abstract

Шум в экспрессии отдельных генов приводит к изменениям активности клеточных путей и генерирует неоднородность в клеточных фенотипах. Фенотипическая неоднородность имеет важные последствия для стойкости антибиотиков, прониения мутаций, роста рака и терапии. Конкретные молекулярные особенности, такие как наличие последовательности коробок TATA и занятость промоторной нуклеосомы, были связаны с шумом. Однако относительная важность этих функций в регулировании шума неясна, и насколько хорошо эти функции могут предсказать шум, еще не оценено. Здесь с помощью интегрированной статистической модели шума экспрессии генов в дрожжах мы обнаружили, что количество регулирующих факторов транскрипции (TF) гена было ключевым предиктором шума, в то время как наличие коробки TATA и занятости промоторной нуклеосом имело плохую прогностической способности. С увеличением числа регулирующих TF увеличилось количество совместно обязательных TF. Кроме того, увеличение числа регулирующих TF означало большее совпадений в сайтах связывания TF, что привело к конкуренции между TF за привязку к одному и тому же региону промоутера. Путем моделирования связывания TF с промотором и применения стохастического моделирования мы продемонстрировали, что конкуренция и сотрудничество между TF могут увеличить шум. Таким образом, наша работа раскрывает процесс шумовой регуляции, который возникает из динамики регуляции генов и не зависит от какого-либо конкретного фактора транскрипции или конкретной последовательности промотора.

Резюме автора

Резюме автора

Уровни экспрессии генов могут варьироваться даже среди генетически идентичных клеток в одинаковых условиях окружающей среды - явление, называемое шумом экспрессии. Шум экспрессии генов был экспериментально измерен в нескольких популяциях клеток, и более ранние исследования связывали наличие определенной последовательности оснований, таких как коробка TATA в области промоторов, уровни занятости нуклеосом и модели модификации гистонов с высоким уровнем экспрессии. Однако, насколько хорошо эти молекулярные особенности гена могут позволить нам предсказать его шум экспрессии, еще не оценено. В текущей работе мы тестируем большое количество молекулярных особенностей, связанных с экспрессией генов, на их способность предсказывать шум. Мы обнаружили, что количество факторов транскрипции гена является ключевым предиктором экспрессии шума. Увеличение количества факторов транскрипции может изменить процесс их связывания с промоторским регионом и привести к большему сотрудничеству или конкуренции. Путем моделирования и моделирования кооперативного и конкурентного связывания мы показываем, что процесс связывания факторов транскрипции в первую очередь управляет шумом экспрессии. Наша работа показывает, что динамика регуляции экспрессии генов является наиболее важной особенностью прогнозирования экспрессического шума, и раскрывает общий механизм регуляции шума.

Знакомство

Случайные колебания в молекулярных событиях, происходящих внутри клетки, генерируют изменения в уровнях экспрессии генов, которые называются шумом экспрессии генов. Экспрессия шума приводит к изменениям активности клеточных путей и генерирует фенотипическую неоднородность среди отдельных клеток изогенной популяции в одинаковых условиях окружающей среды. Шум экспрессии генов играет важную роль в стойкости антибиотиков [[1](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref001)[-5](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref005)] и неполном прониканием мутаций [[6](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref006)[-10](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref010)]. Кроме того, фенотипическая неоднородность играет ключевую роль в росте рака [[11](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref011)[-13](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref013)] и в возникновении терапии резистентности [[14](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref014)[-18](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref018)].

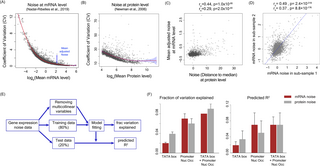
Шум экспрессии генов измерялся в некоторых микробных системах [[19](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref019)[-22](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref022)], и его молекулярное происхождение было широко исследовано [[23](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref023)–[36](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref036)]. Эти исследования показали корреляцию между наличием мотива коробки TATA в промоторной области гена и шумом экспрессии [[20,26,29,37,38](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref020)]. Кроме того, занятость промоторной нуклеосомы, как только, так и в сочетании с наличием мотива коробки TATA и шаблоны модификации гистонов также были связаны с шумом экспрессии [[33,35,39](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref033)–[43](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref043)]. Эти функции могут влиять на размер всплеска транскрипции и частоту всплеска [[43](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref043)–[45](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref045)], что, в свою очередь, может влиять на шум экспрессии [[29,46](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref029)–[48](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref048)]. Однако даже после стольких исследований на протяжении многих лет относительная важность этих молекулярных особенностей в регулировании шума остается неизвестной. Кроме того, в какой степени каждая из этих молекулярных признаков может предсказать шум, еще не определена количественно. То есть, можем ли мы оценить шум экспрессии гена, учитывая наличие или отсутствие последовательности коробки TATA в его промоторе и характер занятости промоторной нуклеосомы, неизвестно. Таким образом, прогнозная модель шума будет чрезвычайно полезна для лучшего понимания регулирования шума в биологических системах.

В текущей работе мы сообщаем о разработке интегрированной статистической модели шума экспрессии генов в дрожжах путем объединения большого количества молекулярных особенностей, которые могут повлиять на экспрессию генов. Мы количественно оценивали относительный вклад каждой из этих признаков в объяснение различий в значениях шума генов и проверили их прогностические способности. Мы заметили, что присутствие коробки TATA и модель занятости нуклеосом промотера были плохими предикторами шума экспрессии. Вместо этого количество регуляторных TF гена стало ключевым предиктором шума. Увеличение числа регулирующих TF было связано с сопутствующим увеличением числа кооперативных TF. Кроме того, увеличение числа регуляторных TF означало переполненность сайтов связывания TF в промоторной области гена. Это привело к большему дубляжению сайтов привязки TF, тем самым усиливая конкуренцию между TF за привязку к одному и тому же сайту промоутера. Математическое моделирование и стохастическое моделирование показали, что простое увеличение числа TF не может объяснить увеличение шума экспрессии, в то время как совместное и конкурентное связывание TF может генерировать более высокий шум экспрессии. В совокупности наша работа показывает, что процесс связывания факторов транскрипции является лучшим предиктором шума в дрожжах. Мы раскрываем динамический механизм регулирования шума, возникающий в результате конкуренции и сотрудничества между факторами транскрипции. Этот механизм не зависит от конкретного фактора транскрипции или конкретной последовательности промотора и, таким образом, может представлять интерес для исследователей, работающих над различными биологическими организмами.

