

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В. ЛОМОНОСОВА»

---

ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ  
КАФЕДРА МЕДИЦИНСКОЙ ФИЗИКИ

БАКАЛАВРСКАЯ РАБОТА

**Молекулярные механизмы ложных генетических  
взаимодействий в библиотеке нокаутных дрожжевых  
штаммов**

Выполнил студент  
440 группы  
Егоров Артём Арменович

---

**Научные руководители:**

Иванов Павел Сергеевич  
Кандидат физико-математических наук  
Старший научный сотрудник  
Кафедры медицинской физики

---

Дмитриев Сергей Евгеньевич  
Кандидат биологических наук  
Старший научный сотрудник  
НИИ физико-химической биологии МГУ

---

Кулаковский Иван Владимирович  
Кандидат физико-математических наук  
Доктор биологических наук  
Ведущий научный сотрудник  
Институт молекулярной биологии РАН

---

Допущена к защите

Зав. кафедрой \_\_\_\_\_

Москва

2020

# Содержание

<b>Список используемых сокращений</b>	<b>3</b>
<b>Введение</b>	<b>4</b>
<b>Обзор литературы</b>	<b>6</b>
<i>Saccharomyces cerevisiae</i> как модельный организм . . . . .	6
Проект SGDP по созданию делеционной коллекции . . . . .	6
Применение нокаутной коллекции . . . . .	8
Генетические взаимодействия . . . . .	9
Глобальная сеть генетических взаимодействий . . . . .	11
Эффект соседствующего гена . . . . .	13
Трансляция у эукариот и её регуляция . . . . .	14
Инициация трансляции . . . . .	14
Терминация трансляции . . . . .	16
Реинициация трансляции и uORF . . . . .	18
Полиаденилирование и уровень экспрессии . . . . .	19
Рибосомный профайлинг . . . . .	20
<b>Цель и задачи исследования</b>	<b>21</b>
<b>Материалы и методы</b>	<b>22</b>
Исходные данные . . . . .	22
Методы . . . . .	22
Обработка результатов секвенирования . . . . .	22
Анализ дифференциальной экспрессии . . . . .	22
Модель для расчёта дифференциальной экспрессии . . . . .	23
Приложение инструмента svist4get для анализа . . . . .	25
<b>Результаты</b>	<b>26</b>
Уровень экспрессии кассеты KanMX . . . . .	26
Кассета KanMX смещает сайты старта транскрипции соседних генов, приводя к появлению блокирующих uORF . . . . .	27
Кассета KanMX нарушает посттранскрипционную модификацию 3' конца транскриптов соседних генов . . . . .	29
Смещения TSS и активация альтернативных PAS, как молекулярный механизм NGE . . . . .	31
<b>Заключение</b>	<b>34</b>
<b>Выводы</b>	<b>36</b>
<b>Список литературы</b>	<b>37</b>



## Список используемых сокращений

bp	– base pairs
FDR	– false discovery rate
FC	– fold change
KO	– knocked out
NB	– negative binomial distribution
NGE	– neighboring gene effect
PAS	– polyadenylation site
puORF	– polymorphic upstream open reading frame
RNA-Seq	– ribonucleic acid sequencing
QL	– quasi-likelihood
Ribo-Seq	– ribosome profiling
RO	– ribosome occupancy
SGDP	– <i>saccharomyces</i> genome deletion project
SGA	– systematic genetic analysis
TSS	– transcription start site
uORF	– upstream open reading frame
WT	– wild type
АТФ	– аденозинтрифосфат
ВИЧ	– вирус иммунодефицита человека
ГТФ	– гуанозинтрифосфат
ДНК	– дезоксирибонуклеиновая кислота
мРНК	– матричная рибонуклеиновая кислота
Мет-тРНК	– инициаторная метионил-тРНК
НТО	– нетранслируемая область
ОРС	– открытая рамка считывания
ПЦР	– полимеразная цепная реакция
тРНК	– транспортная рибонуклеиновая кислота

# Введение

Гаплоидный геном дрожжей *Saccharomyces cerevisiae* содержит около 6000 генов [1]. Два десятилетия назад в рамках программы SGDP (англ. *Saccharomyces Genome Deletion Project*) была создана библиотека нокаутных штаммов [2, 3]. Целью этого проекта было выяснение функций отдельных генов посредством фенотипического анализа соответствующих мутантов. Каждый ген в этой библиотеке был заменен кассетой KanMX [4], содержащей ген устойчивости к антибиотику G418 под контролем сильного TEF-промотора. Терминатор TEF, а также соседние последовательности, в том числе искусственные уникальные теги и участки из исходного вектора, вставляются в хромосому вместе с геном KanMX.

Создание нокаутной библиотеки было революционным в области молекулярной генетики эукариот. В частности, многие гены были функционально охарактеризованы путём анализа фенотипа и генетических взаимодействий соответствующих нокаутных штаммов [5, 6, 7].

Тем не менее, в ходе процедуры производства нокаутного штамма генетический локус претерпевает сильные изменения. Поскольку дрожжи имеют компактный геном с очень короткими расстояниями между генами, введение сильно экспрессирующейся генетической кассеты, в принципе, может существенно изменить профиль транскрипции прилегающих областей, влияя на экспрессию соседних генов. Ранее было проведено исследование влияния соседних генов (англ. *neighboring gene effect*, NGE) на генетические взаимодействия [8, 9, 10]. Однако молекулярные механизмы такого воздействия не анализировались.

В данной работе, используя данные рибосомного профайлинга для нескольких десятков нокаутных штаммов *S. cerevisiae*, мы проанализировали транскрипционные и трансляционные изменения, индуцированные кассетой KanMX в модифицированных геномных локусах. Во многих случаях мы обнаружили существенные изменения в уровне экспрессии соседних генов, происходящие как на уровне транскрипции, так и на уровне трансляции. Эти изменения включают смещение сайтов старта транскрипции и активацию альтернативных сайтов полиаденилирования. Наиболее существенные изменения наблюдались, когда удаляемый ген располагался в направлении “голова к голове” (5'-конец к 5'-концу) относительно соседнего гена. В этом случае смещение старта транскрипции соседнего гена в 5'-концевом направлении удлиняло 5'-нетранслируемую область мРНК и влияло на её трансляцию через появление новых стартовых кодонов, предшествующих основному. При расположении генов “хвост к хвосту” (3'-конец к 3'-концу) была обнаружена активация альтернативных сайтов полиаденилирования в транскрипте сосед-

него гена и, как следствие, укорочение 3'-нетранслируемой области.

Наши наблюдения описывают взаимодействие кассеты KanMX с соседними генами и дают представление о молекулярных механизмах, вовлеченных в феномен NGE. Мы предполагаем, что во многих случаях NGE нокаутных штаммов вызван дефектами на трансляционном уровне, аналогичным тем, которые были выявлены в нашем исследовании.

# Обзор литературы

## *Saccharomyces cerevisiae* как модельный организм

Дрожжи *Saccharomyces cerevisiae* имеют долгую историю взаимодействия с человеком, будучи первым одомашненным организмом [5]. Они давно признаны одной из ключевых моделей эукариотической клетки. “Дрожжи, благодаря легкости, с которой они могут быть культивированы на искусственных средах, и относительно большим размерам клеток, являются особенно благоприятными объектами для экспериментов. ... Они внесли значительный вклад в развитие общей физиологии и биологии.” Это заявление было сделано в далеком 1920 году [11] и до сих пор не потеряло актуальности. В частности, после идентификации гена в любом организме исследователи обычно стараются найти гомолог в базе данных генома *S. cerevisiae*. Благодаря огромной работе «дрожжевого» исследовательского сообщества нахождение гомолога сразу приносит много полезной информации [12]. Успех обусловлен тем, что, в дополнение к известной полной последовательности генома, *S. cerevisiae* – единственный эукариотический организм, для которого существует полногеномная коллекция генетических нокаутов – штаммов с делециями или мутациями каждого из 6000 генов. Важнейшее её применение – понимание функции генов, генетических взаимодействий и их связи с фенотипом [5].

Концепция проекта по созданию коллекции делеций возникла после секвенирования генома *S. cerevisiae*. Проект секвенирования был осуществлён одним из первых подобных и послужил моделью для многих последующих проектов по расшифровке геномов. Два результата данного проекта оказали огромное влияние на научное сообщество. Во-первых, несмотря на десятилетия усилий, большинство белок-кодирующих генов, предсказанных по последовательности ДНК, были новыми (ранее не идентифицированными в каких-либо экспериментах) [13]. Этот результат, в частности, укрепил амбиции проекта “Геном человека”, положив конец многим опасениям по поводу его ценности. Во-вторых, было обнаружено, что очень много генов дрожжей являются эволюционно консервативными, а это, в свою очередь, выступило поддержкой идеи, что сравнительный анализ геномов организмов поможет аннотировать геном человека. И действительно, около 1000 болезней человека ассоциированы с генами, которые имеют функциональные ортологи в геноме *S. cerevisiae* и могут быть изучены в этом организме [5, 14].

## Проект SGDP по созданию делеционной коллекции

К 1996 году геном дрожжей *Saccharomyces cerevisiae* был полностью секвенирован. В результате было определено, что последовательность ДНК размером 12068 тпо (тысяч пар оснований), записанная на 16 хромосомах,

содержит: 5885 потенциально кодирующих белки генов; 140 генов, кодирующих рибосомальные РНК (рРНК); 40 генов малых ядерных РНК (мяРНК) и 275 генов транспортных РНК (тРНК) [1]. Соответственно, следующей задачей было определение биологических функций этих генов.

Чтобы приступить к решению этой задачи, в рамках проекта SGDP была создана коллекция мутантных штаммов. Для этого использовалась стратегия делеции генов, основанная на полимеразной цепной реакции (ПЦР) [15]. Открытые рамки считывания (ОРС) были заменены от стартового кодона до стоп кодона кассетой KanMX, каждая из которых была сконструирована с использованием двух последовательных ПЦР (Рис. 1).

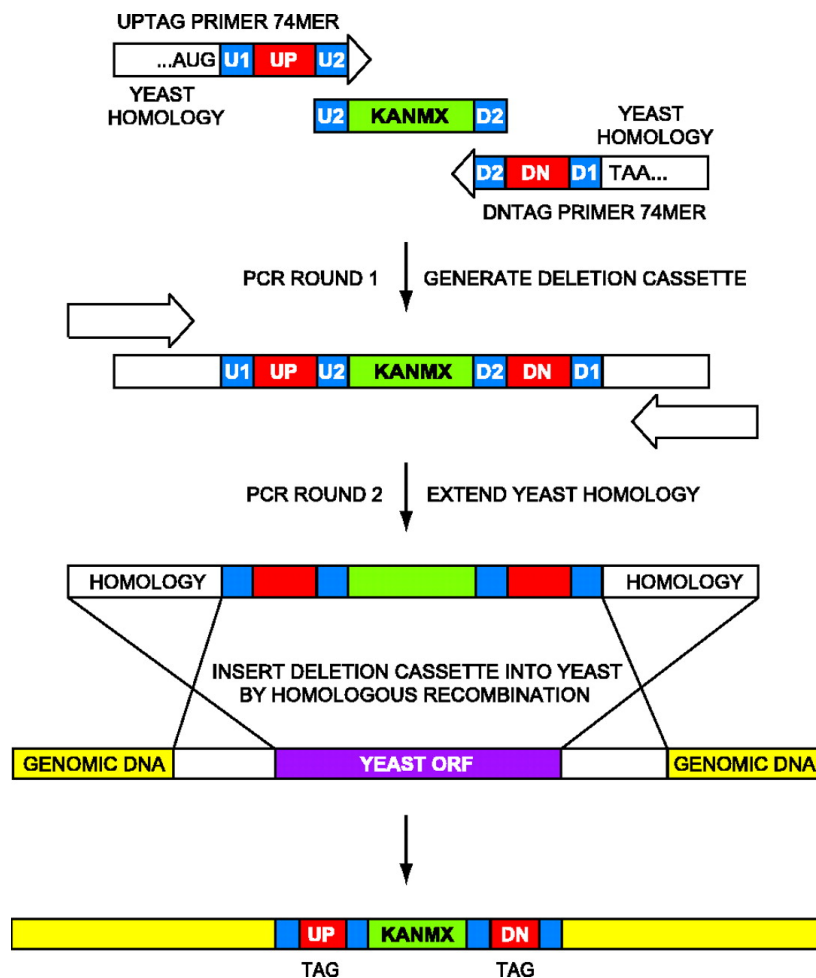


Рис. 1. Стратегия делеции ОРС. U1,UP,U2 – теги, предшествующие кодирующей части вставки (изображена зеленым цветом), в том числе уникальные для каждого нокаута. D1,DN,D2 – аналогичные теги на 3' конце кассеты [4].

В результате каждый ген в этой библиотеке был заменен кассетой KanMX, содержащей ген устойчивости к антибиотику G418 под контролем сильного TEF-промотора. Терминатор TEF, а также соседние последовательности, в том числе искусственные уникальные теги и участки из исходного вектора, вставляются в хромосому вместе с геном KanMX [4] (Рис. 2).



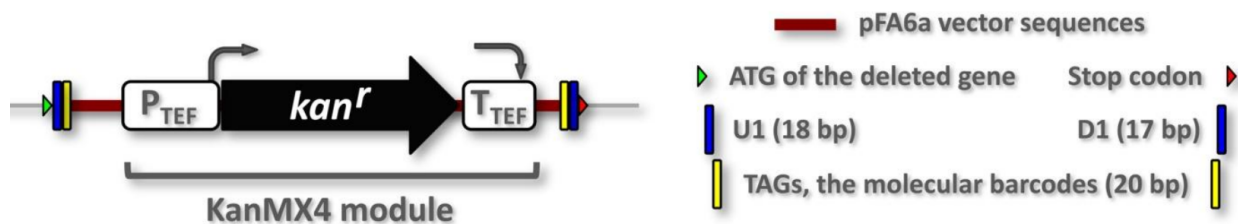


Рис. 2. Строение делеционного модуля.

## Применение нокаутной коллекции

Первое описание проекта появилось в 1999 году [3], когда было получено чуть более трети штаммов.

Основными результатами были следующие:

- 1) 17 % генов являются критически важными (англ. essential), т.е. при их нокауте клетки не растут.
- 2) Только около половины из них были ранее известны.
- 3) Те гены, которые не являются критически важными для роста, с большей вероятностью, по сравнению с критически важными, имеют гомологов.

О завершении проекта было объявлено в 2002 году [2]. В результате стало известно, что 18.7 % из 5916 ОРС являются необходимыми для роста, при этом 15 % жизнеспособных штаммов растут медленно. Определено, что корреляция между генами, необходимыми для приспособленности в определенных условиях, и теми генами, транскрипция которых в этом состоянии повышена, незначительна. В то время это было неожиданным результатом, потому что считалось, что повышенная экспрессия генов сообщает о необходимости гена в данном состоянии. Теперь же у нас есть более глубокое понимание регуляции экспрессии. Например, сейчас известно, что реакция на стресс включает в себя множество посттранскрипционных событий, которые обеспечивают более быстрый ответ. Это осуществляется различными механизмами регуляции, например, с помощью коротких рамок считывания, предшествующих основному стартовому кодону ОРС (англ. upstream ORF)[16], образования стресс-гранул [17], блокирования выхода рибосомальных субъединиц через ядерную пору и многих других [18].

Создание библиотеки позволило проводить полногеномные скрининги для выявления генов, участвующих в клеточных процессах. Несколько ранних применений скрининга отличались своей строгостью и установили стандарт для последующих исследований. Например, работа по выявлению генов, которые регулируют деление клеток [19], а также поиск генов, необходимых для устойчивости к токсину “K1 killer toxin” [20]. Эти два исследования выявили

несколько важных моментов. Во-первых, было обнаружено много генов, ранее не связанных с этими широко изученными фенотипами, что подчеркивает низкую чувствительность и небольшой охват стандартных генетических методов. Во-вторых, многие гены, аннотированные к одной функции, оказались многофункциональными и вовлеченными в несколько разных клеточных процессов [5].

С помощью коллекции также стало доступно обнаружение целевых и нецелевых эффектов конкретных соединений, в первую очередь лекарственных, путем тестирования библиотеки делеционных мутантов на гиперчувствительность. Например, так было обнаружено, что тамоксифен – терапевтический препарат для лечения рака молочной железы – нарушает гомеостаз кальция, а фосфатидилсерин является мишенью для папуамида В – липопептида с активностью против ВИЧ [21].

Одни из наиболее интересных применений коллекции нокаутных штаммов посвящены изучению нейродегенеративных расстройств. В них проводились полногеномные скрининги мутантов, обнаружившие штаммы, чувствительные к действию какого-либо токсического метаболита или белка. Такой подход был применен для изучения действия альфа-синуклеина, вовлеченного в болезнь Паркинсона [22], а также болезней Хантингтона [23], Крейтцфельда – Якоба и других нарушений [24, 25].

