БІОІНФОРМАТИКА

Визначення генів – мішеней транскрипційного фактора ISGF-3

Б. Т. Токовенко, О. О. Драгущенко, А. В. Куклін, М. Ю. Оболенська

Інститут молекулярної біології і генетики НАН України
Вул. Академіка Заболотного, 150, Київ, Україна, 03680

b.t.tokovenko@imbg.org.ua

Мета. *Транскрипційний фактор ISGF-3 активується в результаті передачі сигналу від ІФН домінуючим шляхом Jak-STAT. Мета роботи полягала у пошуку сайтів зв'язування ISGF-3, що дозволяє виявляти гени первинної відповіді на ІФН . **Методи**. COTRASIF – це веб-інструмент для всегеномного пошуку еволюційно консервативних регуляторних ділянок у промоторах генів еукаріотів, що пропонує три методи – позиційно-вагових матриць (ПВМ), гібридний метод на основі прихованих моделей Маркова (ПВМ-ПММ) та філогенетичний футпринтинг. **Результати**. Показано, що метод ПВМ-ПММ має вищу специфічність пошуку. Метод застосовано для виявлення генів – мішеней транскрипційного фактора ISGF-3. Із застосуванням філогенетичного футпринтингу сформовано групу зі 162 генів імовірної первинної відповіді на ІФН . Розроблено та застосовано показник надійності виявлених генів-мішеней. **Висновки**. На основі результатів пошуку сайту зв'язування ISGF-3, статистичного аналізу з використанням Онтології Генів та обчислених показників надійності генів-мішеней 24 білок-кодуючих гени щура визначено як перспективні для вивчення первинної відповіді на ІФН .*

Ключові слова: *інтерферон, ISGF-3, COTRASIF, ІФН .*

Вступ. Розроблений нами раніше веб-інструмент COTRASIF [1] дозволяє проводити пошук генів із заданим користувачем сайтом зв'язування транскрипційного фактора (СЗТФ) у масштабах всього еукаріотного геному (на момент публікації для пошуку доступні 20 геномів), а також дозволяє відібрати з переліку результатів лише ті, що еволюційно консервативні [2]. COTRASIF дає змогу виявляти СЗТФ як за позиційно-частотною матрицею, так і за набором експериментально визначених послідовностей СЗТФ.

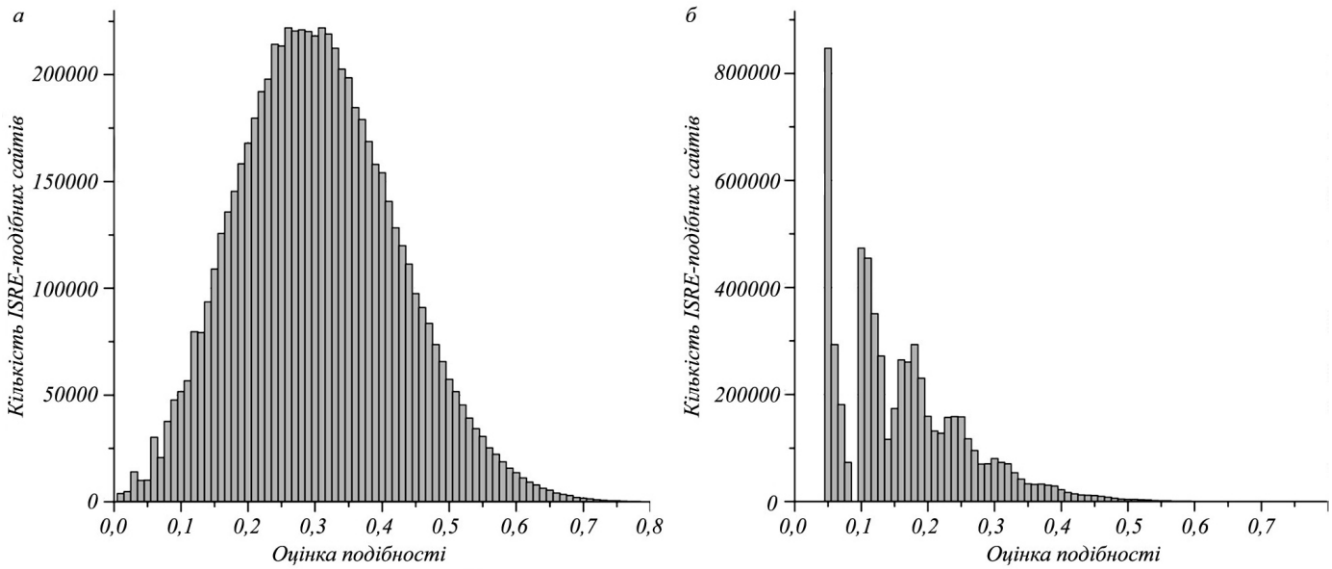
Для інтерферонів першого типу, до яких належить ІФН , достатньо дослідженим є домінуючий шлях передачі сигналу – Jak-STAT-каскад, кінцевим результатом якого є активація транскрипційного фактора ISGF-3, що специфічно зв'язується з сайтом ISRE (interferon-stimulated response element) [3]. Таким чином, пошук ISRE у промоторах білок-кодуючих генів дозволить знайти ймовірні гени первинної відповіді на дію ІФН .