Результаты

**Количественная оценка шума экспрессии на уровне мРНК и белка**

Мы количественно оценивали шум экспрессии генов на уровне мРНК и белка, используя два разных экспериментальных набора данных. Для расчета шума на уровне мРНК мы использовали одноклеточные данные РНК-сек в дрожжах из Надаль-Рибель *и др*. [[49](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref049)] ([Рис. 1А](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g001)). Набор данных содержал значения экспрессии генов в 127 одноклеточных штамме *Saccharomyces cerevisiae*BY4741, выращенных в богатой среде роста (YPD), и профили экспрессии, измеренные на ранней логаментарной фазе. Из этого набора данных мы получили значения экспрессии 5475 генов. Для количественной оценки шума мы использовали меру шума, которая не зависела от среднего уровня экспрессии, подгоняя сплайн к шуму (коэффициент вариации, CV) по сравнению со средним графиком и вычисляя вертикальное расстояние значений шума от подогнанной кривой (рис. [1A](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g001)и [S1)](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s002). Среднескорректированные значения шума из двух подвыборок данных одноклеточного РНК-сека показали значительную корреляцию друг с другом (корреляция Пирсона rp = 0,49, p = 2,4×10−318и корреляция Спирмана rs = 0,37, p = 8,8×10−176; [рис. 1D](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g001)).

[](https://journals.plos.org/plosgenetics/article/figure/image?size=medium&id=10.1371/journal.pgen.1010535.g001)

**Рис. 1. Наличие последовательности TATAbox и уровни занятости нуклеосом промотера являются плохими предикторами шума экспрессии генов.**

**(A)**Значения шума, рассчитанные на уровне мРНК на основе одноклеточных данных РНК-секв в дрожжах [[49](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref049)]. Средний скорректированный шум был рассчитан путем приспособления полиномиальной кривой к графику CV против среднего, показанному красной линией. Каждая точка показывает CV и средний уровень мРНК для гена.**(B)**Значения шума, рассчитанные на уровне белка на основе измерений проточной цитометрии [19]. Красная линия показывает наилучшее полиномиальное посадка, а затененная синяя область показывает 95% доверительный интервал.**(C)**Корреляция между средне-скорректированным шумом на уровне мРНК и шумом (DM) на уровне белка. 'rp' показывает корреляционное значение Пирсона, а 'rs' показывает корреляционное значение Спирмана.**(D)**Корреляция значений шума экспрессии на уровне мРНК, рассчитанных на основе двух подвыборок одноэлементных данных РНК-секв [[49](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref049)].**(E)**Блок-схема, показывающая шаги для примерки модели, расчет разъяснения доли вариации и вывод прогнозируемого R2.**(F)**Доля вариаций, объясненная и предсказанных значений R2по наличию или отсутствию последовательности коробки TATA, средней занятости промоторной нуклеосомы на сайт, связанный с нуклеосомой, и комбинацией присутствия/отсутствия последовательности коробки TATA с заполненностью промоторной нуклеосомы.

<https://doi.org/10.1371/journal.pgen.1010535.g001>

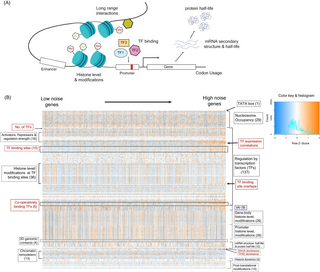
Мы получили значения шума на уровне белка для 2763 генов в штамме *S*. *cerevisiae*S288C, выращенном в богатой среде (YPD) [19] ([рис. 1B](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g001)). Мы использовали их меру "расстояние до медианы" (DM) в качестве меры шума в нашем исследовании ([S1 Рис.](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s002)). Шум на уровне мРНК показал значительную корреляцию с шумом на уровне белка (rp = 0,44, p = 1,7×10−95; rs = 0,29, p = 2,0×10−40; [рис. 1C](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g001)), хотя диапазон абсолютных значений шума был очень разным. Гены показали широкий спектр значений шума экспрессии, при этом очень шумные гены показывали большие положительные значения, а гены с низким уровнем шума показывали большие отрицательные значения.

Чтобы количественно оценить относительную важность каждого молекулярного элемента в регулировании шума и измерить их способность прогнозировать шум, мы случайным образом разделили данные о шуме на обучение (80% от полных данных) и протестированные наборы данных (осталось 20%) ([рис. 1E](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g001)). Для количественной оценки прогнозной способности одного элемента мы подкрепили модель линейной регрессии к тренировочным данным на этом этапе. Для количественной оценки предсказательной способности комбинации признаков мы сначала удалили мультиколлинеарные функции и определили ключевой набор признаков через регрессию Ridge или Lasso на полных данных, а затем установили модель линейной регрессии на тренировочные данные. Это дало нам долю вариаций, объясненную моделью ([рис. 1E](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g001)). Затем мы использовали подходяную модель для прогнозирования тестовых данных и вычисляли прогнозируемые значения R2([рис. 1E](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g001)). Мы провели этот анализ как в мРНК, так и в наборах данных о белковом шуме, чтобы убедиться, что сделанные выводы не были смещены конкретным набором данных.