## Генетические взаимодействия

Отдельно хочется выделить работы, в которых рассматриваются множественные мутации, – такие исследования позволяют глубже понять взаимосвязи генотипа и фенотипа. Еще в далеком 1909 году Уильям Бейтсон приписал эпистаз специфическому типу генетических взаимодействий, при котором одна мутация маскирует эффекты другой мутации, так, что двойной мутант напоминает более экстремальный одиночный мутант, а не ожидаемый комбинированный эффект обеих мутаций. Рональд Фишер обобщил этот термин, чтобы включить любой мультилокусный мутантный эффект, который отклоняется от ожидаемого фенотипа, связанного с объединением соответствующих индивидуальных локусов [26]. Считается, что генетические взаимодействия лежат в основе различных аспектов биологии, включая эволюцию пола, видообразование и сложные заболевания [27].

Генетическое взаимодействие может быть количественно оценено путем измерения какого-либо параметра, характеризующего одно- и двухмутантные фенотипы, и расчёта коэффициента взаимодействия, отражающего любое отклонение от ожидаемого комбинированного эффекта одиночных мутаций. Для фенотипов, связанных с измененной приспособленностью или скоростью

роста, наиболее широко принятая модель объединяет фенотипы, связанные с мутациями в независимых генах, используя мультипликативное уравнение [28]. В случае отсутствия генетического взаимодействия ожидается, что фенотип двойного мутанта будет эквивалентен произведению фенотипов, связанных с соответствующими одиночными мутантами. Следовательно, как показано на Рис. 3, генетическое взаимодействие оценивается путем выявления двойных мутантов, фенотип которых отличается от ожидаемого значения.

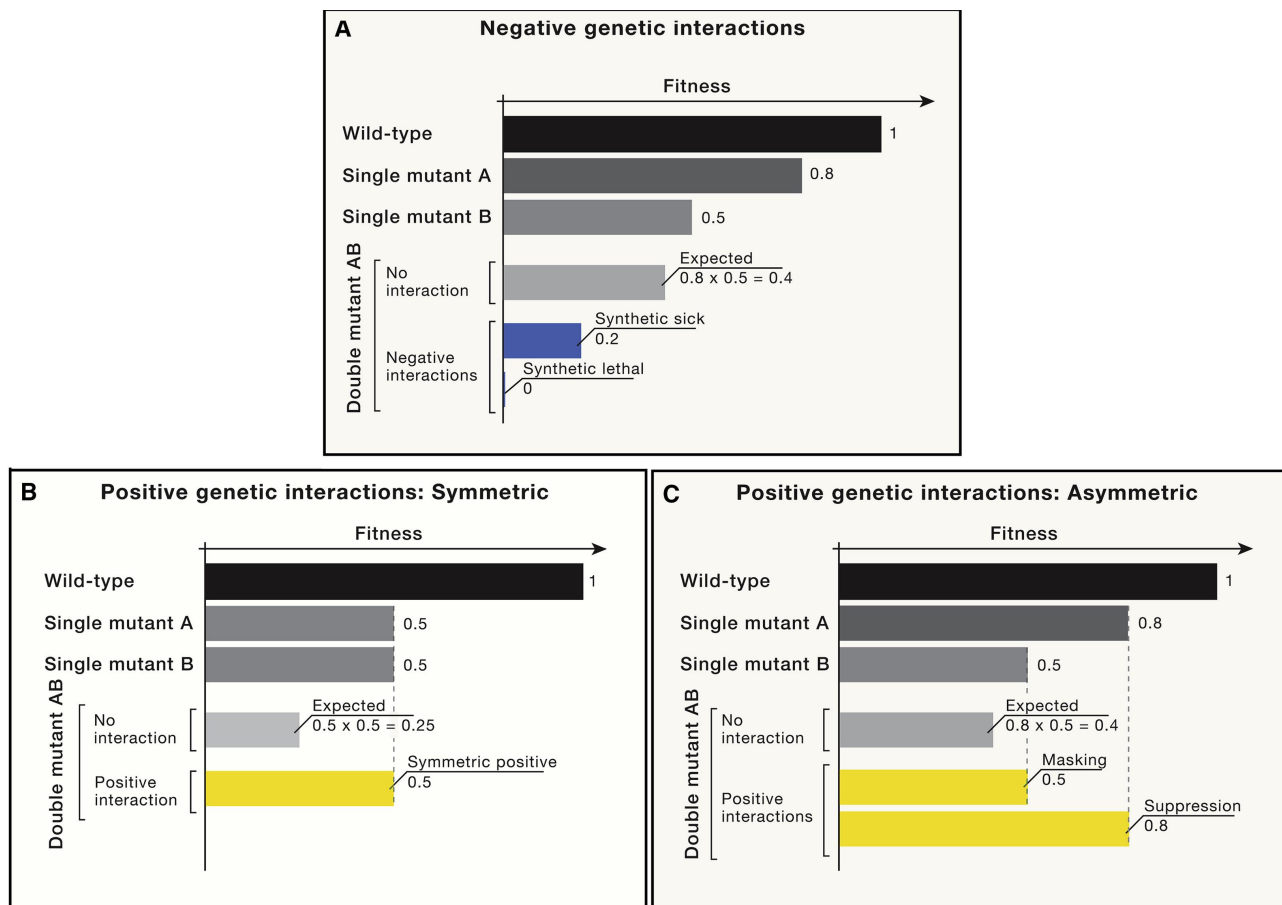


Рис. 3. Графическое представление количественной оценки генетических взаимодействий. А – негативные генетические взаимодействия, В– положительные симметричные, С– положительные ассиметричные [28].

В принципе, любой фенотип может быть использован для обнаружения генетического взаимодействия, но скорость роста клеток, которая может быть легко измерена и количественно определена, интегрирует клеточную физиологию и является функционально информативной для изучения.

Варианты генетического взаимодействия можно классифицировать следующим образом: (Рис. 3) [28]

1) Негативное генетическое взаимодействие:

Генетическое взаимодействие, при котором двойные мутанты демонстри-

руют более низкие показатели роста, чем ожидалось, включая уменьшение или полную нежизнеспособность двойного мутанта

2) Положительное симметричное генетическое взаимодействие:

Генетические взаимодействия, которые характеризуют менее тяжелые последствия двойной мутации, чем ожидалось. Например, делеция дополнительных генов, кодирующих компоненты того же ферментативного или сигнального пути, либо белкового комплекса (который не является критически необходимым для роста), в котором участвует первый делетированный ген, часто не приводит к усилению фенотипа. Таким образом, фенотипы одиночных и двойных мутантов, включающих гены белкового комплекса, будут количественно неразличимы. Это означает, что двойной мутант не соответствует ожидаемой мультипликативной модели, а показывает симметричный положительный вариант генетического взаимодействия.

3) Положительное асимметричное генетическое взаимодействие:

Положительное генетическое взаимодействие также может быть асимметричным – в случае, когда сила фенотипического эффекта варьируется между одиночными и ожидаемым у двойных мутантов. Асимметричный подкласс включает в себя так называемые “маскирующие взаимодействия”, где приспособленность двойного мутанта лучше, чем ожидалось, и напоминает самую низкую приспособленность из одиночных мутантов. А также содержит “супрессирующие взаимодействия”, которые могут быть результатом комбинации двух мутаций с потерей функции, но часто включает специализированные аллельные комбинации и, таким образом, является относительно небольшим подмножеством.

## **Глобальная сеть генетических взаимодействий**

Объединение результатов крупномасштабных исследований по изучению генетических взаимодействий позволяет не только получать информацию о биологических путях и функциях генов, но а также обеспечивать фундаментальное понимание генетической архитектуры, лежащей в основе взаимосвязи генотипа и фенотипа. Для получения таких данных используется SGA (англ. Systematic Genetic Analysis) – автоматизированная методика для получения большого массива данных [29]. Как уже отмечалось, результаты подобного анализа можно разделить на две категории: отрицательные и положительные генетические взаимодействия. Отрицательные взаимодействия представляют особый интерес, поскольку они имеют тенденцию связывать функционально родственные гены, которые воздействуют на один и тот же существенный биологический процесс [30]. И, с другой стороны, положительные взаимодействия могут связывать гены, кодирующие элементы одного и

того же белкового комплекса, поскольку, как только функция комплекса нарушается мутацией одного компонента, фенотип не может уже быть ухудшен удалением дополнительных компонент, но иногда может быть улучшен благодаря снятию токсического эффекта от неполного комплекса. [31]. Однако подавляющее число положительных взаимодействий не связывают гены в рамках одного и того же пути или даже функционально связанных генов, а вместо этого имеют тенденцию отображать более общие регуляторные цепи, связанные с дефектами клеточного цикла или клеточного протеостаза. Кроме того, набор отрицательных и положительных генетических взаимодействий для данного гена обеспечивает количественную меру функции гена, а подмножества генов с очень похожими профилями генетического взаимодействия часто принадлежат к одному и тому же биологическому процессу, пути и/или белковому комплексу [32].

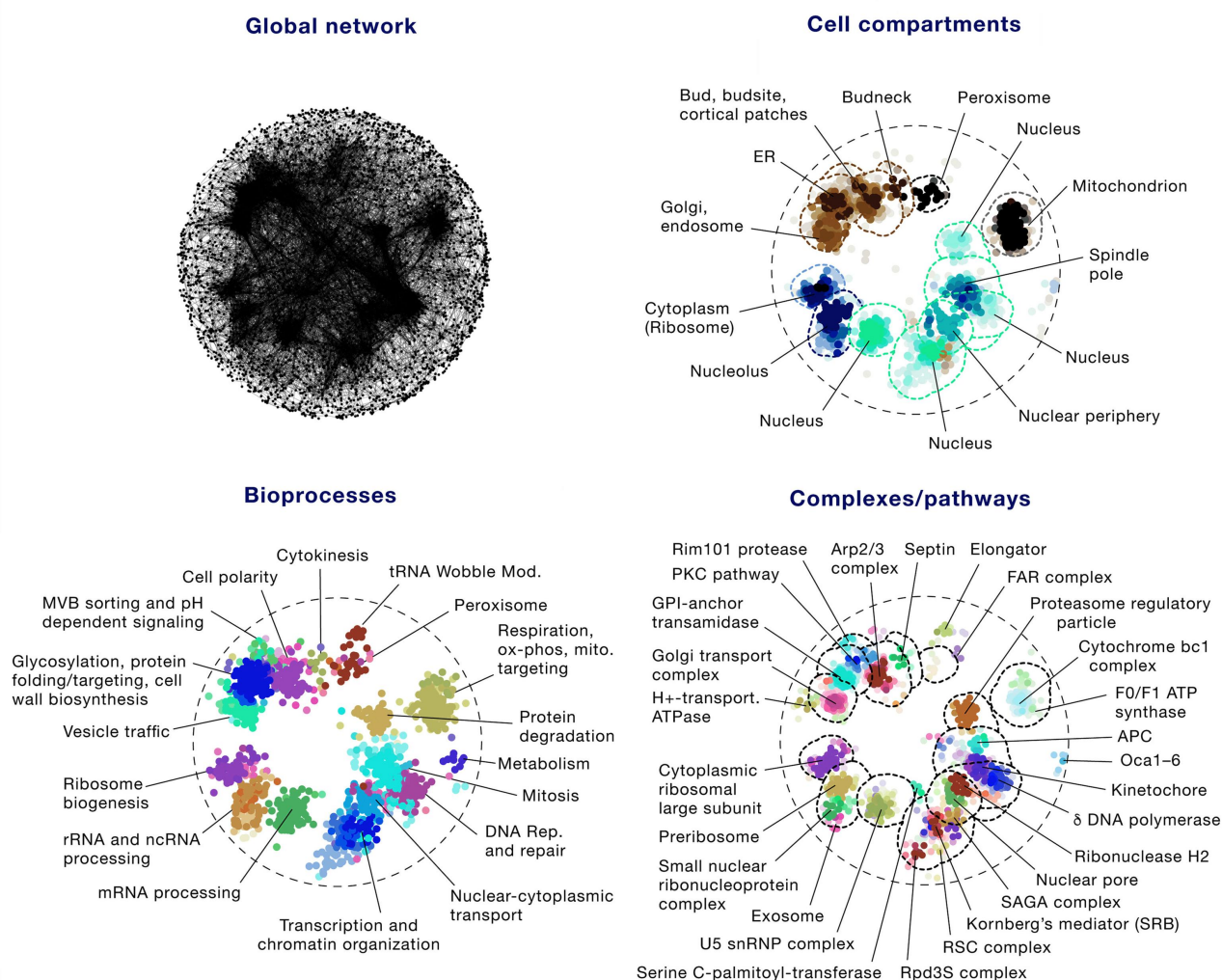


Рис. 4. Кластеризация глобальной сети генетических взаимодействий [32].

На основании этих данных была построена глобальная сеть генетических взаимодействий для *Saccharomyces cerevisiae*, состоящая почти из миллиона связей (Рис. 4). Это исследование не только выявило общие свойства ге-

нетических взаимодействий, но и показало, что глобальная сеть группирует гены в иерархическую модель клеточной функции. Например, при самом высоком разрешении сети подобия профиля, гены группируются в модули, соответствующие белковым процессам и путям. На промежуточном уровне разрешения сети группы объединяются вместе для определения конкретных биологических процессов. Наконец, на самом общем уровне разрешения сети, кластеры генов, представляющие биологические процессы, группируются в более крупные модули, соответствующие конкретным клеточным компонентам. Эта биологическая организация генов обеспечивает мощный ресурс для предсказания функции ранее неаннотированных генов [32, 8] .

## Эффект соседствующего гена

Одно из очевидных предположений при использовании мутантной коллекции состоит в том, что любой фенотип, наблюдаемый в конкретном штамме, проистекает из специфической мутации, которая контролируемо создана исследователем. Однако, как было показано [10], созданная делеция может влиять также и на экспрессию соседних генов. Данный феномен получил название “Эффект соседствующего гена” (англ. Neighboring Gene Effect, NGE).

Прямым результатом NGE является неверная функциональная аннотация гена: мутация в нём ассоциируется с фенотипом, который не связан с этим геном, а связан с геном-соседом. При этом сам соседний ген, ответственный за фенотип, по какой-то причине может оказаться, наоборот, не соотнесён с этим фенотипом. Оба гена также показывают схожий профиль генетических взаимодействий с другими генами, что позволяет делать вывод о том, что они относятся к одному и тому же биологическому процессу, пути и/или белковому комплексу. Очевидно, что в случае NGE данный вывод будет неверным.

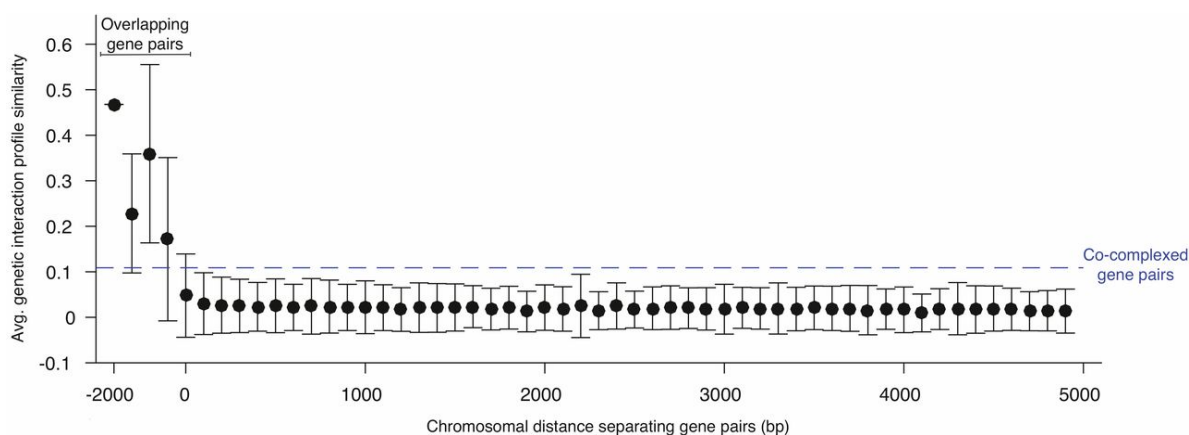


Рис. 5. Зависимость схожести генетического взаимодействия соседних генов от расстояния между ними. Отрицательные значения соответствуют физическому перекрыванию.

В ходе нескольких исследований было определено, что NGE имеет место в 7-15% результатах отрицательных генетических взаимодействий [8, 9, 10]. Кроме того, наблюдается зависимость между расстоянием между генами и вероятностью NGE – наиболее опасны ситуации при расстоянии до 350-400 bp [8, 9] (Рис. 5). При этом в силу компактности генома *S. cerevisiae* такие ситуации довольно часты.

Эти результаты имеют серьезные последствия для системной биологии в связи с широтой использования делеционной коллекции, освещенной выше. При этом молекулярные механизмы NGE оставались неизвестными.

## Трансляция у эукариот и её регуляция

Прежде чем перейти к результатам нашего исследования, описывающим молекулярные механизмы NGE, рассмотрим трансляцию и некоторые способы регуляции экспрессии генов, необходимые для последующего изложения.

