

## МЕХАНІЗМИ ІНТРОН-ОПОСЕРЕДКОВАНОГО ПОСИЛЕННЯ ЕКСПРЕСІЇ: «ЛАСКАВО ПРОСИМО ДО ГОТЕЛЮ КАЛІФОРНІЯ»

ПІДЮРА М.О.<sup>1,2\*</sup>, БЛЮМ Я.Б.<sup>1\*</sup>

<sup>1</sup> Інститут харчової біотехнології та геноміки НАН України, вул. Байди-Вишневецького (Осиповського), 2А, 04123, Київ, Україна

<sup>2</sup> АТ «Фармак», вул. Кирилівська, 63, 04080, Київ, Україна

E-mail: pydiura@gmail.com, cellbio@cellbio.freenet.viaduk.net

*Явище позитивного впливу інтронів на експресію гена, в якому вони знаходяться, що отримало назву інтрон-опосередкованого посилення (ІМЕ), характерне для широкого кола різноманітних організмів, включаючи нематод, комах, ссавців, грибів та рослин, відбувається завдяки досі невизначеному фундаментальному механізму. ІМЕ інтрони вже тривалий час використовуються, зокрема, у біотехнології рослин. Розуміння механізмів даного явища дозволить передбачати та легко генерувати стимулюючі інтрони із заданими властивостями та створювати дуже вигідні фенотипи. Це також увімкне зелене світло для використання ІМЕ у генній терапії та для покращення виробництва фармацевтичних білків. У огляді проаналізовано раніше запропоновані моделі механізмів функціонування ІМЕ і виділено фактори, які можуть напряду або опосередковано визначати ІМЕ за різних умов і на різних рівнях експресії генів, таких як експериментальні методи дослідження ІМЕ, регуляторні РНК, властивості послідовності, позиція та орієнтація інтронів, фактори на рівні ДНК, транскрипції, сплайсингу, мРНК, трансляції, гени, у яких виявляється ІМЕ, тканинна специфічність, репресія та співвідношення за значенням між деякими з факторів. Оскільки не існує єдиного механізму ІМЕ, і ефект може відрізнятися у різних видів, при моделюванні цього процесу слід порівнювати між собою лише випадки ІМЕ, які впливають на експресію на одному і тому ж самому рівні, враховуючи експериментальні умови. Виділення біологічних факторів, які можуть визначати ІМЕ, та співвідношень між ними допоможе в подальшому створити відповідний набір даних, зручний для машинного навчання, та спробувати розгадати таємницю феномену ІМЕ за допомогою машинного навчання.*

**Ключові слова:** *Інтрон, експресія гена, інтрон-опосередковане посилення, ІМЕ, фактори ІМЕ, регуляція експресії, машинне навчання.*

### Вступ

Інтрони є еволюційно консервативними, що вказує на їх фундаментальну та важливу роль у клітині (Carmel and Chorev, 2012; Rogozin et al, 2012). Термін «інтрон-опосередковане посилення» (ІМЕ) було запропоновано для опису позитивного впливу інтронів на експресію генів у невідомий спосіб, який відрізняється від більш відомих механізмів дії промоторів, енхансерів, сайленсерів і модифікацій хроматину (Mascarenhas et al, 1990). У випадку ІМЕ інтрони стимулюють експресію одного єдиного гена, в якому знаходяться, і представляють новий тип регуляторного елемента. Вони відрізняються від енхансерів, які активують транскрипцію з промоторних елементів на великих відстанях в обох напрямках, а також від промоторів, оскільки для того, щоб впливати на експресію, послідовності, які відповідають за ІМЕ, повинні бути розташовані після місця початку транскрипції (Rose, 2019). У зв'язку з цим для позначення ІМЕ інтронів було введено термін «стимулюючий інтрон» (Rose et al, 2011; Emami et al, 2013).

При описанні посилюючого ефекту інтрона 1 алкогільдегідрогенази-1 (*ADHI*) на експресію генів у кукурудзи було узагальнено ознаки ІМЕ, які згодом стали визначальними: 1) декілька різних кодуючих послідовностей (включаючи бактеріальні гени) реагують на інтрони; 2) ІМЕ не обмежується конкретним промотором; 3) кілька різних інтронів мають

здатність стимулювати експресію одного гена; 4) перші інтрони мають тенденцію посилювати експресію більше, ніж наступні інтрони; 5) інтрони підвищують експресію лише в межах транскрибованих послідовностей і в правильній орієнтації; 6) інтрони не мають посилюючого ефекту, якщо розміщені у транскрибованих послідовностях на 3'-кінці гена; 7) ефект інтронів проявляється як збільшення накопичення мРНК (Callis et al, 1987). У той же час було показано, що сплайсинг, хоча і необхідний, але є недостатнім для забезпечення ІМЕ (Rose, 2002; Morello et al, 2011).

Морелло та Брєвіаріо (Morello and Breviario, 2008) узагальнили нові біологічні функції інтронів у світлі даних, накопичених на той час наступним чином:

1) інтрони, особливо перші інтрони, можуть містити сайти зв'язування факторів транскрипції або можуть функціонувати як класичні енхансери транскрипції;

2) у рослин кілька інтронів можуть впливати на експресію власних генів шляхом збільшення рівня транскрипту. Це часто спостерігається для першого або другого інтрона даної одиниці транскрипції;

3) деякі інтрони також відповідають за експресію генів, специфічну для певних тканин або стадій розвитку;

4) деякі інтрони, головним чином перші інтрони, можуть мати властивості промоторів для синтезу альтернативних мРНК. Коли такі інтрони-промотори знаходяться у інтронах-лідерах, альтернативні транскрипти відрізняються лише своїм першим, некодуєчим екзоном;

5) інтрони можуть вивільняти фактори, які працюють у трансположенні, такі, як мікроРНК (miRNA) і малі ядерцеві РНК (snoRNA).

Перелічені пункти 2–4 стосуються саме ІМЕ. Крім того, не можна виключити роль мікроРНК і малих ядерцевих РНК як механізму, пов'язаного з ІМЕ. Ці ж автори зазначають, що вплив ІМЕ на транскрипцію залежить від нуклеотидного складу інтрона та фланкуючих екзонів, використовуваного гена-репортера, використовуваного промотора та положення інтрона (Morello and Breviario, 2008).

Також було показано, що ІМЕ підвищує рівень трансляції. Зокрема, у *Arabidopsis* інтрон 1 гена убіхітину *UBQ10*, інтрон 1 гена серин/

треонінової протеїнкінази (*AtPK1*), інтрони 1 і 6 гена фосфорибозилантранілат ізомерази *TRP1* (*PAT1*) (Rose, 2004), гени субодиниці 5С цитохром с оксидази *COX5c-1* і *COX5c-2* (Curi et al, 2005) та інтрон гена *ADH1* кукурудзи (Mascarenhas et al, 1990) викликають більш виражене підвищення активності ферменту, ніж це можна було б пояснити лише рівнями мРНК (Rose, 2008). Крім того, у деяких випадках спостерігається пов'язана з ІМЕ тканинно-специфічна регуляція експресії, і тому було введено термін «інтрон-залежна просторова експресія» (Morello and Breviario, 2008).

Загалом, стало очевидним, що інтрони можуть функціонувати у декілька різних способів, а також в залежності від різного генного контексту, і тому ІМЕ є складним явищем, яке відображає фундаментальний механізм експресії генів, притаманний широкому ряду різних організмів, включаючи нематод, комах, ссавців, грибів та рослин (Rose, 2008; Morello and Breviario, 2008). Термін «стимулюючий інтрон» було запропоновано використовувати стосовно інтронів, які посилюють експресію генів (Rose et al, 2011; Emami et al, 2013).

Для аналізу інтронів і прогнозування їх ІМЕ впливу на експресію генів було створено першу версію обчислювального алгоритму ІМЕter (Rose et al, 2008; Korf and Rose, 2009; Parra et al, 2011). Передбачувальна здатність ІМЕter була експериментально перевірена для інтронів-лідерів генів *OsTUB6*, *OsTUB4*, *OsCPK2* та інтрона 1 *OsTUA2* і *OsTUA3* рису (Morello et al, 2011). Виявилось, що рівень експресії гена *gus* у тимчасово трансформованих калюсах рису суттєво не корелює з розрахованою оцінкою ІМЕter (Morello et al, 2011). Автори зазначили три моменти, які свідчать про те, що механізм ІМЕ працює на рівні РНК: 1) позиція та орієнтація інтрона важливі; 2) ІМЕ в рослинах є переважно ко- та пост-транскрипційною подією; 3) Властивості ІМЕ є специфічними та не є спільними для всіх інтронів (Morello et al, 2011).

Інтрони ІМЕ значно підвищують стабільний рівень мРНК за допомогою невідомого механізму, не впливаючи істотно на швидкість транскрипції (Dean et al, 1989; Rose and Last, 1997; Rose and Beliakoff, 2000) і стабільність мРНК (Rethmeier et al, 1997). Перший інтрон

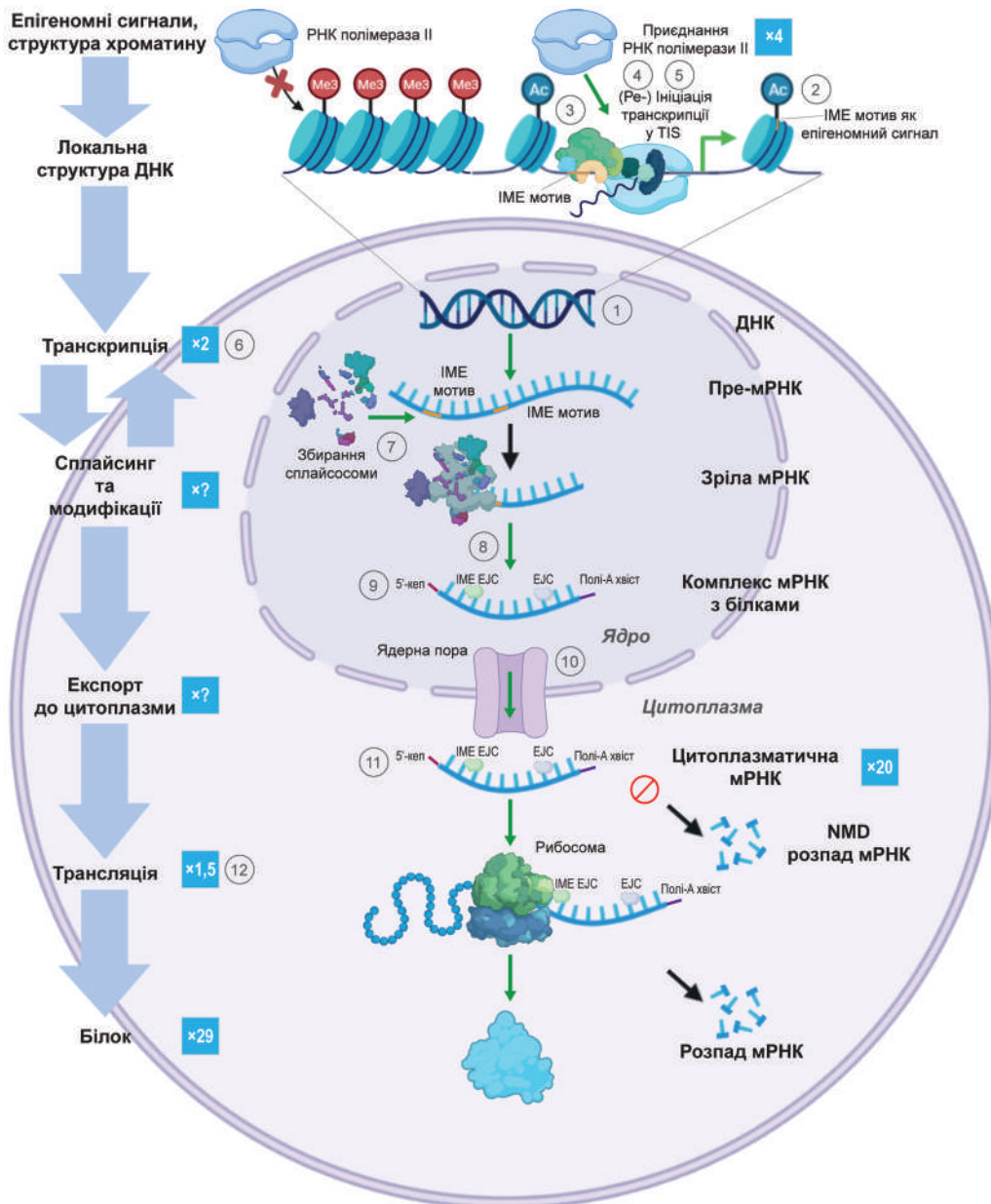
гена *TRP1 Arabidopsis* посилює накопичення мРНК за відсутності сплайсингу (Rose і Beliakoff, 2000). Інтригує відкриття того, що інтрони, які містять частини інтрону 1 гена *UBQ10* або гена *AtPK1*, підвищують експресію до однакового ступеня незалежно від орієнтації, спонукало дослідників зосередитися на моделях ІМЕ, які діють на рівні ДНК, а не РНК (Rose et al, 2011). Такий підхід «розділай і володарюй» для виокремлення різних проявів ІМЕ та вивчення їх окремо може виявитися корисним. Згідно до цього, було проаналізовано лише ті випадки ІМЕ, які відповідають таким трьом характеристикам: 1) інтрон посилює накопичення мРНК; 2) великі частини інтрона можуть бути видалені без втрати його впливу на експресію; і 3) інтрон підвищує рівні мРНК лише тоді, коли знаходиться нижче по послідовності та близько до початку місця транскрипції (Gallegos and Rose 2015). Ці ж автори розглянули кілька моделей на основі ДНК і припустили, що механізм, за допомогою якого інтрони стимулюють ініціацію транскрипту, полягає в створенні сприятливої локальної структури хроматину (Gallegos and Rose, 2015). Крім того, окрему увагу було зосереджено на конкретному типі інтронів, які підвищують накопичення мРНК за допомогою чіткого невідомого механізму, що відіграє важливу роль у регулюванні гена, в якому вони розташовані (Rose, 2019). Відповідно до цього було запропоновано термін «інтрони, що підвищують рівень мРНК» (Rose, 2019), щоб уникнути плутанини з загальним використанням тієї самої фрази, – ІМЕ підсилення, для опису будь-якого посилення експресії, спричиненого інтроном, незалежно від механізму (Mascarenhas et al, 1990; Laxa, 2017).

Феномен ІМЕ, враховуючи його можливий вплив на всіх рівнях експресії генів, знову було розглянуто і іншими авторами (Laxa, 2017; Shaul, 2017). Так, було припущено, що не існує єдиного механізму посилення експресії за допомогою інтрона, натомість кожна комбінація інтрон-ген може характеризуватись власним унікальним набором процесів, які призводять до помітного результату (Shaul, 2017). У загальному вигляді основні вимоги для ІМЕ у рослин можуть бути описані як: 1) інтрон має бути розташований у транс-

крибованій послідовності та поблизу сайту ініціації транскрипції у правильній орієнтації у одно- та дводольних; 2) ІМЕ сильно залежить від сплайсингу в ододольних, але не у дводольних, за винятком інтрона гена тонопластного транспортера металів *Arabidopsis* (*AtMXX*); 3) ІМЕ залежить від довжини та складу фланкуючих послідовностей, промотора та кодуєвої послідовності (Laxa, 2017).