ІФН – цитокін, відомий своїм антивірусним, антипроліферативним та імуномодулюючим ефектом. Його використовують як основний або допоміжний лікувальний засіб у терапії вірусних інфекцій та деяких видів раку. Проте щодо механізмів дії ІФН за відсутності вірусної інфекції відомо мало.

Мета цієї роботи полягала у визначенні білок-кодуючих генів-мішеней фактора ISGF-3 незалежно від їхньої тканиноспецифічної експресії за допомогою інструмента COTRASIF.

Матеріали і методи. Пошук ISRE здійснено з використанням методів позиційно-вагових мат-

Institute of Molecular Biology and Genetics NAS of Ukraine, 2009



Розподіл кількості ISRE-подібних сайтів залежно від значення оцінки подібності для методів ПВМ (*а*) і ПВМ-ПММ (*б*)

риць (ПВМ) та прихованих моделей Маркова (ПВМ-ПММ). Специфічність кожного з методів оцінювали, виявляючи ISRE-подібні сайти у 2-му та 3-му екзонах усіх білок-кодуючих генів щура [4]. Саме ці екзони обрано через низьку ймовірність наявності в них функціональних сайтів ISRE; екзони отримано з бази даних Ensembl випуску 52.

Для пошуку використано всегеномну версію програми COTRASIF з порогом подібності 0,01 (значення порогу 0,00 на момент написання роботи не дозволене). Основою методу ПВМ слугувала матриця M00258 із бази даних Transfac 7.0 Public, методу ПВМ-ПММ – вісім експериментально визначених послідовностей ISRE, взятих із літератури, та матриця M00258. Послідовності та матриця доступні онлайн http://biomed.org.ua/COTRASIF/supplement.html.

У результаті пошуку знайдено 6709322 ISRE-подібних сайтів методом ПВМ та 6163339 ISRE-подібних сайтів методом ПВМ-ПММ. Усю сукупність отриманих оцінок подібності розподілено на 100 груп із кроком 0,01 (з оцінкою подібності від 0,01 до 1,00 включно). Для кожної групи оцінок подібності обчислено кількість виявлених ISRE-подібних сайтів для методів ПВМ і ПВМ-ПММ (рисунок *а*, *б*).

Якщо для методу ПВМ характерним є близький до нормального розподіл із значенням середнього арифметичного 0,3, то для методу ПВМ-ПММ спостерігаємо розподіл Пуассона із максимумом виявлених ISRE-подібних ділянок у зоні з низькими оцінками подібності (менше 0,3), розподілу становить 0,16. З цього випливає, що метод ПВМ-ПММ, як і очікувалося теоретично, має вищу специфічність пошуку.

Для методів ПВМ і ПВМ-ПММ визначено пороги оцінок подібності 0,98 і 0,97 відповідно, які є мінімізаторами невірно-негативних результатів [4] пошуку ISRE, оскільки у 2-му та 3-му екзонах білок-кодуючих генів щура не знайдено жодного ISRE-подібного сайта з оцінкою подібності, вищою за зазначені пороги.

Власне пошук сайтів ISRE здійснено методом ПВМ-ПММ у промоторах генів *Rattus norvegicus* і *Mus musculus*, отриманих з 52-го випуску бази даних Ensembl. Пошук і філогенетичний футпринтинг виконано за допомогою COTRASIF [1].

Достовірне збагачення категорій Онтології Генів (Gene Ontology, GO) проаналізовано з використанням FatiGO [5]. Літературний пошук проведено за допомогою UniGene [6] та ihop-net [7].

