

# *In silico* approach to study and functionally analyze interferon regulated genes

B. T. Tokovenko, A. V. El'skaya, M. Yu. Obolenskaya

Institute of Molecular Biology and Genetics of the National Academy of Sciences of Ukraine

150 Zabolotnoho Str., Kyiv 03143

to.bogdan@gmail.com

---

*Finding genes which have biologically meaningful ISRE (interferon-stimulated response element) is important for better understanding of the Jak-STAT activated cellular IFN response. We used transcription factor binding site (TFBS) search with gene orthology filtering to find putative ISREs in the promoters of protein-coding genes of Rattus norvegicus, and used Gene Ontology (GO) analysis to check the validity of ISRE search results in terms of biological meaning. A total of 23286 promoters of rat genes were analyzed. ISRE search with 80 % threshold produced 5214 sites in 4571 promoters. 850 ISREs in 768 promoters passed orthology filtering. Distribution of ISREs along the promoter in 768-gene set reveals 3 regions of ISRE localization: 0 to -250, -250 to -550, and above -550 relative to TSS (transcription start site). It is not yet known whether ISRE localization has any functional implications. Using BayGO, a total of 84 GO terms were found to be enriched at  $P \leq 0.05$  in the 768-gene set. Among these categories some are directly related to known IFN actions (positive regulation of B cell differentiation, humoral immune response, response to virus, cell differentiation etc.). 768 gene set was compared to the 4571 gene set using GO Tree Machine. Such categories as cell differentiation, cell cycle, regulation of cell cycle, viral life cycle and some others were found to be enriched, and belong to the well-known domains of interferon actions. Their relative enrichment is an indirect indication that the applied orthology filtering does increase the quality of results. Gene orthology-based filtering of the initial TFBS search results was shown to produce viable and expected results. Genes identified in this research as containing ISRE in promoters will be used to seed the construction of the IFN- induced gene regulatory network.*

*Keywords: transcription factor binding site, interferon, ISRE, gene orthology, Gene Ontology.*

---

**Introduction.** Transcription factors play an important role in the regulation of gene expression. Prediction of putative transcription factor binding sites (TFBS) has become an important resource to explore genome organization and predict transcriptional regulation [1]. Computational TFBS prediction methods are also necessary for the efficient annotation of transcriptional regulatory networks [2].

The TFBS consensus sequence motifs are usually represented using either IUPAC (International Union of Pure and Applied Chemistry) nomenclature consensus string, or matrices, the two most common

being PFM (position frequency matrix, also known as position count matrix) and PWM (position weight matrix, or nucleotide weight matrix) [3–6]. PFM is a matrix consisting of nucleotide counts per each position of the identified binding site. PFMs were first used to characterize DNA-binding site specificity in 1982–1986 [7, 8]. Later, quantitative discrimination of sites with calculated site scores using position weight matrices was introduced [5, 9–11]. A weight matrix pattern definition is superior to a simple IUPAC consensus sequence, as it represents the complete nucleotide occurrence probabilities for each position. It also allows the quantification of the similarity between

the weight matrix and a potential TFBS detected in the target sequence. PWM is an estimate of the binding energy of the transcription factor to its specific binding site [5, 6].

High throughput analyses using SELEX (Systematic Evolution of Ligands by Exponential Enrichment) and CHIP-Chip (Chromatin Immunoprecipitation-Microarray), along with computational sampling methods, have generated thousands of PFMs. These data together with the related transcription factor information are curated in online databases – Transfac [12], JASPAR [3], and others.

Computational TFBS prediction provides reliable results in application to prokaryotes and yeast [13–17]. However, in higher eukaryotes accurate and reliable TFBS prediction is an outstanding challenge [18].

Online applications, such as MatInspector [19], MATCH [20] and ConSite [21], have been built to predict TFBS embedded in promoter sequences. However, TFBS search only identifies sites where the transcription factor could bind, but not necessarily will bind.

Taking into account the length of matrices used for TFBS prediction (usually less than 15 nucleotides), no wonder that simple search for IUPAC-denoted binding sites yields numerous false-positive results, occurring by chance. When applying PWM-based methods, matrix-site similarity score threshold can be used to increase specificity (get less false-positives) at the cost of sensitivity (find less true-positives).