В подальшому ІМЕ, а точніше, його вплив на транскрипцію в цілому було розділено на дві категорії: незалежна та залежна від сплайсингу регуляція, яка вимагає наявності функціонального, здатного до сплайсингу інтрона в транскрибованій області гена (Zalabák and Ikeda 2020; Dwyer et al, 2021b). Незалежне від сплайсингу регулювання експресії пояснюється наявністю консенсусних ІМЕ мотивів, якими багаті інтрони, здатні посилювати накопичення мРНК (Rose et al, 2008). У свою чергу, сплайсинг- та інтрон-залежне утворення петлевої структури гена може посилити транскрипцію шляхом сприяння ініціації, реініціації та обумовлення транскрипції у правильному напрямку від промотора в дріжджах (Furger et al, 2002; O’Sullivan et al, 2004; Moabbi et al, 2012; Agarwal and Ansari, 2016; Al-Husini et al, 2020). Таким чином, було запропоновано пов’язувати залежне від сплайсингу ІМЕ з певними інтронними мотивами, які забезпечують посилення транскрипції шляхом сприяння утворенню трьох контактних точок у інтрон-залежній петлевій структурі гена: промотор – термінатор, сайт сплайсингу – промотор-5’ і термінатор – 3’-місце зрощування (Dwyer et al, 2021b). Якщо це так, то таким чином можна пояснити, чому близькість до ділянок промотору та термінатору визначає потенціал посилення транскрипції інтроном (Dwyer et al, 2021a).

Впродовж багатьох років досліджень було показано, що інтрони впливають на експресію генів на багатьох рівнях, включаючи структуру хроматину, транскрипцію, ко-транскрипційний процесинг РНК, стабільність мРНК, експорт мРНК у цитоплазму та ефективність трансляції – рис. 1 (Lu і Cullen, 2003; Le Hir et al, 2003; Rose, 2008; Chorev and Carmel, 2012; Gallegos and Rose, 2015; Laxa, 2017; Shaul, 2017; Rose, 2019; Dwyer et al, 2021b). Однак,



**Рис. 1.** Ключові етапи експресії генів еукаріотів та можливі механізми інтрон-опосередкованого посилення експресії. IME елементи можуть (1) кодувати регуляторні РНК, (2) створювати епігеномні сигнали хроматина, (3) визначати положення сайту початку транскрипції (TIS) та (4) полегшувати ініціацію та реініціацію транскрипції, (5) спрощувати приєднання РНК полімерази II, (6) прискорювати транскрипцію, (7) полегшувати збирання елементів сплайсосоми та (8) модифікації мРНК, (9) обумовлювати структуру і/або склад комплексу мРНК з білками (мРНП), (10) прискорення експорту до цитоплазми через ядерні пори, (11) накопичення мРНК у цитоплазмі та (12) прискорення трансляції. Білим шрифтом на синьому фоні відображено дані про посилення зв'язування РНК-полімерази II (у 4 рази) 5'-UTR інтроном гена *GGT1* (Laha et al, 2016) та посилення на рівні транскрипції (у 2 рази), РНК (у 20 разів), трансляції (на 45 %) та синтезу білка (у 29 разів) 5'-UTR інтроном гена *RUB13* рису (Samadder et al, 2008). Зображення створено за допомогою BioRender та Adobe Illustrator



молекулярні механізми, які лежать в основі модулювання експресії, опосередкованої інтронами, залишаються не зрозумілими.

### Практичне значення інтрон-опосередкованого посилення експресії генів

*Біотехнологія рослин.* Інтрони тривалий час використовувалися для посилення експресії в трансгенних рослинах з метою покращення їх продуктивності. Сильна та конститутивна надекспресія ендогенних або чужорідних генів була і є найпоширенішою стратегією на сьогоднішній день (Basso et al, 2020). Коли такий результат досягається за допомогою сильних конститутивних промоторів або кількох копій трансгена в геномі рослини, ця стратегія схильна призводити до мовчання генів через ко-супресію (Dericker and van Montagu, 1997). Крім того, хоча більшість агрономічно важливих ознак культурних рослин є кількісно успадкованими (полігенними), нові продукти рослинної біотехнології вимагають комбінування набору ознак, також відомого як стекінг ознак, щоб забезпечити привнесення кількох ознак в межах однієї культури (Que et al, 2010). Надмірність послідовності у таких стекінгах ознак була визначена як потенційний фактор ризику нестабільності експресії трансгенів (Vaucheret et al, 1998; Kooter et al, 1999; Fagard and Vaucheret, 2000; Eamens et al, 2008; To et al, 2021). Більш слабкі промотори у поєднанні з сильними ІМЕ-асоційованими інтронами або тканиноспецифічною експресією, що забезпечується ІМЕ інтронами, можуть слугувати інструментом для подолання цієї проблеми мовчання генів (Laha, 2017).

У свою чергу, експресія, залежна від тканини, стадії розвитку рослини або конкретного стану (наприклад, стресу), дозволяє досягти дуже вигідних фенотипів з меншим негативним впливом на врожайність порівняно з використанням сильної конститутивної експресії (Basso et al, 2020). Було висловлено припущення, що експресія, специфічна для тканин і певних стадій розвитку рослини буде корисною для боротьби з комахами-шкідниками, які націлені на конкретну тканину рослини-хазяїна. Крім того, інсектицидні білки не повинні експресуватися в зерні, а специфічна експресія генів у насінні та бульбах спрощує

зберігання та виділення білків (Laha, 2017). Крім того, той факт, що ІМЕ впливає як на накопичення мРНК, так і на трансляцію, може допомогти у тих випадках, коли експресія чужорідного гена в рослинах обмежена низьким рівнем накопичення білка (Laha, 2017).

Загалом, однією з ключових проблем біотехнологічного запровадження нових ознак у рослин є доступність диверсифікованих елементів регуляції експресії для уникнення надмірності послідовностей та підтримання стабільності привнесених ознак (To et al, 2021). Крім того, нові концепції біотехнологічних продуктів можуть вимагати нових режимів контролю експресії для досягнення ефективності прояву тах чи інших ознак.

*Здоров'я та медицина.* Здатність інтронів ІМЕ сильно активувати експресію, але впливати лише на одиницю транскрипції, в якій вони розташовані, може бути використана в генній терапії (Rose, 2019). У свою чергу, здатність інтронів ІМЕ підвищувати рівні експресії в багато разів може бути використана для максимізації виробництва фармацевтичних білків, таких як моноклональні антитіла або терапевтичні ферменти (Rose, 2019).

Нещодавно було розглянуто переваги використання повнорозмірних локусів геномної ДНК (gDNA), які включають інтрони, можливо, ІМЕ інтрони, порівняно з комплементарною ДНК (кДНК) або мінігеном, які зараз використовуються в генній терапії (Simna and Han, 2022). Серед таких переваг слід розглядати більшу тканинну специфічність, наявність генних регуляторних елементів, фізіологічний рівень експресії генів, повторення ендогенних паттернів та вищу терапевтичну ефективність (Simna and Han, 2022).

Загалом, інтрони ІМЕ мають великий потенціал для застосування в генній інженерії, але низька кількість виявлених інтронів, які призводять до ІМЕ, недостатні перевірені дослідження, що підтверджують використання цих послідовностей у поєднанні з типовими промоторами або в конкретних організмах, перешкоджають їх більш широкому використанню. Ця ситуація також ускладнюється тим, що багато інтронів ІМЕ підлягають патентному захисту. Таким чином, здатність передбачати та легко генерувати різноманітні нові

стимулюючі інтрони може бути особливо корисною для подолання мовчання трансгенів (Rose et al, 2016). Короткі інтронні послідовності, асоційовані з ІМЕ, передбачені у *A. thaliana*, можуть бути вставлені або видалені з інтронів за допомогою сайт-спрямованого мутагенезу для ефективного підвищення або зменшення експресії генів (Parra et al, 2011; Rose et al, 2016). Таким чином, передбачення та конструювання ІМЕ інтронів з певними властивостями є перспективним напрямком.

**Можливість застосування машинного навчання для вивчення феномену інтрон-опосередкованого посилення експресії генів**

Машинне навчання – це розділ штучного інтелекту та інформатики, який зосереджується на методах поступового підвищення точності алгоритмів шляхом передачі їм відповідних даних. Машинне навчання використовує два основні підходи: контрольоване та неконтрольоване навчання. У контрольованому навчанні дослідник заздалегідь знає тип результатів, які очікує. У неконтрольованому навчанні алгоритм сам визначає, що відрізняється або є цікавим у наборі даних, дозволяючи отримати знання з великих обсягів нових даних (Chicco, 2017). Глибоке навчання – підгалузь машинного навчання, яка з'явилася нещодавно і дозволяє ефективно використовувати великомасштабні високимірні, неструктуровані та складні набори даних (Zhang et al, 2020). Глибоке навчання можна розглядати як еволюцію звичайних нейронних мереж, оскільки воно складається з кількох штучних нейронних мереж або кількох шарів штучних нейронів. Машинне навчання і глибоке навчання вже революціонізували багато областей досліджень, включаючи обчислювальну біологію, надаючи можливість виводити складні неочевидні взаємодії з величезних і гетерогенних даних і швидко виводити нові біологічні гіпотези за відсутності сильних припущень щодо основних процесів (Auslander et al, 2021).

Наприклад, машинне навчання було успішно застосовано для прогнозування сплайсингу (Rowlands et al, 2019), геноміки (Zou et al, 2019; Shen et al, 2022), молекулярної еволюції, аналізу структури білка, функціональної гено-

міки, регуляції генів (Barshai et al, 2020; Eraslan et al, 2019; Hu et al, 2023, Silva et al, 2019), передбачення сайтів зв'язування РНК- та ДНК-зв'язуючих білків (Alipanahi et al, 2015), аналізу експресії генів і передбачення посттрансляційних модифікацій (Zhang et al, 2020).

Успішність контрольованого навчання сильно залежить від початкових кроків: (i) вибір джерел даних; (ii) виділення ознак із вибраних даних; (iii) оцінка та вибір основних атрибутів для дослідження (Silva et al, 2019). Машинне навчання значною мірою залежить від даних, а збір даних, обробка та анотація є найважливішими кроками в процесі машинного навчання. Результатом підготовки даних є набір даних, придатний для аналізу машинного навчання.

Змінні або характеристики в кожному наборі даних, які вводяться в модель, називаються ознаками. Моделі машинного навчання класифікують дані відповідно до значень ознак у наборі даних. Вибір релевантних ознак з найбільшою варіативністю підвищує точність прогнозування алгоритмів машинного навчання. Показник важливості ознаки вимірює ступінь, до якого зміни цієї ознаки впливають на передбачення. Найбільш релевантні ознаки – це змінні, важливі для досліджуваних явищ. У той час як методи глибокого навчання виділяють характеристики автоматично, традиційне машинне навчання залежить від спеціалізованих статистичних алгоритмів виділення ознак. Для більшості геномних даних не потрібні дуже глибокі мережі (Zou et al, 2019). Для аналізу послідовності ознаки можуть представляти частоту або положення певних нуклеотидних послідовностей у певній області, вміст GC. У метааналітичному підході результати інших інструментів передаються алгоритму в якості ознаки.

Успіх проекту машинного навчання залежить насамперед від набору даних і вибору ознак, а не від самого алгоритму (Chicco, 2017). Кожен набір даних унікальний, має специфічні для домену ознаки та містить дані, пов'язані виключно з його науковою сферою. Крім того, оскільки різні лабораторії можуть працювати над одними і тими самими об'єктами, деякі анотації можуть містити суперечливу інформацію. Розмір набору даних має відповідати кількості

ознак — в ідеалі має бути принаймні в десять разів більше елементів даних, ніж є ознак (Chicco, 2017). Методи машинного навчання часто інтегруються з біоінформаційними методами, а також з курованими базами даних і біологічними мережами, щоб покращити навчання та валідацію, визначити найкращі ознаки, які можна інтерпретувати, а також зробити можливим дослідження ознак і моделей (Auslander et al, 2021).

Так само, як і з набором даних, результат аналізу машинного навчання є унікальним і залежить від домену та потребує обізнаної інтерпретації. Контрольовані алгоритми машинного навчання вирішують два різні типи проблем — регресію та класифікацію. В задачах класифікації розв'язанням є дискретне значення (мітка класу з набору класів), а в задачах регресії — неперервне кількісне значення. Таким чином, результат моделі залежить від її конструкції, типів ознак, які використовуються як вхідні дані, і цілей інструменту прогнозування (Rowlands et al, 2019).

На сьогоднішній день накопичено певний обсяг експериментальних даних щодо явища ІМЕ. Але ці дані несумісні між різними лабораторіями та експериментами. Щоб спробувати розгадати таємницю феномену ІМЕ за допомогою машинного навчання, слід створити відповідний, бажано зручний для машинного навчання набір даних. Такий набір даних повинен враховувати всі можливі біологічні фактори, які визначають ІМЕ. Ці фактори пізніше слугуватимуть ознаками для передачі алгоритму машинного навчання. Саме тому у подальшому нами будуть розглянуті біологічні фактори, які прямо чи опосередковано визначають ІМЕ.

#### **Фактори, які можуть визначати інтрон-опосередковане посилення експресії генів**

*Експериментальні результати дослідження інтрон-опосередкованого посилення експресії генів.* Оскільки ІМЕ модулює експресію генів, його аналізують за допомогою технологій рекомбінантної ДНК у трансформації рослин. Методологія вивчення ІМЕ була вже детально розглянута (Laha, 2017), тому слід зазначити декілька аспектів, важливих для порівняння даних різних досліджень.

