

Поиск и идентификация дифференциально экспрессирующихся генов ответа на тепловой стресс *Drosophila melanogaster* с низкой степенью изменения экспрессии

Дерюженко М.А.*, Груntenко Н.Е.

Институт цитологии и генетики СО РАН, Новосибирск, Россия

* maksimd@bionet.nsc.ru

Ключевые слова: транскриптом; *Drosophila melanogaster*; реконструкция генной сети; потеря информации

Мотивация и цель: Потеря информации – это одна из ключевых проблем современного анализа транскриптомных данных. Основной причиной ее возникновения является влияние множества неучтенных факторов, возникающих при проведении и анализе результатов эксперимента, которые снижают общую достоверность полученных результатов. В попытке избавиться от них, как правило, отсекаются дифференциально экспрессирующиеся гены (ДЭГ) с высоким значением коэффициента ложного обнаружения (FDR). При этом такие гены могут играть важную роль в контроле наблюдаемых изменений функций организма и получить высокие значения FDR ввиду низкой степени изменения экспрессии.

Эту проблему можно решить несколькими способами:

- 1) провести функциональную аннотацию каждого ДЭГ и изучить литературные данные по ним;
- 2) провести метаанализ транскриптомных данных и оценить встречаемость каждого ДЭГ в различных экспериментах.

Оба варианта имеют существенные недостатки: в первом случае анализ будет занимать большой промежуток времени, а во втором необходимо проанализировать большое количество экспериментов, проведенных в сходных условиях, что не представляется возможным для большинства организмов в связи с ограниченным количеством опубликованных транскриптомных данных. Целью нашей работы является оптимизация второго способа и разработка метода поиска и идентификации значимых дифференциально экспрессирующихся генов с низкой степенью изменения экспрессии.

Мы решили остановиться на классическом объекте генетических исследований – *Drosophila melanogaster*. Ранее на *Drosophila* было доказано наличие единого ответа на различные стрессовые факторы у насекомых и описан его нейрогормональный механизм, отличающийся высокой консервативностью [1, 2]. Нами была выбрана высокая температура, как один из классических способов стрессорного воздействия на организм насекомого. Для уменьшения количества факторов, влияющих на результаты эксперимента, были выдвинуты следующие требования к данным для анализа: библиотеки получены из имаго *Drosophila melanogaster*, которые подвергались стрессовому воздействию в 37 °C в течение 5 минут, после чего были зафиксированы в жидком азоте.

Методы и алгоритмы: Метаанализ транскриптомных данных на основе программы InterTransViewer. Функциональная аннотация генов с использованием баз данных NCBI, Ensembl, FlyBase и утилиты ShinyGO [3]. Реконструкция генной сети с использованием базы данных StringDB.

Результаты: В результате поиска транскриптомных экспериментов, посвященных тепловому стрессу, было найдено всего три работы, две из которых удовлетворяли поставленному условию. В отобранных экспериментах использовались целые самки, что также увеличило степень схожести экспериментальных условий. Ввиду низкого числа экспериментов было решено отказаться от использования формулы расчета весов, предлагаемой программой InterTransViewer [4] и сосредоточиться на дифференциально экспрессирующихся генах (ДЭГ) с низким уровнем достоверности ($FDR > 0.05$). За счет функциональной аннотации выбранных генов, экспериментальной верификации и дополнительного анализа литературы удалось расширить существующую генную сеть ответа на тепловой стресс и подтвердить работоспособность предложенного метода.

Финансирование: Исследование поддержано бюджетным проектом FWNR-2022-0019 Министерства науки и высшего образования Российской Федерации.

Search and identification of differentially expressed *Drosophila melanogaster* heat stress response genes with a low degree of expression change

Deryuzhenko M.A.*, Gruntenko N.E.

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

* maksimd@bionet.nsc.ru

Key words: transcriptome; *Drosophila melanogaster*; gene network reconstruction; information loss

Motivation and Aim: Information loss is one of the key problems of modern transcriptome data analysis. The main reason for its occurrence is the influence of many unaccounted factors arising during the conduct and analysis of experimental results, which reduce the overall reliability of the obtained results. In an attempt to get rid of them, as a rule, differentially expressed genes (DEGs) with high false discovery rate (FDR) are cut off. At the same time, such genes may play an important role in controlling the observed changes in body functions and get high FDR values due to the low degree of expression change.

This problem can be addressed in several ways:

- 1) Conduct functional annotation of each DEG and review the literature data on them.
- 2) Conduct a meta-analysis of transcriptomic data and estimate the occurrence of each DEG in different experiments.

Both of these options have significant disadvantages: in the first case, the analysis will take a long period of time, and in the second case it is necessary to analyze a large number of experiments performed under similar conditions, which is not possible for most organisms due to the limited amount of published transcriptomic data. The aim of this work is to optimize the second method and to develop a method for searching and

identifying significant differentially expressed genes with a low degree of expression change.

We decided to focus on a classical object of genetic research – *Drosophila melanogaster*. Previously, a single response to various stress factors in insects was proved in *Drosophila* and its neurohormonal mechanism characterized by high conservativity was described [1, 2]. We chose high temperature as one of the classical ways of stressor influence on the insect organism. To reduce the number of factors affecting the results of the experiment, the following data requirements for analysis were made: libraries were obtained from *Drosophila melanogaster* adults that were stressed at 37 °C for 5 min, after which they were fixed in liquid nitrogen.

Methods and Algorithms: Meta-analysis of transcriptome data, based on InterTransViewer program. Functional annotation of genes using NCBI, Ensembl, FlyBase databases and the ShinyGO utility [3]. Gene network reconstruction using StringDB database.

Results: The search for transcriptome experiments devoted to heat stress found only three papers, two of which fulfilled the condition set. Also, whole females were used in the selected experiments, which also increased the degree of similarity of experimental conditions. Due to the low number of experiments, it was decided not to use the formula for calculating weights proposed by the InterTransViewer program [4] and to focus on differentially expressed genes (DEGs) with a low level of confidence (FDR > 0.05). Due to functional annotation of the selected genes, experimental verification and additional literature analysis, it was possible to expand the existing gene network of response to heat stress and confirm the performance of the proposed method.

Funding: The study is supported by BP #FWNR-2022-0019 of the Ministry of Science and Higher Education of the Russian Federation.

Список литературы/References

1. Bobrovskikh M.A., Gruntenko N.E. Mechanisms of neuroendocrine stress response in drosophila and its effect on carbohydrate and lipid metabolism. *Insects*. 2023;14(5):474. doi 10.3390/insects14050474
2. Rauschenbach I.Y. et al. Stress-like reaction of *Drosophila* to adverse environmental factors. *J Comp Physiol B Biochem Syst Environ Physiol*. 1987;157(4):519-531
3. Ge S.X., Jung D., Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*. 2020;36(8):2628-2629. doi 10.1093/bioinformatics/btz931
4. Tyapkin A.V., Lavrekha V.V., Ubogoeva E.V., Oshchepkov D.Y., Omelyanchuk N.A., Zemlyanskaya E.V. InterTransViewer: a comparative description of differential gene expression profiles from different experiments. *Vavilov J Genet Breed*. 2023;27(8):1042-1052. doi 10.18699/VJGB-23-119