Thus, using only TFBS search is not sufficient, and additional processing is required to refine the results. To avoid the loss of sensitivity, and reduce the number of false-positive binding site predictions, additional analysis can be applied: looking for paired TFBS, TFBS motifs, using gene orthology information, microarray-derived gene co-expression data, applying learning algorithms trained on known transcription factor target genes, *etc.*

Interferons are inducible cytokines with strong antiviral properties and a wide range of gene targets in the cell [22–25]. Type I interferons share common cellular surface receptor, which consists of two subunits (IFNAR-1 and IFNAR-2). Specific interferon binding to the receptor induces hetero-dimerization of

IFNAR-1 and IFNAR-2, with the following activation of receptor-bound kinases Jak1 and Tyk2 by phosphorylation, and the Jak-STAT signal transduction pathway initiation [26, 27]. The ISRE, a conserved regulatory element of interferon-stimulated genes (ISGs), is the target for transcriptional activation by the IFN-stimulated gene factor-3 (ISGF3), which consists of two STAT molecules (STAT1 and STAT2), and IRF-9 (interferon-regulatory factor 9, or p48) [23, 28]. Genes containing ISRE can be considered primary interferon-response genes, with the cautionary note on the ability of not only ISGF3, but also of IRFs to bind to ISRE. Finding genes which have biologically meaningful ISRE sites will provide better understanding the Jak-STAT branch of cellular primary interferon response.

In this study, we searched for putative interferon-stimulated response elements (ISREs) in the promoters of all protein-coding genes of *Rattus norvegicus*, as defined and annotated in the Ensembl database [29], and looked at the validity of using TFBS search in terms of biological meaning.

**Methods.** A total of 23286 protein-coding rat genes were annotated and available in the Ensembl rat genome database release 40. For this set of genes, promoter and exon sequences were extracted and stored in a local database for faster access by sequence processing algorithms. A total of 70844 promoters and 544585 exons (with type = «coding») were obtained for 3 organisms (Table 1). For the purposes of this study, gene promoter was defined as a sequence which starts at the –1000 position relative to the TSS (transcription start site) and ends at position 0. Promoter and exon extraction was carried out using an automated pipeline, built using MySQL 4.1.16 database for data storage and PHP 5.1.6 programming language for querying the Ensembl database. Exons shorter than ISRE matrix (15 nucleotides), and both exons and promoters without the contiguous 15-nucleotide long subsequence, were discarded and not counted. Contiguous 15-nucleotide long subsequence was allowed to have no more than a single unknown (*N*) nucleotide.

ISRE PFM matrix was taken from the database of transcription factors TRANSFAC 7.0 Public [30], accession number M00258. ISRE PFM, compiled from 13 sequences, is shown in Table 2.

Table 1  
Ensembl genomes database, release 40: promoters and coding exons available

| Organism                 | Promoters | Coding exons |
|--------------------------|-----------|--------------|
| <i>Mus musculus</i>      | 24260     | 169400       |
| <i>Rattus norvegicus</i> | 23286     | 196498       |
| <i>Homo sapiens</i>      | 23331     | 178687       |
| Total                    | 70877     | 544585       |

To compute matrix-site similarity scores, PFM was converted into PWM. To obtain ISRE PWM, for each PFM matrix element we applied the following logarithm transformation [31]:

$$w(b,i) = \log_2 \frac{f_{b,i} \sqrt{N/4}}{N \sqrt{N} p(b)}$$

In this formula,  $w(b, i)$  is a calculated PWM matrix element value for nucleotide  $b$  in column  $i$ ;  $f_{b,i}$  is a raw count of nucleotide  $b$  in column  $i$  of the PFM;  $N$  is a number of sequences used to create the PFM ( $N = 13$  for M00258),  $\sqrt{N}$ ,  $\sqrt{N/4}$  are pseudocounts, used in the case of the small  $N$ ;  $p(b)$  is the background frequency of nucleotide  $b$ .

Due to the relatively small number of sequences which are used to compile PFM, the resulting PWM might be inaccurate. For example, in Table 2 nucleotide T has 13 out of 13 possible counts in column 4; when converting this PFM to PWM without the pseudocounts, all other nucleotides at position 4 would have got – as their value, and then any sequence, which does not have «T» in that position, would automatically fail, even if all the other nucleotides gave

the highest similarity score possible. Introduction of pseudocounts into the formula saves from obtaining restrictive base-score values, and thus allows mismatches even in strongly conserved positions. However, mismatch in a conserved position still gets the lowest score.

Most commonly  $p(b)$  is taken equal to 0.25. Another approach is to take  $p(b)$  based on AT/GC content for the whole studied genome. In this study, we compared using  $p(b) = 0.25$  and calculating  $p(b)$  from the background individually for each 1000-nucleotide long promoter before searching for ISRE with generated PWM (Fig. 1). For each target sequence (exon or promoter), full-length comparison of both strands with ISRE matrix was done, and the highest similarity score saved. Then the count of the best matrix-site matches was plotted against the 1–100 % range of similarity scores.

