal gene transfer in *Bdelloid Rotifers**. „Science” 320, 1210-1213.
- Jazdzewski K., Liyanarachchi S., Świerniak M. et al. (2009): *Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer*, *PNAS* 106, 1502-1505.
- Holbrook J.A., G. Neu-Ylik, M.W. Hentze, A.E. Kolozik (2004): *Nonsense-mediated decay approaches the clinic*. „Nature Genetics” 36, 801-808.
- Jura J., Węgrzyn P., Jura J., Koj A. (2006): *Regulatory mechanisms of gene expression: complexity with elements of deterministic chaos*. „Acta Bioch. Pol.” 53, 1-9.
- Jurka J., Kapitonov V.V., Kohany O., Jurka M.V. (2007): *Repetitive Sequences in Complex*

- Genomes: Structure and Evolution*. Ann. Rev. „Genomics Hum. Genet.” 8, 241-259.
- Kawakara Y., Zinshteyn B., Sethupathy P. et al. (2007): *Redirection of silencing targets by adenosine-to-inosine editing of miRNAs*. „Science” 315, 1137-1140.
- Kapitonov V.V., Pavlicek A., Jurka J., *Anthology of human repetitive DNA*. [w:] Meyers, R.A., (ed.), *Encyclopedia of Molecular Cell Biology and Molecular Medicine* (Vol. 1, pp. 251-305), wyd. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2004.
- Keller E.F., *The century of the gene*, Harvard University Press, Cambridge, London, 2002.
- Keller E.F. (2005): *The century beyond the gene*. „J. Biosci.” 30, 3-10.
- Koch Ch.M., Andrews, Flicek F. et al. (2007): *The landscape of histone modifications across 1% of the human genome in five human cell lines*, „Genome Res.” 17, 691-707.
- Kuzmiak H.A., Maquat L.E. (2006): *Applying nonsense-mediated mRNA decay research to the clinic: progress and challenges*, „Trends in Molecular Medicine” 12, 306-316.
- Lander E.S. et al. (2001): *Initial sequencing and analysis of the human genome*. „Nature” 409, 860-921.
- Lewontin R.C. *Biology as ideology – the doctrine of DNA*. (1991), HarperPerennial.
- Lewin B. *Genes V*. Oxford University Press, Oxford New York Tokyo, 1994.
- Łukow P., Żekanowski C. (2005): *Pojęcie genu i genocentryczny paradygmat biologii*. „Przeł. Filozof.” (Nowa Seria) 14, 85-105.
- Madghalchi S.M., Frischmeyer P.A., Mendell J.T. et al. (2001): *Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability*. „Human Molecular Genetics” 10, 99-105.
- Mendell J.T., Sharifi N.A., Meyers J.L. et al. (2004): *Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise*, „Nature Genetics” 36, 1073-1078
- Maquat L. (2005): *Nonsense-mediated mRNA decay in mammals*, J. Cell Science 118, 1773-1776
- Mattick J. S., and Gagen M. J. (2001): *The Evolution of Controlled Multitasked Gene Networks: The Role of Introns and Other Noncoding RNAs in the Development of Complex Organisms*. „Mol. Biol. Evol.” 18(9), 1611-1630.
- Mattick J.S. (2001): *Non-coding RNAs: the architects of eukaryotic complexity*. „EMBO Reports” 2, 986-991.
- Mattick J.S. (2003): *Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms*. „BioEssays” 25, 930-999.
- Morange M.: *A History of Molecular Biology*. Harvard University Press, Cambridge-London, 1998.
- Naumann-Held E. M. (2001): *Let's talk about genes: the process molecular gene concept and its context*. [w:] Oyama S., Griffiths P.E., Grey R.D. (eds.), *Cycles in contingency: developmental systems and evolution*, MIT Press, Cambridge, Mass., 2001, pp. 69-84
- Neves G., Zucker J., Daly M. & Chess A. (2004): *Stochastic yet biased expression of multiple Dscam splice variants by individual cells*. „Nature Genetics” 36, 240-246
- Nirenberg M.W., and Matthaei J.H. (1961): *The dependence of cell-free protein synthesis in E. Coli upon naturally occurring or synthetic poliribonucleotides*. „Proc. Natl. Acad. Sci.” 47, 1588-1602
- Paszewski A. (2005): *Co zdeterminowane, a co przypadkowe w systemach biologicznych – gdzie zaczyna się wolność?* „Nauka” 1/2005, 53-66
- Perrin-Vidoz L., Sinilnikova O.M., Stoppa-Lyonnet D. et al. (2002): *The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature*