Існують деякі докази того, що ефект ІМЕ є на порядок вищим у стабільно трансформованих рослин, ніж у системах транз'єнтної експресії як у дводольних (Jeong et al, 2006; Plesse et al, 2001), так і однодольних (Tanaka et al, 1990). Однак, оскільки дані для систем транз'єнтної експресії сильно варіюють між експериментами, це питання потребує подальшого дослідження. Виявлені ефекти ІМЕ також можуть залежати від методу трансформації (електропорація, трансформація, опосередкована *Agrobacterium* та бомбардування частинками), оскільки кількість копій трансгена залежить від використовуваного методу трансформації (Travella et al, 2005). Різна кількість трансгенних копій призводить до значних варіацій у експресії, тоді як ідентифікація однокопійних ліній вимагає багато часу. Впродовж багатьох років для дослідження ІМЕ використовували п'ять генів експресії: ген *NPTII*, що кодує неоміцинфосфотрансферазу II (NEO), ген *LUC*, що кодує люциферазу світляків (*LUC*), ген *CAT*, що кодує хлорамфенікол ацетилтрансферазу (*CAT*), ген *PAT*, що кодує фосфінотрицин ацетилтрансферазу (*PAT*) і ген *GUSA*, що кодує β-глюкуронідазу (*GUS*). Різні репортерні гени загалом важко порівнювати, оскільки вони відрізняються чутливістю щодо межі виявлення. Для досліджень ІМЕ, зокрема, було показано, що репортерні гени впливають на рівень посилення експресії (Luehrsen and Walbot, 1991; Morita et al, 2012). Ген *GUS* є найбільш використовуваним репортером у дослідженнях ІМЕ, за ним йдуть гени *CAT* і *LUC* (Laha, 2017). Його важливою особливістю є здатність забезпечувати моніторинг тканинно-специфічної експресії (Jefferson et al, 1987). У майбутньому для вивчення ІМЕ можуть бути використані інші нові гени-репортери. Ефект ІМЕ також залежить від виду та системи трансформації, і, крім того, є найвищим у тканинах, що активно діляться (Laha, 2017). Показано, що інтрони опосередковують експресію своїх генів шляхом посилення транскрипції, трансляції або обох цих процесів з необхідністю сплайсингу або без нього (Gallegos and Rose 2015; Shaul, 2017; Laha, 2017). Слід порівнювати між собою лише інтрони, які впливають на експресію на одному і тому ж самому рівні. Нарешті, слід враховувати про-

мотори, що використовуються в трансгенних конструкціях, оскільки швидкості транскрипції, регульовані сильними та слабкими промоторами, можуть відрізнятись на порядок.

**Кодування регуляторних РНК.** Сплайсосомні інтрони можуть впливати на експресію шляхом кодування транс-елементів — регуляторних РНК, таких як інтронна мікроРНК (miRNAs), малі ядерцеві РНК (snoRNA), довгі некодуючі РНК (lncRNA) та інші молекули РНК, які процесуються з прекурсорів інтронів специфічними комплексами РНКаз/білок (Morello and Breviaro, 2008; Chorev and Carmel, 2012). Інтрони та інші некодуючі РНК (нкРНК) утворюють мережу функціональних молекул, які регулюють експресію генів на рівні транскрипції та стабільності РНК, взаємодіючи з іншими молекулами, такими як ДНК, РНК та білки, додаючи ще один рівень генетичної інформації (Carmel and Chorev, 2012; Vogard et al, 2020). Роль цих нкРНК як незамінних гравців у діагностиці, розвитку та терапії генетичних розладів, нейродегенеративних захворювань, раку та вірусних інфекційних захворювань (Bhatti et al, 2021) передбачає широке практичне застосування, по мірі того, як ми отримуємо більше знань про ці процеси.

**Специфічність інтронних послідовностей, відповідальних за посилення експресії.** Ідентифікація специфічних інтронних послідовностей, відповідальних за посилення експресії, є складним завданням, оскільки вони є часто надлишковими, розсіяними та виродженими (Rose and Beliakoff, 2008; Rose et al, 2008; Baier et al, 2020). У той час як делеційний аналіз дозволяє знайти промоторні та енхансерні послідовності, великі внутрішні делеції зазвичай не знижують активності стимулюючих інтронів, і ці інтрони не мають явних консервативних послідовностей. Наприклад, усі чотири фрагменти довжиною 54–113 п.н., які входять до інтрону гена *UBQ10* довжиною 304 п.н., збільшують накопичення мРНК при вставці в нестимулюючий інтрон 2 гена регульованого холодом білку 15A (*COR15a*) (Rose et al, 2008). У свою чергу, внутрішня делеція 85 % інтрон-лідера гена *Shrunken1* кукурудзи не впливає на рівень посилення експресії, тоді як ділянка, багата на тимін, залишається незмінною

(Rose and Beliakoff, 2000). І навпаки, поступова делеція інтрону гена альфа-субодиниці фактору елонгації *AtEF1-A3* свідчить про кореляцію між довжиною інтрону та модулюванням експресії, яку він викликає (Chung et al, 2006).

Статистично незначна кількість інтронів з відомим впливом на посилення експресії, велика різноманітність у довжині та відсутність очевидної гомології робить звичайні обчислювальні методи біоінформатики неефективними для ідентифікації послідовностей, спільних лише для стимулюючих інтронів. Одним з можливих пояснень може бути те, що ІМЕ спричинене фізичними властивостями ДНК та/або РНК, а не взаємодією з білком або РНК, завдяки специфічним послідовностям, що може вимагати менш очевидної консервативності послідовності — не з точки зору ідентичності, а радше складу або регулярності. Наприклад, для інтрону 1 гена сахаросинтази *Sh1* кукурудзи та для інтрону 1 гена *TRP1 (PAT1) Arabidopsis* було припущено, що насиченість U- (T-) нуклеотидом, а не конкретна послідовність, є важливою для ІМЕ (Clancy et al, 2002; Rose et al, 2002), а для інтрону 1 гена гліцеральдегід-3-фосфат дегідрогенази 1 (*GapA1*) кукурудзи — наявність октамеру, багатого нуклеотидами GC (Donath et al, 1995). Проте було показано, що порівняно з цілими інтронами ці елементи мають лише помірний вплив на експресію. Разом з тим, описано декілька випадків ІМЕ, для яких було показано, що модулювання експресії пов'язане з певними фрагментами в інтронах, а послідовність і склад цих фрагментів є одним із вирішальних факторів, що визначають здатність опосередковувати експресію генів (Morello et al, 2010; Parra et al, 2011).

Загалом, пошук нових регуляторних елементів всередині інтронів, особливо перших, може бути перспективним. В еукаріотичних генах перші інтрони є найдовшими, найбільш еволюційно консервативними, містять більш консервативні блоки та виявляють більшу щільність регуляторних елементів або функціональних мотивів, ніж інші інтрони, розташовані далі за послідовністю (Bradnam and Korf 2008; Park et al, 2014; Jo and Choi 2015). Крім того, показано, що в кількох тканинах людини консервативність першого інтрона позитивно



корелює з кількістю екзонів і рівнем експресії гена (Park et al, 2014). Пізніше, та ж група дослідників виявила, що рідкісні алелі були значно збагачені в більшості регуляторних сайтів, розташованих у перших інтронах, що свідчить про те, що сайти в перших інтронах генів можуть виконувати важливі регуляторні функції при експресії генів незалежно від промоторів (Jo and Choi, 2019a). Крім того, різні епігеномні сигнали хроматину частіше розташовані в перших інтронах, ніж у всіх інших інтронах, і що розподіл активуючих і репресуючих хроматинових міток суттєво відрізняється між генами домашнього господарства та тканиноспецифічними генами, а також між основними генами та несуттєвими генами (Jo and Choi, 2019b). Разом це свідчить про те, що інтрони є важливими еволюційно збереженими елементами геному, значення яких ще до кінця не розкрито.

Кілька мотивів інтронів, що контролюють експресію, були передбачені обчислювальним шляхом і перевірені експериментально. Два мотиви TTNGATYTG і CGATT, які відповідають ядру NGATY довшого мотиву, були визначені за допомогою алгоритму IMEter (Rose et al, 2008) як такі, що пов'язані з IME у *Arabidopsis* (Rose et al, 2008; Parra et al, 2011). Подібну до мотиву *Arabidopsis* консенсусну послідовність GATCTG було ідентифіковано за допомогою *in silico* аналізу 12 інтронів IME з рису, що свідчить про те, що загальні мотиви зберігаються у опосередковуючих експресію інтронах рису та *Arabidopsis* (Morita et al, 2012).

Додавання 11-ти копій мотиву CGATT до нестимулюючого інтрону 2 гена *COR15a Arabidopsis* забезпечило йому здатність посилювати накопичення мРНК більше, ніж у 6 разів (Parra et al, 2011). У свою чергу, перестановка нуклеотидів для створення 6-ти або 11-ти копій мотиву TTNGATYTG перетворює той самий інтрон *Arabidopsis* на інтрон, який посилює накопичення мРНК у 14 або 24 рази відповідно (Rose et al, 2016). Серія інтронів, сконструйованих на основі інтрону 2 гена *COR15a* з різною кількістю цього мотиву, показала, що вплив кількості копій мотиву на експресію має напрочуд лінійний характер принаймні до 20, причому кожна копія додає

ще одне 1,5-разове підвищення накопичення мРНК (Gallegos і Rose, 2019).

Навіть незважаючи на те, що мотиви CGATT і TTNGATYTG посилюють експресію, і CGATT відповідає ядру NGATY довшого мотиву, комбінування мотивів і створення послідовності TTCGATTTG спричиняє зменшення стимулюючої здатності вдвічі порівняно з мотивом TTAGATCTG. Цікаво, що перетворення ядра мотиву з GAT на GTA майже повністю усуває його вплив на експресію, тоді як TT на початку та TG наприкінці, очевидно, мають незначний внесок у функцію мотиву (Rose et al, 2016). Оскільки передбачення структури РНК не виявляє очевидних відмінностей між згортанням послідовностей інтрону гена *TRP1*, модифікованих таким чином, щоб містити стимулюючий мотив TTAGATCTG порівняно з нестимулюючим мотивом TTAGTACTG, можна зробити висновок, що стабільність мРНК навряд чи відповідає за вплив на експресію (Gallegos and Rose 2019).

Введення 11-ти копій кожного з мотивів у інтрон 2 гена *COR15a*, який не підсилює експресію, шляхом перегруповання нативних фрагментів з подібним до мотиву нуклеотидним складом, посилює його здатність підвищувати рівні мРНК у 4-и та 13-ть разів відповідно (Parra et al, 2011; Rose et al, 2016). Ці мотиви не виявляють подібності до U1 малих ядерних РНК (snRNA) або відомих сайтів розпізнавання білка SR і навряд чи функціонуватимуть як сайти зв'язування факторів транскрипції. Хоча спосіб, за допомогою якого ці мотиви збільшують вміст мРНК, залишається незрозумілим, було припущено, що IME може спричинятись або фізичною структурою самої ДНК чи РНК, або забезпечуватись взаємодіями, що не залежать від конкретних послідовностей, між нуклеїновою кислотою та білками, такими як гістони або гетерогенні ядерні рибонуклеопротеїни (hnRNP) (Rose et al, 2011; Rose et al, 2016; Gallegos and Rose, 2015, 2017).

Нещодавно за допомогою обчислювального аналізу і глибокого секвенування та численних даних про експресію генів, доступних для *A. thaliana*, було ідентифіковано набір з 16-ти нових гексамерних послідовностей і 5-ти консенсусних мотивів, розташованих у перших

інтронах (Back and Walther, 2021). Два з них подібні до двох раніше описаних ІМЕ мотивів CGATT і TTNGATYTG (Rose et al, 2008; Parra et al, 2011) – GATTCG і ARATCGA, відповідно, – тоді, як три є новими: ACYCYRA, TTTCGA і KCGAGAR. Несподівано класифікація генів на основі алгоритму випадкового лісу (random forest) з огляду на рівень експресії генів показала, що вміст нуклеотидів у послідовності інтронів, передусім, вміст G, виявилися найбільш інформативними та важливими, а не відстань до сайту початку транскрипції, як вважали раніше (Back and Walther, 2021).

Незважаючи на всі зусилля, поки що не видається можливим визначити точний механізм дії ідентифікованих мотивів. Можна сподіватись, що з накопиченням нових даних розуміння основних механізмів біологічної активності цих послідовностей дозволить проводити дизайн нових послідовностей для максимізації експресії. Підтвердження того, що ці мотиви активні в процесі ІМЕ, покращує наші можливості передбачати стимулюючу здатність інтронів, надавати будь-якому інтрону здатності підвищення експресії до бажаного рівня та досліджувати механізм ІМЕ шляхом пошуку факторів, які можуть взаємодіяти з цими послідовностями (Rose et al, 2016).

У той час, як мотиви ІМЕ найчастіше зустрічаються в довших перших інтронах генів, вони також зустрічаються і в 5'-UTR інтрон-лідери і наступних UTR інтронах. Наприклад, показано, що максимальна кількість транскриптів гена *Glu:гліоксилат амінотрансферази 1 (GGT1)* регулюється ефектом ІМЕ 5' інтрон-лідери (Лаха et al, 2016). Даний інтрон насичений фрагментами, багатими на нуклеотиди СТ та містить мотив TGTGATTTG, який дуже схожий на пов'язаний з ІМЕ мотив TTNGATYTG (Лаха et al, 2016). Цей інтрон гена *GGT1* посилює транскрипцію, сприяючи зв'язуванню РНК-полімерази II і, зокрема, забезпечує експресію чужорідних промоторів, специфічну для листа (Лаха et al, 2016). Чи дійсно дані властивості надають названі мотиви, ще не встановлено.

В одному з основоположних досліджень феномену ІМЕ було показано, що інтрони 2 і 6 алкогільдегідрогенази-1 (*ADH1*) кукурудзи підвищують рівні експресії CAT у 12-ть і 20-ть разів, відповідно (Mascarenhas et al, 1990). При-

мітно, що делеція, яка зменшує фланкуючу послідовність 5' екзона, що примикає до інтрона 2, з 54 до 19 нуклеотидів, зменшує посилення в 2,5 рази, а її подальше зменшення до 6-ти нуклеотидів фактично усуває здатність інтрону модулювати посилення експресії гена *CAT*. У випадку інтрона 6 скорочення 5'-екзона з 76 до 29 нуклеотидів також зменшує рівень експресії гена вдвічі (Mascarenhas et al, 1990). Подібний ефект спостерігався і для інтрона 1 цього гена (Luehrsen and Walbot, 1991). Однак, подібна делеція 3'-екзона інтронів 2 і 6 *ADH1* кукурудзи не мала такого ефекту (Mascarenhas et al, 1990). Таким чином, на ІМЕ властивості інтрона впливають послідовності, що фланкують інтрон. Це може бути пов'язано з тим, що ефективність сплайсингу в рослинах залежить від послідовності фланкуючих екзонів (Morello and Breviaro, 2008).

*Позиція та орієнтація інтронів.* У той час, як енхансери можуть функціонувати з нетранскрибованих місць у послідовності, розміщення інтрона, який забезпечує ІМЕ, попереду від промотора або після кодуєчої послідовності – в межах 3'-UTR не посилює експресію гена, що свідчить про те, що ІМЕ працює лише тоді, коли інтрон знаходиться в межах транскрибованих послідовностей (Callis et al, 1987; Jeon et al, 2000; Jeong et al, 2006; Snowden et al, 1996). Крім того, є також численні докази того, що потенціал посилення інтронів обернено залежить від їх відстані від сайтів ініціації транскрипції та трансляції. Вставка інтрона 1 гена *TRP1 (PAT1)* у різні місця транскрибованої послідовності значною мірою впливає на ІМЕ: що більша відстань від 5'-кінця гена, тим менший ефект (Rose, 2004). Подібні результати були отримані для інтрона-лідера гена *Sh1* (Mascarenhas et al, 1990) і для першого інтрона гена триозофосфатізомерази рису *TPI* (Snowden et al, 1996). Три різні інтрони з гена *RPOT* кукурудзи, незважаючи на те, що всі вони ефективно піддаються сплайсингу, мають різний підсилювальний ефект, якщо розмістити їх у різних позиціях у кодуєчій послідовності гену люциферази, причому деякі з цих позицій можуть навіть викликати негативні ефекти (Bourdon et al, 2001). Як загальне правило, запропоновано вважати, що посилюючий інтрон повинен бути розташованим в межах ~1 кб

від початку транскрипції, щоб мати помітний ефект (Rose, 2004).

