

ВЕЛИКИЙ ПРАКТИКУМ З ЗАГАЛЬНОЇ ГЕНЕТИКИ

ЛЕКЦІЯ 1

Тема: Сутність поняття ген

Основним об'єктом генетики є гени (від γένος - рід, походження) - їхня структурно-функціональна організація, закономірності їхнього збереження, зміни та передачі нащадкам, а також принципи реалізації записаної в генах спадкової програми та взаємодій між генами в єдиній складній системі апарату спадковості. У загальному визначенні ген - це окремий елемент спадкової програми.

Термін "ген" було запропоновано Йоханнсенем у 1909 р., невдовзі після відкриття заново менделівських законів спадковості. Із тих часів поняття гена кілька разів піддавалося суттєвій ревізії. Всі вони мають право на існування.

Матеріальною основою гена є молекула дезоксирибонуклеїнової кислоти - ДНК. Як фізичний об'єкт ген - це ділянка полімерної молекули дезоксирибонуклеїнової кислоти (ДНК). Єдиним винятком із цього загального для всіх живих організмів правила є особливий клас вірусів, де роль носія інформації виконує інша нуклеїнова кислота (хімічний аналог ДНК) - рибонуклеїнова кислота (РНК). Численні типи РНК беруть також участь у процесах реалізації генетичної інформації в усіх організмах. Фізико-хімічні властивості молекули ДНК роблять її надійним носієм спадкової інформації та забезпечують простий механізм реплікації - точного відновлення молекули (а отже, і спадкової програми) у дочірніх клітинах при клітинному поділі. Численні взаємодії, в які вступає ДНК з іншими молекулами (насамперед білками), зумовлюють певну схему упаковки ДНК у клітинах, експресію генетичної інформації та регуляцію цього процесу, а також здійснення інших операцій, суттєвих для функціонування спадкового апарату.

Ген - це окрема змістовна ділянка ДНК, у послідовності якої закодована амінокислотна послідовність білка. Проте, крім білкових, існує також велика кількість генів, кінцевими продуктами яких є різноманітні молекули РНК. Але незалежно від типу гена, первинним продуктом його активності (проміжним для білкових генів) є молекула РНК.

Як джерело інформації ген - це ділянка ДНК, у послідовності нуклеотидів якої записано інформацію про певний функціональний продукт. За типом цього продукту всі гени (сукупність генів даного організму називають генотипом) можна поділити на дві групи: гени, кінцевим продуктом яких є певні функціональні молекули РНК (гени РНК), і гени, у послідовності яких відповідно до генетичного коду записано інформацію про послідовність амінокислот у складі білків (білкові гени). Гени РНК кодують різноманітні молекули РНК, що не піддаються трансляції: рРНК – рибосомні РНК (компоненти рибосом); тРНК - транспортні РНК (ключовий елемент системи трансляції); маленькі ядерні РНК; маленькі ядерцеві РНК; мікро-РНК; молекули РНК, що є компонентами деяких ферментів; інші види РНК, із яких ще не для всіх з'ясовані їхні функції.

ВЕЛИКИЙ ПРАКТИКУМ З ЗАГАЛЬНОЇ ГЕНЕТИКИ

На білковому гені синтезується РНК-матриця для наступного синтезу білка - матрична (або інформаційна) РНК - мРНК.

Тож можна сказати, що ген - це ділянка ДНК, яка є необхідною і достатньою для повноцінного синтезу функціональної молекули РНК.

Але важливо розуміти, що кожне з них має в той самий час досить суттєві обмеження. Беззаперечним є лише те, що ген є ділянкою ДНК (і навіть це твердження потребує уточнень стосовно вірусів, що містять РНК як генетичний матеріал, проте таке визначення нічого не говорить про властивості, які мають бути притаманні цій ділянці, щоб її можна було вважати геном.

Від початку існування термін інтерпретувався в дусі Менделя, тобто як дискретна й нероздільна спадкова одиниця, яка не залежить від інших спадкових одиниць і відповідає за прояв певної ознаки.

В окремих випадках така інтерпретація може бути застосована, хоча зрозуміло, що про незалежність спадкування різних генів можна говорити тільки за умови, що гени містяться в різних хромосомах.