It can be seen that calculating target sequence-specific background nucleotide frequencies leads to slightly higher matrix-site similarity scores, especially when comparing matrix to promoter sites. However, performing F-test shows no difference between using  $p(b) = 0.25$  and sequence-specific  $p(b)$ : for exons F-value = 0.04, P = 0.997, for promoters F-value = 0.05, P = 0.995. The reason for promoters to exhibit more noticeable shift towards higher similarity scores can be explained by the differences in promoter and exons lengths: all the promoters were uniformly 1000 nucleotides long (with some exceptions containing long «not know» subsequences), while exons were much shorter, representing the range from 15 (for this study, ISRE-limited) to 19456 nucleotides long, with mean 165.3, and median 123. The two choices of  $p(b)$  were not further compared, but the sequence-specific

Table 2  
Position frequency matrix of 15 bp-long interferon stimulated response element

| Nucleotide | Frequency of nucleotides in each of 15 positions |    |    |    |    |    |    |   |   |    |    |    |    |    |    |
|------------|--|----|----|----|----|----|----|---|---|----|----|----|----|----|----|
|            | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| A          | 1  | 12 | 0  | 0  | 0  | 0  | 0  | 7 | 1 | 1  | 0  | 0  | 0  | 2  | 1  |
| C          | 8  | 0  | 0  | 0  | 0  | 0  | 13 | 1 | 7 | 0  | 0  | 3  | 8  | 7  | 8  |
| G          | 2  | 1  | 13 | 0  | 0  | 0  | 0  | 1 | 2 | 0  | 0  | 0  | 0  | 2  | 3  |
| T          | 2  | 0  | 0  | 13 | 13 | 13 | 0  | 4 | 3 | 12 | 13 | 10 | 5  | 2  | 1  |

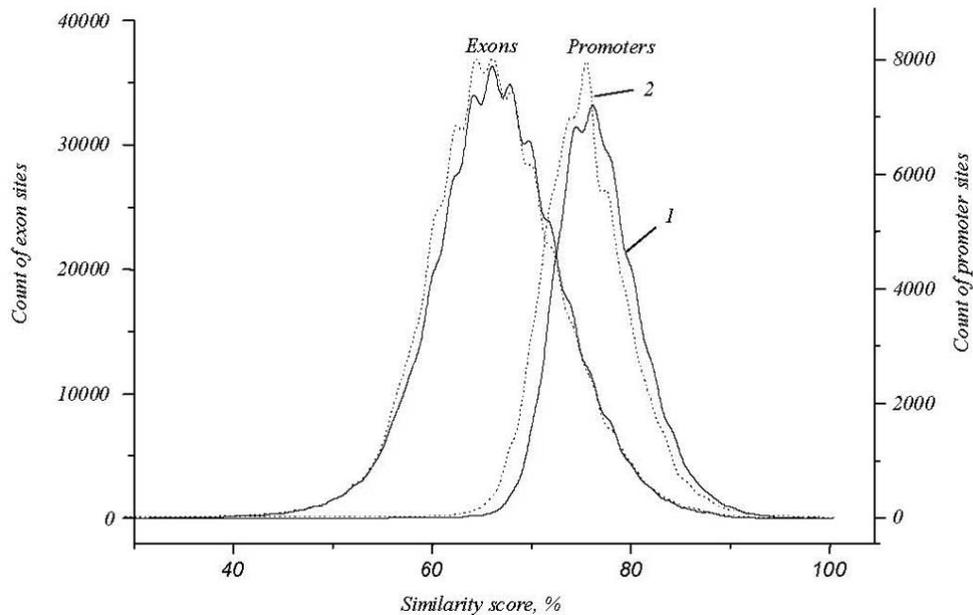


Fig. 1. The number of found binding sites dependence on the calculated similarity scores: 1 –  $p(b)$  sequence-specific; 2 –  $p(b) = 0.25$

$p(b)$  was used, based on the evidence that this would not deteriorate the results quality.

In order to define the threshold to make a presence/absence call for each matrix-site similarity score, we obtained means and standard deviations of the maximal similarity score distributions (Fig. 1). For exons, mean = 66.7%, SD = 6.8; for promoters, mean = 76.7%, SD = 4.3. Assuming that all matches of ISRE matrix in exons were false-positives, and assuming close to normal distribution of the similarity scores, threshold was chosen at the similarity score level of 80%, which includes no more than 2.5% of high-scoring ISRE matches in exons, and includes 16% of potential true positives. Such a stringent threshold was chosen because only one biologically-significant TFBS selection filter (namely, comparison of the promoters of the orthologous rat and mouse genes) was applied. Used in literature threshold level of 75% might be more appropriate in case of more filters applied, as it initially includes over 50% of potential true-positives. Running TFBS search on both DNA strands with 80% threshold in the promoters of all the protein-coding rat genes produced 5214 binding sites in 4571 promoters.