Були запропоновані різні механізми для пояснення таких позиційних обмежень ІМЕ інтронів. Інтрони, близько розташовані до промоторів, можуть сприяти ефективній повторній ініціації у правильному сайті початку транскрипції і полегшувати вибір правильного напрямку промоторів (Shaul, 2017). Запропоновано модель, у якій специфічні послідовності в інтроні сприяють ініціації та реініціації транскрипту в межах окремої ділянки вище у послідовності ДНК (Gallegos and Rose, 2015). Разом з тим, є повідомлення про те, що в деяких генах активаційний потенціал інтрону частково відновлюється за умов його розташування ближче до термінатора (Callis et al, 1987; Kim et al, 2011).

Позитивний вплив інтрона на транскрипцію кількох дріжджових генів було продемонстровано за допомогою методу аналізу ядерного синтезу (nuclear run-on assay) (Furger et al, 2002; Moabbi et al, 2012; Agarwal and Ansari, 2016). Показано, що формування унікальної петлевої архітектури гена за допомогою промотор-проксимального інтрону є одним з механізмів ІМЕ у дріжджів, котрі брунькуються, що призводить до посилення транскрипції мРНК (Moabbi et al, 2012; Agarwal and Ansari, 2016; Dwyer et al, 2021b). Петлева структура гена сприяє рекрутуванню факторів термінації поблизу проксимальної області промотора, і ці фактори термінації пригнічують синтез антимислових РНК (uaRNA) вище у послідовності ДНК, забезпечуючи, таким чином, транскрипцію у правильному напрямку від промотору (Agarwal і Ansari, 2016; Al Husini et al, 2020).

Показано, що за умов вставки інтрону *ACT1* у різні позиції генів інозин-5'-монофосфатдегідрогенази 4 (*IMD4*) та інозитол-3-фосфатсинтази (*INO1*) у дріжджів, які брунькуються, термінатор-проксимальний інтрон також посилює транскрипцію таким же чином, як і промотор-проксимальний інтрон, хоча й меншою мірою (Dwyer et al, 2021a). Однак, хоча більшість промоторів у ссавців ініціюють транскрипцію з обох сторін від себе у протилежних напрямках, що називається дивергентною транскрипцією, рослинні промотори, здається, діють однонаправлено (Zala-

bák and Ikeda, 2020). Крім того, на відміну від дріжджів, які брунькуються, гени вищих еукаріотів містять, зазвичай, по декілька інтронів, розподілених по всій послідовності гена. Отже, результати, отримані для дріжджів, не можуть бути безпосередньо застосовані до вищих еукаріотів, а скоріше служать парадигмою для дослідження потенційної ролі термінатор-проксимальних інтронів в експресії генів (Dwyer et al, 2021a). У гені триозофосфатізомерази людини делеція останнього інтрона в двічі знижує рівень продукту мРНК, що свідчить про те, що послідовності в кінцевому інтроні сприяють правильному утворенню 3'-кінця шляхом розщеплення та поліаденілювання, можливо, завдяки асоціації з компонентами активної сплайсоми (Nesic et al, 1993). У той час, як термінатор-проксимальний інтрон полегшує процесинг 3'-кінця мРНК і припинення транскрипції, є, кілька повідомлень про те, що інтрони, розташовані у 3'-UTR області, негативно впливають на стабільність і трансляцію мРНК у клітинах ссавців і рослин (Dwyer et al, 2021a). У той час, як послідовності у 5'-UTR-області гена в основному сприяють його експресії, було показано, що незвичайно довгі 3'-UTR фрагменти або присутність інтронів у 3'-UTR області може призводити до нонсенс-опосередкованого розпаду мРНК (nonsense-mediated mRNA decay, NMD) (Kertész et al, 2006).

Шість копій «ІМЕ мотиву» TTNGATYTG (Rose et al, 2008; Rose et al, 2016) стимулювали накопичення мРНК до аналогічного рівня при розміщенні як всередині інтрону, так і при введенні у 5'-UTR фрагмент або кодуєчі послідовності безінтронного конструкту, демонструючи, що сплайсинг не потрібен для посилення експресії даною послідовністю (Gallegos and Rose, 2019). Це свідчить про те, що роташування мотивів, пов'язаних з незалежним від сплайсингу ІМЕ у рослин, не обмежується інтронами. Але загалом, положення інтрону в гені є важливим визначальним фактором його потенціалу активації транскрипції (Dwyer et al, 2021b).

Узагальнення накопичених даних (Laha, 2017) вказує на те, що інтрон повинен бути розташований у транскрибованій послідовності та у правильній орієнтації, близько до сайту

ініціації транскрипції як у однодольних (Callis et al, 1987; Vasil et al, 1989; McElroy et al, 1990; Maas et al, 1991; Clancy et al, 1994; Snowden et al, 1996), так і у дводольних (Gidekel et al, 1996; Chaubet-Gigot et al, 2001; Mun et al, 2002; Rose, 2002, 2004; Jeong et al, 2006, 2007; Akua et al, 2010). Особливості структури гена та послідовності ДНК, які сприяють ІМЕ, відображено на рис. 2, а.

*Можливі механізми дії ІМЕ на рівні ДНК.* Варто також обговорити загальну фізичну структуру хроматину та модифікації ДНК, які полегшують ініціацію та реініціацію транскрипції і експресію генів загалом, а вже потім розглянути властивості інтронів як промоторів (див. наступний розділ). Послідовне накопичення доказів механізму ІМЕ на основі властивостей ДНК свідчить про те, що деякі інтрони підвищують рівні накопичення мРНК за відсутності сплайсингу (Rose and Beliakoff, 2000). Так, інтрон 1 гена *UBQ10 A. thaliana* посилює накопичення мРНК майже в 15 разів (Rose, 2002). Даний інтрон несе надлишкові та дисперговані послідовності, що посилюють транскрипцію (Rose et al, 2008), підвищує експресію в однаковій мірі у будь-якій орієнтації (Rose et al, 2011), визначає тканинну специфічність експресії у поєднанні з різними промоторами, і наявність промотору не потрібна для активації експресії даним інтроном (Emami et al, 2013). Однак, один ІМЕ інтрон може впливати на кілька рівнів експресії генів, діючи, як на рівні ДНК, так і на наступних рівнях, як було показано для інтрону 5'-UTR гена убіхтинового 40S рибосомного білка *UBI3* рису (Samadder et al, 2008; Lu et al, 2008) і інтрона 2 *ADH1* кукурудзи (Mascarenhas et al, 1990), який збільшує накопичення як мРНК, так і білку (Laha, 2017).

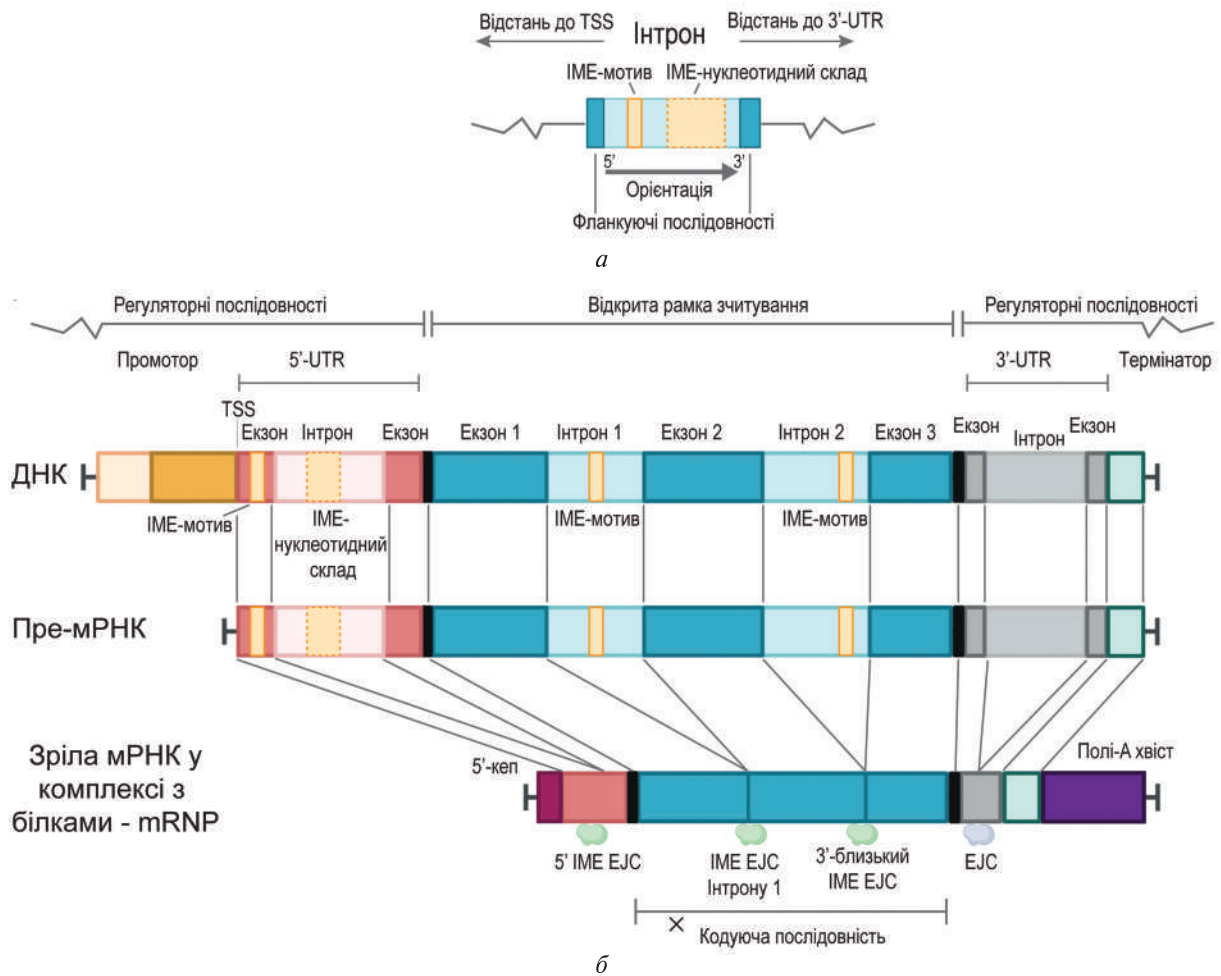
На користь функціонування механізму ІМЕ на рівні ДНК можуть свідчити результати двох інтригуючих спостережень: 1) 5'-кінцеві області перших інтронів у генах людини багаті на активуючі модифікації гістонів (Vieberstein et al, 2012) і 2) інтрони мають тенденцію формувати меншу кількість нуклеосом порівняно з екзонами (Spies et al, 2009). З огляду на це, спочатку було встановлено вражаючу кореляцію між розподілом теоретичних балів ІМЕter та активуючими модифікаціями гісто-

нів уздовж послідовностей інтронів у *Arabidopsis* (Gallegos and Rose, 2015). Надалі, експериментально було показано, що принаймні 2-разове ІМЕ підсилення 5'-UTR інтроном гена *GGT1* проявляється на рівні транскрипції, і його присутність призводить до підвищеного майже в 4 рази рівня зв'язування РНК-полімерази II порівняно з безінтронною конструкцією (Laha et al, 2016). Однак, в трансгенних рослинах, у яких відсутній 5' інтрон-лідер гена *GGT1*, не спостерігалось значних змін активуючих модифікацій гістонів (Laha et al, 2016).

У свою чергу, інтрон гена *UBQ10 A. thaliana* специфічно активує ініціацію транскрипції на кілька сотень пар основ вище свого місця розташування у послідовності за допомогою раніше невідомого механізму, відмінного від сайтів зв'язування факторів транскрипції та ехансерів (Gallegos and Rose, 2017). Видалення з послідовності гена *UBQ10* промотору, включаючи всі відомі сайти початку транскрипції, не впливає на рівень експресії гена, якщо у послідовності присутній інтрон, що містить стимулюючі послідовності. Стимулюючі інтрони можуть підвищувати синтез мРНК шляхом створення локальної структури хроматину, котра сприяє ініціації, зменшуючи частку молекул РНК-полімерази II, які транскрибують поза межами гена, і позитивного зворотного зв'язку, який посилює накопичення мРНК через реініціацію (Gallegos and Rose, 2015, 2017).

Було запропоновано ряд можливих механізмів дії ІМЕ на рівні ДНК. Припускається, що ІМЕ може включати неспецифічні взаємодії послідовності між ДНК і білками, такими як гістони або гетерогенні ядерні рибонуклеопротеїни (hnRNP) чи інші ще невизначені фактори, або через фізичну структуру самої ДНК або РНК (Gallegos and Rose, 2015). Зокрема, можливо, що інтронна ДНК сприяє відокремленню ланцюгів для формування транскрипційної бульбашки та розсіює торсійну деформацію, створювану транскрипцією (Rose et al, 2016; Niu and Yang 2011). Крім того, запропонована гіпотеза про те, що сигнали ІМЕ, розташовані будь-де в першій тисячі нуклеотидів транскрибованої ДНК, здатні змінювати стан хроматину в усіх клітинах, що сприяє





**Рис. 2.** *a* – фактори структури гена та послідовності ДНК, які сприяють ІМЕ: відстань до сайту початку транскрипції (TSS), відстань до 3'-UTR та термінатору, орієнтація інтрону відповідно до 3'- і 5'-напрямків послідовності, сенс- або антисенс-послідовність, послідовності екзонів, які фланкують інтрон; *б* – структура гена еукаріотів у складі ДНК у вигляді пре-мРНК та зрілої мРНК з приєднаними комплексами зрощування екзонів (EJC) на місці зшивання екзонів. Елементи ІМЕ мають бути у транскрибованій області гена, як у складі інтронів, так і у складі екзонів 5'-UTR і екзонів кодуючої області. Вважається, що після сплайсингу ефект ІМЕ передається на наступні етапи експресії за рахунок того, що ІМЕ елементи обумовлюють певну структуру і/або склад комплексу мРНК з білками (mRNP), зокрема, – елементів комплексу зрощування екзонів. Зображення створено за допомогою Adobe Illustrator

високому рівню нерегульованої транскрипції, і що послідовності, з яких може ініціюватись транскрипція, є напрочуд варіабельними (Rose, 2019). Існує кілька інших механізмів, за допомогою яких інтронна послідовність ДНК може впливати на експресію генів (Gallegos and Rose, 2015; Zalabák and Ikeda, 2020).