Молекулярные особенности, которые ранее считались влияющими на шум экспрессии, такие как наличие последовательности коробок TATA в промоторе [[20,29,37,38](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref020)] и занятость промоторных нуклеосом [[39](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref039)–[41](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref041)], показали значительную связь с шумом ([S2 Рис](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s003)), но были плохими предикторами ([рис. 1F](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g001)). В частности, последовательность коробок TATA, занятость промоторных нуклеосом только и в сочетании с последовательностью коробок TATA могла объяснить только ~2–4%, ~6–7% и ~8–9% колебаний шума, соответственно, и имела низкую предсказативную способность (прогнозируемое значение R20,02 0,03 только для коробки TATA, 0,05–0,07 только для занятости промоторной нуклеосомы и 0,07 для коробки TATA + занятость промоторной нуклеосомы; [Рис. 1F](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g001)). Это говорит о том, что эти функции в значительной степени связаны с шумом и не были предсказуемыми.

**Молекулярные особенности, связанные с связыванием TF, были лучшими предикторами шума**

Чтобы определить молекулярные особенности, которые могли бы объяснить наблюдаемые изменения в значениях шума и предсказать шум, мы создали интегрированную статистическую модель, учитывая большое количество признаков, которые были известны или могут повлиять на экспрессию генов, поскольку это могут быть потенциальные регуляторы шума ([рис. 2](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g002)и [таблица S1)](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s012). Цели интегрированной статистической модели заключались в том, чтобы проверить прогностическое значение каждого молекулярного элемента индивидуально и определить наилучший набор признаков для прогнозирования шума из большого количества возможных комбинаций.

[](https://journals.plos.org/plosgenetics/article/figure/image?size=medium&id=10.1371/journal.pgen.1010535.g002)

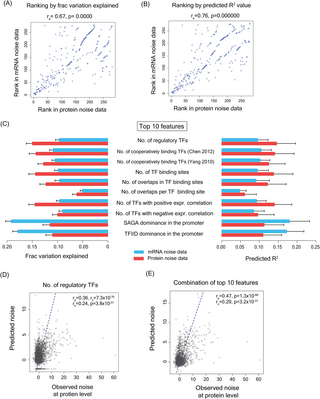
**Рис. 2.**

**Интегрированная статистическая модель шума экспрессии генов (A)**Схематическая диаграмма, изображающая молекулярные особенности, которые могут повлиять на экспрессию генов и, таким образом, может играть ключевую роль в регуляции шума экспрессии.**(B)**Интегрированная модель шума, построенная с учетом последовательности коробок TATA, абсолютного уровня занятости нуклеосом, регуляции генов TF, индекса адаптации тРНК, моделей модификации гистонов в областях тела генов и промотора, 3D-геномных контактов, структуры мРНК и периода полураспада, периода полураспада белка, активности ремоделеров хроматина, динамики связывания гистонов и посттрансляционных модификаций. Тепловая карта показывает значения всех этих признаков (масштабированных и центрированных) в генах (представленных в столбцах), отсортированные в соответствии с их значениями шума на уровне белка. Количество элементов, для которых данные отображаются в тепловой карте, указано в скобках. Особенности, выделенные красным цветом, отличаются по своим значениям между генами с низким и высоким уровнем шума. На панели справа показана ключ цвета для тепловой карты вместе с распределением значений всех элементов (гистограммы).

<https://doi.org/10.1371/journal.pgen.1010535.g002>

Молекулярные особенности, включенные в интегрированную модель, включали в себя количество регулирующих TF, расположение их сайтов связывания, их среднюю экспрессию и уровни шума, зависимость генов SAGA/TFIID от их экспрессии [[50](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref050)], был ли ген избыточным или зависим от TFIID [[51](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref051)], связывающую активность нескольких широко действующих TF, таких как TBP, ABF1 и RAP1 [[52](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref052)–[54](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref054)], шаблоны связывания ремоделеров хроматина [[55](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref055)–[57](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref057)], уровни гистонов, модели модификации гистонов и динамика связывания гистонов [[58,59](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref058)], трехмерные геномные контакты [[60](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref060)], индекс адаптации тРНК [[61](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref061)], вторичная структура мРНК и период полураспада белка [[62](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref062)–[64](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref064)], посттрансляционные модификации [[65](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref065)], в дополнение к нуклеосомной модели занятости [ наличие/отсутствие последовательности коробок TATA в промоторе [[67](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref067)]. Для гена мы рассмотрели только те TF, для которых были получены экспериментальные доказательства связывания ДНК или экспериментально наблюдалось изменение экспрессии при нокаутировании TF. Для занятости нуклеосом мы не только рассмотрели количество сайтов, связанных с нуклеосомами, но и включили абсолютную схему занятости нуклеосом [[66](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref066)]. В общей сложности мы рассмотрели 329 функций в нашей интегрированной модели (см. [Методы](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#sec008), [Таблица S1)](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s012).

Мы протестировали каждую функцию индивидуально на ее способность объяснять изменения в данных о шуме и прогнозировать шум как в мРНК, так и в наборах данных о белковом шуме. Затем мы ранжировали эти функции в соответствии с объясненной долей вариаций и предсказанных значений R2. Рейтинги функций, будь то на объясненной доли вариаций или на прогнозируемом значении R2, существенно коррелировали между наборами данных мРНК и белкового шума со значениями корреляции Спирмана 0,67 и 0,76 соответственно ([рис. 3A и 3B](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g003)).

[](https://journals.plos.org/plosgenetics/article/figure/image?size=medium&id=10.1371/journal.pgen.1010535.g003)

**Рис. 3.**

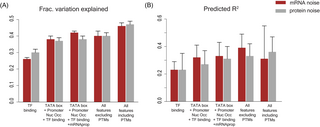
**Особенности с самыми высокими прогнозными способностями были в значительной степени связаны с процессом связывания факторов транскрипции (A)**Ранжирование признаков в соответствии с долей вариаций, объясненной в наборе данных о мРНК-шуме и в наборе данных о белковом шуме, были сильно коррелированы **(B)**Рейтинги признаков в соответствии с прогнозируемым значением R2в мРНК и наборы данных о белковом шуме были высококоррелированы **(C)**Доля вариаций объяснила и предсказала значение R2для 10 основных функций как для мРНК, так и для наборов данных о белкового шума.**(D-E)**Корреляция между наблюдаемыми значениями шума и значениями шума, предсказанное моделью линейной регрессии с учетом одного признака (количество регуляторных TF) (D) и комбинацией 10 основных признаков (E).

