

Пунько Анастасия (мАДБМ-22)

Обзор научной статьи

‘Процесс связывания транскрипционных факторов как главный предиктор шума в экспрессии генов’

Parab L, Pal S, Dhar R (2022) Transcription factor binding process is the primary driver of noise in gene expression. *PLoS Genet* 18(12): e1010535. <https://doi.org/10.1371/journal.pgen.1010535>

Данный обзор посвящен результатам исследования, которое было опубликовано в журнале *Plos Genetics* в конце 2022 года. Статья раскрывает актуальную на сегодняшний день тему — шум в экспрессии генов. Авторы разработали статистическую модель шума экспрессии генов у дрожжей для того, чтобы оценить вклад большого числа молекулярных признаков в объяснение вариаций шума и проверить предсказательные способности этих признаков. Данное исследование показывает, что процесс связывания транскрипционных факторов (ТФ) является лучшим предиктором шума у дрожжей. Также показан механизм регуляции шума, возникающий в результате конкуренции и кооперации между транскрипционными факторами. Важно, что этот механизм не зависит от конкретного транскрипционного фактора или конкретной последовательности промотора, следовательно, может представлять интерес для исследователей, работающих с различными биологическими организмами.

Первое, о чем сообщают авторы в результатах, — это **количественная оценка шума экспрессии на уровне мРНК и белка**. Шум на уровне мРНК показал значительную корреляцию с шумом на уровне белка, хотя диапазон абсолютных значений шума был очень разным. Также была проведена количественная оценка относительной важности каждого молекулярного признака в регуляции шума и измерение их способности предсказывать шум. И выяснилось, что *молекулярные признаки, которые ранее считались влияющими на шум экспрессии*, такие как наличие последовательности TATA-box в промоторе и занятость нуклеосом промотора, показали значительную связь с шумом, но *оказались плохими предикторами*.

Далее авторы сообщают о молекулярном признаке, который показал себя как **лучший предиктор шума**. Для определения признаков, которые могли бы объяснить наблюдаемые вариации значений шума и предсказать шум, авторы построили интегрированную статистическую модель. Данная модель учитывала 329 признаков, которые влияют на экспрессию генов. 8 лучших признаков были

ассоциированы со связыванием ТФ, что указывает на *ключевую роль ТФ в регуляции шума*. К ним относятся число регуляторных ТФ гена и число сайтов связывания ТФ. Еще два признака были ассоциированы с комплексами SAGA и TFIIID, от которых зависит транскрипция генов стрессового ответа и генов домашнего хозяйства. Также авторы отметили, что прогнозируемые значения, полученные с использованием числа регуляторных ТФ в качестве единственного признака и с использованием комбинации 10 лучших признаков, показали значительную корреляцию с наблюдаемыми значениями шума на уровне белка. Из всех признаков в статистической модели *процесс связывания ТФ мог объяснить большую часть вариаций в данных и имел самую высокую предсказательную силу*. Кроме того, интегрированная модель смогла показать, что только процесс связывания ТФ мог предсказать шум как на уровне мРНК, так и на уровне белка.

Следующее, о чем сообщают авторы, это то, что **гены с высоким уровнем шума экспрессии регулируются большим числом ТФ**. Интересно, что гены, регулируемые большим числом ТФ, демонстрировали сопутствующее увеличение числа кооперативно связывающихся ТФ. Соответственно, *шум значительно коррелировал с числом кооперативно связывающихся ТФ* как для мРНК, так и для белкового шума. Также было показано, что увеличение числа регуляторных ТФ и соответствующее увеличение сайтов их связывания привело к значительному увеличению перекрытия сайтов связывания ТФ в промоторной области. Это нашло отражение в значительной корреляции между шумом на уровне мРНК и белка с количеством перекрытий сайтов связывания ТФ (среднее число перекрытий увеличилось для генов с самым высоким уровнем шума по сравнению с генами с самым низким уровнем шума). Сотрудничество и конкуренция между ТФ могут происходить только тогда, когда ТФ экспрессируются в клетке в одно и то же время (гены с самым высоким уровнем шума в среднем имели на 90 % больше коэкспрессирующих ТФ, чем гены с самым низким уровнем шума).

Последнее, что упомянули авторы, это то, что **кооперативное и конкурентное связывание ТФ может вызвать высокий шум экспрессии**. Оказалось, что *простое увеличение числа регуляторных ТФ не может объяснить более высокий уровень шума*, наблюдаемый в генах с большим числом регуляторных ТФ. Как кооперативное, так и конкурентное связывание

ТФ приводило к более высокому шуму по сравнению с регуляцией одним ТФ или двумя независимыми ТФ, что говорит о том, что *динамика процесса связывания ТФ в случае регуляции генов несколькими ТФ играет важную роль в генерации шума экспрессии*. Также авторы показали, что при вариации параметров модели (скорость транскрипции и трансляции, скорость деградации, количество кооперативных и конкурентных ТФ) конкурентное и кооперативное связывание ТФ показали более высокий уровень шума по сравнению с регуляцией одиночными ТФ.

С моей точки зрения, такого рода исследования очень важны. Ведь шум в экспрессии генов приводит к вариациям в активности клеточных путей, и как результат, возникает фенотипическая гетерогенность клеток. Фенотипическая гетерогенность играет важную роль в устойчивости к антибиотикам, неполной пенетрантности мутаций, в росте опухолевых клеток и устойчивости к терапии.

Также данная статья важна с точки зрения применения математических моделей. Моделирование помогает эффективно планировать исследования и интерпретировать их результаты с использованием всех доступных данных и знаний. Например, на модель из данного исследования можно опираться при построении моделей шума экспрессии генов в других организмах. Однако не стоит забывать, что ни одна модель не может быть совершенна в предсказании. Например, возможно, что некоторые признаки, которые могут регулировать шум, не были учтены в модели. Расчет шума на данных одной временной точки также может повлиять на предсказательную силу (т.к. молекулярным процессам присуща случайность, и шум экспрессии также может меняться со временем). Условия роста, скорость роста клеток и клеточный цикл влияют на шум экспрессии генов. Таким образом, объединение данных о молекулярных особенностях во многих наборах данных без учета этих переменных может потенциально повлиять на предсказательную силу.

Стохастические вариации в молекулярных процессах повсеместно распространены в клетках всех биологических систем и имеют серьезные последствия для заболеваний человека. Таким образом, способность точно предсказывать вариации в биологических процессах будет чрезвычайно полезна для количественной оценки степени гетерогенности клеточных признаков и фенотипов.