**Посилення транскрипції.** Роль інтронів у посиленні транскрипції зберігалась впродовж ево-

люції, що продемонстровано на ряді еукаріотичних систем, включаючи дріжджі, черви, дрозофілу, людину, водорості та рослини (Mc Kenzie and Brennan, 1996; Juneau et al, 2006; Rose, 2008; Shabalina et al, 2010; Gallegos and Rose, 2015; Laxa, 2017; Shaul, 2017; Baier et al, 2020).

Відомо, що інтрони посилюють транскрипцію, впливаючи на структуру хроматину у промотор-проксимальній області та сприяючи при-

єднанню апарату транскрипції до промоторної області. Зокрема, у клітинах ссавців інтрон сприяє НЗК4-триметилуванню і НЗК9-ацетилюванню гістонів у промотор-проксимальних областях (Bieberstein et al, 2012). Ці обидві модифікації гістонів допомагають приєднанню апарату транскрипції до промоторної області. Також припускається, що у ссавців деякі гени можуть потребувати для транскрипції сигналів сплайсингу поблизу промотору (Furger et al, 2002).

IME впливає на транскрипцію як незалежно, так і залежно від сплайсингу (Dwyer et al, 2021b; Zalabák and Ikeda, 2020). Для регуляції, що залежить від сплайсингу, необхідний інтрон, придатний для сплайсингу, у транскрибованій області гена у сенс-орієнтації та з інтактними 5'- і 3'-сайтами сплайсингу і точкою розгалуження (Furger et al, 2002; Samadder et al, 2008; Moabbi et al, 2012; Laha et al, 2016; Agarwal and Ansari, 2016). Навпаки, незалежні від сплайсингу інтрони можуть впливати на транскрипцію, навіть якщо їх функція сплайсингу порушена. Можливим механізмом ІМЕ, незалежного від сплайсингу, вважається елемент, подібний за властивостями до промотору.

Показано, що сплайсинг є котранскрипційним процесом (Custudio and Carmo-Fonseca, 2016; Chaudhary et al, 2019; Zalabák and Ikeda, 2020). Здатність інтрону регулювати транскрипцію залежить від котранскрипційної природи сплайсингу, коли фактори сплайсингу хоч і не контактують безпосередньо з ДНК, але близькість сайтів сплайсингу на РНК до відповідної ділянки ДНК під час котранскрипційного сплайсингу призводить також до зшивання факторів сплайсингу зі сайтами сплайсингу ДНК (Dwyer et al, 2021b). Наприклад, у гені *TRP1 (PAT1)* арабідопсису посилення експресії інтроном 1 було усунуто шляхом одночасного видалення точок розгалуження та 5'-сайту сплайсингу, структур, залучених до перших двох етапів збирання сплайсосоми, але не 3'-сайту сплайсингу (Rose, 2002). Це свідчить про те, що розпізнавання інтронів механізмами сплайсингу необхідне для посилення накопичення мРНК, а не сплайсинг як такий.

Інтрони можуть безпосередньо впливати на стадію ініціації та термінації, сприяючи зв'я-

зуванню загальних факторів транскрипції з промотором, або опосередковано, впливаючи на всі стадії транскрипції, включаючи елонгацію, за рахунок впливу на структуру хроматину в області промотора (Dwyer et al, 2021b). Результати спостережень, що стимулюючий інтрон під контролем тканинспецифічного промотора може викликати сильну експресію в тканинах, у яких промотор зазвичай неактивний, свідчать про те, що інтрони можуть мати сильніший вплив на експресію генів, ніж промотори, і дозволяють припустити, що інтрони в першу чергу впливають на ініціацію транскрипції (Gallegos and Rose, 2015; Emami et al, 2013; Jeong et al, 2006; Gianì et al, 2009).

Сплайсинг не є необхідною вимогою для перших інтронів *Arabidopsis ACT1* (Vitale et al, 2003) і гена *UBQ10* для підвищення накопичення мРНК в обох орієнтаціях (Rose et al, 2011, Emami et al, 2013). Крім того, інтрон 1 гена *UBQ10* своїм розташуванням у гені визначає сайт початку транскрипції (Gallegos and Rose, 2017) і забезпечує такий рівень тканинної специфічності, що перевищує ці ж показники ряду промоторів (Emami et al, 2013). Іншим прикладом посилення транскрипції, незалежного від сплайсингу, є два ІМЕ мотиви TTNGATYTG і CGATT з *Arabidopsis* (Rose et al, 2008; Parra et al, 2011; Rose et al, 2016), які стимулюють накопичення мРНК в однаковій мірі – як всередині інтрону, так і при вставленні у 5'-UTR область, або кодуєчі послідовності безінтронної конструкції (Gallegos and Rose, 2019).

Іншим механізмом впливу інтронів на експресію генів може бути сплайсинг та інтрон-залежне утворення унікальної петлевої архітектури гена, яка об'єднує 3'-кінець і промотор гена у дріжджах, що брунькуються (Moabbi et al, 2012; Agarwal and Ansari, 2016; Dwyer et al, 2021a). Подібним чином гени ссавців також демонструють внутрішньогенну взаємодію промотору з тілом гена, особливо екзонами. Такі промотор-екзон взаємодії можуть відповідати за залежну від сплайсингу регуляцію транскрипції. Таким чином, обумовлені сплайсингом зміни в архітектурі гена можуть відігравати вирішальну роль у регуляції транскрипції у дріжджів, а також у вищих еукаріотів (Dwyer et al, 2021b).

Один інтрон ІМЕ може впливати на кілька етапів експресії генів, і на кожній наступній стадії ці ефекти можуть накопичуватися, помножуючись один на одного. Найбільш виражений ефект ІМЕ, зазвичай, спостерігається для стадії накопичення мРНК та трансляції, тоді як, загалом, вплив ІМЕ на транскрипцію не перевищує зміни у 2–3 рази (Лаха, 2017).

Деякі інтрони здатні забезпечувати принаймні слабку експресію гена, позбавленого промотору, і тому вважається, що вони містять послідовності, яким притаманні властивості промотору, і дані послідовності можуть бути частково відповідальні за посилення експресії (Salgueiro et al, 2000; Morello et al, 2002, 2006; Weise et al, 2006; Kim et al, 2006; Liao et al, 2013; Rose, 2008; Лаха, 2017).

*Залежність ІМЕ від сплайсингу.* Результатом транскрипції є пре-мРНК, яка надалі піддається ряду модифікацій – процесингу. Ключовими етапами процесингу є сплайсинг, 5'-кепування та поліаденілювання. Результатом процесингу є зріла мРНК. На рис. 2, б представлена схематична структура гена еукаріотів у складі ДНК, у вигляді пре-мРНК та зрілої мРНК.

У однодольних рослин ефект ІМЕ інтроном 2 *ADHI* (Mascarenhas et al, 1990; Luehrsen and Walbot, 1991; Sinibaldi і Mettler, 1992), інтроном 1 гена *SHI* (Clancy et al, 1994; Clancy і Hannah, 2002) та інтроном 1 гена молекулярного шаперону *HSP82* (Sinibaldi and Mettler, 1992) кукурудзи та 5'-UTR інтрон-лідера *OsTUB6* рису (Morello et al, 2011) сильно залежить від сплайсингу. Це також стосується принаймні одного випадку у дводольних – інтрона-лідера гена *MHX Arabidopsis* (Akua et al, 2010).

Інтрони, як правило, є неактивними, якщо розташовані в області 3'-UTR (Snowden et al, 1996; Rose, 2002). У деяких випадках механізм ІМЕ тісно пов'язаний зі сплайсингом посилюючого інтрону (Лаха, 2017). Дуже важливе питання, порушене цими експериментами, стосується можливої залежності ІМЕ від сплайсингу (Morello and Breviario, 2008). Наявність інтрону призводить до зв'язування факторів сплайсингу. Показано, що сплайсинг позитивно впливає на поліаденілювання, кепінг, стабільність транскриптів і трансляцію (Proudfoot et al, 2002; Le Hir et al, 2003). Взаємодія між компонентами механізму сплайсингу та

поліаденілювання може призвести до підвищення рівнів мРНК шляхом підвищення ефективності поліаденілювання деяких транскриптів (Le Hir et al, 2003).

У кількох дослідженнях було показано, що для посилення експресії достатньо присутності інтронів, навіть якщо можливість сплайсингу було усунуто (оглянуто Shaul, 2017). У клітинах ссавців введення в безінтронну конструкцію 5'-сайту сплайсингу (SS) або повного інтрону підвищувало рівні мРНК у 2 та 8 разів, відповідно, порівняно з безінтронною конструкцією (Damgaard et al, 2008). У кукурудзи фрагменти 5'-UTR інтрону гена *SHI* обумовлюють 25–44-разове посилення експресії, але лише 2-разове підвищення у тому разі, коли їхні сайти сплайсингу є мутованими (Clancy and Hannah, 2002). Мутагенез сайтів розпізнавання сплайсингу, який повністю порушує сплайсинг, усуває ІМЕ для деяких інтронів, таких, як інтрони генів кукурудзи *ADHI* (Luehrsen and Walbot, 1991), *HSP82* і *SHI* (Sinibaldi and Mettler, 1992). Інтрон гена *TRP1 (PAT1)* арабідопсису зумовлює приблизно 5-та 2,5-разове посилення експресії інтронами, здатними до сплайсингу, та з мутованими 5'-сайтами сплайсингу, відповідно (Rose, 2002; Rose and Beliakoff, 2000). 5'-UTR інтрон гена *AtMHX Arabidopsis* значно посилює експресію гена (David-Assael et al, 2006). Нативний інтрон або інтрон з мутованими 5'- і 3'-сайтами сплайсингу зумовлюють 270- і 5-разове посилення експресії, відповідно, порівняно з конструкцією без інтрону (Akua et al, 2010). Загалом, ці дані свідчать про те, що в клітинах ссавців, а також однодольних (кукурудза) і дводольних (*Arabidopsis*) рослин сплайсинг є важливим для істотного посилення експресії, але за відсутності сплайсингу незначне посилення можна отримати при наявності у гені послідовностей інтронів.

Було припущено, що можуть існувати два окремі ефекти на рівні мРНК, – незалежне від сплайсингу збільшення у 2–5 разів і більш істотне підвищення завдяки залежному від сплайсингу механізму, який характерний лише для деяких інтронів (Akua et al, 2010). Схоже, що сплайсинг сам по собі недостатній для виникнення ІМЕ, оскільки ступінь посилення експресії значно змінюється, залежно від ви-

користовуваного інтрону, а деякі інтрони, які успішно вирізаються з послідовності, не справляють жодного ефекту на експресію (Morello and Breviaro, 2008).

*Вплив ІМЕ на накопичення мРНК.* Невелика різниця сигналу, отримана в тестах аналізу ядерного синтезу (nuclear run-on) з інтроном і без нього, свідчить про те, що інтрони переважно підвищують накопичення мРНК за допомогою посттранскрипційного механізму (Rose and Last 1997). Є докази того, що на додаток до помірного посилення транскрипції існує сильне посттранскрипційне посилення. У багатьох випадках присутність інтронів підвищує рівні стаціонарного стану зрілої мРНК у цитозолі (Akua et al, 2010; Brinster et al, 1988; Callis et al, 1987; Curi et al, 2005; Dean et al, 1989; Morello et al, 2011; Neuberger і Williams, 1988; Nott et al, 2003; Rethmeier et al, 1997; Rose, 2004, 2008). Це може бути наслідком впливу на швидкість транскрипції, ядерний експорт і стабільність транскрипту (Shaul, 2017). Подробиці щодо посттранскрипційної регуляції експресії генів за допомогою інтрону висвітлені в багатьох нещодавно опублікованих оглядах (Gallegos і Rose, 2015; Laha, 2017; Shaul, 2017).

Практично всі етапи синтезу та розпаду мРНК можуть бути основоположним рівнем, на якому ІМЕ впливає на накопичення мРНК (Gallegos and Rose, 2015). Сплайсинг може сприяти експорту мРНК у цитозоль (Shaul, 2017). Швидкість експорту мРНК у 6–10 разів вища для мРНК, яка була піддана сплайсингу, порівняно з її кДНК аналогами (Valencia et al, 2008). мРНК експортується з ядра до цитозолу у вигляді комплексу інформаційного рибонуклеопротеїну (mRNP), який містить багатобілковий комплекс зрощування екзонів. Під час сплайсингу сплайсосома приєднує комплекс зрощування екзонів на 20–24 нуклеотиди вище від з'єднань ексона з екзоном (exon-exon junction) (Boehm and Gehring, 2016; Le Hir et al, 2016), що пізніше сприяє експорту мРНК з ядра до цитозолу (Le Hir et al, 2001) шляхом взаємодії з коротким мотивом компоненту ALYREF транскрипційно-експортного комплексу (transcription-export complex) (Gromadzka et al, 2016). ALYREF взаємодіє з 5'-кепом і компонентами комплексу зрощуван-

ня екзонів або зв'язуючими білками. Можливо, близькість інтрону до сайту початку транскрипції і, отже, комплекс зрощування екзонів до 5'-кепа, може сприяти цій взаємодії (Shaul, 2017). ІМЕ може проявляти себе через полегшення будь-якого з названих процесів.

У свою чергу, описано низку механізмів, за допомогою яких інтрони можуть позитивно впливати на стабільність мРНК, включаючи більшу кількість інтронів у гені, посилення процесингу 3'-кінця пре-мРНК та поліаденілювання, пригнічення передчасного розщеплення та поліаденілювання шляхом зв'язування U1 snRNA з 5' SS (Shaul, 2017). Крім того, у ссавців розташування інтрону на  $\geq 50$ –55 п.н. нижче кодону термінації знижує стабільність мРНК проти механізму розпаду антисенсових мРНК. Хоча основні характеристики цього механізму зберігаються схожими у рослин та ссавців, існують деякі механістичні аспекти, в яких вони можуть відрізнятися (оглянуто Shaul, 2017). Наразі виглядає так, що жоден з цих механізмів не пов'язаний з ІМЕ.

*Вплив ІМЕ на ефективність трансляції.* Вплив ІМЕ на посттранскрипційні та трансляційні стадії експресії було нещодавно оглянуто (Laha, 2017; Shaul, 2017). ІМЕ ефект 5'-UTR інтрона гена *RUB13* рису було кількісно оцінено на різних рівнях регуляції генів у суспензії трансгенних клітин рису. Як результат, було виявлено майже дворазове підвищення на рівні транскрипції, приблизно 20-разове підвищення накопичення мРНК, посилення приблизно на 45 % на трансляційному рівні та 29-разове посилення на рівні білку, що було визначено шляхом оцінки активності ферменту GUS (Sattadde et al, 2008). Таким чином, ІМЕ ефект 5'-UTR інтрону гена *RUB13* рису має помірний ефект як на транскрипційному, так і на трансляційному рівнях, але значно посилюється на посттранскрипційному рівні, що базується на посиленні на попередніх рівнях експресії. Такий же ефект був показаний для цього інтрону для трансгенних рослин рису; крім того, ІМЕ вплив гена *RUB13* на трансляцію залежить від тканини (Lu et al, 2008). Подібним чином спостерігали підвищення кількості мРНК гена *CAT* у 3,9 раза та активності CAT у 12,1 раза, що вказує на те, що ІМЕ, опосередкований інтроном 2 *ADH1* кукурудзи, та-



кож націлений на два різні регуляторні рівні в протопластах кукурудзи (Mascarenhas et al, 1990). Збільшення пулу мРНК також не корелювало з підвищенням рівня або активності білку у випадку інтрона 4 гена *RROT* кукурудзи (Bourdon et al, 2001), першого інтрона гена актин-деполімеризуючого фактору 1 *PhADF1* петунії (Jeong et al, 2007) та інтрона-лідера двох генів *A. thaliana*, що кодує субодиницю 5С цитохром с-оксидази (Cugi et al, 2005).

