

DOI: 10.25205/978-5-4437-1691-6-18

**ВЫЯВЛЕНИЕ ГЕНОВ *DROSOPHILA MELANOGASTER*,
ДИФФЕРЕНЦИАЛЬНО ЭКСПРЕССИРУЮЩИХСЯ С НИЗКИМ УРОВНЕМ ДОСТОВЕРНОСТИ,
НО ВНОСЯЩИХ ЗНАЧИМЫЙ ВКЛАД В ОТВЕТ НА ТЕПЛОВОЙ ШОК**

**IDENTIFICATION OF *DROSOPHILA MELANOGASTER* GENES THAT ARE DIFFERENTIALLY
EXPRESSED WITH LOW CONFIDENCE BUT CONTRIBUTE MEANINGFULLY
TO THE HEAT SHOCK RESPONSE**

М. А. Дерюженко, Н. Е. Грунтенко

Институт цитологии и генетики СО РАН, Новосибирск

M. A. Deryuzhenko, M. G. Gruntenko

Institute of Cytology and Genetics SB RAS, Novosibirsk

✉ maksimd@bionet.nsc.ru

Аннотация

Потеря информации — это одна из ключевых проблем современного анализа транскриптомных данных. Одной из причин ее возникновения является отсечение данных по порогу достоверности. Частично эту проблему можно решить за счет увеличения выборки. Однако — только частично. В этой работе на примере метаанализа ответа на тепловой шок у *Drosophila melanogaster* показан новый подход к изучению генов, находящихся на разных уровнях достоверности.

Abstract

Information loss is one of the key problems of modern transcriptome data analysis. One of the reasons for its occurrence is data cutoff by confidence threshold. Partially, this problem can be solved by increasing the sample size. However, it is a partial solution. In this paper, a new approach to the study of genes at different confidence levels is demonstrated using a meta-analysis of the response to heat shock in *Drosophila melanogaster* as an example.

Потеря информации — это одна из ключевых проблем современного анализа транскриптомных данных. Одной из причин ее возникновения является отсечение данных по порогу достоверности. Делают это из-за влияния множества неучтенных факторов, возникающих при проведении и анализе результатов эксперимента, которые снижают общую достоверность полученных результатов. При таком подходе отбрасываются 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) увеличить достоверность результатов за счет увеличения размера выборки или числа исследований;
- 2) разработать или оптимизировать методы получения и анализа данных.

Оба варианта имеют свои преимущества и недостатки, но, применяя их одновременно, можно добиться хороших результатов. Таким комбинированным подходом является метаанализ. Однако при его проведении, как правило, также отсекаются гены, имеющие низкую достоверность. Из-за этого получаемые результаты не охватывают все нюансы ответа на изучаемый фактор. Целью нашей работы является проведение метаанализа ответа на тепловой шок *Drosophila melanogaster* и изучение дифференциально экспрессирующихся генов, находящихся на разных уровнях достоверности.

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 мин, после чего были зафиксированы в жидком азоте.

В результате поиска транскриптомных экспериментов, посвященных тепловому шоку, было найдено всего три работы, две из которых удовлетворяли поставленному условию. При этом суммарное количество экспериментов равно трем. В отобранных экспериментах использовались целые самки, что также увеличило степень схоже-

сти экспериментальных условий. Полученные данные были предобработаны с использованием утилиты FastQC и Trimmomatic. Выравнивание и квантификация проводились с использованием Hisat2 и featureCounts. Функциональная аннотация генов была проведена на основе баз данных NCBI, Ensembl, FlyBase и утилиты ShinyGO [3]. Реконструкция геновой сети была сделана с использованием базы данных StringDB.

Ввиду низкого числа экспериментов было решено отказаться от использования формулы расчета весов, предлагаемой программой InterTransViewer [4], и сосредоточиться на дифференциально экспрессирующихся генах (ДЭГ) с разным уровнем достоверности. Были взяты гены с $FDR < 0,95$. Всего было найдено 7650 общих генов. Были выделены 4 основные группы генов: участвующие в процессах катаболизма липидов, укладки белков, регуляции трансляции в ответ на стресс и — неожиданно — реакции на холод. За счет функциональной аннотации выбранных генов, экспериментальной верификации и дополнительного анализа литературы удалось расширить существующую геновую сеть ответа на тепловой стресс и подтвердить работоспособность предложенного метода.

Литература

1. Rauschenbach I. Y. et al. Stress-like reaction of *Drosophila* to adverse environmental factors // *J. Comp. Physiol. B. Biochem. Syst. Environ. Physiol.* 1987. Vol. 157 (4). P. 519–531.
2. Bobrovskikh M. A., Gruntenko N. E. Mechanisms of neuroendocrine stress response in *drosophila* and its effect on carbohydrate and lipid metabolism // *Insects.* 2023. Vol. 14 (5). P. 474.
3. Ge S. X., Jung D., Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants // *Bioinformatics.* 2020. Vol. 36 (8). P. 2628–2629.
4. Туаркин А. В., Лавреखा В. В., Убогоева Е. В. et al. InterTransViewer: a comparative description of differential gene expression profiles from different experiments // *Vavilov J. Genet. Breed.* 2023. Vol. 27 (8). P. 1042–1052.