Синтез белка – один из ключевых процессов в клетке, состоящий из нескольких стадий – инициации, элонгации, терминации и рециклинга. Каждая стадия регулируется множеством факторов. В нашем случае особого внимания заслуживают стадии инициации и терминации.

### Инициация трансляции

Инициация трансляции мРНК происходит по механизму “сканирующей рибосомы” [33]. Данный процесс можно также разбить на этапы [34, 35, 36]:

- 1) Формирование тройственного комплекса:  
Происходит активация фактора инициации трансляции eIF2 путем связывания ГТФ и последующее связывание инициаторной метионил-тРНК (Met-тРНК) с ГТФ-формой eIF2.
- 2) Формирование 43S преинициаторного комплекса:  
С 40S субъединицей рибосомы связывается тройственный комплекс, а также факторы инициации eIF1, eIF1A, eIF5 и мультисубъединичный фактор eIF3.
- 3) Образование 48S инициаторного комплекса:  
43S комплекс связывается с 5' концом мРНК, ассоциированным с группой факторов eIF4, а также, вероятно, с поли-А связывающим белком РABP, который может замыкать мРНК в кольцо благодаря взаимодействию с поли-А хвостом и кэп-связывающим комплексом eIF4F.
- 4) Сканирование 5' нетранслируемой области мРНК:  
Малая субъединица рибосомы вместе с факторами движется вдоль мРНК

в направлении от 5' к 3', сканируя последовательность на наличие стартового кодона. На этом этапе факторы eIF1 и eIF1A связаны с Р- и А-сайтами рибосомы, соответственно. Фактор eIF1 обеспечивает точность узнавания, а фактор eIF5 является стимулятором ГТФазной активности eIF2, однако до момента узнавания AUG-кодона он неактивен (точнее, гидролиз ГТФ, по-видимому, происходит ещё в начале сканирования, но высвобождение фосфата возможно только после узнавания AUG).

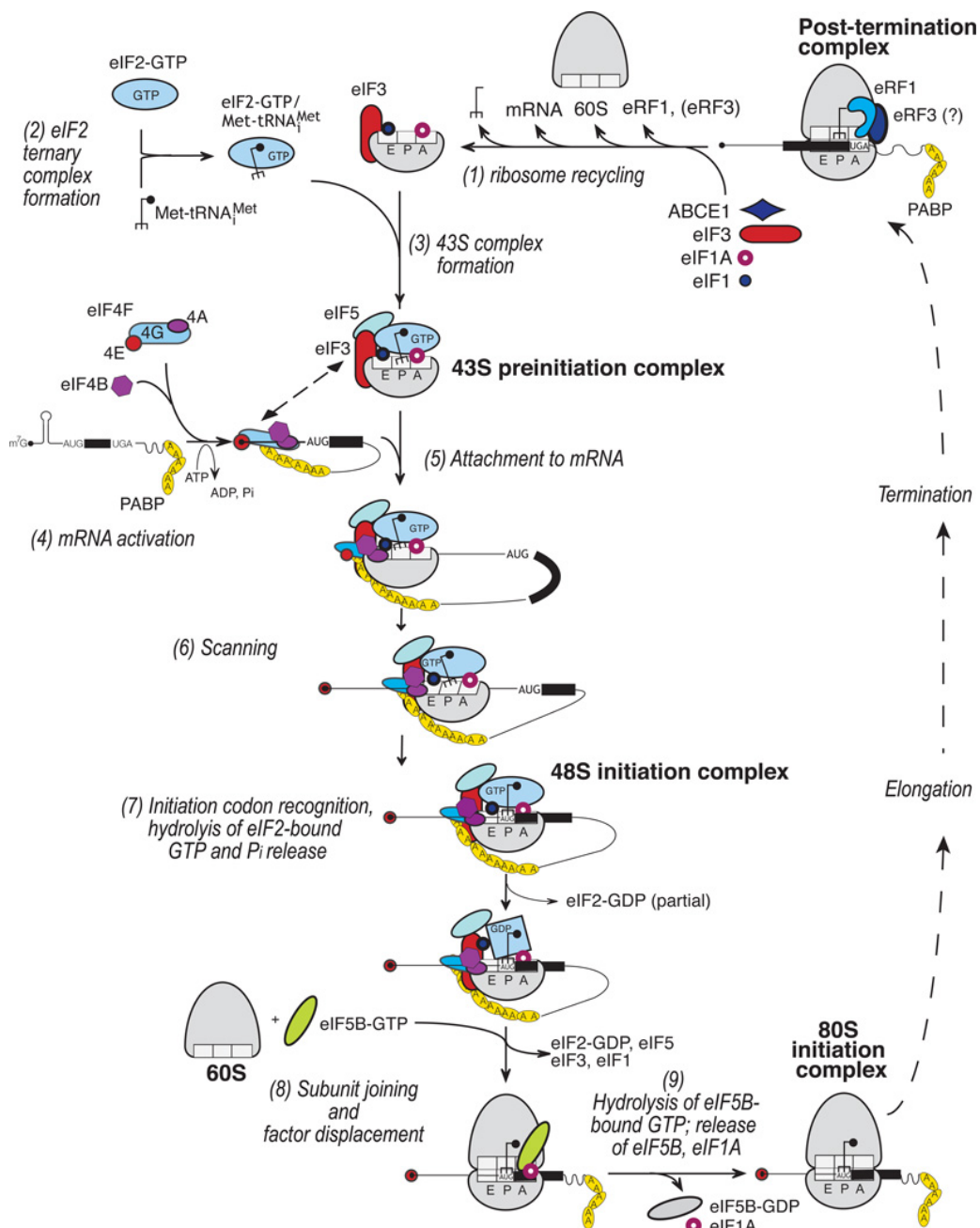


Рис. 6. Модель канонической эукариотической инициации трансляции [35].

### 5) Формирование 80S рибосомы:

Когда в Р-сайте рибосомы оказывается AUG кодон в подходящем нуклеотидном контексте, кодон-антикодоновое взаимодействие между стар-



товым кодоном и антикодоном Мет-тРНК приводит к вытеснению eIF1 из Р-сайта и изменению конформации рибосомы. Далее из связанной с eIF2 молекулы ГТФ, ранее гидролизованной под действием eIF5, освобождается фосфат, после чего ГДФ-форма eIF2, а также eIF5 освобождаются из комплекса. С комплексом связывается фактор eIF5B, который способствует присоединению 60S субъединицы рибосомы. При этом происходит сначала диссоциация всех оставшихся факторов, кроме eIF1A и самого eIF5B, а по завершении формирования 80S рибосомы из комплекса уходят и они.

Свободный А-сайт связывает первую элонгационную аминоксил-тРНК в форме её тройственного комплекса с eEF1A и ГТФ. Далее происходит расщепление ГТФ, освобождение eEF1A из рибосомы, что дает возможность осуществлению реакции транспептизации. Образование первой пептидной связи – это уже первый акт элонгации полипептида, за которым следует первая транслокация. Далее рибосома продолжает элонгационный цикл, состоящий из следующих друг за другом стадий: связывания аминоксил-тРНК, транспептизации и транслокации.

### **Терминация трансляции**

Синтез продолжается до тех пор, пока в А-сайте рибосомы не окажется стоп-кодон. Далее происходит высвобождение синтезированного полипептида, осуществляемое работой двух факторов терминации – eRF1 и eRF3 [37].

Фактор eRF1 состоит из 3 доменов и имеет пространственное сходство с тРНК. N-концевой домен eRF1 в комплексе с рибосомой отвечает за специфичное распознавание стоп-кодонов [38]. Средняя часть eRF1 осуществляет гидролиз пептидил-тРНК связи, в результате которого происходит высвобождение синтезированного полипептида [39]. С-концевой домен eRF1 содержит участки связывания с фактором eRF3 и фактором рециклинга ABCE1 [40]. Фактор eRF3 является ГТФазой, необходимой, по-видимому, в первую очередь для доставки eRF1 [41].

Считается, что посадка фактора eRF1 осуществляется в составе тройственного комплекса eRF1·eRF3·ГТФ [42]. После распознавания стоп-кодона происходит гидролиз ГТФ, приводящий к конформационным изменениям в структуре комплекса, в результате которых становится возможной нуклеофильная атака эфирной связи пептидил-тРНК молекулой воды. После гидролиза ГТФ происходит диссоциация eRF3·ГДФ из терминационного комплекса, что делает возможным посадку фактора рециклинга ABCE1, связывание с которым стимулирует каталитическую активность eRF1 и стабилизирует комплекс 80S·eRF1 [43, 40]. Далее происходит эффективное расщепление

эфирной связи пептидил-тРНК и синтезированный полипептид высвобождается (Рис. 7). Однако точный механизм терминации еще не до конца изучен. Например, в недавней работе было выдвинуто предположение, что фактор ABCE1 участвует во всех стадиях терминации трансляции: после завершения элонгации сначала происходит присоединение комплекса ABCE1·eRF3, а только потом посадка eRF1 [44].

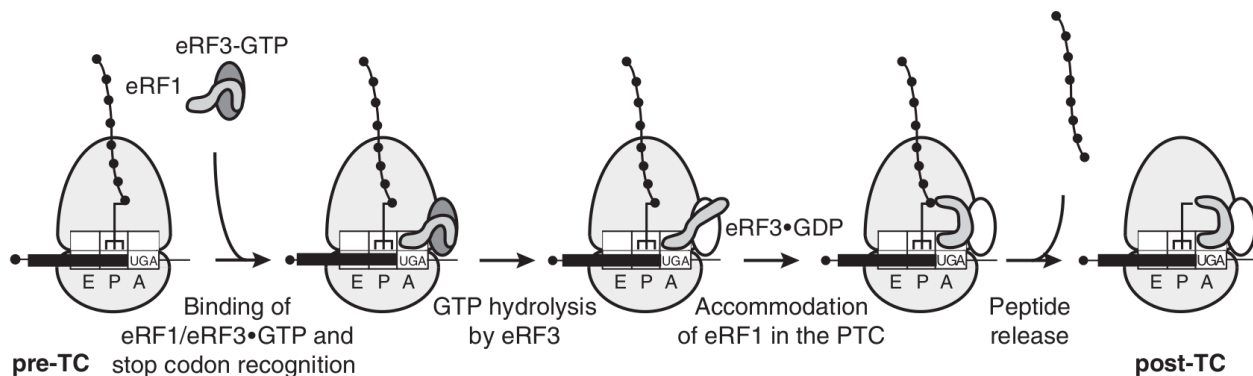


Рис. 7. Модель эукариотической терминации трансляции [42].

В результате терминации трансляции мы имеем 80S пост-терминационный комплекс, состоящий из 80S рибосомы, мРНК и диацелированной тРНК. В А-сайте рибосомы остается фактор eRF1, С-домен которого связан с фактором ABCE1 [42].

Следующий этап – рециклинг рибосом: высвобождение субъединиц рибосом из пост-терминационного комплекса для того, чтобы заново быть вовлеченными в цикл трансляции. На первом шаге рециклинга происходит высвобождение большой субъединицы рибосомы, осуществляемое фактором ABCE1, при этом eRF1 используется как «молекулярный рычаг».

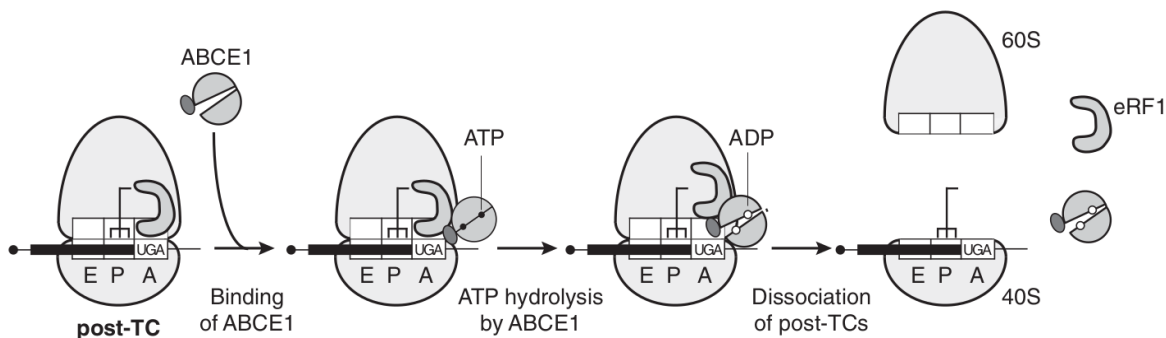


Рис. 8. Модель эукариотического рециклинга рибосом [42].

В результате происходит диссоциация 60S субъединицы и высвобождение eRF1 (Рис. 8). Считается, что после этого шага рециклинга фактор ABCE1

может оставаться связанным с пост-терминационной 40S субъединицей, препятствуя обратному присоединению 60S субъединицы и способствуя привлечению факторов инициации [45]. По видимому, он продолжает находиться на 40S и при инициации – вплоть до узнавания стартового кодона.

После освобождения 60S субъединицы на стоп-кодоне остается 40S субъединица, связанная с мРНК и содержащая в Р-сайте деацелированную тРНК. Дальнейший разбор оставшегося комплекса изучен пока недостаточно, однако в этом процессе, скорее всего, принимают участие белки MCT-1/DENR, eIF2D и, возможно eIF3 [46, 47, 48].

## **Реинициация трансляции и uORF**

Большое значение имеют случаи, когда трансляционный цикл завершается не высвобождением пост-терминационных рибосом, а вторым раундом инициации на той же мРНК – реинициацией.

Около 13% дрожжевых мРНК и 49% мРНК человека перед основной рамкой имеют короткие рамки считывания в 5' нетранслируемой области, называемые uORF (англ. upstream ORF) [49, 50]. uORF обычно предшествует основной рамке, либо перекрываются с ней. Ранее считалось, что uORF чаще всего выполняет регуляторную функцию, влияя на уровень трансляции основной рамки и не кодирует смысловой пептид, однако в последнее время появляется всё больше указаний на функциональность кодируемых ими пептидов, а в недавней работе с помощью систематического скрининга были выявлены десятки таких случаев, а также было показано, что некоторые из этих микропептидов образуют комплексы с белками, закодированными в основной рамке той же мРНК [51]. Но всё же основной функцией uORF является, видимо, регуляция трансляции основной рамки – такие случаи показаны для мРНК, кодирующих факторы, вовлеченные в клеточный цикл, стресс, точную дифференцировку и целый ряд других процессов [52, 53, 50].

Основной механизм этой регуляции следующий: трансляция uORF приводит к тому, что инициаторный комплекс не доезжает до основной рамки считывания, располагающейся дальше. В результате число рибосом, транслирующих основную рамку, намного ниже числа тех, что связались с 5'-концом мРНК. Однако из-за того, что факторы инициации и элонгации медленно диссоциируют с рибосомы после начала элонгации, после терминации на стоп-кодоне короткой uORF они могут быть заново вовлечены, чтобы произвести реинициацию на следующей открытой рамке [54]. Потенциал uORF для такой регуляции зависит от нескольких факторов: близости uORF к основному старт-кодону, её длины, аминокислотного состава и контекста старт-кодона, а также количества таких рамок в лидере [55].

Однако важно отметить, что вклад uORF не ограничивается этими простыми эффектами. Было показано, что редкие мутации и полиморфизм uORF транскриптов человека могут быть причиной заболеваний и иметь сильное влияние на фенотип. В одной из работ, посвященной этому феномену, было идентифицировано 509 человеческих генов с полиморфными uORF (puORF) и описано несколько интересных проявлений [56]. Например, puORF в хемокиновом рецепторе CCR5 имеет связь с восприимчивостью к ВИЧ-1 и прогрессированием СПИДа [57], puORF в рецепторах горького вкуса TAS2R1 и TAS2R3 могут приводить к изменениям в восприятии вкуса, а puORF в рецепторах адренокортикотропного гормона, серотонина и окситоцина могут изменять нейрогормональные реакции.

В дополнение к распространенным полиморфизмам известны редкие мутации uORF способные вызывать серьезные заболевания. К примеру, наследственная форма тромбоцитемии, вызванная мутацией в мРНК гена THPO, устраняет uORF из-за дефектов сплайсинга и, таким образом, вызывает повышенную трансляцию тромбопоэтина [58]. Иная мутация, вводящая uORF в CDKN2A, вызывает семейную предрасположенность к меланоме [59]. Список, что содержит наследственный гипотрихоз Мари Унна, дисгенез гонад, наследственный панкреатит и другие заболевания [56], можно продолжить, но вышеперечисленных фактов достаточно, чтобы продемонстрировать высокую чувствительность процесса трансляции к огромному числу факторов, а также ключевое влияние его регуляции на уровне каждого транскрипта на фенотип и работу целого организма.