Було також показано, що наявність інтронів підвищує ефективність трансляції мРНК у ссавців, дріжджів і *Xenopus* (Shaul, 2017). Підвищення ефективності трансляції в цитозолі завдяки інтронам, які піддаються сплайсингу в ядрі, залежить від комплексу зрощування екзонів, який видаляється з мРНК під час першого раунду трансляції (Boehm і Gehring, 2016; Le Hir et al, 2016; Nott et al, 2004). Слід зазначити, що можливий механізм залучення комплексу зрощування екзонів до ІМЕ може бути різним у рослин і тварин (Samadder et al, 2008). Проте, було показано, що кількість лідерів комплексів зрощування екзонів та відкритих рамок зчитування (uORF) зворотно корелює з трансляцією основного білку у дрозофіли, рибики данію, миші, людини та *A. thaliana*. Серед п'яти досліджених видів найнижчі рівні трансляції спостерігалися для мРНК як з лідерами екзон-екзонних з'єднань, так і з uORF (29 %) (Lim et al, 2018).

При вивченні ІМЕ за допомогою 5'-UTR інтрона гена *MHX A. thaliana* показано, що для серії протестованих конструкцій показники ІМЕter корелювали з їхньою здатністю підвищувати рівні мРНК, але не з їхнім впливом на ефективність трансляції (Akua and Shaul, 2013). Це вказує на те, що здатність покращувати трансляцію не опосередковується мотивами ІМЕter. Натомість внутрішній елемент довжиною 118 п.н. зі специфічною здатністю підвищувати ефективність трансляції, не впливаючи на сплайсинг, був ідентифікований за допомогою делеційного аналізу в 5'-UTR інтрона довжиною 416 п.н. гена *AtMHX*. Введення цього елемента в ген *SAT1* призвело до 19-разового підвищення ефективності трансляції. Крім того, перенесення посилюючих інтронних конструкцій з 5'-UTR у кодуючу послідовність зменшувало їх здатність

посилювати трансляцію (Akua and Shaul, 2013). Цікаво, що 5'-UTR інтрон гена *AtMHX* також містить короткий елемент, багатий на U (73 % залишків U), який, незважаючи на позитивний внесок у сплайсинг, знижує здатність першого елемента посилювати трансляцію (Akua and Shaul, 2013). Таким чином, існують інтронні елементи, здатні позитивно чи негативно впливати на трансляцію.

Можливим поясненням здатності ІМЕ інтронів підвищувати ефективність трансляції може бути здатність певних інтронів або інтронних елементів по-різному впливати на відкладення, локалізацію та/або склад білків mRNP, особливо периферичних білків комплексів зрощування екзонів (оглянуто Bono and Gehring, 2011; Shaul, 2017).

*Гени, у яких виявляється інтрон-опосередковане посилення експресії.* Перші інтрони генів домашнього господарства або основних генів містять більшу частку активних хроматинових міток, ніж тканиноспецифічні гени або несуттєві гени, а гени з високим рівнем експресії демонструють більшу щільність хроматинових регуляторних модифікацій, ніж гени з низькими рівнями експресії (Jo and Choi, 2019b). Більше того, гени, що несуть численні, притаманні першим інтронам, однонуклеотидні поліморфізми, асоційовані з певними ознаками (trait-associated single-nucleotide polymorphisms), взаємодіють один з одним у великій мережі білок-білкових взаємодій (Jo and Choi, 2019b). Таким чином, більшою є імовірність знайти явище ІМЕ в першу чергу у генах домашнього господарства та генах з високим рівнем експресії, а також у генах, пов'язаних з ними у мережах білок-білкових взаємодій.

*Тканинна специфічність ІМЕ.* Після виявлення чітких закономірностей експресії ізотипів тубуліну рису за допомогою ІМЕ (Fiume et al, 2004; Giani et al, 2009) для опису впливу регуляторних ІМЕ інтронів як на рівень, так і на фактичне місце експресії це явище було означено, що вже зазначалось вище, як інтрон-залежна просторова експресія (Morello and Breviaro, 2008). Тканиноспецифічну експресію та експресію за певних умов, модульовану ІМЕ інтронами, було також раніше розглянуто в літературі (Morello et al, 2011; Laha, 2017).

Інтрони можуть повністю перевизначати конституційну або тканиноспецифічну регуляцію експресії промотором. У *Arabidopsis* перший інтрон гена *UBQ10* порушує тканиноспецифічну експресію генів *CNGC2* та *YAB3* і призводить до експресії обох генів у коренях (Emami et al, 2013). Аналогічним чином даний інтрон забезпечує в коренях експресію генів *GAE1*, *ROP10*, *ADL1A*, *MSBP1* і *ULI3* (Emami et al, 2013).

Деякі інтрони містять послідовності, яким притаманні властивості промотору, що забезпечує експресію генів за відсутності мінімальної промоторної послідовності. Наприклад, інтрон гена *UBQ1* кукурудзи у однодольних (Salgueiro et al, 2000) і *mad-box* ген *FBP11* (floral binding protein 11) петунії у дводольних (Liao et al, 2013). Продемонстровано, що тканиноспецифічна експресія часто асоціюється з промоторною активністю інтрону. 5'-UTR інтрон гена *FAD2* кунжуту, який кодує десатуразу жирних кислот, на додаток до тканинної специфічності має низьку промоторну активність. Даний інтрон здатен перевизначити конститутивний профіль експресії промотору 35S і зумовлювати експресію GUS специфічно у насінні *Arabidopsis*, що розвивається, де він також здатен забезпечувати низький рівень експресії за відсутності промотору (Kim et al, 2006). У свою чергу, інтрон 5'-UTR гена *FAD2 Brassica napus*, якому притаманна незначна промоторна активність у трансгенному арабідопсисі посилює експресію в усіх проаналізованих тканинах (Xiao et al, 2014). Припускається, що ця невідповідність пояснюється тим фактом, що інтрони, які походять з інших видів, можуть по-різному впливати на тканинну специфічність (Лаха, 2017).

Інтрони, пов'язані з ІМЕ, можуть забезпечувати диференціальну експресію генів різних ізотипів білків з однієї родини у репродуктивних або вегетативних тканинах і на різних стадіях розвитку. Показано, що конститутивна експресія багатьох генів, що кодують білки цитоскелету, регулюється інтронами. Наприклад, виявлено, що експресія генів *PRF1* і *PRF2*, що кодують вегетативні профіліни, визначається виключно першим інтроном (Jeong et al, 2006). Експресія гена *PRF5* у вегетативних

тканинах також може бути модульована першим інтроном петунії *ADF1* (фактор деполімеризації актину 1) у стабільно трансформованих рослинах *Arabidopsis* (Jeong et al, 2007). Як і для гена *PRF2*, промотор гена *PhADF1* забезпечує експресію в судинній системі, тоді як інтрон активує експресію у вегетативних тканинах *Arabidopsis* (Mun et al, 2002). Експресія гена актину *ACT1* у пилку сильно посилюється його 5'-UTR інтроном-лідером, тоді як при його заміні на L2 5'-UTR інтрон вегетативного гена актину *ACT2* експресія у пилку не зберігається (Vitale et al, 2003).

Існують і інші приклади конститутивної експресії, опосередкованої інтронами. Експресія залежного від реплікації гістону H4 є меристемно-специфічною та обмежена S-фазою клітинного циклу. Введення у послідовність даного гена першого інтрону гістону H3 нижче місця розташування промотору H4 призводить до його конститутивної експресії в арабідопсисі (Chaubet-Gigot et al, 2001). Таким чином, інтрон здатний перезаписувати тканинну специфічність, яка визначається промотором. Високий рівень експресії гена *COX5c-2 Arabidopsis*, який кодує субодиницю 5С мітохондріальної цитохром с оксидази в меристемах і активно зростаючих тканинах, зумовлена його 5'-UTR інтроном, тоді як промотор сам по собі стимулює експресію гена *GUSA* лише в пилку (Curi et al, 2005). Подібне спостереження було зроблено для *AtPUX7*, який кодує рослинний регулятор убіхітину X. Перший інтрон зумовлює сильну експресію гена *GUS* у всіх тканинах проростків, тоді як промотор — лише в ранньому чоловічому гаметофіті (Gallois et al, 2013). Показано, що інтрон 5'-UTR гена *SUVH3*, гомологу Su(var)3-9, що кодує білковий домен SET з активністю метилтрансферази H3K9, є необхідним для надання тканиноспецифічної експресії в коренях, листі та квітах (Casas-Mollano et al, 2006). Тканиноспецифічна експресія *mad-box* гена *FBP11* у петунії модулюється як промотором, так і першим інтроном. Промотор сам по собі зумовлює експресію як у вегетативних тканинах, так і у квітках, тоді як інтрону притаманна промоторна активність і він опосередковує експресію в таких частинах квітки, як чашо-

листок, пелюстки, тичинки та плодолисток. Однак, комбінація промотору та інтрону лише стимулює експресію GUS у тканинах жіночих статевих органів та плодолистка (Liao et al, 2013).

Є також дослідження, результати яких свідчать про обумовлену інтронами експресію відповідно до стадій розвитку. 5'-UTR інтрон-лідер гену *OsTUB4* рису модулює експресію у вузлах, міжвузлях і листі (Giani et al, 2009). Подібне спостереження було зроблено для гена *TUA1* рису, перший інтрон якого спрямовує експресію генів до тканин, що активно діляться, як-от кінчики коренів (Jeon et al, 2000). 5'-UTR інтрон гена *PHT1;4 Arabidopsis*, що кодує високоафінний транспортер фосфату, є важливим для експресії в кінчиках коренів і для збільшення експресії під час фосфатного голодування (Karthikeyan et al, 2009). Таким чином, є докази того, що опосередкована інтронами експресія може брати участь у реакції на стрес і у механізмі трансляції сигналу середовища в експресію гена.

Для рослин залишається відкритим питання, чи є тканиноспецифічна активація інтрону наслідком справжньої ініціації транскрипції, посиленням експресії порівняно з фоновими рівнями чи комбінацією обох (Laha, 2017). У разі справжньої ініціації транскрипції, цис-елементи та активність інтрону як промотору можуть бути необхідними, щоб уможливити залучення апарату транскрипції до промотору. У разі простого посилення передбачається, що експресія в певній тканині вже зумовлена, але вона надто низька для виявлення.

*Репресія, негативна регуляція різних інтронних елементів.* Інтрон 1 гена *PhADF1* необхідний для експресії цього гена у трансгенному тютюні на достатньому для виявлення рівні. Якщо його замінити на інтрон 2 того ж гена або на цей же підсилюючий інтрон 1 у зворотній орієнтації, рівень експресії падає нижче рівня, який спостерігається для конструкції, у якій інтрони взагалі відсутні (Mun et al, 2002). Ген *MHX A. thaliana* включає інтронний елемент, який посилює трансляцію, коли він локалізований у 5'-UTR інтрона (Akua and Shaul, 2013). Вплив ідентифікованого інтронного елемента на ефективність трансляції може бути посилено шляхом видалення сусідніх інтрон-

них елементів. Наразі немає інших свідчень взаємодії між різними інтронними елементами. Така інтерференція може підтвердити біоінформатично обґрунтоване припущення того, що деякі додаткові послідовності інтронів також необхідні для розділення різних функціональних інтронних елементів з тим, щоб білки або РНК, які з ними зв'язуються, не заважали один одному (Bradnam and Korf, 2008).

### Висновки

Таким чином, можна констатувати, що не існує єдиного механізму ІМЕ. Цей ефект може відрізнятися у різних видів, а один інтрон ІМЕ може впливати на кілька етапів експресії генів. Зазвичай, найбільш виражений ефект ІМЕ спостерігається для стадії накопичення мРНК та трансляції. Перелік факторів, які можуть визначати ІМЕ за різних умов і на різних рівнях експресії генів, включає, але не обмежується такими обставинами, як довжина інтрону, його положення та орієнтація в гені, послідовність та нуклеотидний склад інтрону та фланкуючих послідовностей, наявність певних модифікацій ДНК та ІМЕ мотивів, наявність промоторної активності, фізичні властивості і структура ДНК та РНК, наявність сигналів сплайсингу та здатність до сплайсингу, забезпечення впливу на різні посттранскрипційні механізми. Крім того, мотиви пов'язаних з незалежним від сплайсингу ІМЕ у рослинах, не обмежується інтронами, а можуть розташовуватись і в 5'-UTR фрагментах або кодуючих послідовностях. В свою чергу, при аналізі даних, отриманих в різних експериментах, необхідно враховувати використаний організм, систему експресії, тканини та стадію розвитку, вплив зовнішніх факторів, використаний промотор та особливості генетичної конструкції. При моделюванні процесу ІМЕ слід порівнювати між собою лише випадки ІМЕ, які впливають на експресію на одному і тому ж самому рівні експресії, враховуючи вид організму та умови експресії. Виділення біологічних факторів, які можуть визначати ІМЕ, та співвідношень між ними допоможе в подальшому створити відповідний набір даних, зручний для машинного навчання, та спробувати розгадати таємницю феномену ІМЕ за допомогою машинного навчання.



MECHANISMS OF INTRON-MEDIATED  
ENHANCEMENT OF EXPRESSION:  
WELCOME TO HOTEL CALIFORNIA

M. O. Pydiura, Ya. B. Blume

Institute of Food Biotechnology and Genomics NAS  
of Ukraine, Bayda Vyshnevetskyi street (Osipovskogo),  
2A, 04123, Kyiv, Ukraine  
JSC «Farmak», Kyrylivska Street, 63, 04080, Kyiv,  
Ukraine

E-mail: pydiura@gmail.com,  
cellbio@cellbio.freenet.viaduk.net

The phenomenon of the positive influence of introns on the expression of a corresponding gene, which is called intron-mediated enhancement (IME), is characteristic of a wide variety of organisms, including nematodes, insects, mammals, fungi, and plants, and occurs due to an as-yet-undefined fundamental mechanism. IME introns have been used for a long time, in particular, in plant biotechnology. Understanding the mechanisms of this phenomenon allows predicting and easily generating stimulatory introns with the given properties and creating highly advantageous phenotypes. It will also greenlight the use of IME in gene therapy and to improve the production of pharmaceutical proteins. In this review, we analyzed previously proposed models of IME functioning mechanisms and identified factors that can directly or indirectly determine IME under different conditions and at different levels of gene expression, such as experimental methods of IME research, regulatory RNAs, sequence properties, intron position and orientation, factors at the levels of DNA, transcription, splicing, mRNA, translation, genes in which IME is detected, tissue specificity, repression and how some factors relate to each other by importance. Since there is no single mechanism of IME, and the effect may differ in different species, when modeling this process, only the cases of IME affecting the same level of expression should be compared with each other, taking into account the experimental conditions. Identifying the biological factors that can determine IME and the relationship between them will help in the future to create a corresponding dataset suitable for machine learning and to try to solve the mystery of the IME phenomenon with the help of machine learning.