In order to filter away biologically insignificant binding sites, we compared promoters of the orthologous rat and mouse genes. The assumption behind this procedure was that if the binding site has no

biological meaning, then it is more likely not to be preserved in the process of evolution, and vice versa for meaningful binding sites. The procedure of identifying potential ISRE binding sites in the promoters of 24260 mouse protein-coding genes was identical to that used for rat genes, with the same threshold of 80%. The list of orthologous genes was obtained from Ensembl. This database provides similarity percentage for each pair of genes, and orthology type flag, which can be one of: «ortholog\_one2one» (reciprocal best hit genes), «ortholog\_one2many», «ortholog\_many2many» and «apparent\_ortholog\_one2one» (when no similarities were found, or annotation system error occurred). The following criteria were used to choose genes considered orthologous for the purposes of this study:

- 1) for «ortholog\_one2one» entries, identity of over 70% was required;
- 2) for «ortholog\_one2many», 75%;
- 3) for «ortholog\_many2many» and «apparent\_ortholog\_one2one», 80% of identity were required.

Two orthology-based approaches were tried: looking for occurrence of ISRE both in rat and mouse orthologous genes with no other constraints, and looking for ISRE occurrence in orthologs with an additional constraint for the start positions of found sites to be no more than 25 bp apart (distances measured relative to the TSS in the genes of each organism).

**Table 3**  
*GO Slims representation of the enriched GO categories from the «biological\_process» ontology in 768 gene set (single-count categories not included)*

| GO id             | Definition   | GO term counts |
|-------------------|--|----------------|
| <b>GO:0008150</b> | <b>Biological_process</b>                                      | <b>84</b>      |
| GO:0009987        | Cellular process   | 56             |
| GO:0007582        | Physiological process  | 54             |
| GO:0050875        | Cellular physiological process                                 | 41             |
| GO:0007275        | Development  | 22             |
| GO:0050789        | Regulation of biological process                               | 20             |
| GO:0008152        | Metabolism   | 19             |
| GO:0050896        | Response to stimulus   | 10             |
| GO:0006810        | Transport  | 9              |
| GO:0006139        | Nucleobase, nucleoside, nucleotide and nucleic acid metabolism | 9              |
| GO:0043170        | Macromolecule metabolism                                       | 9              |
| GO:0030154        | Cell differentiation   | 8              |
| GO:0007154        | Cell communication   | 7              |
| GO:0006928        | Cell motility  | 6              |
| GO:0008219        | Cell death   | 5              |
| GO:0046903        | Secretion  | 4              |
| GO:0009056        | Catabolism   | 2              |

Simple co-occurrence in orthologous genes decreased the numbers to 1722 TFBS in 1419 promoters, and position constraint brought that number down to 850 TFBS in 768 promoters. This means that out of 4571 genes 768 (16.8 %) have between-species position-conserved ISRE sites in promoters. The set of 768 genes was used for further analysis.

Gene Ontology (GO) categories enrichment analysis was performed using BayGO [32], with 100 iterations and at least a single gene within a category to consider it during analysis. Categories with  $P > 0.05$  were considered insignificant.

**Results and Discussion.** In the final 768-gene set, 697 genes had single putative ISRE, 62 genes had two putative ISREs, and 9 genes had three or more ISREs in their promoters.

Graph of the distribution of found ISRE start sites along the length of the promoter (Fig. 2) reveals three characteristic regions of ISRE localization: 0 to -250, -250 to -550, and above -550 nucleotides relative to the TSS. It is not yet known whether ISRE localization has any functional implications.

GO categories enrichment analysis was conducted for the 768 gene set against all the rat protein-coding genes (see Table 3 for GO Slims representation). GO has three main categories: biological process, molecular function, and cellular component. We analyzed GO terms enrichment within the biological process category.

A total of 84 GO terms in the biological\_process category were found to be enriched at  $P \leq 0.05$ . GO Slims representation includes GO terms which are hierarchically linked, *e. g.* cellular process, physiological process, and cellular physiological process (Fig. 3). Thus in Table 3 GO Slims terms are overlapping, and do not add up to the total of 84 enriched categories.

To identify the effects of orthology-based filtering, 768 gene set was compared to the 4571 gene set using GO Tree Machine (Fig. 3). Categories in bold and underlined in Fig. 3 are enriched in 768 set as compared to 4571 set with  $P < 0.01$  (no multiple-testing correction was applied). Such categories as cell differentiation, cell cycle, regulation of cell cycle, viral life cycle and some others belong to the well-known domains of interferon actions. Their relative enrichment is an indirect indication that the applied orthology filtering does increase the quality of results.