## **Полиаденилирование и уровень экспрессии**

В каноническом сценарии для участия в трансляции пре-мРНК должна пройти ряд посттранскрипционных модификаций. Одна из них – добавление поли(А)-хвоста к 3' концу пре-мРНК. В 3' конце пре-мРНК располагаются сайты полиаденилирования (обычно ААТААА или похожий участок), вскоре после которого ко-транскрипционно происходит разрез пре-мРНК (преимущественно после “СА” динуклеотида) и дальнейшее добавление поли(А)-хвоста, содержащего от 50 до нескольких тысяч аденинов, с помощью поли(А)-полимеразы. Затем поли(А)-последовательность узнается упомянутым ранее белком РАВР (англ. poly(A)-binding protein), который при связывании закрывает около 27 аденинов [60]. Полиаденилирование важно для защиты 3' конца от экзонуклеаз, для экспорта мРНК в цитоплазму и для полноценной трансляции [61].

Один из механизмов контроля экспрессии генов – использование альтернативных сайтов полиаденилирования, а не основного [62]. Интересно, что бы-

ла продемонстрирована тесная связь между транскрипцией и выбором сайта полиаденилирования – высоко экспрессирующиеся гены предпочитают более короткие 3'-НТО (нетранслируемые области) по сравнению с длинными изоформами 3'-НТО, наблюдаемыми в генах с низким уровнем экспрессии [63].

## Рибосомный профайлинг

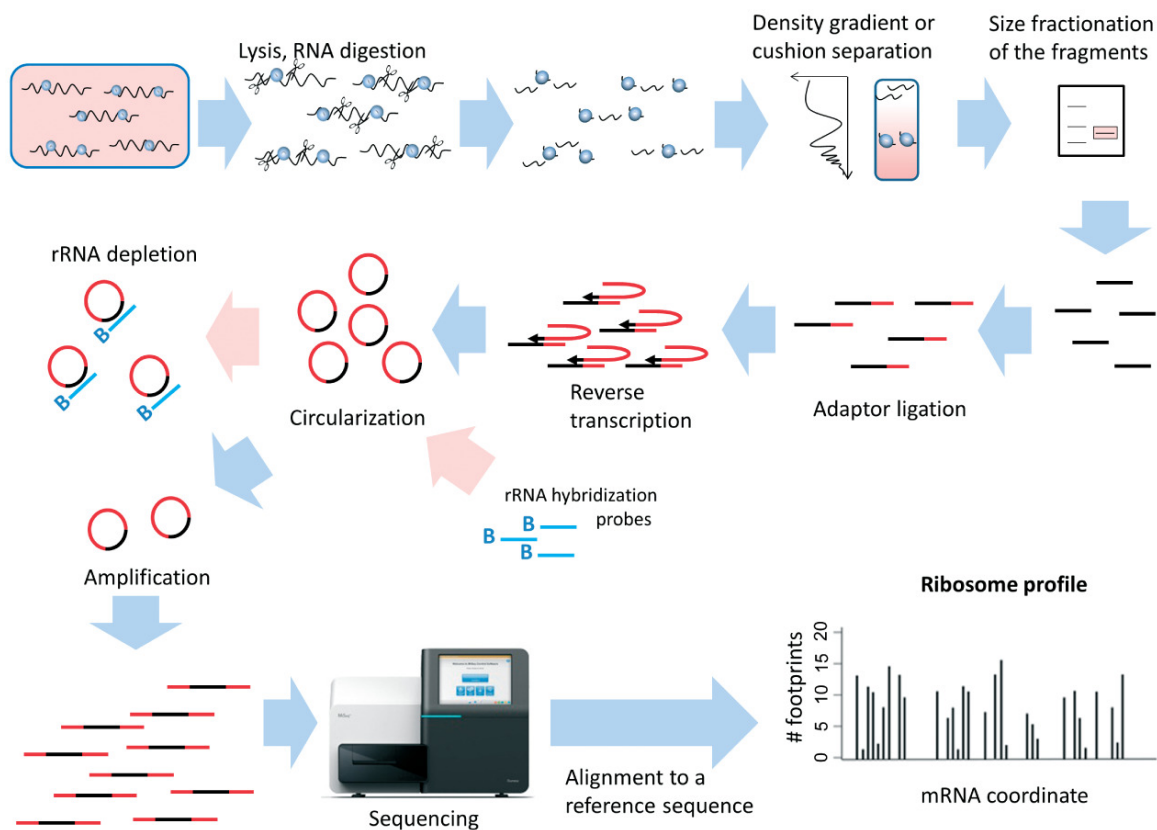


Рис. 9. Основные шаги эксперимента по рибосомному профайлингу [64].

Главным инструментом системного изучения трансляции является метод рибосомного профайлинга (Ribo-Seq). Он основан на глубоком секвенировании защищенных рибосомами фрагментов мРНК, позволяющем измерить плотность рибосом вдоль всех молекул РНК, присутствующих в клетке. Высокая разрешающая способность метода позволяет проводить детальный анализ положения рибосом на отдельных РНК. Основными шагами метода являются: иммобилизация рибосом с помощью антибиотика (как правило циклогексимида), обработка РНК-азами, выделение фрагментов, защищенных рибосомами и их секвенирование (Рис. 9) [64].

## Цель и задачи исследования

**Цель** – выявить влияние нокаутующей KanMX-кассеты на экспрессию гена, соседнего с нокаутуемым, и определить молекулярные механизмы этого процесса.

### Задачи:

- 1) Проанализировать доступные данные рибосомного профайлинга нокаутных штаммов дрожжей *S. cerevisiae*.
- 2) На основе анализа выявить случаи, когда KanMX-кассета оказывает непосредственное влияние на экспрессию соседствующих с ней генов.
- 3) Количественно охарактеризовать влияние этих случаев на уровень транскрипции и трансляции соседнего гена.
- 4) Предложить модель механистического объяснения феномена NGE в найденных случаях.

# Материалы и методы

## Исходные данные

Для нашего анализа был взят уникальный набор данных, содержащий результаты рибосомного профайлинга и РНК-секвенирования для 57 нокаутных штаммов (GEO: GSE100626)[65], а также данные, полученные ранее в нашей лаборатории для трёх других штаммов (GEO: GSE122039)[66]. Использовалась геномная последовательность и аннотация, опубликованная проектом Ensembl (версии R64-1-1) [67]. Координаты альтернативных сайтов полиаденилирования были загружены с GEO: (GSE95139)[68].

## Методы

### Обработка результатов секвенирования

Аннотация и геномная последовательность нокаутных штаммов была создана с помощью инструмента *reform* [69], путем модификации исходной последовательности генома штамма BY4741 дикого типа, опубликованной проектом Ensembl (версии R64-1-1). Сырые данные секвенирования формата SRA были загружены и преобразованы в *fastq* формат с помощью *SRA toolkit v2.9.4* [70]. Удаление адаптеров и тримминг прочтений по качеству было произведено инструментом *cutadapt v1.18* [71].

Выравнивания на оригинальный и модифицированные аннотированные геномы, а также число уникальных на каждый ген выровненных прочтений были получены программой *STAR v2.7.0* [72]. С помощью инструментов *samtools* [73] и *bedtools v2.27.1* [74] результаты выравниваний формата BAM были преобразованы в *BedGraph* профили. Каждый *BedGraph* профиль был нормирован с помощью размеров библиотек и нормировочных факторов, полученных из анализа дифференциальной экспрессии. Визуализация профилей была произведена с помощью программы *svist4get* [75].

### Анализ дифференциальной экспрессии

Обработка и статистический анализ числа уникальных на каждый ген выровненных прочтений было произведено в среде языка R с использованием пакета *edgeR Bioconductor* [76]. Гены с покрытием менее, чем 2 прочтения на миллион, как минимум в 3 образцах РНК-секвенирования из 11 и 4 образцах рибосомного профайлинга из 16, были исключены из анализа (для каждый пары: нокаут–контроль). Для подсчета дифференциальной экспрессии РНК-секвенирования и RO (рибосомный профайлинг, нормированный на количество транскриптов) была использована обобщенная линейная модель (*glmQLFit*, *glmQLFTest* пакета *edgeR*).

## Модель для расчёта дифференциальной экспрессии

В результате выравнивания данных секвенирования строится таблица числа прочтений фрагментов – каунтов (от англ. counts) со строками, соответствующими генам, и столбцами для образцов. Для каждой пары ген-образец мы имеем число, которое соответствует уровню экспрессии данного гена в образце. Для рибосомного профайлинга это число рибосомных футпринтов на транскрипте данного гена. Наша задача – определить уровень изменения активности транскрипции и трансляции каждого гена между группами образцов, а также значимость этого изменения [77].

Данные с подсчётом каких-либо дискретных величин возникают во многих биологических приложениях и часто моделируются распределением Пуассона, когда среднее и дисперсия одинаковы. Однако в ситуациях, когда существуют положительные корреляции между значениями, наблюдаемая дисперсия значительно больше среднего значения. Для корректировки edgeR использует негативное биномиальное распределение (NB) [78]. NB в данном случае идеально, поскольку оно дискретно и обеспечивает высокую точность при небольших значениях рассматриваемых величин. Оно также позволяет учитывать изменчивость между биологическими репликами внутри одной группы образцов, используя параметры дисперсии [79]. Модель NB в современных версиях также расширена методами квази-правдоподобия (англ. quasi-likelihood, QL) для учета генно-специфической изменчивости как по биологическим, так и по техническим причинам [80]. Таким образом:

$$Y_{gi} \sim \text{NB}(M_i p_{gj}, \phi_g)$$

А соответствующая функция вероятности NB распределения имеет вид [78]:

$$f(y; \mu_{gi}, \phi) = P(Y_{gi} = y) = \frac{\Gamma(y + \phi^{-1})}{\Gamma(\phi^{-1})} \left( \frac{1}{1 + \mu_{gi}\phi} \right)^{\phi^{-1}} \left( \frac{\mu_{gi}}{\phi^{-1} + \mu_{gi}} \right)^y$$

где  $Y_{gi}$  – число прочтений для гена  $g$  в образце  $i$ ,  $M_i$  – размер библиотеки (общее число выровненных прочтений) в образце  $i$ ,  $\phi_g$  – дисперсия, а  $p_{gj}$  – относительное покрытие гена  $g$  в экспериментальной группе  $j$ , к которой относится образец  $i$ . Для NB распределения среднее значение  $\mu_{gi} = M_i p_{gj}$ . Модифицированная дисперсия рассматривается, как квадратичная функция среднего значения:

$$\text{var}(Y_{gi}) = \sigma_g^2(\mu_{gi} + \mu_{gi}^2\phi)$$

где  $\phi$  – параметр дисперсии NB и  $\sigma_g^2$  – параметр дисперсии QL.

Любое увеличение наблюдаемой дисперсии  $Y_{gi}$  будет моделироваться увеличением  $\phi$  и/или  $\sigma_g^2$ . В данной модели NB дисперсия  $\phi$  – глобальный параметр,



в то время как QL – генно-специфичный. Таким образом два параметра дисперсии играют совершенно разные роли. NB дисперсия описывает общую биологическую изменчивость по всем генам (разница между биологическими образцами). Дисперсия же QL улавливает любую генно-специфическую изменчивость (на уровне отдельного гена дисперсия может быть как больше, чем общая, так и меньше). Главная сложность подхода заключается в точном расчёте дисперсии при небольшом числе реплик. Решение этой проблемы – отдельная большая алгоритмическая задача. В последних версиях edgeR для этого используется эмпирический байесовских подход (EB) [81].

Для средних значений edgeR использует обобщенную линейную модель (GLM) [79]. Среднее число каунтов для гена  $g$  в образце  $i$  представляется, как

$$\log(\mu_{gi}) = x_{i1}\beta_{g1} + x_{i2}\beta_{g2} + \dots + x_{in}\beta_{gn} + o_i$$

где  $\beta_{gj}$  – ген-специфичное значение коэффициента  $j$ ,  $x_{ij}$  – образец-специфичный предиктор для  $j$ , а  $o_i$  – образец специфичное смещение. Всего  $n$  коэффициентов, где каждый должен описывать некоторый фактор дизайна эксперимента (например, различные биологические условия или batch эффект). Предикторы могут быть дискретными или непрерывными, рассмотрим дискретные. Выставление  $x_{ij} = 1$  означает, что коэффициент  $j$  влияет на экспрессию образца  $i$ , т.е. фактор дизайна эксперимента  $j$  оказывает влияние на этот образец. Величина  $\beta_{gj}$  в свою очередь характеризует влияние этого фактора на экспрессию гена  $g$  в образце, когда  $x_{ij} \neq 0$ , то есть это влияние присутствует.

Наконец, смещение определяет размер библиотеки – общее число каунтов по всем генам в образце. Рассмотрим два образца  $i = 1$  и  $i = 2$ , где размер библиотеки  $i = 1$  в два раза больше, чем для  $i = 2$ . Это означает, что ожидаемое значение  $\mu_{g1}$  будет также в два раза больше, чем  $\mu_{g2}$  для одного и того же гена  $g$  с одинаковым в действительности уровнем экспрессии. Смещение гарантирует, что различия в размере библиотек не способствуют неверному расчёту:

$$\log(\mu_{g1}) - o_1 = \log(\mu_{g2}) - o_2$$

Искомые значения  $\log(FC)_{ij}^g$  (FC от англ. fold change), характеризующее во сколько раз изменилась экспрессия гена  $g$  между сравниваемыми группами образцов  $i$  и  $j$ , определяется отношением логарифмов средних. Значимость же определяется характеристикой  $p\text{-value}_{ij}^g$  – вероятностью получить данное  $\log(FC)_{ij}^g$  при условии, что верна нулевая гипотеза (отсутствие в действительности изменений в экспрессии между генами).  $P\text{-value}_{ij}^g$  определяется значениями средних и дисперсий, рассчитывается через t-критерий Стьюдента. На основе  $p\text{-value}$  делается поправка на множественное тестирование генов – процедура Бенджамини-Хохберга, в результате которой получаем параметр FDR (англ. False Discovery Rate) [79].

## Приложение инструмента svist4get для анализа

Один из ключевых этапов анализа данных был произведен с помощью программы svist4get. Это инструмент командной строки для настраиваемой визуализации данных высокопроизводительного секвенирования (RNA-Seq, Ribo-Seq и др.). Он был разработан мной и опубликован в 2019 году [75], а его тестовая версия была представлена в виде курсовой работы на 2-ом курсе.

В качестве входных данных svist4get поддерживает следующие форматы: bedGraph для геномных профилей (таблица, где для каждой геномной координаты ставится в соответствие число, как правило характеризующее уровень экспрессии), GTF формат для геномной аннотации (описание структуры и расположения генов, транскриптов, экзонов и других структурных элементов генома), fasta формат для геномной последовательности каждой хромосомы/контига.

Получив конкретное геномное окно, id гена или транскрипта для визуализации, svist4get автоматически выполняет сглаживание с помощью скользящего среднего сигнальных треков, если это необходимо, учитывая ширину изображения и размер геномного окна. В результате на выходе представляя файлы pdf и png формата с оптимальными параметрами для каждого масштаба: от десятков, до сотен тысяч нуклеотидов и для любого размера изображения. Однако svist4get – это инструмент визуализации, поэтому данные требуют предварительной обработки с помощью внешних инструментов, описанных выше.

В данной работе была использована уникальная способность svist4get непрерывно визуализировать сотни заданных геномных участков в различных масштабах, находясь на сервере, где хранятся обработанные данные.

# Результаты

## Уровень экспрессии кассеты KanMX

Первым делом мы решили изучить общую структуру транскрипционных и трансляционных изменений в модифицированном геномном локусе. Для этого мы реконструировали геном дикого типа для нокаутных штаммов и выровняли на них соответствующие данные РНК-секвенирования и рибосомного профайлинга.

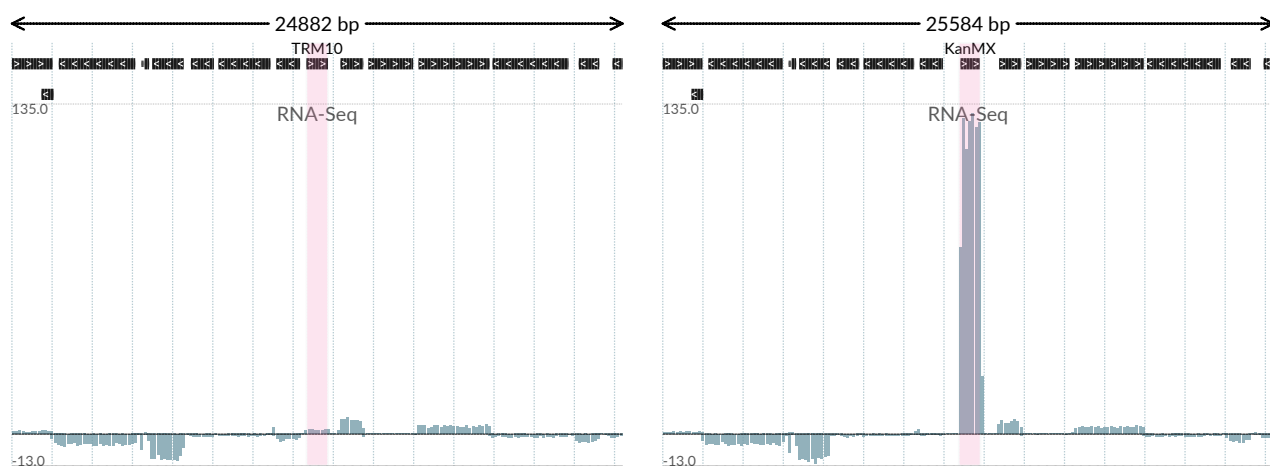


Рис. 10. Экспрессия KanMX кассеты в модифицированном локусе. Левая панель – дикий тип (wt), правая – нокаутный штамм (КО). Треки (ось ординат) отображают нормированное покрытие прочтениями РНК-секвенирования. Положительные значения соответствуют прямой цепи ДНК, отрицательные – обратной.