<https://doi.org/10.1371/journal.pgen.1010535.g003>

10 лучших функций для объяснения вариаций, существующих в данных о шуме, и для прогнозирования значений шума содержали одни и те же функции, хотя их рейтинг немного отличался. Распределение значений некоторых из этих признаков показано на [рис](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s004). Интересно, что восемь из этих функций были связаны с связыванием TF, что свидетельствует о ключевой роли TF в регулировании шума ([рис. 3C](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g003)). К ним относятся количество регуляторных TF гена (дробь вариации объяснила ~0,1–0,15 и предсказанное R2~0,1–0,15) и количество сайтов связывания TF (фракция вариации объяснила ~0,1–0,14, предсказанное R2~0,1–0,14). Две особенности из 10 основных признаков были связаны с SAGA-зависимостью и TFIID-зависимостью генов для их транскрипции. Известно, что гены стрессовой реакции в дрожжах шумнее, чем гены домашнего хозяйства [19]. В то время как гены домашнего хозяйства зависят от комплекса TFIID по своей экспрессии, гены стрессовой реакции обычно зависят от комплекса SAGA. Зависимость SAGA и зависимость TFIID могут объяснить 0,11–0,19 и 0,11–0,17 доли вариации соответственно с прогнозируемыми значениями R20,11–0,18 и 0,11–0,17 соответственно ([рис. 3C](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g003)).

Мы дополнительно подтвердили прогностические способности этих признаков, соподав наблюдаемые и прогнозируемые значения шума. Прогнозируемые значения, полученные с использованием количества регуляторных TF в качестве единственной функции и с использованием комбинации 10 лучших признаков, показали значительные корреляции с наблюдаемыми значениями шума на уровне белка ([рис. 3D и 3E](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g003)).

Из всех особенностей нашей модели процесс связывания TF мог объяснить большую часть доли вариаций данных и имел самую высокую прогностические способности ([рис. 4)](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g004). Интегрированная модель, состоящая из всех функций, смогла объяснить 0,46 доли изменения шума на уровне мРНК и 0,47 доли изменения шума на уровне белка ([рис. 4A](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g004)). Только связывание TF объяснило 0,26 доли изменения шума на уровне мРНК и 0,30 доли шума на уровне белка ([рис. 4A](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g004)). Кроме того, интегрированная модель смогла предсказать шум на уровне мРНК с прогнозируемым значением R20,31 и на уровне белка с прогнозируемым значением R20,36 ([рис. 4B](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g004)). Как и раньше, только процесс связывания TF может предсказать шум как на мРНК, так и на уровне белка с прогнозируемым значением R20,23 ([рис. 4B](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g004)).

[](https://journals.plos.org/plosgenetics/article/figure/image?size=medium&id=10.1371/journal.pgen.1010535.g004)

**Рис. 4. Объяснение фракции вариаций и прогнозная способность комбинаций молекулярных признаков.**

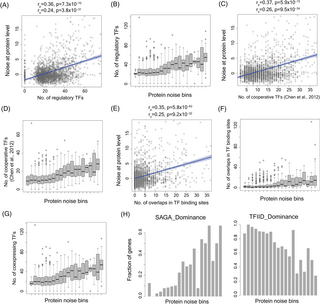
**(A)**Доля вариаций, объясненная в данных о шуме экспрессии генов, и **(B)**прогностическая способность (данная предсказанное значением R2) функциями, связанными с связыванием TF; комбинация связывания TF с последовательностью коробки TATA и занятостью промоторных нуклеосом; сочетание связывания TF с последовательностью коробок TATA, занятость промоторных нуклеосом и свойства мРНК; сочетание всех функций, исключающих посттрансляционные модификации (PTM), и сочетание всех функций, включая PTM.

<https://doi.org/10.1371/journal.pgen.1010535.g004>

Несколько генов в геноме дрожжей были сохранены в результате дублирования всего генома [[68](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref068)] и, таким образом, имеют много молекулярных особенностей, включая промоторные и кодирующие последовательности областей с их дубликатами. Это может смещение нашего анализа и привести к завышению предсказательных значений R2. Таким образом, чтобы оценить влияние дубликатов генов в нашем анализе, мы удалили дубликаты из наших наборов данных и повторили весь анализ. Доля вариаций, объясненных и предсказанных значений R2по отдельным признакам и комбинациям признаков, была сопоставима между наборами данных с повторяющимися генами и без них ([S4](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s005)и [S5](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s006)Figs).

**Гены с высоким уровнем экспрессии регулировались большим количеством TF**

Наша модель выявила значительную корреляцию между количеством регулирующих TF гена и шума как на уровне мРНК, так и на уровне белка (для шума белка корреляция Пирсона rp = 0,36, p = 7,3×10−70и корреляция Спирмана rs = 0,24, p = 3,8×10−31; [Рис. 5A](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g005); для шума мРНК rp = 0,26, p = 5,7×10−85; rs = 0,19,19, p = 7,0×10−47; [S6A Рис](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s007)). Мы также классифицировали гены на 20 равномерно расположенных шумовых бункеров (за исключением первого и последнего бункеров), отсортированных в соответствии с их значениями шума. Первый бункер имел открытый нижний предел значений шума, чтобы включать гены, показывающие очень низкий уровень шума. Последний бункер имел открытый верхний предел значений шума, чтобы включить гены, показывающие очень высокий уровень шума. Это помогло нам избежать бункеров с очень небольшим количеством генов. Затем мы рассмотрели распределение количества регуляторных TF генов в этих бункерах (рис. [5B](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g005)и [S6B](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s007)). Гены в самых высоких шумовых бункерах в среднем имели на 75% больше регуляторных TF по сравнению с генами в самых низких шумовых бункерах (рис. [5B](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g005)и [S6B](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s007)).