Among the 84 enriched categories, a number of categories are directly related to known IFN actions (*e. g.* positive regulation of B cell differentiation, humoral immune response, response to virus, cell differentiation, positive regulation of transcription factor activity *etc.*). Some of the GO terms are found both in the set of 84 categories (in 768 versus all genes enrichment test), and in the 20 categories, shown in bold and underlined in Fig. 3 (in 768 versus 4571 test): development, cell differentiation, and nervous system development.

Some of the categories are closely related, namely as cell adhesion and regulation of cell adhesion, viral life cycle and response to virus, myeloid cell

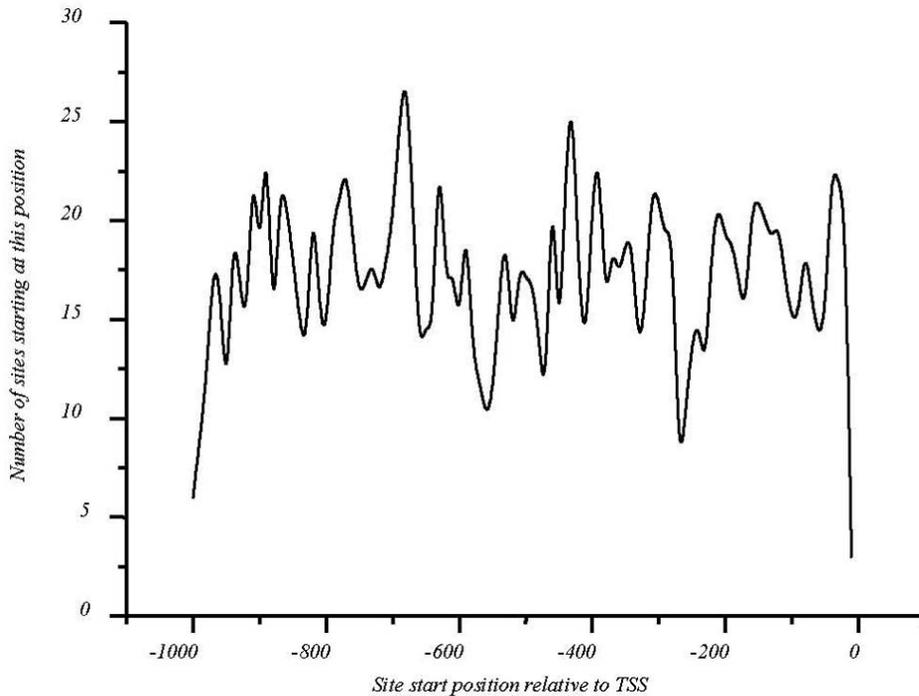


Fig. 2. Distribution of found ISRE start sites along the length of the promoter

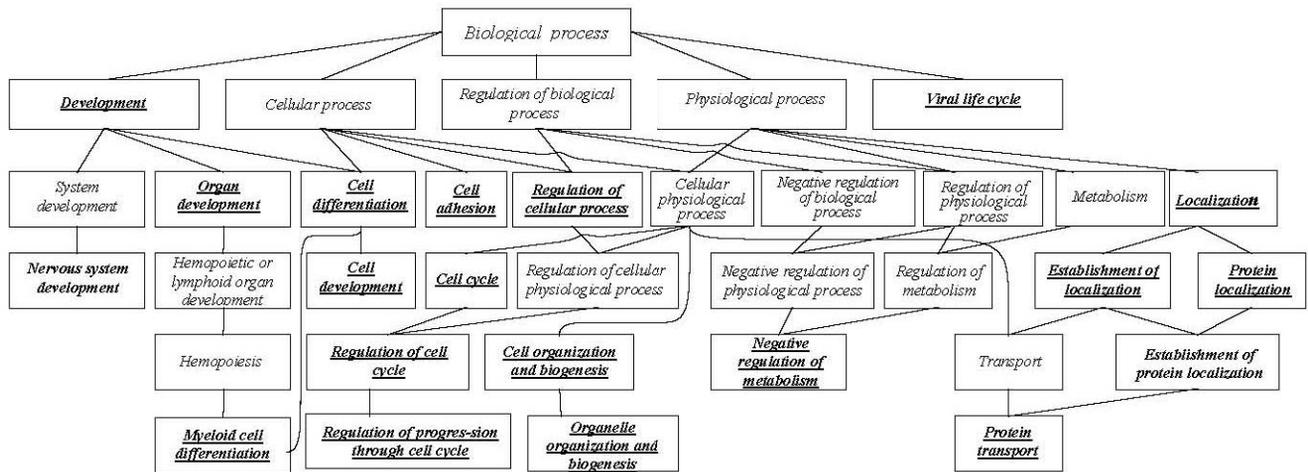


Fig. 3. Gene Ontology hypergeometric test for enrichment of the biological\_process ontology tree for 768 genes set versus 4571 gene set

differentiation and positive regulation of myeloid cell differentiation (in 768 versus 4571 and 768 versus all gene sets, respectively).