Рисунок 10 демонстрирует результаты такого сравнения для пары штаммов:  $\Delta trm10$  и  $WT$ . Мы наблюдали непропорционально высокую экспрессию гена *kan* в мутантном штамме, по сравнению с эффективностью транскрипции других генов, расположенных в пределах локуса. Высокий уровень экспрессии был ожидаем, однако мы впервые показали, как неадекватно высок этот уровень. Экспрессия на уровне трансляции аналогичным образом дерегулирована – наблюдается неадекватная нагрузка транскрипта *kan* на клеточный механизм трансляции (Приложение, Рис. 18). Очень похожие результаты экспрессии кассеты были получены и для других штаммов.

Чрезмерная транскрипция искусственного модуля может исказить общую активность всего локуса на эпигенетическом уровне. Действительно, даже на Рис. 10, несмотря на мелкий масштаб, можно заметить достаточно сильное падение транскрипции соседнего слева гена *RFC4* (*YOL094C*). При анализе результатов дифференциальной экспрессии было установлено, что его экспрессия на уровне транскрипции упала в 4 раза (Приложение, Рис. 19).

## Кассета KanMX смещает сайты старта транскрипции соседних генов, приводя к появлению блокирующих uORF

Для дальнейшего анализа изменений внутри мутантных локусов мы дополнительно стали изучать экспрессию непосредственно близлежащих генов. В случае штамма  $\Delta pus3$  мы наблюдали существенное снижение экспрессии соседнего гена *SPB4* на трансляционном уровне, тогда как его экспрессия на транскрипционном уровне была понижена незначительно (Рис. 11, Приложение: Рис. 20). Сопутствовал данным изменениям явный сдвиг сайта старта транскрипции (TSS) гена *SPB4* в сторону кассеты KanMX в нокаутном штамме. Из-за этого сдвига AUG кодон, предшествующий основному, был включен в 5' нетранслируемую область транскрипта *SPB4*, тем самым образуя ингибирующую uORF (нижняя панель Рис. 11). Также стоит отметить, что приобретенный uAUG кодон имеет оптимальный нуклеотидный контекст (AggAUGA), и поэтому должен быть эффективно распознан трансляционным комплексом на этапе инициации [82]. Действительно, мы наблюдаем наличие рибосомного покрытия соответствующей uORF и сильное падение его на основной рамке транскрипта. Снижение же количества мРНК *SPB4* (Рис. 11, Приложение: Рис. 20), вероятно, может быть объяснено понижением её стабильности, так как транскрипты, содержащие uORF, в клетках дрожжей подвержены NMD (англ. nonsense-mediated decay) распаду [83].

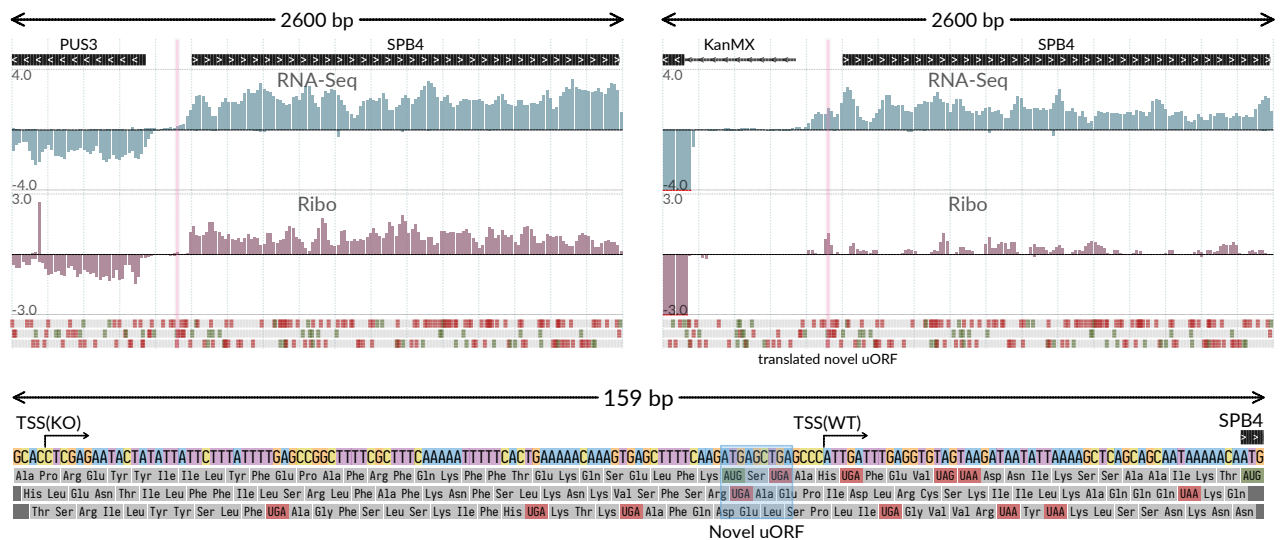


Рис. 11. Визуализация данных РНК-секвенирования и рибосомного профайлинга *PUS3-SPB4* локуса в штаммах *wt* (слева) и  $\Delta pus3$  (справа). Треки (ось ординат) отображают нормированное покрытие, положительные значения соответствует прямой цепи ДНК, отрицательные – обратной. Нижняя панель демонстрирует сдвиг TSS в нокаутном штамме (KO) в сравнении с TSS дикого штамма (WT). Трансляция новой uORF (подсвечена на всех панелях) приводит к критическому падению трансляции основной рамки гена *SPB4*.

Еще более драматический эффект на трансляционном уровне наблюдался в случае нокаута гена *TMA64*, приводящего к дерегуляции соседнего гена *APC4*, также расположенного в ориентации 5' конец к 5' концу (“голова к голове”) (Рис. 12, Приложение: Рис. 21). В этом случае 5' нетранслируемая область транскрипта *APC4*, удлинненная в результате смещения TSS, содержит два AUG кодона (нижняя панель Рис. 12). Трансляция появившихся в результате этого uORF, продемонстрированная на результатах рибосомного профайлинга, оказывается блокирующей для основной рамки транскрипта.

Любопытно, что предыдущие анализы штаммов нокаута  $\Delta tma64$  выявили сильные отрицательные генетические взаимодействия *TMA64* с генами, кодирующими компоненты комплекса стимуляции анафазы, циклосомы (APC/C) [32], а также аномалии клеточного цикла [84]. Действительно, 7 из 10 генетических взаимодействий *TMA64*, перечисленных в базе данных SGD [85] и задокументированные по крайней мере в 2 источниках данных, связаны с mitotic checkpoint (переходом от метафазы к анафазе). Наши данные предполагают, что эти взаимодействия являются ложно-положительными, вызванными дефектом трансляции мРНК *APC4* в штамме нокаута  $\Delta tma64$ , который мы обнаружили в нашем исследовании.

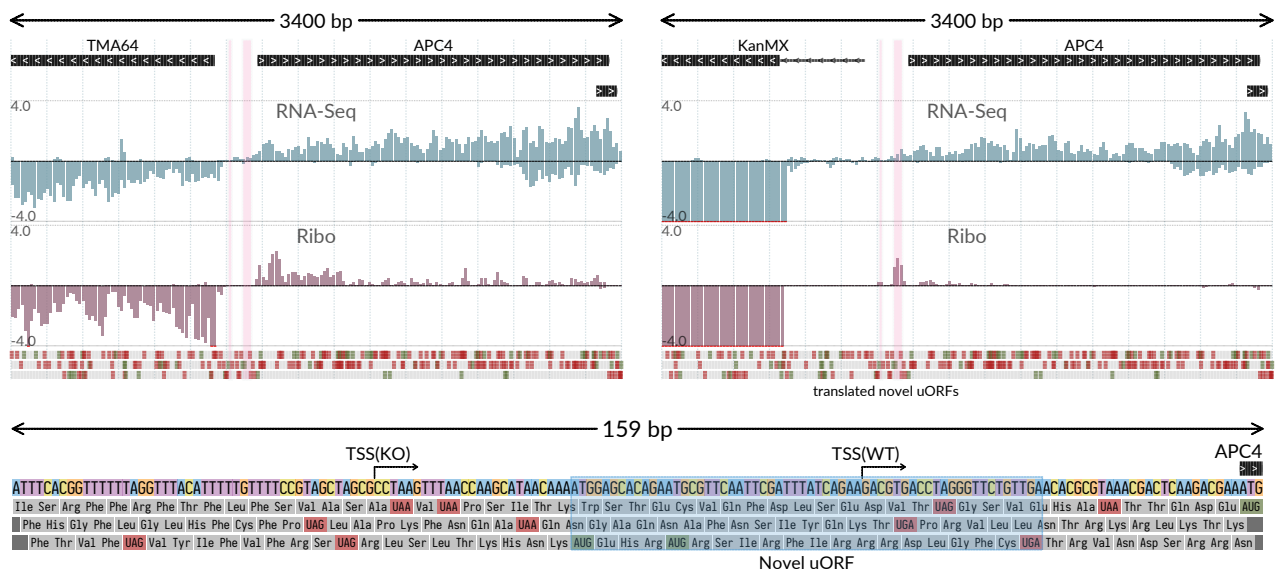


Рис. 12. Визуализация данных РНК-секвенирования и рибосомного профайлинга *TMA64-APC4* локуса в штаммах *wt* (слева) и  $\Delta tma64$  (справа). Треки (ось ординат) отображают нормированное покрытие, положительные значения соответствуют прямой цепи ДНК, отрицательные – обратной. Нижняя панель демонстрирует сдвиг TSS в нокаутном штамме (KO) в сравнении с TSS дикого штамма (WT). Трансляция новых uORF (подсвечены на всех панелях) приводит к катастрофическому падению трансляции основной рамки гена *APC4*.

## Кассета KanMX нарушает посттранскрипционную модификацию 3' конца транскриптов соседних генов

Вышеприведенные случаи иллюстрировали изменения, когда нокаутный ген и соседний, в котором произошли изменения, находились в расположении “голова к голове”. Однако модуль KanMX также может нарушать правильную терминацию транскрипции и корректное формирование 3' конца соседнего транскрипта при взаимном расположении 3' конец к 3' концу (“хвост к хвосту”). Чтобы изучить это влияние, мы проанализировали ряд нокаутных штаммов с таким взаимным расположением и, действительно, обнаружили смещение сайтов расщепления (англ. cleavage site) и полиаденилирования (PAS).

Смещение сайта полиаденилирования может быть проиллюстрировано нокаутом *TRM12* с резким изменением положения активного PAS соседнего гена *GLO1* (Рис. 13). В данном конкретном случае изменение, вероятно, было неизбежным, поскольку исходный транскрипт *GLO1* своим 3' концом пересекается с кодирующей областью *TRM12*, где его PAS обычно и находится. Однако в нокаутном штамме  $\Delta trm12$  эта область удалена и заменена кассетой KanMX. Картирование результатов РНК-секвенирования на геном  $\Delta trm12$  выявило укорочение 3' нетранслируемой области более, чем в 2 раза. Данное изменение, вероятно, является результатом активации одного из альтернативных сайтов полиаденилирования, расположенных в этом локусе (нижняя панель, Рис. 13).

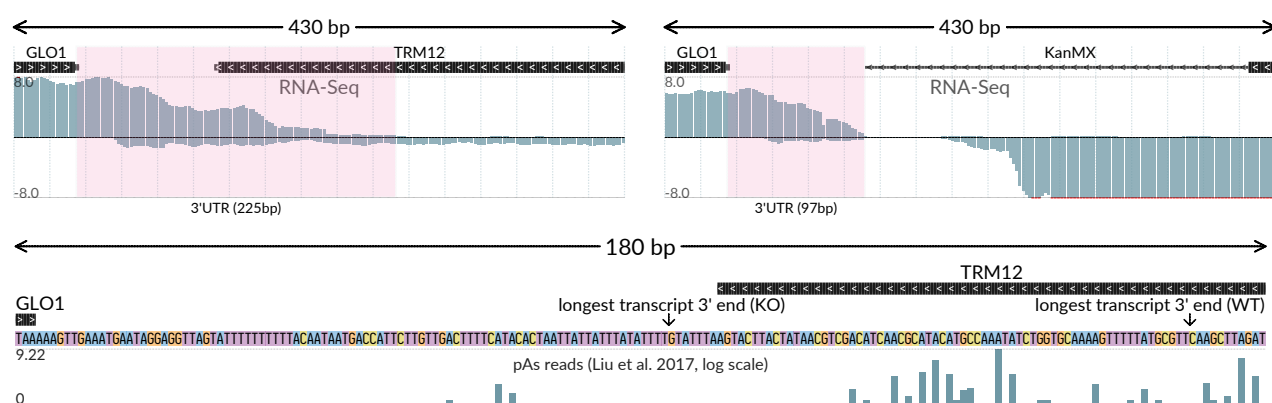


Рис. 13. Визуализация данных РНК-секвенирования *GLO1-TRM12* локуса в штаммах *wt* (слева) и  $\Delta trm12$  (справа). Треки (ось ординат) отображают нормированное покрытие, положительные значения соответствуют прямой цепи ДНК, отрицательные – обратной. Подсвеченной рамкой выделена 3' нетранслируемая область транскрипта *GLO1*. На нижней панели стрелками отмечены длины 3'-НТО в обоих штаммах. Вертикальная ось в логарифмической шкале отображает частоту активаций альтернативных сайтов полиаденилирования [68].

Поскольку 3' нетранслируемые области играют важную роль в регуляции трансляции и контроле стабильности мРНК [61], в некоторых случаях изменения в них могут влиять как на количество и стабильность транскрипта, так и на уровень его трансляции. Действительно, в случае другой пары генов, имеющей взаимное расположение “хвост к хвосту”, мы наблюдаем значительное снижение покрытия рибосомами основной кодирующей области гена *SET3* в штамме, где соседний ген *SAP190* был заменен кассетой KanMX (Рис.14, Приложение: Рис.22). Данные изменения сопровождаются резким укорочением 3' нетранслируемой области *SET3* в нокаутном штамме практически на 100 нуклеотидов. Важно отметить, что в этом случае отсутствовало физическое пересечение основного сайта полиаденилирования с удаленной кодирующей областью соседнего гена, но его активация всё равно была нарушена.

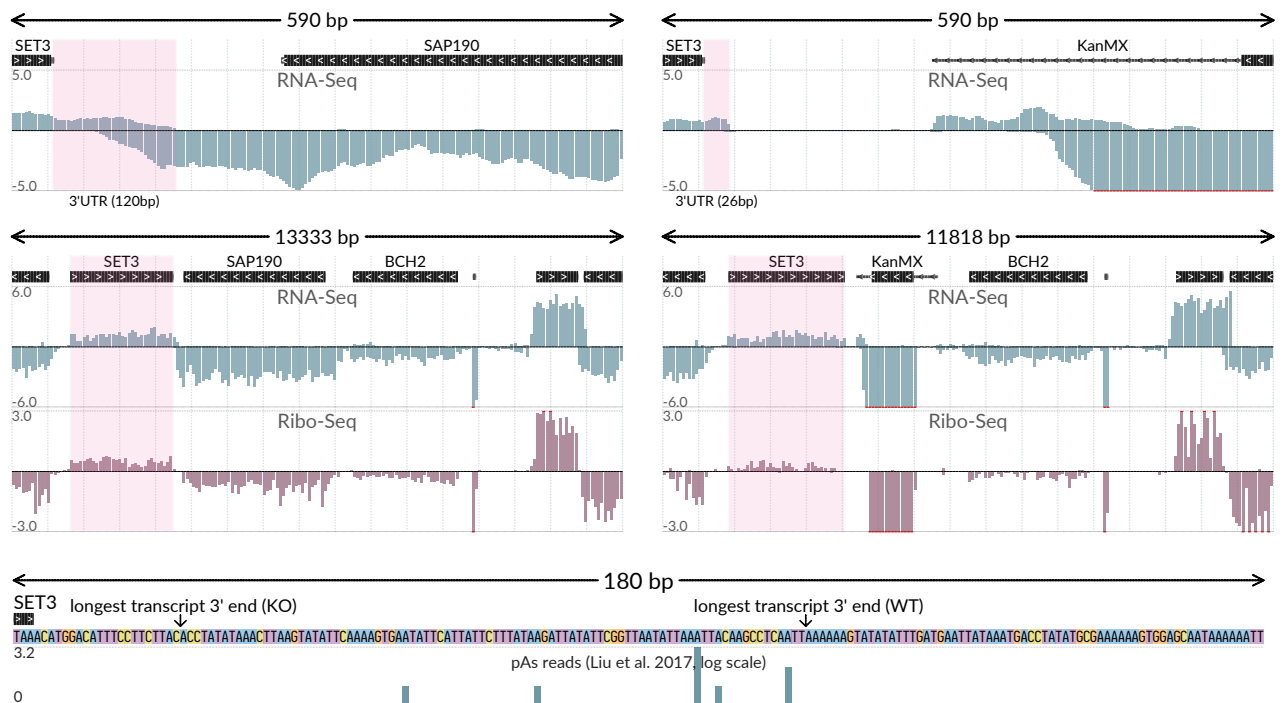


Рис. 14. Визуализация данных РНК-секвенирования и рибосомного профайлинга *SET3-SAP190* локуса в штаммах *wt* (слева) и  $\Delta sap190$  (справа). Треки (ось ординат) отображают нормированное покрытие, положительные значения соответствуют прямой цепи ДНК, отрицательные – обратной. Подсвеченной рамкой на верхней панели выделена 3' нетранслируемая область транскрипта *SET3*. На средней панели рамкой подсвечен транскрипт *SET3* и соответствующие ему профили покрытия. На нижней панели стрелками отмечены длины 3' НТО в обоих штаммах. Вертикальная ось в логарифмической шкале отображает частоту активации альтернативных сайтов полиаденилирования [68].