Результати і обговорення. Ми виявили 743 сайти ISRE у 707 ймовірних білок-кодуючих генах первинної відповіді щура на ІФН та 1292 ISRE у 1163 генах миші. За допомогою філогенетичного футпринтингу отримано групу із 162 генів щура.

Таблиця 1
Ймовірні новітні гени – мішені інтерферонів першого типу з «неканонічною» функцією

№	Ідентифікатор гена	Назва гена	РН
1	ENSRNOG00000011507	PRKCA, Protein kinase C-alpha binding protein	0,5
2	ENSRNOG00000005723	N-methyl-D-aspartate (NMDA) receptor subunit 3A precursor	0,84
3	ENSRNOG00000002349	Gabra2, gamma-aminobutyric acid receptor subunit б2	0,84

Повні списки генів доступні онлайн за адресою http://biomed.org.ua/COTRASIF/supplement.html.

З використанням FatiGO [5] проведено аналіз збагачення категорій GO для 162 генів щура. Єдиною збагаченою категорією виявився біологічний процес *імунна відповідь* (GO:0006955) із коригованим $p = 8{,}52\ \ 10^{-3}$. До цієї категорії увійшли дев'ять генів, з них п'ять – відомі мішені ІФН ; ще у двох нових генів щура (GBP4_predicted, LOC687510) знайдено гени-ортологи миші, для яких показано регуляцію ІФН . Для двох генів (PF-4, Mbl1) у літературі відсутні підтвердження регуляції їхньої експресії ІФН .

Одержання нами єдиної достовірно збагаченої категорії GO *імунна відповідь* відповідає очікуванням, оскільки найкраще охарактеризовано саме імуностимулювальні ефекти ІФН . Для визначення серед 162 генів таких, для яких існують експериментальні докази їхньої регуляції ІФН , проведено літературний пошук за кожним із генів. У результаті пошуку встановлено, що для 61 гена даними літератури доведено їхню регуляцію ІФН , а для 101 гена подібних свідчень не знайдено.

Функціональним аналізом з використанням Онтології Генів виявлено три збагачені категорії у групі з 61 гена (*імунна відповідь, відповідь на віруси*, *ГТФазна активність*) і одну категорію *компонент синапсу* в групі з 101 гена (табл. 1).

Для зазначених генів нам не вдалося знайти в літературі експериментальних підтверджень їхньої регуляції інтерфероном.

Ми пропонуємо просту систему ранжування генів-мішеней первинної відповіді на транскрипційний фактор від найнадійніших до найменш надійних на основі таких факторів:

1. Кількість знайдених однотипних СЗТФ в одному промоторі [8]. Якщо ймовірність біологічної функції одного СЗТФ становить p_b, то для N СЗТФ імовірність P_b визначається як

$P_b = 1 - (1 - p_b)^N$.

Результат знаходиться в межах $[p_b; 1]$.

2. Збереження відстані від еволюційно консервативних СЗТФ до точок початку транскрипції відповідних генів-ортологів [9]. Ми використали п'ять повних обертів ДНК як максимальне відхилення відстані між СЗТФ і точкою початку транскрипції [9]. У межах одного оберту (10 нуклеотидів) збереження позиції СЗТФ вважаємо ідеальним [10]. Кожен сайт одержує оцінку еволюційної консервативності ECscore в межах [0; 1] для варіації відстані L у межах [10; 50] нуклеотидів за формулою:

$Ecscore = 1{,}25 - 0{,}025L$.

3. Оцінка подібності, вища за мінімізатор невірно-позитивних результатів [4], додає 0,25 до рейтингу надійності гена (фактор високої оцінки подібності враховується з ваговим коефіцієнтом 0,25).

У результаті оцінювання кожний ген набирає рейтинг у теоретичних межах [0,5; 2,25], який і визначає його надійність: чим вищий рейтинг – тим вища надійність. Тут і далі за текстом ми вживаємо термін «надійність гена» винятково для позначення обчисленої на основі наведених вище критеріїв різниці між характеристиками та кількістю сайтів зв'язування у кожному з промоторів досліджуваної групи генів.

Фактичний максимум обчислених рейтингів 162 генів, знайдених COTRASIF (див. табл. http://biomed.org.ua/COTRASIF/content/162.html), становить 1,5. У групі зі 101 невідомого гена 17 генів мають найвищий рейтинг надійності та є кандидатами на першочергове дослідження.