It was interesting to observe the «nervous system development» category to be found enriched both in 768 gene set versus all genes (with  $P < 0.01$  and  $\gamma = 0.4$ , using BayGO), and in 768 versus 4571 ( $P = 0.0099$ , using GOTM). This observation needs further investigation, especially in the light of known interferon side effects, which include headache, increased irritability and some other.

**Conclusions.** We identified 768 rat genes which contain ISRE in their promoters, and are the potential targets of transcriptional regulation by type I interferons. Functional analysis of these genes, conducted using Gene Ontology, had shown the enrichment of the GO categories related to already known IFN effects.

Additional step, based on the comparison of promoters of the orthologous mouse and rat genes, was applied after the TFBS search. GO analysis of the selected genes revealed 20 categories, with more than 8

of them related to known IFN effects (cell differentiation, regulation of progression through cell cycle, viral life cycle and others).

As the GO analysis shows, a simple yet effective additional step in TFBS search data post-processing would be to concentrate on the genes, known to be expressed in the tissue/cells of interest. This would allow focusing on the tissue-specific factor effects, and seeing a clearer picture of GO terms enrichment.

Another improvement would be to refine the TFBS matrix before use. In the case of interferon-regulated genes, one could use public microarray data from the interferon-stimulation experiments to build the list of early interferon-response genes, and then build a new matrix, taking into account actual base frequencies in the promoters of the up/down-regulated genes. With this approach it is also possible to take into account paired and dependent changes of the matrix nucleotides, using hidden Markov models or learning algorithms, such as support vector machine.

Genes identified in this research as ISRE-containing will be used to seed the construction of the IFN- $\alpha$ -induced gene regulatory network.

*Б. Т. Токовенко, Г. В. Єльська, М. Ю. Оболенська*

Підхід до вивчення та функціонального аналізу регульованих інтерфероном генів методами біоінформатики

Резюме

Пошук генів, які містять у промоторі біологічно значущий сайт ISRE (interferon-stimulated response element), необхідний для вдосконалення сучасних уявлень про опосередковану Jak-STAT стимуляцію клітин інтерфероном. Для виявлення імовірних ISRE у промоторах кодуєвих блоків генів *Rattus norvegicus* використано метод пошуку сайтів зв'язування із додатковим відбором результатів за наявністю сайту в промоторах мишиачих генів-ортологів. Придатність методу пошуку ISRE з точки зору біологічної значущості перевіряли за допомогою функціонального аналізу виявлених генів з використанням GO (Gene Ontology). Проаналізовано 23286 промоторів генів щура. Пошук ISRE з порогом подібності 80% виявив 5214 сайтів у 4571 промоторі. Після відбору за ортологією отримано 850 елементів ISRE у 768 промоторах. У розташуванні знайдених ISRE можна виділити три основні ділянки: від 0 до -250, від -250 до -550 та вище за -550-ту позицію відносно точки початку транскрипції. Поки незрозуміло, чи пов'язані між собою розташування ISRE та його функції. Використовуючи BayGO, у групі із 768 генів виявлено 84 відносно збагачені категорії GO з  $P \leq 0,05$ . Деякі з цих категорій належать до відомих ефектів інтерферону (позитивна регуляція диференціації В-клітин, гуморальна імунна відповідь, відповідь на

вірус, клітинна диференціація тощо). За допомогою GO Tree Machine порівняно функціональні категорії в групах із 768 та 4571 гена. Такі функціональні категорії, як клітинна диференціація, клітинний цикл, регуляція клітинного циклу, вірусний життєвий цикл та деякі інші, належать до відомих мішеней інтерферону. Їхнє відносне збагачення після відбору за ортологією є непрямим доказом того, що застосування зазначеного підходу дає змогу підвищити якість результатів. У цілому пошук сайту зв'язування із наступним відбором за ортологією дозволив отримати значущі та очікувані результати. Гени, у промоторах яких знайдено ділянку ISRE, стануть підґрунтям для створення мережі генної регуляції, стимульованої інтерфероном.

*Ключові слова:* сайт зв'язування транскрипційного фактора, інтерферон, ISRE, ортологія генів, Gene Ontology.