## Смещения TSS и активация альтернативных PAS, как молекулярный механизм NGE

В ходе анализа имеющихся в доступности данных рибосомного профайлинга и РНК-секвенирования для нокаутных штаммов было выявлено, что смещение TSS и активация альтернативных сайтов полиаденилирования представляют особый интерес. Чтобы систематизировать полученные результаты, мы построили график изменения длин нетранслируемых областей соседних транскриптов, индуцируемых касетой KapMX, в зависимости от их взаимной ориентации (Рис. 15).

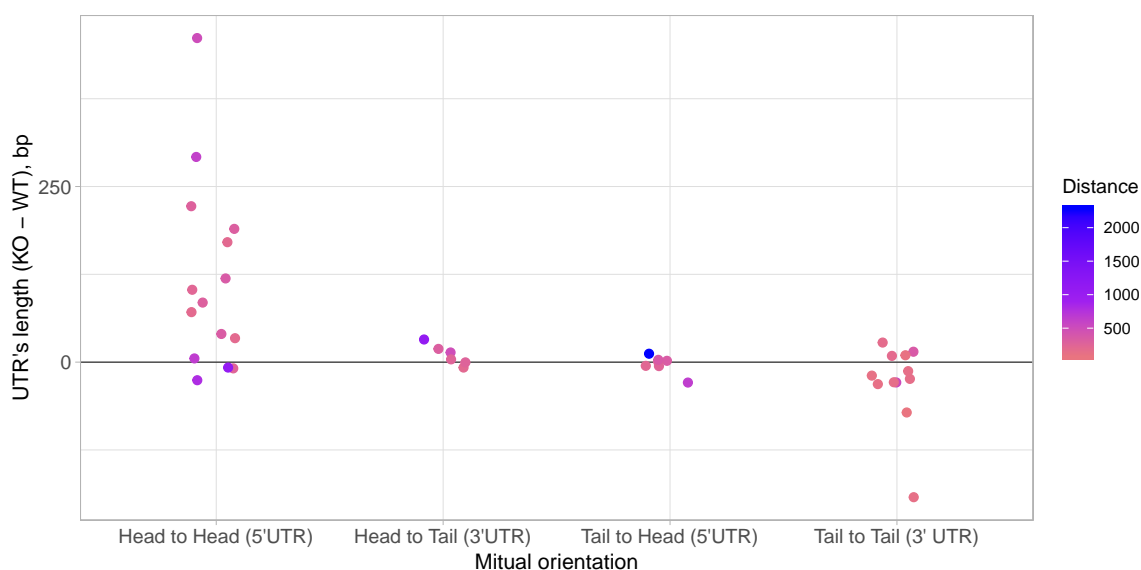


Рис. 15. Изменения длин нетранслируемых областей соседних генов, индуцируемые касетой KapMX, в зависимости от взаимной ориентации.

Анализ показывает, что в случае ориентации “голова к голове” происходят самые серьезные изменения: в 11 из 12 рассмотренных случаев, когда расстояние между кодирующими частями генов менее 1000 нуклеотидов, мы наблюдали увеличение 5' UTR вследствие сдвига TSS. Стоит отметить, что смещение TSS имело место даже при расстояниях между кодирующими частями, превышающих 350-400 нуклеотидов. В случае взаимного расположения “хвост к хвосту” наблюдается тренд к укорочению 3' нетранслируемой области в результате активации альтернативного сайта полиаденилирования, однако сильным этот эффект, в сравнении со смещением TSS, назвать нельзя. В случае взаимных ориентаций “голова к хвосту” и “хвост к голове” серьезных изменений в длинах нетранслируемых областей соседних генов не наблюдалось. Однако изучение изменений в некодирующих областях таких пар генов затруднено из-за того, что покрытие РНК-секвенирования соседнего и нокаутного генов находятся на одной цепи ДНК и в случае близкого расположения генов являются неразличимыми и недоступными для включения в рассмотрение.



Для результатов дифференциальной экспрессии РНК-секвенирования и RO (англ. ribosome occupancy, отношение рибосомного покрытия к уровню транскрипции, характеризующее эффективность трансляции) мы провели аналогичный анализ изменений для каждой взаимной ориентации (Рис. 16).

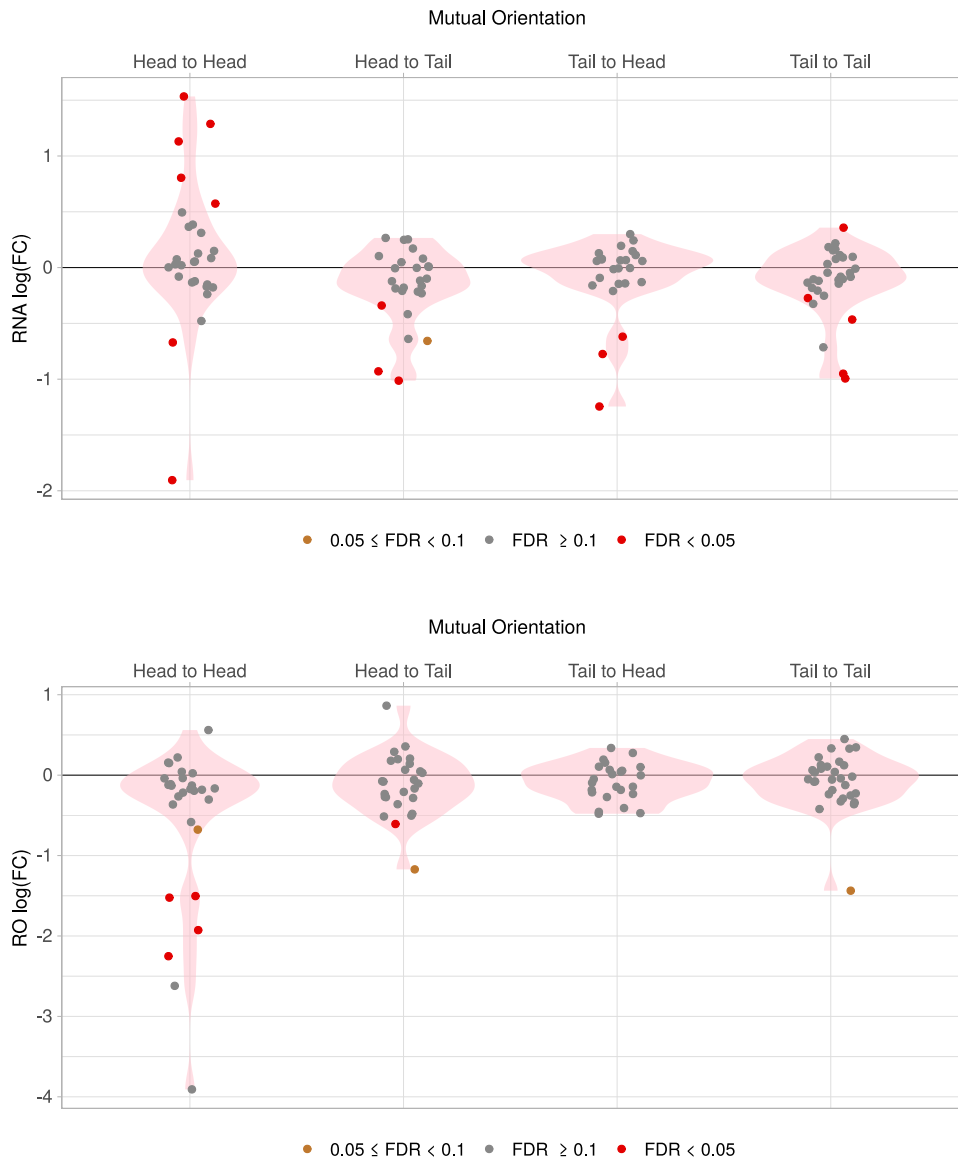


Рис. 16. Изменения уровня транскрипции (верхняя панель) и RO (нижняя панель) соседних генов, индуцируемые кассетой KanMX в зависимости от взаимной ориентации. FDR (англ. False Discovery Rate) – коэффициент ложного обнаружения.

На уровне изменения транскрипции (верхняя панель Рис. 16) снова выделяется ориентация “голова к голове”, в случае которой наблюдаются сильные статистически значимые выбросы, как в сторону повышения, так и в сторону падения транскрипции соседнего гена в несколько раз. Для других взаимных

ориентаций, при наличии отклонений, чаще наблюдается снижение уровня мРНК соседних транскриптов. В случае дифференциальной экспрессии RO при ориентации “голова к голове” ожидаемо наблюдается повышенная вероятность многократного значимого падения уровня трансляции. Случаи иной ориентации имеют однократные выбросы, но в целом показывают колебание изменения значений вокруг нулевой отметки (в логарифмической шкале).

## Заключение

Библиотека нокаутных штаммов SGDP является мощным и незаменимым инструментом в изучении функции генов эукариот [5, 6, 7]. Однако из-за компактности генома дрожжей замена каждого отдельного гена сильно экспрессирующейся кассетой KanMX потенциально может нарушать экспрессию соседних генов. Важно отметить, что нежелательные эффекты кассеты могут выходить за пределы транскрипционного уровня, влияя на трансляцию транскриптов. Во многих случаях эти изменения могут привести к неправильной функциональной аннотации целевого гена, в то время как на самом деле эффекты вызваны проблемами экспрессии соседних генов. Это явление известно как NGE. Его следует принимать во внимание при проведении высокопроизводительных исследований и анализа генетических взаимодействий. Несколько систематических исследований, посвященных этому важному вопросу, показали, что NGE может быть вовлечен в 7-15% идентифицированных взаимодействий между генами [8, 9, 10]. Однако никаких механистических объяснения этих эффектов предложено не было.

В нашем исследовании мы использовали преимущества данных рибосомного профайлинга для изучения этой проблемы. Хотя этот мощный метод был изначально разработан на *S. cerevisiae* [16, 86], на сегодняшний день было проведено небольшое число исследований, содержащих данные рибосомного профайлинга для нокаутных дрожжевых штаммов. Мы использовали один из наиболее полных наборов данных, содержащих Ribo-Seq и RNA-Seq для 57 мутантных штаммов [65], а также данные, полученные в нашей лаборатории (лаборатория системной биологии старения НИИ физико-химической биологии им. А.Н. Белозерского МГУ) [66]. Мы задокументировали несколько типов изменений генной экспрессии, включая транскрипционную активацию или репрессию соседних генов; смещение TSSs соседнего гена в сторону KanMX-кассеты, вызывающее появление ингибирующих uAUG в удлиненном 5' НТО, снижение эффективности трансляции и, вероятно, стабильности мРНК; а также активацию альтернативных сайтов полиаденилирования пре-мРНК, приводящих к дерегулированию экспрессии мРНК на трансляционном уровне.

Хотя наше исследование было ограничено количеством доступных наборов данных рибосомного профайлинга дрожжей, даже эти несколько примеров ясно показывают, что во многих случаях замена гена кассетой KanMX резко влияет на экспрессию соседних генов как через транскрипционные, так и через трансляционные механизмы. Наиболее яркими результатами нашего исследования были сдвиги TSS (в случае расположения генов “голова к голове”) и активация альтернативных PAS (при ориентации генов “хвост к хвосту”), приводящие к снижению трансляции мРНК. Эффекты сдвигов TSS и изме-

нений PAS в нокаутирующих штаммах могут быть еще более важными при различных условиях, клеточном стрессе или споруляции, когда правильный выбор TSS и PAS наиболее важен [87, 88].

Выявленные закономерности изменения экспрессии соседних генов (Рис. 17) могут объяснить феномен NGE в рассмотренных нами случаях, а также, вероятно, во многих других.

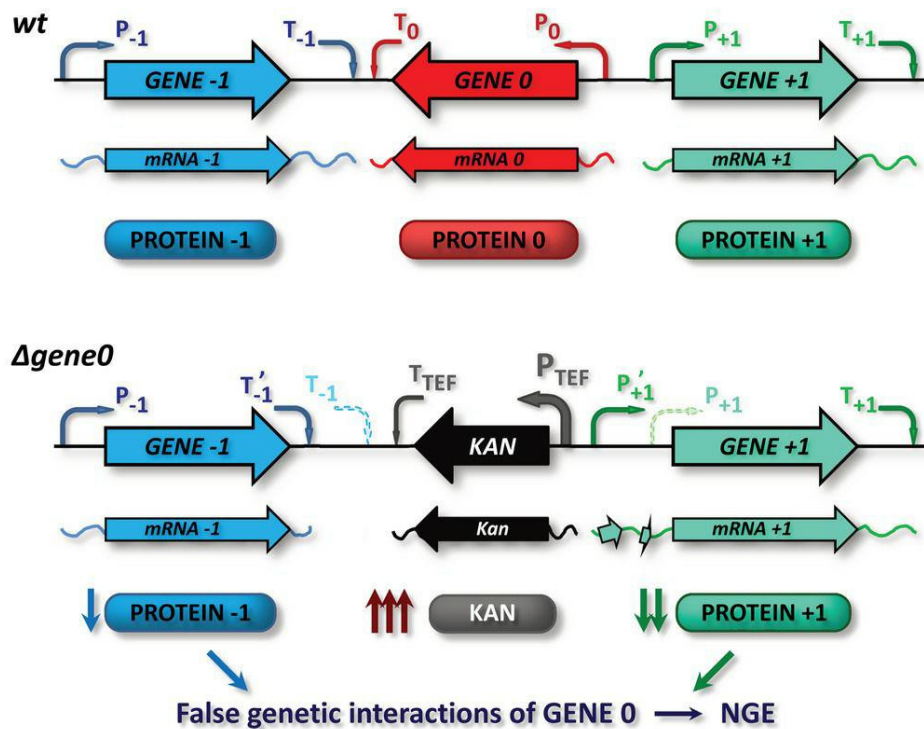


Рис. 17. Модель, объясняющая NGE в нокаутных штаммах.

В целом мы предполагаем, что во многих случаях NGE в нокаутных штаммах вызывается дефектами на трансляционном уровне, аналогичными тем, которые были выявлены в данном исследовании. Это следует учитывать при использовании коллекции дрожжевых штаммов для функциональной аннотации генов и анализа генетических взаимодействий, а также при интерпретации результатов, основанных на аналогичных методах целенаправленной манипуляции генами в любом другом модельном эукариотическом организме.

## Выводы

- 1) В результате анализа данных рибосомного профайлинга нокаутных штаммов дрожжей *S. cerevisiae* обнаружены случаи влияния нокаутующей KanMX-кассеты на экспрессию соседнего гена.
- 2) Составлена классификация таких случаев:
  - а) транскрипционная активация или репрессия соседних генов;
  - б) смещение сайта старта транскрипции соседнего гена в сторону кассеты KanMX, вызывающее появление ингибирующих uAUG в удлиненной 5'-НТО и падение уровня трансляции основной рамки;
  - в) активация альтернативных сайтов полиаденилирования пре-мРНК, приводящая к изменению длины 3'-НТО и дерегулированию экспрессии мРНК на трансляционном уровне.
- 3) Количественно охарактеризовано влияние этих случаев на уровень транскрипции и трансляции соседнего гена для 59 нокаутных штаммов дрожжей.
- 4) Предложена модель механистического объяснения феномена NGE (эффекта соседствующего гена).