[](https://journals.plos.org/plosgenetics/article/figure/image?size=medium&id=10.1371/journal.pgen.1010535.g005)

**Рис. 5. Гены с шумом высокой экспрессии регулировались большим количеством TF, имели большее количество совместно связывающих TF и демонстрировали большее совпадений в сайтах связывания TF по сравнению с генами с низким уровнем шума.**

**(A)**Корреляция между шумом на уровне белка и количеством регуляторных TF.**(B)**Количество регуляторных TF генов в 20 белковых шумовых бункерах.**(C)**Корреляция между шумом на уровне белка и количеством кооперативных TF [[70](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref070)] **(D)**Количество кооперативных TF генов через белковые шумовые бункеры.**(E)**Корреляция между шумом на уровне белка и количеством перекрытий в местах связывания TF.**(F)**Количество совпадений между сайтами связывания TF для генов в белковых шумовых бункерах.**(G)**Количество соэкспрессирующих регуляторных TF в белковых шумовых бункерах.**(H)**Доля генов, показывающих доминирование SAGA и TFIID в белковых шумовых бункерах.

<https://doi.org/10.1371/journal.pgen.1010535.g005>

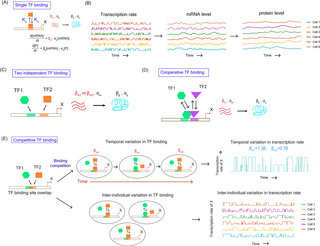
Это подняло ключевой вопрос - как увеличение числа регулирующих TF может привести к увеличению шума экспрессии. Интересно, что гены, регулируемые большим количеством TF, показали сопутствующий рост числа TF, демонстрирующий кооперативное связывание [[69,70](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref069)]. Ожидается, что шум был значительно коррелирован с количеством совместно связывающих TF как для мРНК, так и для белкового шума (рис. [5C](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g005)и [S6C](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s007). Гены в самых высоких шумовых бункерах в среднем имели более 66% кооперативных TF, чем гены в самых низких шумовых бункерах (рис. [5D](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g005)и [S6D](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s007)).

Кроме того, увеличение числа регулирующих TF и соответствующее увеличение их сайтов связывания привели к значительному увеличению дублирования сайтов связывания TF в промоторском регионе. Это нашло отражение в значительной корреляции между шумом на мРНК и уровнем белка с количеством перекрытий сайтов связывания TF (рис. [5E](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g005)и [S6E](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s007)). Медиана перекрытий увеличилась более чем в 4 раза для генов в самых высоких шумовых бункерах по сравнению с генами в самых низких шумовых бункерах (рис. [5F](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g005)и [S6F](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s007)). Сотрудничество и конкуренция между ФТ могут происходить только тогда, когда ФТ экспрессируются одновременно внутри клетки. Интересно, что гены в самых высоких шумовых бункерах в среднем увеличили количество коэкспрессивных TF на более чем на 90%, чем гены в самых низких шумовых бункерах (рис. [5G](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g005)и [S6G](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s007)). Кроме того, гены в самых высоких шумовых бункерах имели более чем в четыре раза большую долю генов, зависящих от SAGA, и имели примерно в три раза меньшее количество зависимых от TFIID генов по сравнению с самыми низкими шумовыми бунками, учитывая шум как на мРНК, так и на уровне белка (рис. [5H](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g005)и [S6H](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s007)).

**Совместное и конкурентное связывание TF может генерировать высокий уровень шума**

Чтобы лучше понять, как увеличение числа регуляторных TF может привести к более высокому шуму экспрессии, мы построили математическую модель регуляции генов и провели стохастическое моделирование в популяции клеток. В частности, мы спросили, может ли простое увеличение числа регулирующих TF объяснить более высокий шум экспрессии и может ли кооперативное и конкурентное связывание TF сыграть какую-либо роль в создании шума более высокой экспрессии.

Сначала мы изучили регуляцию гена одним TF ([рис. 6A](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g006)). TF binding - это динамический процесс, состоящий из шагов быстрого связывания и отвязывания [[71,72](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref071)]. Таким образом, мы использовали двухстоятельную модель экспрессии генов, где ген может существовать в состояниях on-, так и off с определенными темпами перехода между этими двумя состояниями. Связывание TF привело к переходу в состояние, что привело к производству мРНК и белков. Мы количественно оценивали изменения уровней экспрессии генов с течением времени с помощью стохастического моделирования с использованием алгоритма Гиллеспи ([рис. 6B](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g006)). Мы смоделировали динамику экспрессии генов в 10000 клетках и количественно четали средний уровень экспрессии и шум.

[](https://journals.plos.org/plosgenetics/article/figure/image?size=medium&id=10.1371/journal.pgen.1010535.g006)

**Рис. 6.**

**Математическое моделирование и стохастическое моделирование связывания TF и влияния на шум экспрессии генов (A)**Математическое представление модели, описывающей регуляцию одной схемой TF **(B),**показывающей изменение скорости транскрипции, уровней мРНК и уровня белка среди отдельных клеток, полученных в результате математического моделирования и стохастического моделирования **(C)**Схематическая диаграмма, показывающая регуляцию генов двумя TF, связывающихся независимо с промотором.**(D)**Схематическая диаграмма, показывающая кооперативное связывание двух TF с промотором гена и индукцию транскрипции **( E)**Перекрытие сайтов связывания TF приводит к конкуренции между TF. Это может привести к временным изменениям связывания TF в той же промоторной области в клетке. Кроме того, асинхронность в связывании ТФ между отдельными клетками может привести к межиндивидуальным изменениям в связывании TF и скорости транскрипции.