*Б. Т. Токовенко, А. В. Ельская, М. Ю. Оболенская*

Подход к изучению и функциональному анализу регулируемых интерфероном генов методами биоинформатики

Резюме

Поиск генов, содержащих в промоторе биологически значимый сайт ISRE (interferon-stimulated response element), является важной частью дальнейшего изучения опосредованной Jak-STAT стимуляции клеток интерфероном. Для определения вероятных ISRE в промоторах кодирующих белок генов *Rattus norvegicus* мы использовали метод поиска сайтов связывания с дополнительным отбором результатов по признаку наличия сайта в промоторе мышиных генов-ортологов. Пригодность метода поиска ISRE с точки зрения биологического значения проверяли с помощью функционального анализа найденных генов с использованием онтологии генов (GO, Gene Ontology). При анализе 23286 промоторов генов крысы с порогом сходства с матрицей ISRE 80% выявлены 5214 сайтов в 4571 промоторе. Отбор по ортологии прошли 850 ISRE в 768 промоторах. В расположении обнаруженных элементов ISRE можно выделить три основных участка: от 0 до -250, от -250 до -550 и выше -550 относительно точки начала транскрипции. Пока не понятно, есть ли связь между расположением ISRE и его функцией. Используя BayGO, в группе из 768 генов вычислены 84 обогащенные категории GO при  $P \leq 0,05$ . Некоторые из этих категорий непосредственно связаны с известными эффектами интерферона (позитивная регуляция дифференциации В-клеток, гуморальный иммунный ответ, ответ на вирус, клеточная дифференциация и др.). При помощи GO Tree Machine мы сравнили функциональные категории в группах из 768 и 4571 гена. Такие функциональные категории, как клеточная дифференциация, клеточный цикл, регуляция клеточного цикла, жизненный цикл вирусов и некоторые другие, принадлежат к известным мишеням интерферона. Их относительное обогащение после отбора по признаку ортологии является косвенным доказательством того, что применение отбора по ортологии позволяет улучшить качество полученных результатов. В целом поиск сайта связывания с последующим отбором по признаку ортологии позволил получить значимые и ожидаемые результаты. Гены, в промоторах которых найден сайт ISRE, будут использованы для создания сети генной регуляции, стимулируемой интерфероном.

*Ключевые слова:* сайт связывания транскрипционного фактора, интерферон, ISRE, ортология генов, Gene Ontology.

