

Genom ludzki ma głównie niekodujących charakter, co oznacza, że 99% sekwencji nie ma zdolności do wytwarzania białek – złożonych cząsteczek, które są niezbędne z punktu widzenia struktury i funkcji komórek. Co ciekawe, niekodujące DNA wytwarza różne typy RNA, mających zdolność kontrolowania aktywności genów kodujących białka (~19,000) na różnych poziomach. Długie niekodujące RNA (ang. *long noncoding RNAs*, lncRNAs) – cząsteczki RNA dłuższe niż 200 nukleotydów, które nie wytwarzają żadnego funkcjonalnego białka – to najbardziej zagadkowa i jednocześnie najbardziej intrygująca klasa niekodujących RNA. Chociaż wiele lncRNA odgrywa rolę w kluczowych procesach fizjologicznych i patologicznych, wciąż niewiele o nich wiemy. Nawet nie mamy pewności, jaka jest ich całkowita liczba w naszym genomie. Brakuje nam także sposobu na precyzyjne określenie, które lncRNA są funkcjonalne. W rezultacie, choć w naszym genomie zidentyfikowano już ponad 100,000 genów produkujących niekodujące RNA, jedynie <2% z nich jest scharakteryzowanych funkcjonalnie. Stąd też, ujawnienie ich pełnego, biologicznego potencjału wymaga odpowiedzi na kluczowe pytanie: „*Jak funkcje lncRNA są przechowywane w ich pierwotnej sekwencji?*”.

Najpopularniejsza w biologii definicja „funkcji” odnosi się do aktywności molekuł, narządów i układów. Jednak w biologii ewolucyjnej „funkcja” jest przyczyną obecności jakiegoś obiektu lub procesu w systemie. Ewolucja to proces, który zachowuje kluczowe funkcje biologiczne przekazując je jako fragmenty sekwencji DNA z jednego genomu do drugiego na przestrzeni milionów lat. Dlatego też zachowawczość ewolucyjna często oznacza obecność podobnych genów, ich fragmentów lub nawet dużych fragmentów DNA w genomach różnych gatunków. Zatem zachowawczość ewolucyjna odzwierciedla nie tylko wspólne pochodzenie gatunków, ale także funkcjonalne znaczenie fragmentów DNA, które zostały zachowane. Porównania międzygatunkowe okazały się być skuteczną metodą identyfikacji funkcjonalnych regionów w genomie. Przez lata badania oparte na zachowawczości ewolucyjnej znacznie pogłębiły naszą wiedzę na temat genów kodujących białka. Dlatego też oczekuje się, że dostarczą one również informacji na temat funkcjonalności lncRNA. Jednakże w przeciwieństwie do większości silnie zachowanych genów kodujących białka, lncRNA ewoluują szybko i jedynie <100 z nich ma wykrywalną zachowawczość sekwencji pomiędzy człowiekiem, a danio pręgowanym – małą tropikalną rybką. Jednocześnie tysiące lncRNA wykazują zachowawczość pozycyjną w tych gatunkach bez wykrywalnych podobieństw na poziomie sekwencji, co oznacza, że natura ma tendencję do utrwalania lokalizacji i kolejności lncRNA, a nie ich sekwencji. Niektóre z zachowanych pozycyjnie lncRNA zawierają krótkie fragmenty sekwencji (<20 nt), które także są utrwalone ewolucyjnie. Nie są one jednak wykrywane przez standardowe narzędzia, które zaprojektowane zostały do analizy porównawczej genów kodujących białka, posiadających długie odcinki DNA o wysokim poziomie zachowawczości. Inną ważną różnicą pomiędzy lncRNA, a genami kodującymi białka jest sposób ich przetwarzania przez komórkę. Droga od genu do białka jest niemalże identyczna dla wszystkich genów kodujących białka. Najpierw informacja zakodowana w DNA przepisywana jest na RNA w jądrze, a następnie transportowana do cytoplazmy, gdzie wytwarzane jest białko. lncRNA mają znacznie bardziej różnorodne sposoby przetwarzania przez komórkę, ponieważ specyficzna lokalizacja komórkowa jest podstawą ich funkcji. W przeciwieństwie do genów kodujących białka, końcowym produktem aktywności genu lncRNA jest dojrzała cząsteczka RNA. Ze względu na zmniejszoną stabilność, RNA musi być natychmiast dostarczone do określonego przedziału komórkowego – miejsca jego działania. W rezultacie populacje lncRNA charakteryzują się różnorodnymi lokalizacjami komórkowymi, obejmującymi jądro, cytoplazmę, mitochondria, itp. Co ciekawe odstępnie badania pokazują, że zachowane cząsteczki lncRNA są także inaczej przetwarzane przez komórki różnych gatunków. Dany lncRNA jest eksportowany do cytoplazmy przez ludzkie, a zatrzymywany w jądrze przez mysie komórki przez co charakteryzuje się obniżonym potencjałem funkcjonalnym. Nasuwa to kolejne ważne pytanie: „*Czy zachowane lncRNA są naprawdę funkcjonalne?*”.

Celem tego projektu jest zbadanie potencjału funkcjonalnego zachowanych pozycyjnie lncRNA u kręgowców. Najpierw zastosujemy nasz nowy program ConnectOR (Connecting RNA Orthologues) to wykrycia zachowanych pozycyjnie lncRNA w genomach człowieka i myszy. Następnie opracujemy nową metodę katalogowania lncRNA w rozdzielczości subkomórkowej w celu ich identyfikacji w wybranych typach komórek. Porównanie składu sekwencji na poziomie subkomórkowym pozwoli zidentyfikować ciągi sekwencji, które stanowią podstawę funkcji lncRNA. Zbadamy nie tylko ich wkład w napędzanie konkretnych lokalizacji subkomórkowych, ale po raz pierwszy ujawnimy w jaki sposób potencjał funkcjonalny lncRNA zmienia się pomiędzy blisko i daleko spokrewnionymi gatunkami kręgowców dzięki włączeniu do tych porównań danio pręgowanego. Na koniec eksperymentalnie zbadamy potencjał funkcjonalny zachowanych ewolucyjnie lncRNA za pomocą modeli komórkowych i mysich.

Podsumowując, jest to innowacyjny i ambitny projekt, który pomoże lepiej zrozumieć, w jaki sposób funkcje lncRNA przechowywane są w ich pierwotnej sekwencji. Co więcej, pozwoli nam badać funkcje lncRNA z nowej perspektywy – „*dlaczego tam jest?*”, co jest dokładniejszym uzasadnieniem obecności elementów funkcjonalnych w genomie, niż obecnie stosowana strategia „*co robi?*”.