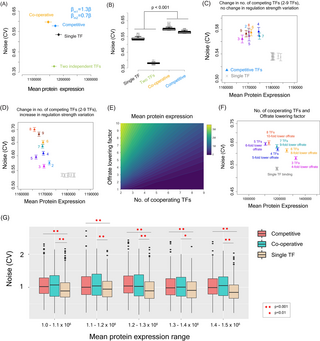
<https://doi.org/10.1371/journal.pgen.1010535.g006>

На следующем этапе мы проверили, может ли простое увеличение количества TF повлиять на шум экспрессии. Для этого мы смоделировали регуляцию гена двумя TF, связывающимися независимо с промоторной областью (без какого-либо сотрудничества или конкуренции) ([рис. 6C](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g006)). Здесь мы предположили, что связывание любого из TF с промотором привело к состоянию on и привело к производству мРНК и белка. Когда оба TF были привязаны к промотору, скорость транскрипции увеличивалась и была равна сумме скорости транскрипции для отдельных TF.

В кооперативном связывании двух TF мы смоделировали скорость транскрипции функцией Хилла и предположили, что транскрипция как процесс "все или ни один" независимо от значения коэффициента Хилла. Это означало, что при совместном связывании двух TF существенная транскрипция происходила только тогда, когда оба TF были одновременно связаны с областью промотора ([рис. 6D](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g006)). Этот процесс может изменить частоту всплесков транскрипции, тем самым влияя на общую экспрессию мРНК и белка ([S7 рис.](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s008)). Однако кооперативная связывание TF может продлить продолжительность пребывания в штате и предотвратить переход в негосударство [[73](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref073)]. Мы смоделировали это путем снижения темпов перехода в оф-государство. Это позволило нам выполнить все сравнения шума экспрессии на аналогичных средних уровнях экспрессии ([S7 рис.](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s008)).

Конкуренция между двумя TF за привязку к перекрывающимся сайтам в промоторном регионе может вызвать шум двумя возможными способами. Во-первых, конкуренция между TF может привести к сценарию, когда ген будет регулироваться различными подмножествами TF в разные моменты времени, тем самым создавая временные колебания ([рис. 6E](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g006)). При наличии ФТ, которые отличаются сильные стороны регуляции, это может привести к временным изменениям скорости транскрипции внутри клетки. Во-вторых, асинхронные временные изменения в связывании ТФ между отдельными клетками в популяции могут генерировать межиндивидуальную вариацию экспрессии ([рис. 6E](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g006)).

Интересно, что при аналогичных средних уровнях экспрессии белка регуляция двумя независимыми TF имела более низкий уровень шума, чем одна регуляция TF ([рис. 7A и 7B](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g007)), так как целевой ген чаще находился в состоянии под действием одного из двух TF и, следовательно, имел меньшую временную и межиндивидуальную вариацию уровня белка. Это показало, что простое увеличение числа регуляторных ТФ не может объяснить более высокий шум, наблюдаемый в генах с большим количеством регуляторных ТФ. Для сравнения, как кооперативное, так и конкурентное связывание TF привело к более высокому шуму по сравнению с регуляацией одним TF или двумя независимыми TF ([рис. 7A и 7B](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g007)), что свидетельствует о том, что динамика процесса связывания TF в случае регуляции генов несколькими TF играет важную роль в генерации экспрессивного шума.

[](https://journals.plos.org/plosgenetics/article/figure/image?size=medium&id=10.1371/journal.pgen.1010535.g007)

**Рис. 7. Средняя экспрессия и шум в случае регуляции генов одним TF, двумя независимыми TF, двумя кооперативными TF и двумя конкурентными TF.**

**(A)**Значения средней экспрессии и шума, полученные в результате моделирования и моделирования регуляции генов одним TF (черный), двумя TF связываются независимо (зеленый), двумя TF, связываются совместно (оранжевый) и двумя TF, связываются конкурентно (синий).**(B)**Распределение шума в случаях регуляции генов одним TF и двумя TF, связывающихся независимо, совместно и конкурентно. Шум был рассчитан в нескольких временных точках в 10 000 смоделированных ячеек.**(C-D)**Изменения в шуме экспрессии с увеличением количества конкурентоспособных TF с изменениями и без изменений в силе регулирования.**(E)**Увеличение числа сотрудничающих TF может снизить среднюю экспрессию белка без каких-либо изменений в включенных и выключенных атрифах. Тем не менее, экспрессия остается неизменной, если связывание более кооперативных TF может увеличить время, в течение которого ген остается в состоянии on, снизив откоэф.**(F)**Шум экспрессии генов в случае 3-9 кооперативных TF на аналогичном среднем уровне экспрессии белка. Параметр нескопаратной ставкой был скорректирован для достижения одинакового диапазона среднего выражения во всех случаях.**(G)**Boxplots, показывающие значения шума в одиночном TF, конкурентном TF и кооперативном связывании TF из стохастического моделирования с выборкой параметров математической модели Markov-Chain Monte-Carlo. Целевая средняя экспрессия белка была установлена между 1×106и 1,1×106, между 1,1×106и 1,2×106, между 1,2×106и 1,3×106, между 1,3×106и 1,4×106, между молекулами 1,4×106и 1,5×106.

<https://doi.org/10.1371/journal.pgen.1010535.g007>

Мы также изучили, могут ли различия в параметрах модели, такие как скорость транскрипции и перевода, показатели деградации, количество кооперативных и конкурентоспособных TF, повлиять на наш вывод ([текст С1)](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s001). Сначала мы провели математически контролируемое сравнение между одинарным, конкурентным и кооперативным связыванием TF, где мы сохранили все параметры модели одинаковыми в этих различных сценариях и изменили только процесс связывания TF. Мы сделали это, чтобы понять вклад процесса связывания TF в шум экспрессии и избежать путаницы наших результатов вариациями других параметров, которые также могут влиять на шум экспрессии. Мы провели стохастическое моделирование с выбором параметров модели по широкому диапазону значений параметров математически контролируемым образом в отдельных сценариях TF, конкурентных TF и кооперативных сценариях связывания TF. В этом широком диапазоне значений параметров конкурентное и кооперативное связывание TF показало более высокий шум по сравнению с одним регулированием TF ([S8](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s009)и [S9](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s010)Figs).