Нами проведено мікромасив-експеримент зі стимуляції культури первинних гепатоцитів щура

Таблиця 2
Гени – нові мішені ІФН

№	Ідентифікатор гена	Назва гена	РН
1	ENSRNOG00000007899	Transcription termination factor mTERF	1,5
2	ENSRNOG00000008593	Receptor-like tyrosine kinase	1,5
3	ENSRNOG00000009491	Polyhomeotic-like 3	1,5
4	ENSRNOG00000011883	–	1,5
5	ENSRNOG00000014037	DCN1	1,5
6	ENSRNOG00000014326	Calcium channel gamma-6 subunit	1,5
7	ENSRNOG00000016583	Bcl-2-like protein	1,5
8	ENSRNOG00000017053	F-box only protein 36	1,5
9	ENSRNOG00000018012	Similar to tubby super-family protein	1,5
10	ENSRNOG00000018129	NADH dehydrogenase (ubiquinone) 1	1,5
11	ENSRNOG00000018511	39S ribosomal protein L30mt	1,5
12	ENSRNOG00000018841	SRY-box containing gene 8	1,5
13	ENSRNOG00000019936	–	1,5
14	ENSRNOG00000020178	Coatomer protein complex, subunit epsilon	1,5
15	ENSRNOG00000030374	–	1,5
16	ENSRNOG00000030464	Olfactory receptor Olr711	1,5
17	ENSRNOG00000032163	Olfactory receptor Olr575	1,5
18	ENSRNOG00000004744	–	0,5
19	ENSRNOG00000017613	Polymerase (DNA directed) sigma	0,5
20	ENSRNOG00000028768	–	0,5
21	ENSRNOG00000029191	–	1,1
22	ENSRNOG00000011507	PRKCA	0,5
23	ENSRNOG00000005723	NMDA receptor subunit 3A precursor	0,84
24	ENSRNOG00000002349	Gabra2	0,84

ІФН протягом 3 і 6 год (неопубліковані дані). Контрольні клітини культивували протягом такого ж часу, але без ІФН . При достовірності 0,005 43 і 148 генів (після 3 і 6 год культивування відповідно) змінили рівень експресії. Сім із 43 генів (один з них також присутній у переліку 148 генів) є в знайденій за допомогою COTRASIF групі 162 генів, і поряд з відомими генами відповіді на ІФН включають два невідомих – ген ENSRNOG00000004744 з невизначеною функцією та ген ENSRNOG00000017613, який кодує ДНК-полімеразу сигма (Pols).

У табл. 2 представлено 24 гени, які є перспективними новими мішенями ІФН

Висновки. Показано, що розроблений нами раніше гібридний метод ПВМ-ПММ пошуку регуляторних ділянок у промоторах має вищу специфічність порівняно з методом ПВМ. Обидва методи реалізовано в інструменті COTRASIF.

За допомогою методу ПВМ-ПММ та філогенетичного футпринтингу виявлено 162 гени ймовірної первинної відповіді на дію ІФН , серед яких 61 ген має опубліковані експериментальні підтверд-

ження їхньої регуляції ІФН , а 101 ген – це новітні гени відповіді на дію ІФН .

Розроблено експрес-метод ранжування генів за надійністю для визначення оптимального порядку їхньої експериментальної перевірки.

На основі рейтингу надійності, виявлення збагаченої категорії GO *компонент синапсу*, та профілів експресії генів у гепатоцитах після обробки їх ІФН виявлено 24 гени, що є перспективними новими мішенями ІФН .

Роботу підтримано грантом УНТЦ #4381 «Нові технології у вивченні функціональної активності ІФН- ».

B. T. Tokovenko, O. O. Dragushchenko, A. V. Kuklin, M. Yu. Obolenskaya

Identification of gene targets of ISGF-3 transcription factor

Summary

Aim. *Transcription factor ISGF-3 is activated as a result of the signal transduction from IFN via the dominating Jak-STAT pathway. The discovery of ISGF-3 binding sites will assist in identifying genes of primary response to IFN .* ***Methods***. *COTRASIF is a web-based tool for the genome-wide identification of the evolutionary-conservative regulatory sites in the promoters of eukaryotic genes. It offers 3 search methods: based on position-weight matrices (PWM), hybrid method based on hidden Markov models (PWM-HMM), and phylogenetic footprinting.* ***Results***. *We have demonstrated that PWM-HMM method has higher search specificity, and used this method to identify the gene targets of ISGF-3 transcription factor. After applying phylogenetic footprinting, we have obtained a list of 162 genes of putative primary response to IFN . The reliability metrics to these gene targets has been developed and applied.* ***Conclusions***. *Based on the search results, Gene Ontology over-representation analysis, and reliability metrics, we have identified 24 rat protein-coding genes as promising targets for further studies on the primary response to IFN .*

Keywords: interferon, ISGF-3, COTRASIF, IFN .