## Список литературы

- [1] A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. Life with 6000 genes. *Science*, 274(5287):546–567, October 1996.
- [2] Guri Giaever, Angela M. Chu, Li Ni, Carla Connelly, Linda Riles, Steeve Véronneau, Sally Dow, Ankuta Lucau-Danila, Keith Anderson, Bruno André, Adam P. Arkin, Anna Astromoff, Mohamed El Bakkoury, Rhonda Bangham, Rocio Benito, Sophie Brachat, Stefano Campanaro, Matt Curtiss, Karen Davis, Adam Deutschbauer, Karl-Dieter Entian, Patrick Flaherty, Francoise Foury, David J. Garfinkel, Mark Gerstein, Deanna Gotte, Ulrich Güldener, Johannes H. Hegemann, Svenja Hempel, Zelek Herman, Daniel F. Jaramillo, Diane E. Kelly, Steven L. Kelly, Peter Kötter, Darlene LaBonte, David C. Lamb, Ning Lan, Hong Liang, Hong Liao, Lucy Liu, Chuanyun Luo, Marc Lussier, Rong Mao, Patrice Menard, Siew Loon Ooi, Jose L. Revuelta, Christopher J. Roberts, Matthias Rose, Petra Ross-Macdonald, Bart Scherens, Greg Schimmack, Brenda Shafer, Daniel D. Shoemaker, Sharon Sookhai-Mahadeo, Reginald K. Storms, Jeffrey N. Strathern, Giorgio Valle, Marleen Voet, Guido Volckaert, Ching yun Wang, Teresa R. Ward, Julie Wilhelmy, Elizabeth A. Winzeler, Yonghong Yang, Grace Yen, Elaine Youngman, Kexin Yu, Howard Bussey, Jef D. Boeke, Michael Snyder, Peter Philippsen, Ronald W. Davis, and Mark Johnston. Functional profiling of the *saccharomyces cerevisiae* genome. *Nature*, 418(6896):387–391, July 2002.
- [3] Elizabeth A. Winzeler, Daniel D. Shoemaker, Anna Astromoff, Hong Liang, Keith Anderson, Bruno Andre, Rhonda Bangham, Rocio Benito, Jef D. Boeke, Howard Bussey, Angela M. Chu, Carla Connelly, Karen Davis, Fred Dietrich, Sally Whelen Dow, Mohamed El Bakkoury, Françoise Foury, Stephen H. Friend, Erik Gentalen, Guri Giaever, Johannes H. Hegemann, Ted Jones, Michael Laub, Hong Liao, Nicole Liebundguth, David J. Lockhart, Anca Lucau-Danila, Marc Lussier, Nasiha M'Rabet, Patrice Menard, Michael Mittmann, Chai Pai, Corinne Rebischung, Jose L. Revuelta, Linda Riles, Christopher J. Roberts, Petra Ross-MacDonald, Bart Scherens, Michael Snyder, Sharon Sookhai-Mahadeo, Reginald K. Storms, Steeve Véronneau, Marleen Voet, Guido Volckaert, Teresa R. Ward, Robert Wysocki, Grace S. Yen, Kexin Yu, Katja Zimmermann, Peter Philippsen, Mark Johnston, and Ronald W. Davis. Functional characterization of

the *s. cerevisiae* genome by gene deletion and parallel analysis. *Science*, 285(5429):901–906, 1999.

- [4] Saccharomyces Genome Deletion Project. [http://www-sequence.stanford.edu/group/yeast\\_deletion\\_project/project\\_desc.html#cassettes](http://www-sequence.stanford.edu/group/yeast_deletion_project/project_desc.html#cassettes).
- [5] Guri Giaever and Corey Nislow. The yeast deletion collection: A decade of functional genomics. *Genetics*, 197(2):451–465, jun 2014.
- [6] Michael Costanzo, Elena Kuzmin, Jolanda van Leeuwen, Barbara Mair, Jason Moffat, Charles Boone, and Brenda Andrews. Global genetic networks and the genotype-to-phenotype relationship. *Cell*, 177(1):85–100, mar 2019.
- [7] Siew Loon Ooi, Xuewen Pan, Brian D. Peyser, Ping Ye, Pamela B. Meluh, Daniel S. Yuan, Rafael A. Irizarry, Joel S. Bader, Forrest A. Spencer, and Jef D. Boeke. Global synthetic-lethality analysis and yeast functional profiling. *Trends in Genetics*, 22(1):56–63, jan 2006.
- [8] Matej Usaj, Yizhao Tan, Wen Wang, Benjamin VanderSluis, Albert Zou, Chad L. Myers, Michael Costanzo, Brenda Andrews, and Charles Boone. Thecellmap.org: A web-accessible database for visualizing and mining the global yeast genetic interaction network. *G3: Genes, Genomes, Genetics*, 7(5):1539–1549, 2017.
- [9] Nir Atias, Martin Kupiec, and Roded Sharan. Systematic identification and correction of annotation errors in the genetic interaction map of *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 44(5):e50–e50, nov 2015.
- [10] Taly Ben-Shitrit, Nir Yosef, Keren Shemesh, Roded Sharan, Eytan Rupp, and Martin Kupiec. Systematic identification of gene annotation errors in the widely used yeast mutation collections. *Nature Methods*, 9(4):373–378, feb 2012.
- [11] Alexandre Guilliermond. The yeasts, translated from the french by f. w. tanner. *The Plimpton Press, Norwood, MA.*, 1920.
- [12] Alan G. Hinnebusch and Mark Johnston. YeastBook: An encyclopedia of the reference eukaryotic cell. *Genetics*, 189(3):683–684, November 2011.
- [13] Bernard Dujon. The yeast genome project: what did we learn? *Trends in Genetics*, 12(7):263–270, July 1996.
- [14] Sven Heinicke, Michael S. Livstone, Charles Lu, Rose Oughtred, Fan Kang, Samuel V. Angiuoli, Owen White, David Botstein, and Kara Dolinski. The

princeton protein orthology database (p-POD): A comparative genomics analysis tool for biologists. *PLoS ONE*, 2(8):e766, August 2007.

- [15] A. Baudin, O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute, and C. Cullin. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 21(14):3329–3330, 1993.
- [16] N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, and J. S. Weissman. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 324(5924):218–223, April 2009.
- [17] Nancy Kedersha and Paul Anderson. Chapter 4 regulation of translation by stress granules and processing bodies. In *Progress in Molecular Biology and Translational Science*, pages 155–185. Elsevier, 2009.
- [18] Michael Altmann and Patrick Linder. Power of yeast for analysis of eukaryotic translation initiation. *Journal of Biological Chemistry*, 285(42):31907–31912, August 2010.
- [19] Li Ni and Michael Snyder. A genomic study of the bipolar bud site selection pattern in *Saccharomyces cerevisiae*. *Molecular Biology of the Cell*, 12(7):2147–2170, July 2001.
- [20] Nicolas Pagé, Manon Gérard-Vincent, Patrice Ménard, Maude Beaulieu, Masayuki Azuma, Gerrit J. P. Dijkgraaf, Huijuan Li, José Marcoux, Thuy Nguyen, Tim Dowse, Anne-Marie Sdicu, and Howard Bussey. A *saccharomyces cerevisiae* genome-wide mutant screen for altered sensitivity to k1 killer toxin. *Genetics*, 163(3):875–894, 2003.
- [21] Ainslie B. Parsons, Andres Lopez, Inmar E. Givoni, David E. Williams, Christopher A. Gray, Justin Porter, Gordon Chua, Richelle Sopko, Renee L. Brost, Cheuk-Hei Ho, Jiyi Wang, Troy Ketela, Charles Brenner, Julie A. Brill, G. Esteban Fernandez, Todd C. Lorenz, Gregory S. Payne, Satoru Ishihara, Yoshikazu Ohya, Brenda Andrews, Timothy R. Hughes, Brendan J. Frey, Todd R. Graham, Raymond J. Andersen, and Charles Boone. Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell*, 126(3):611–625, August 2006.
- [22] Esti Yeger-Lotem, Laura Riva, Linhui Julie Su, Aaron D Gitler, Anil G Cashikar, Oliver D King, Pavan K Auluck, Melissa L Geddie, Julie S Valastyan, David R Karger, Susan Lindquist, and Ernest Fraenkel. Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. *Nature Genetics*, 41(3):316–323, February 2009.



- [23] S. Willingham. Yeast genes that enhance the toxicity of a mutant huntingtin fragment or  $\alpha$ -synuclein. *Science*, 302(5651):1769–1772, December 2003.
- [24] Anita L. Manogaran, Viviana M. Fajardo, Robert J. D. Reid, Rodney Rothstein, and Susan W. Liebman. Most, but not all, yeast strains in the deletion library contain the [PIN+] prion. *Yeast*, 2009.
- [25] Zhihui Sun, Zamia Diaz, Xiaodong Fang, Michael P. Hart, Alessandra Chesi, James Shorter, and Aaron D. Gitler. Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. *PLoS Biology*, 9(4):e1000614, April 2011.
- [26] R. A. Fisher. XV.—the correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52(2):399–433, 1919.
- [27] Patrick C. Phillips, Sarah P. Otto, and Michael C. Whitlock. Beyond the average. *Epistasis and the evolutionary process*, pages 20–38, 2000.
- [28] Ramamurthy Mani, Robert P St Onge, John L Hartman, Guri Giaever, and Frederick P Roth. Defining genetic interaction. *Proceedings of the National Academy of Sciences*, 105(9):3461–3466, 2008.
- [29] Amy Hin Yan Tong, Marie Evangelista, Ainslie B Parsons, Hong Xu, Gary D Bader, Nicholas Pagé, Mark Robinson, Sasan Raghizadeh, Christopher WV Hogue, Howard Bussey, et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, 294(5550):2364–2368, 2001.
- [30] Michael Costanzo, Anastasia Baryshnikova, Jeremy Bellay, Yungil Kim, Eric D Spear, Carolyn S Sevier, Huiming Ding, Judice LY Koh, Kiana Toufighi, Sara Mostafavi, et al. The genetic landscape of a cell. *Science*, 327(5964):425–431, 2010.
- [31] Michael Costanzo, Anastasia Baryshnikova, Chad L Myers, Brenda Andrews, and Charles Boone. Charting the genetic interaction map of a cell. *Current opinion in biotechnology*, 22(1):66–74, 2011.
- [32] Michael Costanzo, Benjamin VanderSluis, Elizabeth N Koch, Anastasia Baryshnikova, Carles Pons, Guihong Tan, Wen Wang, Matej Usaj, Julia Hanchard, Susan D Lee, et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science*, 353(6306), 2016.
- [33] O. M. Alekhina and K. S. Vassilenko. Translation initiation in eukaryotes: Versatility of the scanning model. *Biochemistry (Moscow)*, 77(13):1465–1477, December 2012.

- [34] Alan G. Hinnebusch. The scanning mechanism of eukaryotic translation initiation. *Annual Review of Biochemistry*, 83(1):779–812, June 2014.
- [35] Richard J. Jackson, Christopher U. T. Hellen, and Tatyana V. Pestova. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, 11(2):113–127, February 2010.
- [36] William C. Merrick and Graham D. Pavitt. Protein synthesis initiation in eukaryotic cells. *Cold Spring Harbor Perspectives in Biology*, 10(12):a033092, May 2018.
- [37] Elena Z. Alkalaeva, Andrey V. Pisarev, Lyudmila Y. Frolova, Lev L. Kisselev, and Tatyana V. Pestova. In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell*, 125(6):1125–1136, June 2006.
- [38] Sichen Shao, Jason Murray, Alan Brown, Jack Taunton, V. Ramakrishnan, and Ramanujan S. Hegde. Decoding mammalian ribosome-mRNA states by translational GTPase complexes. *Cell*, 167(5):1229–1240.e15, November 2016.
- [39] Ludmila Y. Frolova, Ruslan Y. Tsivkovskii, Galina F. Sivolobova, Nina Y. Oparina, Oleg I. Serpinsky, Vladimir M. Blinov, Sergey I. Tatkov, and Lev L. Kisselev. Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA*, 5(8):1014–1020, August 1999.
- [40] C. J. Shoemaker and R. Green. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proceedings of the National Academy of Sciences*, 108(51):E1392–E1398, December 2011.
- [41] Vasili Hauryliuk, Andrey Zavialov, Lev Kisselev, and Måns Ehrenberg. Class-1 release factor eRF1 promotes GTP binding by class-2 release factor eRF3. *Biochimie*, 88(7):747–757, July 2006.
- [42] Christopher U.T. Hellen. Translation termination and ribosome recycling in eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 10(10):a032656, May 2018.
- [43] Anne Preis, Andre Heuer, Clara Barrio-Garcia, Andreas Hauser, Daniel E. Eyler, Otto Berninghausen, Rachel Green, Thomas Becker, and Roland Beckmann. Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. *Cell Reports*, 8(1):59–65, July 2014.

- [44] Christian Beiel, Bettina Neumann, Simon Uhse, Irene Hampe, Prajwal Karki, and Heike Krebber. Translation termination depends on the sequential ribosomal entry of eRF1 and eRF3. *Nucleic Acids Research*, 47(9):4798–4813, March 2019.
- [45] Andr Heuer, Milan Gerovac, Christian Schmidt, Simon Trowitzsch, Anne Preis, Peter Ktter, Otto Berninghausen, Thomas Becker, Roland Beckmann, and Robert Tamp. Structure of the 40s–ABCE1 post-splitting complex in ribosome recycling and translation initiation. *Nature Structural & Molecular Biology*, 24(5):453–460, April 2017.
- [46] Andrey V. Pisarev, Christopher U.T. Hellen, and Tatyana V. Pestova. Recycling of eukaryotic posttermination ribosomal complexes. *Cell*, 131(2):286–299, October 2007.
- [47] M. A. Skabkin, O. V. Skabkina, V. Dhote, A. A. Komar, C. U. T. Hellen, and T. V. Pestova. Activities of ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes & Development*, 24(16):1787–1801, August 2010.
- [48] David J. Young, Desislava S. Makeeva, Fan Zhang, Aleksandra S. Anisimova, Elena A. Stolboushkina, Fardin Ghobakhlou, Ivan N. Shatsky, Sergey E. Dmitriev, Alan G. Hinnebusch, and Nicholas R. Guydosh. Tma64/eIF2d, tma20/MCT-1, and tma22/DENR recycle post-termination 40s subunits in vivo. *Molecular Cell*, 71(5):761–774.e5, September 2018.
- [49] Guo-Liang Chew, Andrea Pauli, and Alexander F. Schier. Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nature Communications*, 7(1), May 2016.
- [50] Craig Lawless, Richard D Pearson, Julian N Selley, Julia B Smirnova, Christopher M Grant, Mark P Ashe, Graham D Pavitt, and Simon J Hubbard. Upstream sequence elements direct post-transcriptional regulation of gene expression under stress conditions in yeast. *BMC Genomics*, 10(1):7, 2009.
- [51] Jin Chen, Andreas-David Brunner, J. Zachery Cogan, James K. Nuez, Alexander P. Fields, Britt Adamson, Daniel N. Itzhak, Jason Y. Li, Matthias Mann, Manuel D. Leonetti, and Jonathan S. Weissman. Pervasive functional translation of noncanonical human open reading frames. *Science*, 367(6482):1140–1146, March 2020.
- [52] A. G. Hinnebusch, I. P. Ivanov, and N. Sonenberg. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science*, 352(6292):1413–1416, June 2016.

- [53] Peggy Janich, Alaaddin Bulak Arpat, Violeta Castelo-Szekely, Maykel Lopes, and David Gatfield. Ribosome profiling reveals the rhythmic liver translome and circadian clock regulation by upstream open reading frames. *Genome Research*, 25(12):1848–1859, October 2015.
- [54] Stanislava Gunišová, Vladislava Hronová, Mahabub Pasha Mohammad, Alan G Hinnebusch, and Leoš Shivaya Valášek. Please do not recycle! translation reinitiation in microbes and higher eukaryotes. *FEMS Microbiology Reviews*, 42(2):165–192, December 2017.
- [55] Joanna Somers, Tuija Pöyry, and Anne E. Willis. A perspective on mammalian upstream open reading frame function. *The International Journal of Biochemistry & Cell Biology*, 45(8):1690–1700, August 2013.
- [56] Sarah E. Calvo, David J. Pagliarini, and Vamsi K. Mootha. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proceedings of the National Academy of Sciences*, 106(18):7507–7512, 2009.
- [57] Zdenka Navratilova et al. Polymorphisms in *ccl2* & *ccl5* chemokines/chemokine receptors genes and their association with diseases. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*, 150(2):191–204, 2006.
- [58] Adrian Wiestner, Ronald J Schlemper, Anthonie PC van der Maas, and Radek C Skoda. An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. *Nature genetics*, 18(1):49–52, 1998.
- [59] Ling Liu, David Dilworth, Luzhang Gao, Jose Monzon, Ann Summers, Norman Lassam, and David Hogg. Mutation of the *cdkn2a* 5'utr creates an aberrant initiation codon and predisposes to melanoma. *Nature genetics*, 21(1):128–132, 1999.
- [60] RA Steinman. mRNA stability control: a clandestine force in normal and malignant hematopoiesis. *Leukemia*, 21(6):1158–1171, 2007.
- [61] Eva Matoulkova, Eva Michalova, Borivoj Vojtesek, and Roman Hrstka. The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biology*, 9(5):563–576, May 2012.
- [62] Pea Carninci, T Kasukawa, S Katayama, J Gough, MC Frith, Norihiro Maeda, Rieko Oyama, T Ravasi, B Lenhard, C Wells, et al. The transcriptional landscape of the mammalian genome. *science*, 309(5740):1559–1563, 2005.