Зрозуміло також, що ген не є дискретним і нероздільним: ген як ділянка ДНК має довжину, ця ділянка може бути розділеною на фрагменти. У процесі своєї експресії гени взаємодіють на різних рівнях: активація транскрипції та шлях сплайсингу залежать від активності генів транскрипційних факторів і сплайсинг-регуляторів, трансляція мРНК - від активності генів білків-регуляторів трансляції; білкові продукти різних генів взаємодіють тощо. У результаті активність одного гена може підсилити чи пригнітити експресію іншого, і взагалі, для прояву ознаки, як правило, є необхідною активність кількох генів. Останнім часом набула популярності комп'ютерна метафора, згідно з якою спадковий апарат (геном і систему його експресії) можна розглядати як "операційну систему", що керує організмом, а ген - як "підпрограму" цієї системи.

Після того, як було з'ясовано, що гени знаходяться в хромосомах, ген став розглядатися як хромосомний локус, а сама хромосома - як лінійна комбінація генів, що не перекриваються. Часто так воно і є, але, з іншого боку, один локус (одна ділянка хромосомної ДНК) може містити кілька генів - гени перекриваються або за рахунок перекриття рамок зчитування (деякі бактеріофаги та окремі гени еукаріотів), або за рахунок розташування генів (в еукаріотичних геномах) у межах інтрону іншого гена, або за рахунок розташування двох кодуючих послідовностей на одній ділянці ДНК на двох різних ланцюгах. Крім того, концепція гена як локусу не може бути застосована для мобільних елементів - ділянок ДНК (які часто містять один або кілька генів), що можуть змінювати свою локалізацію в геномі.

Розвиток молекулярної біології спочатку привів до розуміння того, що ген - це ділянка молекули ДНК, яка відповідає за синтез молекули білка ("один ген - один білок"). Іноді це справді так, проте сьогодні вже ясно, що, по-перше, не менш важливими є гени, що кодують різноманітні РНК, які не піддаються трансляції. По-друге, одна ділянка ДНК (сукупність екзонів еукаріотичного гена) часто дає кілька білкових продуктів за рахунок альтернативного

ВЕЛИКИЙ ПРАКТИКУМ З ЗАГАЛЬНОЇ ГЕНЕТИКИ

сплайсингу. Різниця між біологічними видами часто зумовлена не тільки і не стільки різницею в наборах кодуючих послідовностей (екзонів), скільки різними комбінаціями цих екзонів. Причому таке перекомбінування можливо як на рівні ДНК, так і на рівні кінцевих транскриптів. Отже, якщо трактувати ген як спадковий фактор, то ген - це не тільки ділянка ДНК, що містить певну інформацію, а й система експресії цієї інформації.

Інтенсивний розвиток нової дисципліни - геноміки, що спрямована на встановлення та аналіз нуклеотидних послідовностей цілих геномів, зумовив тенденцію розглядати ген як анотовану геномну ділянку з певними властивостями. Згідно з визначенням міжнародного консорціуму онтології послідовностей (Sequence Ontology Consortium), ген - це певна визначена зона геномної послідовності, яка відповідає одиниці спадковості й містить регуляторні ділянки та ділянки, що транскрибуються. Під словами "одиниця спадковості" слід розуміти той факт, що ген кодує певні (один або кілька) функціональні продукти (білки або молекули РНК, що не транслюються). Під "ділянкою, що транскрибується" мається на увазі певна група екзонів, з'єднаних інтронами, яка транскрибується як одне ціле. При цьому за правилами анотації послідовностей, прийнятими сучасними базами даних геномних послідовностей, первинні транскрипти, котрі піддаються альтернативному сплайсингу, вважаються такими, що належать одному гену, навіть якщо кінцеві білки є різними. Тобто ген - це група екзонів, що транскрибуються разом, або ген - це ділянка геному, що дає набір кінцевих транскриптів, які містять хоча б один спільний екзон. Нарешті, важливим моментом наведеного визначення гена є те, що до складу цієї елементарної одиниці спадковості організму прийнято включати регуляторні ділянки, які контролюють його активність.

З тим уточненням, що в прокаріотичних системах (у випадку оперонів) регуляторні ділянки можуть контролювати групу генів, таке трактування гена до самого останнього часу залишалося загальноприйнятим.

З метою ретельного аналізу інформації, яка записана в геномі й реалізується внаслідок транскрипції, Національним інститутом США з вивчення геному людини (National Human Genome Research Institute) було започатковано міжнародний проект ENCODE (Encyclopedia of DNA Elements), одним з основних завдань якого є тотальний аналіз транскриптому людини. С початку було завершено перший етап цієї роботи - проаналізовано функціонування 1 % геному людини (приблизно 30 млн пар основ), і перші результати принесли низку несподіванок.