Мы расширили наш анализ, чтобы количественно оценить шум в случаях сотрудничества или конкуренции между более чем двумя TF, так как многие комплексы инициирования транскрипции могут содержать несколько TF. Конкурентное связывание TF показало более высокий шум по сравнению с одним регулированием TF независимо от количества TF и независимо от изменения силы регулирования среди TF ([рис. 7C и 7D](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g007)). Для кооперативного связывания ТФ увеличение количества ТФ привело к снижению частоты взрывов и снижению среднего уровня белка ([S7 рис.](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s008)). Тем не менее, увеличение числа сотрудничающих TF может пропорционально увеличить время, проводимое в состоянии on, которое мы смоделировали путем снижения отсуточного курса, чтобы сохранить аналогичный средний уровень белка даже при увеличении количества сотрудничающих TF ([рис. 7E](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g007)). Во всех этих сценариях шум был выше при регулировании кооперативными TF по сравнению с одним регулированием TF ([рис. 7F](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g007)). Это говорит о том, что выводы, сделанные из наших моделей, имеют место независимо от количества TF, участвующих в совместной и конкурентной связи.

Перекрытие в сайтах связывания различных TF также может привести к вырождению последовательностей связывания TF для размещения различных мотивов связывания консенсуса различных TF. Такое вырождение может изменить сродство TF к ДНК и привести к шумной транскрипции. Однако мы не видели никакой разницы в вырождении сайтов связывания между промоторными областями генов с высоким и низким уровнем шума экспрессии ([S10 Рис.](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.s011)).

Чтобы дополнительно проверить, соответствуют ли эти результаты какую-либо комбинации значений параметров, при этом единственное условие средней экспрессии белка одинаково в одиночном, конкурентном и кооперативном связывании TF, мы провели выборку пространства параметров Markov-Chain Monte-Carlo (MCMC). Короче говоря, мы сделали инициализацию модели с набором случайных значений параметров для всех параметров модели с последующей оптимизацией параметров, чтобы достичь целевого среднего уровня выражения. На каждом этапе мы проводили стохастическое моделирование более 1000 ячеек в каждой из отдельных, конкурентоспособных и кооперативных связываний TF и рассчитанных средних экспрессий. Как только моделирование достигло целевого среднего уровня выражения, мы также рассчитали значения шума. Мы выбрали пять различных целевых средних уровней выражения для сравнения ([рис. 7G](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g007)), которые были в разумном диапазоне размера и частоты всплесков. Это позволило избежать сравнения шума в состояниях экспрессии, когда ген всегда был включен или всегда выключен. При всех этих средних уровнях экспрессии конкурентное и кооперативное привязка TF показало значительно более высокое значение шума экспрессии по сравнению с одним привязкой TF ([рис. 7G](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen-1010535-g007)).

Обсуждение

Таким образом, благодаря комплексному статистическому анализу мы показали, что процесс связывания факторов транскрипции является наиболее важным фактором шума экспрессии генов. Хотя многие более ранние исследования исследовали молекулярное происхождение экспрессивного шума, большинство из них были сосредоточены на роли последовательности коробок TATA, моделях занятости нуклеосом промоторов и модификациях гистонов. Хотя более ранние работы и наш анализ обнаружили значительную связь между наличием последовательности коробок TATA и шумом экспрессии в дрожжах, такая связь не наблюдалась Ву *и др*. [[34](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref034)] в эмбриональных стволовых клетках человека. Здесь мы показываем, что, несмотря на сильную ассоциацию, наличие/отсутствие последовательности коробок TATA и занятость промоторной нуклеосомы не являются хорошими предикторами экспрессивного шума в дрожжах. Вместо этого наша работа раскрывает важную роль TF в регулировании шума. Мы показываем, что шумные гены, как правило, контролируются большим количеством TF. К ним относится значительная часть TF, которые совместно связываются с промоторским регионом. Кроме того, увеличение числа регулирующих TF может привести к увеличению совпадения между сайтами связывания TF, что может привести к конкуренции между TF за связывание с одним и тем же регионом промоутера. Это может привести к временным, а также к межиндивидуальным изменениям в связывании TF, тем самым увеличивая шум.

Более ранняя работа показала, что увеличение количества сайтов фактора транскрипции может увеличить шум экспрессии генов [32]. Это исследование показало, что количество сайтов связывания TF и их интервал могут влиять на шум, как и вставка неблагоприятного элемента нуклеосомы. Они также отметили, что более крупные и плотные кластеры сайтов связывания TF приводят к более высокому шуму, что говорит о том, что конкуренция между TF может привести к более высокому шуму экспрессии. Однако это исследование было сосредоточено только на промоторах, содержащих TATA, и сайтах связывания двух активаторов, GCN4 и LEU3. Таким образом, выводы, сделанные из их работы, могут быть неприменимы к промоторам, отличным от TATA, или к широкому спектру факторов транскрипции, присутствующих в дрожжах. Тем не менее, их работа предоставила некоторые экспериментальные доказательства влияния конкурентного связывания TF на шум экспрессии. Кроме того, их экспериментальный дизайн может стать шаблоном для дальнейших экспериментов, чтобы понять, как кооперативное связывание TF может привести к более высокому шуму экспрессии, который до сих пор не исследовался.

Другое исследование показало, что конкуренция между взаимодействующими партнерами белка, связывающего TATA, влияет на шум [[38](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref038)], но это было ограничено белком связывания TATA (TBP). Напротив, наш анализ рассмотрел все возможные промоторные последовательности и факторы транскрипции и показал, что процесс связывания факторов транскрипции является ключевым фактором шума экспрессии. Таким образом, мы описываем общий молекулярный механизм генерации шума, который не зависит от какого-либо конкретного TF или какой-либо конкретной последовательности промоторов.