- termination codons*. „Human Molecular Genetics” 11, 2805-2814.
- Pheasant M., Mattick J.S. (2007): *Raising the estimate of functional human sequences*. „Genome Res.” 17, 1245-1253.
- Piatigorsky J: *Gene sharing and evolution – The diversity of protein function*. Harvard University Press, Cambridge – Massachusetts, London, 2007).
- Schmuckler D., Clemes J.C., Shu H. et al. (2000): *Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity*. „Cell” 101, 671-684.
- Schrödinger E. *Czym jest życie?* Pruszyński i Sk-a, Warszawa, 1998.
- Smith H.C., Gott J.M., Hanson M. R. (1997): *A guide to RNA editing*, „RNA” 3, 1106-1123
- Stuard K., Panigrahi A.K. (2002): „Molecular Microbiology” 45, 591-696.
- Szymański M., Barciszewski J. (2002): *Beyond the proteome: non-coding regulatory RNAs*, *Genome Biology*, <<http://genomebiology.com/2002/3/5/reviews/0005>>
- Szymański M., Barciszewska M.Z., Żywicki M., Barciszewski J. (2003): *Noncoding RNA transcripts*. „J. Appl. Genet.” 44, 1-19).
- Taylor J.H., (red). *Selected Papers on Molecular Genetics*, Academic Press, New York and London, 1965 (replika 55 wybranych oryginalnych. prac dotyczących, genetyki, poszukiwań materiału genetycznego, struktury i replikacji DNA, funkcji materiału genetycznego itp.)
- Tress M.L., Martelli P.L., Frankish A. in. (2007): *The implications of alternative splicing in the ENCODE protein complement*. „PNAS” 104, 5495-5500.
- Wain H.M., Bruford E.A., Lovering R.C. et al. (2002). *Guidelines for human gene nomenclature*. „Genomics” 79, 464-470.
- Wang D. (2005): „*Molecular gene*”: *interpretation in the right context*. „Biology and Philosophy” 20, 453-464.
- Watson J.D., Crick F.H.C. (1953): *Molecular structure of nucleic acids: a structure of deoxy-ribose nucleic acid*. „Nature” 171, 737-738.
- Watson J.D. *The double helix*, Weidenfeld and Nicholson, London, 1968.
- West-Eberhard, M.J. *Developmental plasticity and evolution*. Oxford University Press, Oxford – New York, 2003.
- Wierzbicki A.T. (2004) *Dziedziczenie epigenetyczne*, „Kosmos”, 53, 271-280).
- Zalewski Z. (2001): *Mapa ludzkiego genomu – złudne(?) nadzieje i realne(?) zagrożenia*. „Sztuka Leczenia” 7, 65-75.
- Zheng D., Gerstein M.B. (2007): *The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they?* „Trends in Genetics” 23, 219-224.
- Venter J.C. (2001): *The sequence of the human genome*. „Science” 291, 1304-1351).
- Xu Q., Modrek B., Lee C. (2002): *Genome-wide detection of tissue-specific alternative splicing in the human transcriptome*. „Nucleic Acids Res.” 30, 3754-3766.

A structural gene – evolving term and dilemmas

The term “gene” was originally used as a purely theoretical concept. After discovery of DNA structure, and understanding the genetic code, the gene acquired a form of a distinct physical entity with its borders and specific signal sequences, having rather simple (as it was thought at that time) functions and relation to phenotype outcome. The term “structural gene” has been coined. The unique gene structure, and several unusual and omnipotent traits have been ascribed to the gene that resulted in the formulation of a “genocentric” theory as a basic expla-

nation of all features of living organisms. However, recent discoveries reveal a complex structure and functions of eukaryotic genes. It appears now that the coding sequences (exons) are spread out over extended regions (hundreds of thousands of kilobases) of DNA. The role of protein-non coding DNA sequences were recognized, and the new mechanisms controlling gene functions have been discovered. In addition, we acquired the knowledge about a powerful ability of the cell to interpret the information potential of genes, accordingly to the needs of a cell/organ or actual "context" and status of the dynamic systems operating within the cell. All these discoveries undermine the genocentric view of life. At this time any definition of "gene" seems to be inadequate with present knowledge, and one may ask again: what is a gene?

Key words: DNA, structural gene, RNA transcripts, non-protein coding RNA, "secondary" information