## REFERENCES

1. *Su G., Mao B., Wang J.* A web server for transcription factor binding site prediction // *Bioinformatics.*–2006.–**1.**–P. 156–157.
2. *Zhao F., Xuan Z., Liu L., Zhang M. Q.* TRED: a transcriptional regulatory element database and a platform for *in silico* gene regulation studies // *Nucl. Acids Res.*–2005.–**33**, suppl. 1.–P. D103–D107.
3. *Sandelin A., Alkema W., Engstrom P., Wasserman W. W., Lenhard B.* JASPAR: an open-access database for eukaryotic transcription factor binding profiles // *Nucl. Acids Res.*–2004.–**32**, suppl. 1.–P. D91–D94.
4. *Lenhard B., Wasserman W. W.* TFBS: Computational framework for transcription factor binding site analysis // *Bioinformatics.*–2002.–**18.**–P. 1135–1136.
5. *Schones D. E., Sumazin P., Zhang M. Q.* Similarity of position frequency matrices for transcription factor binding sites // *Bioinformatics.*–2005.–**21.**–P. 307–313.
6. *Stormo G. D.* DNA binding sites: representation and discovery // *Bioinformatics.*–2000.–**16.**–P. 16–23.
7. *Staden R.* Computer methods to locate signals in nucleic acid sequences // *Nucl. Acids Res.*–1984.–**12.**–P. 505–519.
8. *Stormo G. D., Schneider T. D., Gold L., Ehrenfeucht A.* Use of the «Perceptron» algorithm to distinguish translational initiation sites in *E. coli* // *Nucl. Acids Res.*–1982.–**10.**–P. 2997–3011.
9. *Hertz G. Z., Stormo G. D.* Identifying DNA and protein patterns with statistically significant alignments of multiple sequences // *Bioinformatics.*–1999.–**15.**–P. 563–577.
10. *Hertz G. Z., Hartzell G. W., III, Stormo G. D.* Identification of consensus patterns in unaligned DNA sequences known to be functionally related // *Bioinformatics.*–1990.–**6.**–P. 81–92.
11. *Stormo G. D., Hartzell G. W.* Identifying protein-binding sites from unaligned DNA fragments // *Proc. Nat. Acad. Sci. USA.*–1989.–**86.**–P. 1183–1187.
12. *Matys V., Kel-Margoulis O. V., Fricke E., Liebich I., Land S., Barre-Dirrie A., Reuter I., Chekmenev D., Krull M., Hornischer K., Voss N., Stegmaier P., Lewicki-Potapov B., Saxel H., Kel A. E., Wingender E.* TRANSFAC(R) and its module TRANSCOMP(R): transcriptional gene regulation in eukaryotes // *Nucl. Acids Res.*–2006.–**34**, suppl. 1.–P. D108–D110.
13. *Liu X. S., Brutlag D. L., Liu J. S.* An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments // *Nat. Biotech.*–2002.–**20.**–P. 835–839.
14. *van Helden J., Andre B., Collado-Vides J.* Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies // *J. Mol. Biol.*–2006.–**281.**–P. 827–842.
15. *Pilpel Y., Sudarsanam P., Church G. M.* Identifying regulatory networks by combinatorial analysis of promoter elements // *Nat. Genet.*–2001.–**29.**–P. 153–159.
16. *Bussemaker H. J., Li H., Siggia E. D.* Regulatory element detection using correlation with expression // *Nat. Genet.*–2001.–**27.**–P. 167–174.
17. *Furman I., Pilpel Y.* Promoting human promoters // *Mol. Syst. Biol.*–2006.–**2.**–P. E1–E2.
18. *Zhang M. Q.* Computational prediction of eukaryotic protein-coding genes // *Nat. Revs Genet.*–2002.–**3.**–P. 698–709.
19. *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.**–P. 2933–2942.
20. *Kel A. E., Gossling E., Reuter I., Chermushkin E., Kel-Margoulis O. V., Wingender E.* MATCHTM: a tool for searching transcription factor binding sites in DNA sequences // *Nucl. Acids Res.*–2003.–**31.**–P. 3576–3579.
21. *Sandelin A., Wasserman W. W., Lenhard B.* ConSite: web-based prediction of regulatory elements using cross-species comparison // *Nucl. Acids Res.*–2004.–**32**, suppl. 1.–P. W249–W252.
22. *Pestka S., Langer J. A., Zoon K. C., Samuel C. E.* Interferons and their actions // *Ann. Rev. Biochem.*–1987.–**56.**–P. 727–777.
23. *Pestka S., Krause C. D., Walter M. R.* Interferons, interferon-like cytokines, and their receptors // *Immunol. Revs.*–2004.–**202.**–P. 8–32.
24. *Sen G. C.* Viruses and interferons // *Annu. Rev. Microbiol.*–2001.–**55.**–P. 255–281.
25. *Stark G. R., Kerr I. M., Williams B. R. G., Silverman R. H., Schreiber R. D.* How cells respond to interferons // *Ann. Rev. Biochem.*–1998.–**67.**–P. 227–264.
26. *Horvath C. M.* The Jak-STAT Pathway stimulated by interferon {alpha} or interferon {beta} // *Sci. STKE.*–2004.–**260.**–P. 10.
27. *Fu X., Schindler C., Improta T., Aebersold R., Darnell J. E., Jr.* The proteins of ISGF-3, the interferon {alpha}-induced transcriptional activator, define a gene family involved in signal transduction // *Proc. Nat. Acad. Sci. USA.*–1992.–**89.**–P. 7840–7843.
28. *Kessler D. S., Veals S. A., Fu X. Y., Levy D. E.* Interferon-alpha regulates nuclear translocation and DNA-binding affinity of ISGF3, a multimeric transcriptional activator // *Genes and Develop.*–1990.–**4.**–P. 1753–1765.
29. *Curwen V., Eyraes E., Andrews T. D., Clarke L., Mongin E., Searle S. M. J., Clamp M.* The Ensembl automatic gene annotation system // *Genome Res.*–2004.–**14.**–P. 942–950.
30. *Levy D. E., Kessler D. S., Pine R., Reich N., Darnell J. E., Jr.* Interferon-induced nuclear factors that bind a shared promoter element correlate with positive and negative transcriptional control // *Genes and Develop.*–1988.–**2.**–P. 383–393.
31. *Jin V. X., Leu Y. W., Liyanarachchi S., Sun H., Fan M., Nephew K. P., Huang T. H. M., Davuluri R. V.* Identifying estrogen receptor {alpha} target genes using integrated computational genomics and chromatin immunoprecipitation microarray // *Nucl. Acids Res.*–2004.–**32.**–P. 6627–6635.
32. *Vencio R., Koide T., Gomes S., de B. Pereira C.* BayGO: Bayesian analysis of ontology term enrichment in microarray data // *BMC Bioinformatics.*–2006.–**7.**–P. 86.
33. *Zhang B., Schmoyer D., Kirov S., Snoddy J.* GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies // *BMC Bioinformatics.*–2004.–**5.**–P. 16.