СПИСОК ЛІТЕРАТУРИ

Agarwal NY, Ansari A (2016) Enhancement of transcription by a splicing-competent intron is dependent on promoter directionality. *PLoS Genet* 12(5): e1006047. <https://doi.org/10.1371/journal.pgen.1006047>  
Akua T, Shaul O (2013) The *Arabidopsis thaliana* MHX gene includes an intronic element that boosts translation when localized in a 5' UTR intron. *J Exp Bot* 64(14): 4255–4270. <https://doi.org/10.1093/jxb/ert235>

Akua T, Berezin I, Shaul O (2010) The leader intron of AtMHX can elicit, in the absence of splicing, low-level intron-mediated enhancement that depends on the internal intron sequence. *BMC Plant Biol* 10:93. <https://doi.org/10.1186/1471-2229-10-93>  
Al-Husini N, Medler S, Ansari A (2020) Crosstalk of promoter and terminator during RNA polymerase II transcription cycle. *Biochim Biophys Acta Gene Regul Mech* 1863(12):194657. <https://doi.org/10.1016/j.bbagr.2020.194657>  
Alipanahi B, DeLong A, Weirauch MT et al (2015) Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol* 33(8):831–838. <http://dx.doi.org/10.1038/nbt.3300> PMID: 26213851  
Auslander N, Gussow AB, Koonin EV (2021) Incorporating machine learning into established bioinformatics frameworks. *Int J Mol Sci* 22(6):2903. <https://doi.org/10.3390/ijms22062903>  
Baier T, Jacobebbinghaus N, Einhaus A et al (2020) Introns mediate post-transcriptional enhancement of nuclear gene expression in the green microalga *Chlamydomonas reinhardtii*. *PLoS Genet* 16(7): e1008944. <https://doi.org/10.1371/journal.pgen.1008944>  
Barshai M, Tripto E, Orenstein Y (2020) Identifying regulatory elements via deep learning. *Ann Rev Biomed Data Sci* 3:315–338. <https://doi.org/10.1146/annurev-bio-datasci-022020-021940>  
Basso MF, Arraes FBM, Grossi-de-Sa M et al (2020) Insights into genetic and molecular elements for transgenic crop development. *Front. Plant Sci* 11: 509. <https://doi.org/10.3389/fpls.2020.00509>  
Bhatti GK, Khullar N, Sidhu IS et al (2021) Emerging role of non-coding RNA in health and disease. *Metab Brain Dis* 36(6):1119–1134. <https://doi.org/10.1007/s11011-021-00739-y>  
Bieberstein NI, Carrillo Oesterreich F, Straube K et al (2012) First exon length controls active chromatin signatures and transcription. *Cell Rep* 2(1):62–68. <https://doi.org/10.1016/j.celrep.2012.05.019>  
Boehm V, Gehring NH (2016) Exon junction complexes: supervising the gene expression assembly line. *Trends Genet* 32(11):724–735. <https://doi.org/10.1016/j.tig.2016.09.003>  
Bogard B, Francastel C, Hubé F (2020) Multiple information carried by RNAs: total eclipse or a light at the end of the tunnel? *RNA Biol* 17(12):1707–1720. <https://doi.org/10.1080/15476286.2020.1783868>  
Bourdon V, Harvey A, Lonsdale DM (2001) Introns and their positions affect the translational activity of mRNA in plant cells. *EMBO Rep* 2(5):394–398. <https://doi.org/10.1093/embo-reports/kve090>  
Bradnam KR, Korf I (2008) Longer first introns are a general property of eukaryotic gene structure. *PLoS*



- One 3(8):e3093. <https://doi.org/10.1371/journal.pone.0003093>
- Callis J, Fromm M, Walbot V (1987) Introns increase gene expression in cultured maize cells. *Genes Dev* 1(10):1183–1200. <https://doi.org/10.1101/gad.1.10.1183>
- Casas-Mollano JA, Lao NT, Kavanagh TA (2006) Intron-regulated expression of SUVH3, an Arabidopsis Su(var)3-9 homologue. *J Exp Bot* 57(12):3301–3311. <https://doi.org/10.1093/jxb/erl093>
- Chaubet-Gigot N, Kapros T, Flenet M et al (2001) Tissue-dependent enhancement of transgene expression by introns of replacement histone H3 genes of Arabidopsis. *Plant Mol Biol* 45(1):17–30. <https://doi.org/10.1023/a:1006487023926>
- Chaudhary S, Khokhar W, Jabre I et al (2019) Alternative splicing and protein diversity: plants versus animals. *Front Plant Sci* 10:708. <https://doi.org/10.3389/fpls.2019.00708>
- Chicco D (2017) Ten quick tips for machine learning in computational biology. *BioData Min* 10:35. <https://doi.org/10.1186/s13040-017-0155-3>
- Chorev M, Carmel L (2012) The function of introns. *Front Genet* 3:55. <https://doi.org/10.3389/fgene.2012.00055>
- Chung BY, Simons C, Firth AE et al (2006) Effect of 5'UTR introns on gene expression in Arabidopsis thaliana. *BMC Genomics* 7:120. <https://doi.org/10.1186/1471-2164-7-120>
- Clancy M, Hannah LC (2002) Splicing of the maize Sh1 first intron is essential for enhancement of gene expression, and a T-rich motif increases expression without affecting splicing. *Plant Physiol* 130(2):918–929. <https://doi.org/10.1104/pp.008235>
- Clancy M, Vasil V, Hannah CL et al (1994) Maize Shrunken-1 intron and exon regions increase gene expression in maize protoplasts. *Plant Sci* 98:151–161. [https://doi.org/10.1016/0168-9452\(94\)90005-1](https://doi.org/10.1016/0168-9452(94)90005-1)
- Curi GC, Chan RL, Gonzalez DH (2005) The leader intron of Arabidopsis thaliana genes encoding cytochrome c oxidase subunit 5c promotes high-level expression by increasing transcript abundance and translation efficiency. *J Exp Bot* 56(419):2563–2571. <https://doi.org/10.1093/jxb/eri250>
- Custodio N, Carmo-Fonseca M (2016) Co-transcriptional splicing and the CTD code. *Crit Rev Biochem Mol Biol* 51(5):395–411. <https://doi.org/10.1080/10409238.2016.1230086>
- Damgaard CK, Kahns S, Lykke-Andersen S et al (2008) A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell* 29(2):271–278. <https://doi.org/10.1016/j.molcel.2007.11.035>
- David-Assael O, Berezin I, Shoshani-Knaani N et al (2006) AtMHX is an auxin and ABA-regulated transporter whose expression pattern suggests a role in metal homeostasis in tissues with photosynthetic potential. *Funct Plant Biol* 33(7):661–672. <https://doi.org/10.1071/FP05295>
- Dean C, Favreau M, Bond-Nutter D et al (1989) Sequences downstream of translation start regulate quantitative expression of two petunia rbcS genes. *Plant Cell* 1(2):201–208. <https://doi.org/10.1105/tpc.1.2.201>
- Depicker A, Montagu MV (1997) Post-transcriptional gene silencing in plants. *Curr Opin Cell Biol* 9(3):373–382. [https://doi.org/10.1016/s0955-0674\(97\)80010-5](https://doi.org/10.1016/s0955-0674(97)80010-5)
- Donath M, Mendel R, Cerff R et al (1995) Intron-dependent transient expression of the maize GapA1 gene. *Plant Mol Biol* 28(4):667–676. <https://doi.org/10.1007/BF00021192>
- Dwyer K, Agarwal N, Gega A et al (2021a) Proximity to the promoter and terminator regions regulates the transcription enhancement potential of an intron. *Front Mol Biosci* 8:712639. <https://doi.org/10.3389/fmolb.2021.712639>
- Dwyer K, Agarwal N, Pile L et al (2021b) Gene architecture facilitates intron-mediated enhancement of transcription. *Front Mol Biosci* 8:669004. <https://doi.org/10.3389/fmolb.2021.669004>
- Eamens A, Wang MB, Smith NA et al (2008) RNA silencing in plants: yesterday, today, and tomorrow. *Plant Physiol* 147(2):456–468. <https://doi.org/10.1104/pp.108.117275>
- Emami S, Arumainayagam D, Korf I et al (2013) The effects of a stimulating intron on the expression of heterologous genes in Arabidopsis thaliana. *Plant Biotechnol J* 11(5):555–563. <https://doi.org/10.1111/pbi.12043>
- Eraslan G, Avsec Ž, Gagneur J et al (2019) Deep learning: new computational modelling techniques for genomics. *Nat Rev Genet* 20(7):389–403. <https://doi.org/10.1038/s41576-019-0122-6>
- Fagard M, Vaucheret H (2000) (Trans)gene silencing in plants: how many mechanisms? *Ann Rev Plant Physiol Plant Mol Biol* 51:167–194. <https://doi.org/10.1146/annurev.arplant.51.1.167>
- Fiume E, Christou P, Gianm S et al (2004) Introns are key regulatory elements of rice tubulin expression. *Planta* 218(5):693–703. <https://doi.org/10.1007/s00425-003-1150-0>
- Furger A, O'Sullivan JM, Binnie A et al (2002) Promoter proximal splice sites enhance transcription. *Genes Dev* 16(21):2792–2799. <https://doi.org/10.1101/gad.983602>
- Gallegos JE, Rose AB (2019) An intron-derived motif strongly increases gene expression from transcribed

- sequences through a splicing independent mechanism in *Arabidopsis thaliana*. *Sci Rep* 9(1):13777. <https://doi.org/10.1038/s41598-019-50389-5>
- Gallegos JE, Rose AB (2017) Intron DNA Sequences can be more important than the proximal promoter in determining the site of transcript initiation. *Plant Cell* 29(4):843–853. <https://doi.org/10.1105/tpc.17.00020>
- Gallegos JE, Rose AB (2015) The enduring mystery of intron-mediated enhancement. *Plant Sci* 237:8–15. <https://doi.org/10.1016/j.plantsci.2015.04.017>
- Gallois JL, Drouaud J, Lécureuil A et al (2013) Functional characterization of the plant ubiquitin regulatory X (UBX) domain-containing protein AtPUX7 in *Arabidopsis thaliana*. *Gene* 526(2):299–308. <https://doi.org/10.1016/j.gene.2013.05.056>
- Giani S, Altana A, Campanoni P et al (2009) In transgenic rice, alpha- and beta-tubulin regulatory sequences control GUS amount and distribution through intron mediated enhancement and intron dependent spatial expression. *Transgenic Res* 18(2):151–162. <https://doi.org/10.1007/s11248-008-9202-7>
- Gidekel M, Jimenez B, Herrera-Estrella L (1996) The first intron of the *Arabidopsis thaliana* gene coding for elongation factor 1 beta contains an enhancer-like element. *Gene* 170(2):201–206. [https://doi.org/10.1016/0378-1119\(95\)00837-3](https://doi.org/10.1016/0378-1119(95)00837-3)
- Gromadzka AM, Steckelberg AL, Singh KK et al (2016) A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. *Nucl Acids Res* 44(5):2348–2361. <https://doi.org/10.1093/nar/gkw009>
- Hu X, Fernie AR, Yan J (2023) Deep learning in regulatory genomics: from identification to design. *Curr Opin Biotechnol* 79:102887. <https://doi.org/10.1016/j.copbio.2022.102887>
- Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J* 6(13):3901–3907. <https://doi.org/10.1002/j.1460-2075.1987.tb02730.x>
- Jeon JS, Lee S, Jung KH et al (2000) Tissue-preferential expression of a rice alpha-tubulin gene, OsTubA1, mediated by the first intron. *Plant Physiol* 123(3):1005–1014. <https://doi.org/10.1104/pp.123.3.1005>
- Jeong YM, Mun JH, Kim H et al (2007) An upstream region in the first intron of petunia actin-depolymerizing factor 1 affects tissue-specific expression in transgenic *Arabidopsis thaliana*. *Plant J* 50(2):230–239. <https://doi.org/10.1111/j.1365-3113.2007.03053.x>
- Jeong YM, Mun JH, Lee I et al (2006) Distinct roles of the first introns on the expression of *Arabidopsis* profilin gene family members. *Plant Physiol* 140(1):196–209. <https://doi.org/10.1104/pp.105.071316>
- Jo BS, Choi SS (2015) Introns: the functional benefits of introns in genomes. *Genomics Inform* 13(4):112–118. <https://doi.org/10.5808/GI.2015.13.4.112>
- Jo SS, Choi SS (2019a) Enrichment of rare alleles within epigenetic chromatin marks in the first intron. *Genomics Inform* 17(1):e9. <https://doi.org/10.5808/GI.2019.17.1.e9>
- Jo SS, Choi SS (2019b) Analysis of the functional relevance of epigenetic chromatin marks in the first intron associated with specific gene expression patterns. *Genome Biol Evol* 11(3):786–797. <https://doi.org/10.1093/gbe/evz033>
- Karthikeyan AS, Ballachanda DN, Raghothama KG (2009) Promoter deletion analysis elucidates the role of cis elements and 5'UTR intron in spatiotemporal regulation of AtPht1, vol. 4 expression in *Arabidopsis*. *Physiol Plant* 136(1):10–18. <https://doi.org/10.1111/j.1399-3054.2009.01207.x>
- Kertész S, Kerényi Z, Mérai Z et al (2006) Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. *Nucl Acids Res* 34(21):6147–6157. <https://doi.org/10.1093/nar/gkl737>
- Kim MJ, Kim H, Shin JS et al (2006) Seed-specific expression of sesame microsomal oleic acid desaturase is controlled by combinatorial properties between negative cis-regulatory elements in the SeFAD2 promoter and enhancers in the 5'-UTR intron. *Mol Genet Genomics* 276(4):351–368. <https://doi.org/10.1007/s00438-006-0148-2>
- Kim S, Kim H, Fong N et al (2011) Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc Natl Acad Sci USA* 108(33):13564–13569. <https://doi.org/10.1073/pnas.1109475108>
- Kooter JM, Matzke MA, Meyer P (1999) Listening to the silent genes: transgene silencing, gene regulation and pathogen control. *Trends Plant Sci* 4(9):340–347. [https://doi.org/10.1016/s1360-1385\(99\)01467-3](https://doi.org/10.1016/s1360-1385(99)01467-3)
- Korf IF, Rose AB (2009) Applying word-based algorithms: the IMEter. *Methods Mol Biol* 553:287–301. [https://doi.org/10.1007/978-1-60327-563-7\\_14](https://doi.org/10.1007/978-1-60327-563-7_14)
- Laxa M (2017) Intron-mediated enhancement: a tool for heterologous gene expression in plants? *Front Plant Sci* 7:1977. <https://doi.org/10.3389/fpls.2016.01977>
- Laxa M, Müller K, Lange N et al (2016) The 5'UTR intron of *Arabidopsis* GGT1 aminotransferase enhances promoter activity by recruiting RNA polymerase II. *Plant Physiol* 172(1):313–327. <https://doi.org/10.1104/pp.16.00881>
- Le Hir H, Gatfield D, Izaurralde E et al (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-