Б. Т. Токовенко, Е. О. Драгущенко, А. В. Куклин, М. Ю. Оболенская

Определение генов – мишеней транскрипционного фактора ISGF-3

Цель. *Транскрипционный фактор ISGF-3 активируется в результате передачи сигнала от ИФН доминирующим Jak-STAT-путем. Цель работы состояла в поиске сайтов связывания ISGF-3, что позволяет определять гены первичного ответа на ИФН .* ***Методы***. *COTRASIF – это веб-инструмент для всегеномного поиска эволюционно консервативных регуляторных участков в промоторах генов эукариотов, который предлагает три метода – позиционно-весовых матриц (ПВМ), гибридный метод на основе скрытых моделей Маркова (ПВМ-СММ) и филогенетический футпринтинг.* ***Результаты***. *Мы показали, что метод ПВМ-СММ имеет более высокую специфичность поиска и использовали его для определения генов – мишеней транскрипционного фактора ISGF-3. С применением филогенетического футпринтинга к результатам поиска сформирована группа из 162 генов вероятного первичного ответа ИФН . Разработан и использован показатель надежности выявленных генов-мишеней.* ***Выводы***. *На основе результатов поиска сайта связывания ISGF-3, статистического анализа с использованием Онтологии Генов и вычисленных показателей надежности генов-мишеней 24 белок-кодирующих генов крысы определены как перспективные для изучения первичного ответа на ИФН .*

Ключевые слова: *интерферон, ISGF-3, COTRASIF, ИФН .*

ПЕРЕЛІК ЛІТЕРАТУРИ

1. *Tokovenko B., Golda R., Protas O., Obolenskaya M., El'skaya A*. COTRASIF: conservation-aided transcription-factor-binding site finder // Nucl. Acids Res.–2009.–**37**, N 7.–P. e49.
2. *Cartharius K., Frech K., Grote K., Klocke B., Haltmeier M., Klingenhoff A., Frisch M., Bayerlein M., Werner T*. MatInspector and beyond: promoter analysis based on transcription factor binding sites // Bioinformatics.–2005.–**21**, N 13.–P. 2933–2942.
3. *Horvath C. M*. The Jak-STAT pathway stimulated by interferon alpha or interferon beta // Sigtrans.–2004.–2004.–**260**.–P. tr10.
4. *Kel A. E., Gossling E., Reuter I., Cheremushkin E., Kel-Margoulis O. V., Wingender E*. MATCH: a tool for searching transcription factor binding sites in DNA sequences // Nucl. Acids Res.–2003.–**31**, N 13.–P. 3576–3579.
5. *Al-Shahrour F., Minguez P., Tarraga J., Montaner D., Alloza E., Vaquerizas J. M., Conde L., Blaschke C., Vera J., Dopazo J*. BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments // Nucl. Acids Res.–2006.–**34**, suppl. 2.–P. W472–W476.
6. *Wheeler D. L., Church D. M., Federhen S., Lash A. E., Madden T. L., Pontius J. U., Schuler G. D., Schriml L. M., Sequeira E., Tatusova T. A., Wagner L*. Database resources of the national center for biotechnology // Nucl. Acids Res.–2003.–**31**, N 1.–P. 28–33.
7. *Hoffmann R., Valencia A*. A gene network for navigating the literature // Nat. Genet.–2004.–**36**, N 7.–P. 664.
8. *Li X., Leung S., Burns C., Stark G. R*. Cooperative binding of Stat1-2 heterodimers and ISGF3 to tandem DNA elements // Biochimie.–1998.–**80**, N 8–9.–P. 703–710.
9. *Hannenhalli S., Levy S*. Predicting transcription factor synergism // Nucl. Acids Res.–2002.–**30**, N 19.–P. 4278–4284.
10. *Bluyssen H. A., Vlietstra R. J., van der Made A., Trapman J*. The interferon-stimulated gene 54 K promoter contains two adjacent functional interferon-stimulated response elements of different strength, which act synergistically for maximal interferon-alpha inducibility // Eur. J. Biochem.–1994.–**220**, N 2.–P. 395–402.

УДК 577.218
Надійшла до редакції 10.07.09