По-перше, відносна кількість ДНК, що піддається транскрипції, виявилася неочікувано високою - на рівні 80 %. Зважаючи на частку геному, яку становлять екзони разом з інтронами, виникає питання, чи всі ці транскрипти відповідають генам? Напевно, частина цих первинних транскриптів є просто наслідком неспецифічної хаотичної активності РНК-полімераз. Проте значна частка транскриптів містить консервативні (для ссавців і серед популяцій людини) елементи послідовності - близько 60 % таких елементів знаходиться поза межами відомих раніше білкових генів чи

ВЕЛИКИЙ ПРАКТИКУМ З ЗАГАЛЬНОЇ ГЕНЕТИКИ

регуляторних ділянок. Імовірно, у багатьох випадках ці транскрипти є невідомими раніше РНК, які не піддаються трансляції, для більшості з них їхнє функціональне значення ще має бути з'ясовано. Крім того, було встановлено, що часто піддаються транскрипції регуляторні ділянки (промотори, енхансери). Імовірно, така транскрипція є просто одним із способів підтримувати регуляторну ділянку в доступному стані деконденсованої хроматинової фібрили.

Для відомих білкових генів (399 у дослідженій частині геному) було продемонстровано наявність великої кількості невідомих раніше стартових точок транскрипції. Приблизно половина генів мають альтернативні стартові точки, які відстоять на 100 тис. пар основ від анованих раніше стартів транскрипції цих генів. Деякі з цих стартових точок використовують промотори інших генів: одна така точка може бути спільною для двох чи трьох генів, і первинний транскрипт іноді містить кілька генних локусів - груп екзонів.

Крім того, для більшості білкових генів аналіз їхніх транскриптів показує наявність невідомих раніше екзонів. Деякі з цих екзонів розташовані на відстані до кількох тисяч пар основ від усіх інших екзонів гена, іноді опиняючись у межах іншого гена. Більшою за очікувану виявилася кількість різних ізоформ мРНК, що виникають як унаслідок альтернативного сплайсингу, так і трансплайсингу.

Результати проекту ENCODE вказують на диспергований характер розподілу регуляторних елементів по всьому геному: багато регуляторних елементів розташовано всередині екзонів та інтронів, при цьому вони можуть бути елементами системи регуляції зовсім іншого гена.

Отже, поняття гена знову потребує певної ревізії. Відповідно до одного з нещодавно запропонованих визначень, ген - це об'єднання (union) геномних послідовностей, що кодують зв'язний (coherent) набір функціональних продуктів, які можуть частково перекриватися. Під когерентністю набору продуктів мається на увазі, що у випадку білкових генів кожен екзон є спільним хоча б для двох продуктів даного набору. Головний акцент у цьому визначенні робиться на кінцеві продукти активності генів - перекриття між проміжними транскриптами ігнорується. Якщо виходити з перекриття між первинними транскриптами (інтерпретувати ген як кластер екзонів, що можуть спільно транскрибуватися), то, інколи, на геномній ділянці слід було б визначити лише два гени. Якщо ж виходити з перекриття між кінцевими продуктами, то на тій самій ділянці міститься принаймні шість генів.

Тобто ген не обов'язково складається з екзонів, що розташовані поряд; група екзонів одного гена може бути диспергованою по геномній зоні, і окремі екзони групи можуть одночасно належати до інших генів. Зрозуміло, що у простому випадку, коли немає альтернативного сплайсингу (або взагалі немає інтронів), визначення зводиться до класичного: ген - це ділянка ДНК, що кодує молекулу білка або РНК.

Крім того, згідно з визначенням, що обговорюється, регуляторні елементи не є компонентами гена (пропонується називати їх "елементами,

ВЕЛИКИЙ ПРАКТИКУМ З ЗАГАЛЬНОЇ ГЕНЕТИКИ

асоційованими з генами" - gene-associated): система регуляції є складнішою, ніж просте співвідношення "один до одного" між регуляторними елементами та кодуючими ділянками. Наведене визначення, звичайно, не може вважатися остаточним. Очевидно, що поняття гена є надто складним, щоб його можна було чітко сформулювати: різні визначення, які акцентують увагу на різних властивостях генів різних типів, є, відповідно до принципу додатковості, одночасно справедливими.