- mediated mRNA decay. *EMBO J* 20(17):4987–4997. <https://doi.org/10.1093/emboj/20.17.4987>
- Le Hir H, Nott A, Moore MJ (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* 28(4):215–220. [https://doi.org/10.1016/S0968-0004\(03\)00052-5](https://doi.org/10.1016/S0968-0004(03)00052-5)
- Le Hir H, Saulière J, Wang Z (2016) The exon junction complex as a node of post-transcriptional networks. *Nat Rev Mol Cell Biol* 17(1):41–54. <https://doi.org/10.1038/nrm.2015.7>
- Liao L, Ning G, Liu C et al (2013) The intron from the 5'-UTR of the *FBP11* gene in *petunia* displays promoter- and enhancer-like functions. *Sci Hort* 2013:154:96–101. <https://doi.org/10.1016/j.scienta.2013.02.009>
- Lim CST, Wardell SJ, Kleffmann T et al (2018) The exon-intron gene structure upstream of the initiation codon predicts translation efficiency. *Nucl Acids Res* 46(9):4575–4591. <https://doi.org/10.1093/nar/gky282>
- Lu J, Sivamani E, Li X et al (2008) Activity of the 5' regulatory regions of the rice polyubiquitin *rubi3* gene in transgenic rice plants as analyzed by both GUS and GFP reporter genes. *Plant Cell Rep* 27(10):1587–1600. <https://doi.org/10.1007/s00299-008-0577-y>
- Lu S, Cullen BR (2003) Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *RNA* 9(5):618–630. <https://doi.org/10.1261/rna.5260303>
- Luehrsen KR, Walbot V (1991) Intron enhancement of gene expression and the splicing efficiency of introns in maize cells. *Mol Gen Genet* 225(1):81–93. <https://doi.org/10.1007/BF00282645>
- Maas C, Laufs J, Grant S et al (1991) The combination of a novel stimulatory element in the first exon of the maize *Shrunken-1* gene with the following intron 1 enhances reporter gene expression up to 1000-fold. *Plant Mol Biol* 16(2):199–207. <https://doi.org/10.1007/BF00020552>
- Mascarenhas D, Mettler IJ, Pierce DA et al (1990) Intron-mediated enhancement of heterologous gene expression in maize. *Plant Mol Biol* 15(6):913–920. <https://doi.org/10.1007/BF00039430>
- McElroy D, Zhang W, Cao J et al (1990) Isolation of an efficient actin promoter for use in rice transformation. *Plant Cell* 2(2):163–171. <https://doi.org/10.1105/tpc.2.2.163>
- Meagher RB, McKinney EC, Vitale AV (1999) The evolution of new structures: clues from plant cytoskeletal genes. *Trends Genet* 15(7):278–284. [https://doi.org/10.1016/S0168-9525\(99\)01759-x](https://doi.org/10.1016/S0168-9525(99)01759-x)
- Moabbi AM, Agarwal N, El Kaderi B et al (2012) Role for gene looping in intron-mediated enhancement of transcription. *Proc Natl Acad Sci USA* 109(22):8505–8510. <https://doi.org/10.1073/pnas.1112400109>
- Morello L, Bardini M, Cricm M et al (2006) Functional analysis of DNA sequences controlling the expression of the rice *OsCDPK2* gene. *Planta* 223(3):479–491. <https://doi.org/10.1007/s00425-005-0105-z>
- Morello L, Bardini M, Sala F et al (2002) A long leader intron of the *Ostub16* rice beta-tubulin gene is required for high-level gene expression and can autonomously promote transcription both in vivo and in vitro. *Plant J* 29(1):33–44. <https://doi.org/10.1046/j.0960-7412.2001.01192.x>
- Morello L, Breviario D (2008) Plant spliceosomal introns: not only cut and paste. *Curr Genomics* 9(4):227–238. <https://doi.org/10.2174/138920208784533629>
- Morello L, Gianm S, Troina F et al (2011) Testing the IMeter on rice introns and other aspects of intron-mediated enhancement of gene expression. *J Exp Bot* 62(2):533–544. <https://doi.org/10.1093/jxb/erq273>
- Morita S, Tsukamoto S, Sakamoto A et al (2012). Differences in intron-mediated enhancement of gene expression by the first intron of cytosolic superoxide dismutase gene from rice in monocot and dicot plants. *Plant Biotechnol J* 29:115–119. <https://doi.org/10.5511/plantbiotechnology.11.1207a>
- Mun JH, Lee SY, Yu HJ et al (2002) *Petunia* actin-depolymerizing factor is mainly accumulated in vascular tissue and its gene expression is enhanced by the first intron. *Gene* 292(1–2):233–243. [https://doi.org/10.1016/S0378-1119\(02\)00646-7](https://doi.org/10.1016/S0378-1119(02)00646-7)
- Nesic D, Cheng J, Maquat LE (1993) Sequences within the last intron function in RNA 3'-end formation in cultured cells. *Mol Cell Biol* 13(6):3359–3369. <https://doi.org/10.1128/mcb.13.6.3359-3369.1993>
- Niu DK, Yang YF (2011) Why eukaryotic cells use introns to enhance gene expression: splicing reduces transcription-associated mutagenesis by inhibiting topoisomerase I cutting activity. *Biol Direct* 6:24. <https://doi.org/10.1186/1745-6150-6-24>
- O'Sullivan JM, Tan-Wong SM, Morillon A et al (2004) Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* 36(9):1014–1018. <https://doi.org/10.1038/ng1411>
- Park SG, Hannehalli S, Choi SS (2014) Conservation in first introns is positively associated with the number of exons within genes and the presence of regulatory epigenetic signals. *BMC Genomics* 15(1):526. <https://doi.org/10.1186/1471-2164-15-526>
- Parra G, Bradnam K, Rose AB et al (2011) Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. *Nucl Acids Res* 39(13):5328–5337. <https://doi.org/10.1093/nar/gkr043>
- Plesse B, Criqui MC, Durr A et al (2001) Effects of the polyubiquitin gene *Ubi*. U4 leader intron and first ubiquitin monomer on reporter gene expression in

- Nicotiana tabacum. *Plant Mol Biol* 45(6):655–667. <https://doi.org/10.1023/a:1010671405594>
- Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* 108(4): 501–512. [https://doi.org/10.1016/s0092-8674\(02\)00617-7](https://doi.org/10.1016/s0092-8674(02)00617-7)
- Que Q, Chilton MD, de Fontes CM et al (2010) Trait stacking in transgenic crops: challenges and opportunities. *GM Crops* 1(4):220–229. <https://doi.org/10.4161/gmcr.1.4.13439>
- Rethmeier N, Seurinck J, Van Montagu M et al (1997) Intron-mediated enhancement of transgene expression in maize is a nuclear, gene-dependent process. *Plant J* 12(4):895–899. <https://doi.org/10.1046/j.1365-313x.1997.12040895.x>
- Rogozin IB, Carmel L, Csuros M et al (2012) Origin and evolution of spliceosomal introns. *Biol Direct* 7:11. <https://doi.org/10.1186/1745-6150-7-11>
- Rose AB (2002) Requirements for intron-mediated enhancement of gene expression in Arabidopsis. *RNA* 8(11):1444–1453. <https://doi.org/10.1017/s1355838202020551>
- Rose AB (2004) The effect of intron location on intron-mediated enhancement of gene expression in Arabidopsis. *Plant J* 40(5):744–751. <https://doi.org/10.1111/j.1365-313x.2004.02247.x>
- Rose AB (2008) Intron-mediated regulation of gene expression. *Curr Top Microbiol Immunol* 326:277–290. [https://doi.org/10.1007/978-3-540-76776-3\\_15](https://doi.org/10.1007/978-3-540-76776-3_15)
- Rose AB (2019) Introns as gene regulators: a brick on the accelerator. *Front Genet* 9:672. <https://doi.org/10.3389/fgene.2018.00672>
- Rose AB, Beliakoff JA (2000) Intron-mediated enhancement of gene expression independent of unique intron sequences and splicing. *Plant Physiol* 122(2):535–542. <https://doi.org/10.1104/pp.122.2.535>
- Rose AB, Last RL (1997) Introns act post-transcriptionally to increase expression of the Arabidopsis thaliana tryptophan pathway gene PAT1. *Plant J* 11(3):455–464. <https://doi.org/10.1046/j.1365-313x.1997.11030455.x>
- Rose AB, Carter A, Korf I et al (2016) Intron sequences that stimulate gene expression in Arabidopsis. *Plant Mol Biol* 92(3):337–346. <https://doi.org/10.1007/s11103-016-0516-1>
- Rose AB, Elfersi T, Parra G et al (2008) Promoter-proximal introns in Arabidopsis thaliana are enriched in dispersed signals that elevate gene expression. *Plant Cell* 20(3):543–551. <https://doi.org/10.1105/tpc.107.057190>
- Rose AB, Emami S, Bradnam K et al (2011) Evidence for a DNA-based mechanism of intron-mediated enhancement. *Front Plant Sci* 2:98. <https://doi.org/10.3389/fpls.2011.00098>
- Rowlands CF, Baralle D, Ellingford JM (2019) Machine learning approaches for the prioritization of genomic variants impacting pre-mRNA splicing. *Cells* 8(12):1513. <https://doi.org/10.3390/cells8121513>
- Salgueiro S, Pignocchi C, Parry MA (2000) Intron-mediated gusA expression in tritordeum and wheat resulting from particle bombardment. *Plant Mol Biol* 42(4):615–622. <https://doi.org/10.1023/a:1006331831858>
- Samadder P, Sivamani E, Lu J et al (2008) Transcriptional and post-transcriptional enhancement of gene expression by the 5' UTR intron of rice rubi3 gene in transgenic rice cells. *Mol Gene. Genomics* 279(4): 429–439. <https://doi.org/10.1007/s00438-008-0323-8>
- Shaul O (2015) Unique aspects of plant nonsense-mediated mRNA decay. *Trends Plant Sci* 20(11):767–779. <https://doi.org/10.1016/j.tplants.2015.08.011>
- Shaul O (2017) How introns enhance gene expression. *Int J Biochem Cell Biol* 91(Pt B):145–155. <https://doi.org/10.1016/j.biocel.2017.06.016>
- Shen X, Jiang C, Wen Y et al (2022) A Brief review on deep learning applications in genomic studies. *Front Syst Biol* 2:877717. <https://doi.org/10.3389/fsysb.2022.877717>
- Silva JCF, Teixeira RM, Silva FF et al (2019) Machine learning approaches and their current application in plant molecular biology: A systematic review. *Plant Sci* 284:37–47. <https://doi.org/10.1016/j.plan tsci.2019.03.020>
- Simna SP, Han Z (2022) Prospects of non-coding elements in genomic dna based gene therapy. *Curr Gene Ther* 22(2):89–103. <https://doi.org/10.2174/1566523221666210419090357>
- Sinibaldi RM, Mettler IJ (1992) Intron splicing and intron-mediated enhanced expression in monocots. *Prog Nucleic Acid Res Mol Biol* 42:229–257. [https://doi.org/10.1016/s0079-6603\(08\)60577-2](https://doi.org/10.1016/s0079-6603(08)60577-2)
- Snowden KC, Buchholz WG, Hall TC (1996) Intron position affects expression from the tpi promoter in rice. *Plant Mol Biol* 31(3):689–692. <https://doi.org/10.1007/BF00042241>
- Spies N, Nielsen CB, Padgett RA et al (2009) Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* 36(2):245–254. <https://doi.org/10.1016/j.molcel.2009.10.008>
- Tanaka A, Mita S, Ohta S et al (1990) Enhancement of foreign gene expression by a dicot intron in rice but not in tobacco is correlated with an increased level of mRNA and an efficient splicing of the intron. *Nucl Acids Res* 18(23):6767–6770. <https://doi.org/10.1093/nar/18.23.6767>
- To JPC, Davis IW, Marengo MS et al (2021) Expression elements derived from plant sequences provide effective gene expression regulation and new op-



- portunities for plant biotechnology traits. *Front Plant Sci* 12:712179. <https://doi.org/10.3389/fpls.2021.712179>
- Travella S, Ross SM, Harden J et al (2005) A comparison of transgenic barley lines produced by particle bombardment and *Agrobacterium*-mediated techniques. *Plant Cell Rep* 23(12):780–789. <https://doi.org/10.1007/s00299-004-0892-x>
- Valencia P, Dias AP, Reed R (2008) Splicing promotes rapid and efficient mRNA export in mammalian cells. *Proc Natl Acad Sci USA* 105(9):3386–3391. <https://doi.org/10.1073/pnas.0800250105>
- Vasil V, Clancy M, Ferl RJ et al (1989) Increased gene expression by the first intron of maize shrunken-1 locus in grass species. *Plant Physiol* 91(4):1575–1579. <https://doi.org/10.1104/pp.91.4.1575>
- Vaucheret H, Béclin C, Elmayan T et al (1998) Transgene-induced gene silencing in plants. *Plant J* 16(6):651–659. <https://doi.org/10.1046/j.1365-313x.1998.00337.x>
- Vitale A, Wu RJ, Cheng Z et al (2003) Multiple conserved 5' elements are required for high-level pollen expression of the *Arabidopsis* reproductive actin ACT1. *Plant Mol Biol* 52(6):1135–1151. <https://doi.org/10.1023/b:plan.0000004309.06973.16>
- Weise A, Rodriguez-Franco M, Timm B et al (2006) Use of *Physcomitrella patens* actin 5' regions for high transgene expression: importance of 5' introns. *Appl Microbiol Biotechnol* 70(3):337–345. <https://doi.org/10.1007/s00253-005-0087-6>
- Xiao G, Zhang ZQ, Yin CF et al (2014) Characterization of the promoter and 5'-UTR intron of oleic acid desaturase (FAD2) gene in *Brassica napus*. *Gene* 545(1):45–55. <https://doi.org/10.1016/j.gene.2014.05.008>
- Zalabók D, Ikeda Y (2020) First come, first served: sui generis features of the first intron. *Plants* 9(7):911. <https://doi.org/10.3390/plants9070911>
- Zhang Y, Yan J, Chen S et al (2020) A Review on the application of deep learning in bioinformatics. *Curr Bioinformatics* 15(8). <https://doi.org/10.2174/1574893615999200711165743>
- Zou J, Huss M, Abid A et al (2019) A primer on deep learning in genomics. *Nat Genet* 51(1):12–18. <https://doi.org/10.1038/s41588-018-0295-5>

Надійшла в редакцію 12.04.23  
Після доопрацювання 23.05.23  
Прийнята до друку 18.09.23