Более раннее исследование Фора *и др*. [[35](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref035)] проанализировало шум экспрессии в эмбриональных стволовых клетках мыши и рассмотрело роль нескольких молекулярных особенностей в регуляции шума. Они проанализировали роль паттернов модификации гистонов, областей суперусиления, а также функции промоторной последовательности, такие как сайты инициирования транскрипции и наличие мотива коробки TATA. Путем количественной оценки размеров эффектов они наблюдали связь некоторых из этих признаков с шумом экспрессии. Кроме того, они оценили относительную важность этих молекулярных особенностей, как по отдельности, так и в комбинациях, при классификации генов по категориям высокого или низкого уровня шума. Однако авторы не сообщили о объясненной доле вариаций или прогнозируемых значениях R2.Напротив, наша интегрированная статистическая модель конкретно сообщала о прогностических возможностях молекулярных признаков, как индивидуально, так и в сочетании. Это позволило нам количественно оценить, насколько хорошо наличие или отсутствие молекулярной функции по отдельности или в сочетании с другими признаками гена или его промотора может предсказать шум экспрессии этого гена.

Интересное наблюдение из наших интегрированных моделей заключалось в том, что производительность модели прогнозирования мРНК-шумов аналогична производительности модели прогнозирования белкового шума. И это несмотря на то, что измерение экспрессии мРНК в отдельных клетках технически сложнее и менее точно по сравнению с измерением экспрессии белка в отдельных клетках, так как количественная оценка уровней мРНК в отдельных клетках страдает от низкой эффективности захвата и эффектов выборки. Мы считаем, что есть две возможные причины, которые могут объяснить сопоставимую производительность модели прогнозирования мРНК-шумов и модели прогнозирования белкового шума. Во-первых, у нас было значительно больше данных в наборе данных о шуме мРНК (~5500 генов) по сравнению с набором данных о белках (~2800 генов). Это означает, что прогнозная модель шума мРНК имела значительно больший набор тренировочных данных, который помог построить модель с производительностью, аналогичной модели прогнозирования белкового шума. Во-вторых, наши функции в значительной степени сосредоточены на синтезе мРНК и скорости распада, стабильности мРНК, которые непосредственно влияют на шум экспрессии на уровне мРНК, но только косвенно влияют на шум белка.

Хотя наша интегрированная модель могла предсказать значительную долю изменения шума, тем не менее, все еще существовала большая доля шума, которая не могла быть объяснена нашей моделью. Это может быть связано с несколькими причинами. Во-первых, возможно, что в нашей модели не было учтено несколько других молекулярных особенностей, которые могут регулировать шум. Некоторые из этих молекулярных особенностей могут быть до сих пор неизвестны. Во-вторых, существует случайность в молекулярных процессах, происходящих внутри клетки, и шум экспрессии также может меняться со временем. Таким образом, наш расчет шума за один момент времени также может повлиять на прогностическое влияние. В-третьих, экспериментальные данные о молекулярных особенностях, рассмотренные в этом исследовании, были получены из разных исследовательских групп и в разных условиях роста. Это может повлиять на прогнозную способность нашей модели. В-четвертых, мы понимаем, что некоторые особенности, такие как уровень занятости нуклеосом и модели модификации гистонов, являются динамичными по своей природе и могут меняться со временем. Поскольку мы моделировали эти функции с использованием наборов данных, полученных за один момент времени, мы, возможно, полностью упустили вклад динамического характера этих функций в регулирование шума. Кроме того, наблюдалось, что условия роста, скорость роста клеток и клеточный цикл влияют на шум экспрессии генов [[74](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref074)-[76](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref076)]. Таким образом, объединение данных о молекулярных особенностях во многих наборах данных без учета этих переменных может потенциально повлиять на прогностическое влияние. Наконец, что касается моделирования, мы использовали модель линейной регрессии для нашего анализа, которая способна фиксировать линейные тенденции в данных, но может пропустить нелинейные ассоциации, присутствующие в данных. Это может повлиять на производительность модели. Однако для противодействия этому недостатку мы также использовали случайную модель леса, которая способна фиксировать нелинейные тенденции в наборе данных.

Таким образом, наши выводы дают шаг вперед для прогнозирования шума экспрессии. Недавний взрыв геномных данных привел к геномной характеристике сайтов связывания ТФ в различных организмах. Кроме того, с увеличением доступности карт занятости нуклеосом в масштабах всего генома, моделей модификации гистонов и трехмерных данных о конфигурации генома наше исследование обеспечивает основу для построения интегрированных моделей шума экспрессии генов в других организмах в будущем. Стохастические вариации молекулярных процессов повсеместно распространены в клетках в биологических системах и имеют серьезные последствия для заболеваний человека. Таким образом, повышенная способность прогнозировать изменения в биологических процессах будет чрезвычайно полезна для количественной оценки степени неоднородности клеточных признаков и фенотипов.

Способы

**Расчет шума экспрессии для отдельных генов**

Значения шума для отдельных генов в дрожжах на уровне белка были получены от Ньюмана *и др*. [19], а значения DM в среде YPD использовались для анализа экспрессивного шума. Значения DM в среде YPD сильно коррелировали со значениями DM, полученными в среде SD. Значения шума всех генов на уровне мРНК были рассчитаны на основе одноклеточных данных РНК-сек, предоставленных Nadal-Ribelles *et al*.[[49](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010535#pgen.1010535.ref049)] следующим образом. Короче говоря, для каждого гена коэффициент вариации (CV) был рассчитан на основе его средней экспрессии и значения стандартного отклонения. Были сделаны различные полиномиальные соответствия для CV против логарифмических преобразованных средних значений выражения, и было выбрано наилучшее соответствие. Было обнаружено, что многочлен порядка 5 дает наилучшую посадку. Среднее скорректированное значение шума гена было получено путем расчета вертикального расстояния между значением CV и наилучшим образом подходящим кривой. Для оценки влияния выбросов на фитинг также были оценены 95% доверительных интервалов для посадок, которые были построены вместе с установленной линией.