- [63] Zhe Ji, Ju Youn Lee, Zhenhua Pan, Bingjun Jiang, and Bin Tian. Progressive lengthening of 3' untranslated regions of mrnas by alternative polyadenylation during mouse embryonic development. *Proceedings of the National Academy of Sciences*, 106(17):7028–7033, 2009.
- [64] Audrey M. Michel and Pavel V. Baranov. Ribosome profiling: a hi-def monitor for protein synthesis at the genome-wide scale. *WIREs RNA*, 4(5):473–490, May 2013.
- [65] Hsin-Jung Chou, Elisa Donnard, H. Tobias Gustafsson, Manuel Garber, and Oliver J. Rando. Transcriptome-wide analysis of roles for tRNA modifications in translational regulation. *Molecular Cell*, 68(5):978–992.e4, December 2017.
- [66] Desislava S. Makeeva, Andrey S. Lando, Aleksandra Anisimova, Artyom A. Egorov, Maria D. Logacheva, Alexey A. Penin, Dmitry E. Andreev, Pavel G. Sinitcyn, Ilya M. Terenin, Ivan N. Shatsky, Ivan V. Kulakovskiy, and Sergey E. Dmitriev. Translatome and transcriptome analysis of TMA20 (MCT-1) and TMA64 (eIF2d) knockout yeast strains. *Data in Brief*, 23:103701, April 2019.
- [67] Daniel R Zerbino, Premanand Achuthan, Wasiu Akanni, M Ridwan Amode, Daniel Barrell, Jyothish Bhai, Konstantinos Billis, Carla Cummins, Astrid Gall, Carlos García Girón, Laurent Gil, Leo Gordon, Leanne Haggerty, Erin Haskell, Thibaut Hourlier, Osagie G Izuogu, Sophie H Janacek, Thomas Juettemann, Jimmy Kiang To, Matthew R Laird, Ilias Lavidas, Zhicheng Liu, Jane E Loveland, Thomas Maurel, William McLaren, Benjamin Moore, Jonathan Mudge, Daniel N Murphy, Victoria Newman, Michael Nuhn, Denye Ogeh, Chuang Kee Ong, Anne Parker, Mateus Patricio, Harpreet Singh Riat, Helen Schuilenburg, Dan Shepard, Helen Sparrow, Kieron Taylor, Anja Thormann, Alessandro Vullo, Brandon Walts, Amonida Zadissa, Adam Frankish, Sarah E Hunt, Myrto Kostadima, Nicholas Langridge, Fergal J Martin, Matthieu Muffato, Emily Perry, Magali Ruffier, Dan M Staines, Stephen J Trevanion, Bronwen L Aken, Fiona Cunningham, Andrew Yates, and Paul Flicek. Ensembl 2018. *Nucleic Acids Research*, 46(D1):D754–D761, November 2017.
- [68] Xiaochuan Liu, Mainul Hoque, Marc Larochele, Jean-François Lemay, Nathan Yurko, James L. Manley, François Bachand, and Bin Tian. Comparative analysis of alternative polyadenylation in *s. cerevisiae* and *s. pombe*. *Genome Research*, 27(10):1685–1695, September 2017.
- [69] Reform. <https://github.com/gencorefacility/reform>.
- [70] R. Leinonen, H. Sugawara, and M. Shumway and. The sequence read archive. *Nucleic Acids Research*, 39(Database):D19–D21, November 2010.

- [71] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10, May 2011.
- [72] Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, October 2012.
- [73] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin and. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, June 2009.
- [74] Aaron R. Quinlan and Ira M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, January 2010.
- [75] Artyom A. Egorov, Ekaterina A. Sakharova, Aleksandra S. Anisimova, Sergey E. Dmitriev, Vadim N. Gladyshev, and Ivan V. Kulakovskiy. svist4get: a simple visualization tool for genomic tracks from sequencing experiments. *BMC Bioinformatics*, 20(1), March 2019.
- [76] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, November 2009.
- [77] Aaron T. L. Lun, Yunshun Chen, and Gordon K. Smyth. It’s DE-licious: A recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. In *Methods in Molecular Biology*, pages 391–416. Springer New York, 2016.
- [78] Mark D. Robinson and Gordon K. Smyth. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*, 9(2):321–332, 08 2007.
- [79] Davis J. McCarthy, Yunshun Chen, and Gordon K. Smyth. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40(10):4288–4297, 01 2012.
- [80] Steven P. Lund, Dan Nettleton, Davis J. McCarthy, and Gordon K. Smyth. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Statistical Applications in Genetics and Molecular Biology*, 11(5), January 2012.
- [81] Mark D Robinson and Alicia Oshlack. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3):R25, 2010.

- [82] A. G. Hinnebusch. Molecular mechanism of scanning and start codon selection in eukaryotes. *Microbiology and Molecular Biology Reviews*, 75(3):434–467, September 2011.
- [83] Christophe Malabat, Frank Feuerbach, Laurence Ma, Cosmin Saveanu, and Alain Jacquier. Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *eLife*, 4, April 2015.
- [84] Machika Watanabe, Daisuke Watanabe, Satoru Nogami, Shinichi Morishita, and Yoshikazu Ohya. Comprehensive and quantitative analysis of yeast deletion mutants defective in apical and isotropic bud growth. *Current Genetics*, 55(4):365–380, May 2009.
- [85] SGD. <https://www.yeastgenome.org>.
- [86] G. A. Brar, M. Yassour, N. Friedman, A. Regev, N. T. Ingolia, and J. S. Weissman. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science*, 335(6068):552–557, December 2011.
- [87] Karl Waern and Michael Snyder. Extensive transcript diversity and novel upstream open reading frame regulation in yeast. *G3: Genes, Genomes, Genetics*, 3(2):343–352, February 2013.
- [88] Ze Cheng, George Maxwell Otto, Emily Nicole Powers, Abdurrahman Keskink, Philipp Mertins, Steven Alfred Carr, Marko Jovanovic, and Gloria Ann Brar. Pervasive, coordinated protein-level changes driven by transcript isoform switching during meiosis. *Cell*, 172(5):910–923.e16, February 2018.
- [89] Rickard Sandberg, Joel R Neilson, Arup Sarma, Phillip A Sharp, and Christopher B Burge. Proliferating cells express mrnas with shortened 3'untranslated regions and fewer microrna target sites. *Science*, 320(5883):1643–1647, 2008.

# Приложение

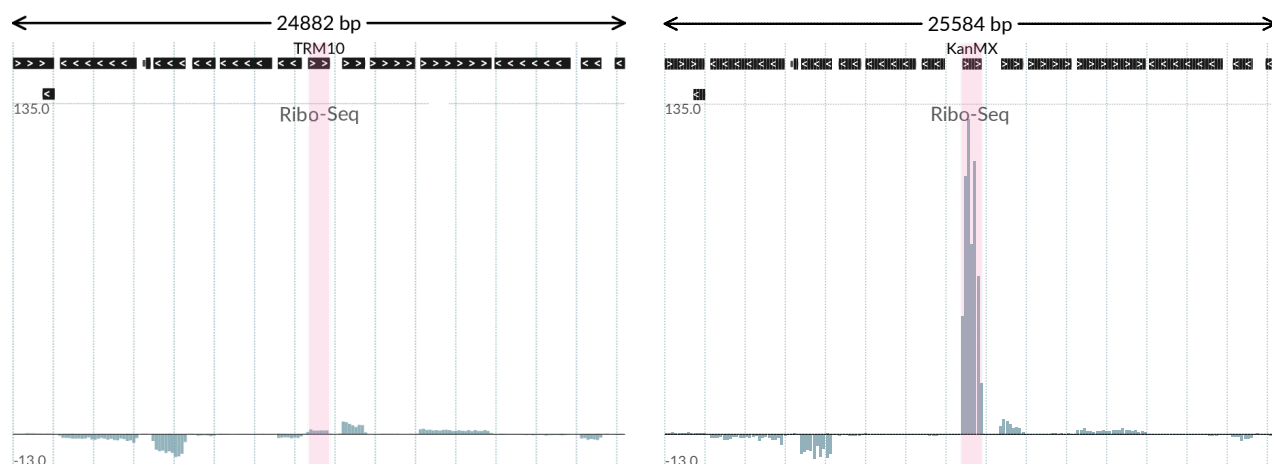


Рис. 18. Экспрессия KanMX кассеты в модифицированном локусе на уровне трансляции

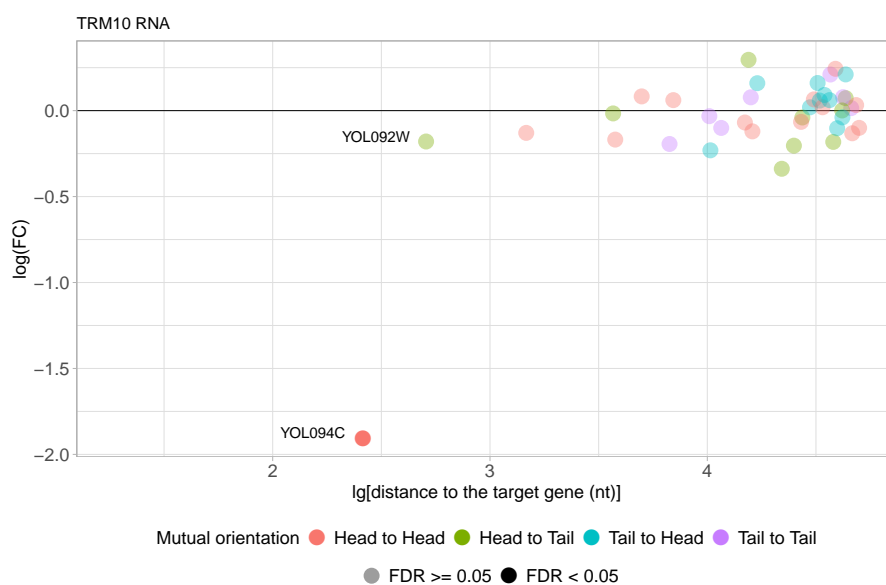


Рис. 19. Анализ дифференциальной экспрессии на уровне транскрипции для генов, находящихся в одном локусе с геном *TRM10*. Сравнение произведено для нокаутного штамма  $\Delta trm10$ , контроль – *WT*. По оси ординат логарифм диапазона изменения (англ. fold change), по оси абсцисс десятичный логарифм расстояния до соседнего гена от *TRM10*. FDR (англ. False Discovery Rate) - коэффициент ложного обнаружения.



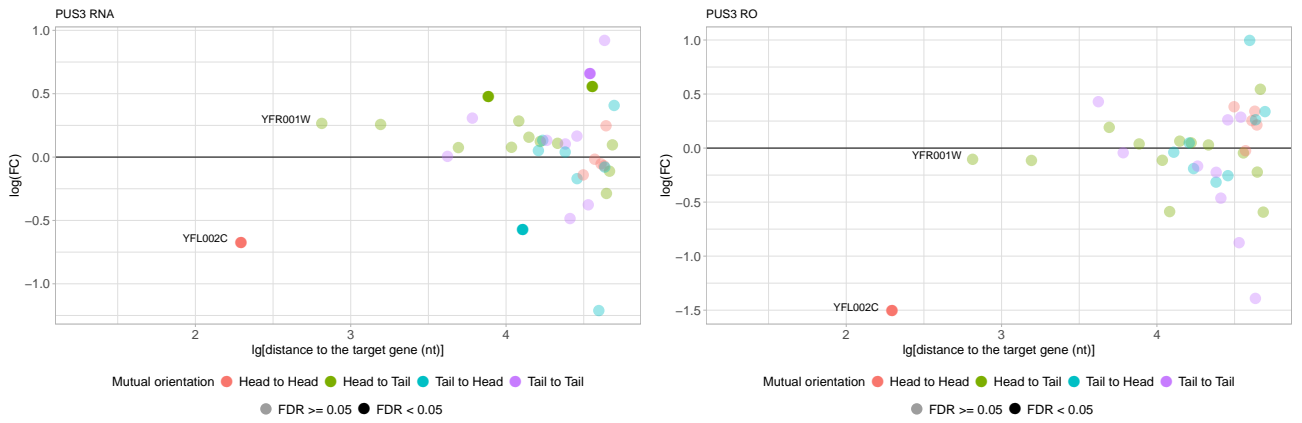


Рис. 20. Анализ дифференциальной экспрессии на уровне транскрипции (левая панель) и RO (правая панель) для генов, находящихся в локусе *PUS3*. Сравнение произведено для нокаутного штамма  $\Delta pus3$ , контроль – *WT*. По оси ординат логарифм диапазона изменения (англ. fold change), по оси абсцисс десятичный логарифм расстояния до соседнего гена от *PUS3*. FDR (англ. False Discovery Rate) - коэффициент ложного обнаружения.

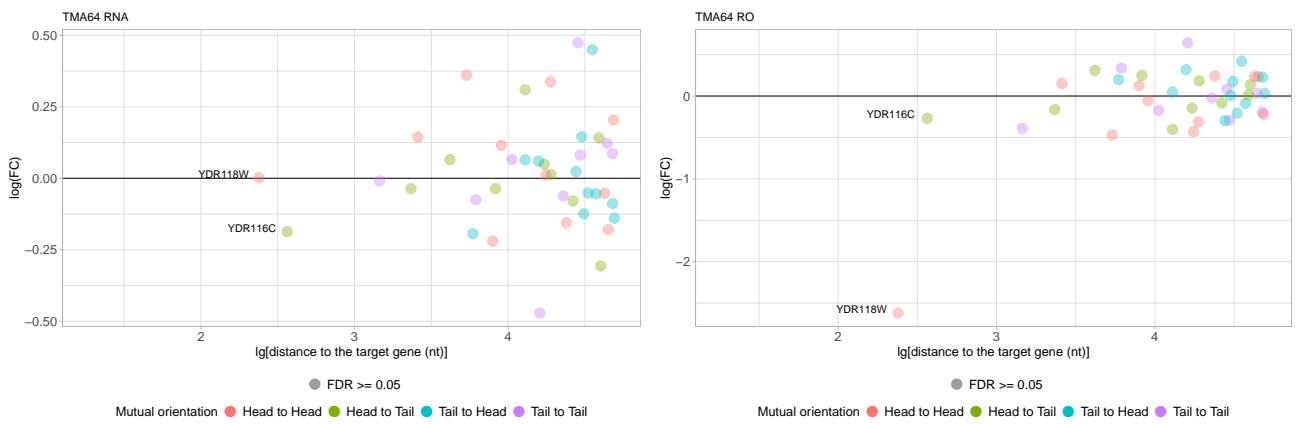


Рис. 21. Для локуса *TMA64*, аналогично описанию выше

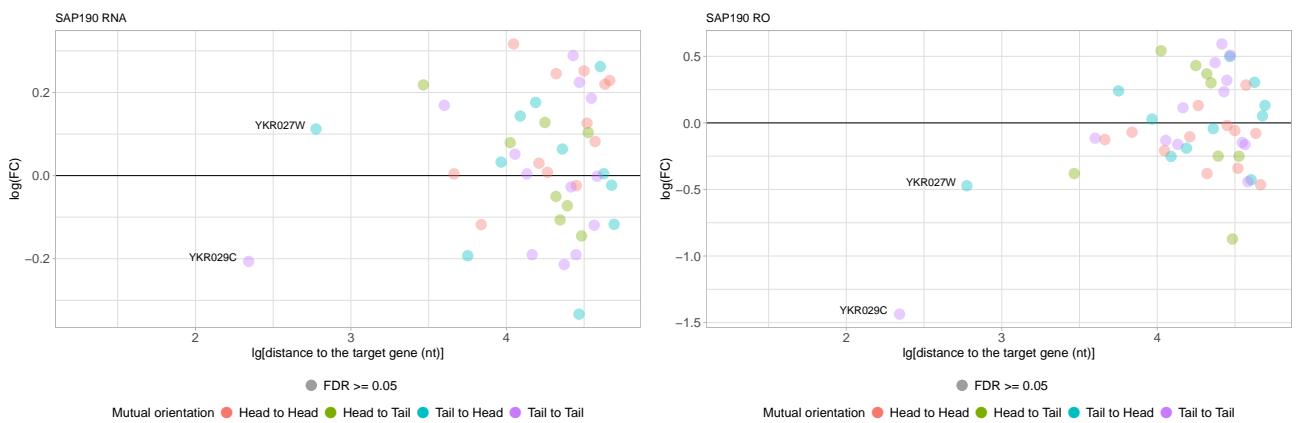


Рис. 22. Для локуса *SAP190*, аналогично описанию выше