Роль структуры хроматина в регуляции экспрессии у дрожжей

Саша Галицына

Кристина Перевощикова

Оля Кузьмич

Введение

Дрожжи - модельный организм, который позволил изучить множество аспектов молекулярной биологии эукариот, однако в довольно молодой области структурной биологии хроматина этот организм по-прежнему остается мало исследованным. Так, показано образование структурных доменов в геноме дрожжей,[1] которые являются относительно компактными в сравнении с другими эукариотами и включают в себя небольшое количество генов. При этом у дрожжей образуются небольшие корегуляторные кластеры соседних генов.[2] Не известно, соответствуют ли структурные домены хроматина дрожжей и регуляторные домены генов. К примеру, для человека было продемонстрировано разбиение генома на корегуляторные кластеры и показано, что корегуляторные кластеры слабо сопоставимы со структурными доменами.[3] В данном проекте мы рассматриваем подобную взаимосвязь структурной организации генома и экспрессии генов у дрожжей.

Материалы и методы

Мы провели анализ нескольких типов данных из набора статей. Так, с помощью Hi-C карт с разрешением в 500 нуклеотидов[4] были предсказаны структурные домены путем подсчета показателя инсуляции[5] (размер рамки 2000*2000 нуклеотидов, граница принималась, если она была сильнее, чем 40% границ).

I. Для подсчета корреляции экспрессии генов был собран агрегированный датасет RNA-seq из нескольких источников.[6] Значения экспрессий генов были нормализованы и представлены в виде Transcripts Per Million (TPM), таким образом было возможно производить сравнение экспрессий между датасетами. Всего было получено 26 состояний экспрессии. За среднюю корреляцию экстрессий в домене было взято среднее от всех попарных спирмановских корреляций генов, принадлежащих домену. Также были подсчитаны корреляции генов по принципу все-со-всеми в пределах одной хромосомы. Корреляции ген-ген были переведены в матрицу бин-бин для отрисовки теплокарт и их сопоставления с Hi-C картами. Выделение корегуляторных доменов производилось глобально методом k-means[7] (cluster number = 2).

II. Для изучения влияния вставки кассеты KanMX[8] были отобраны только делеции тех генов, которые не приводили к значимому изменению экспрессии генов на других хромосомах (log2(FC) > 1.5). Таким образом, были исключены глобальные регуляторы, а значит был сделан фокус на изучении локальных изменений, которые вносит кассета. Для изучения осталось 1054 делеции. Далее были подсчитаны все попарные расстояния между стартом кассеты и стартами генов в пределах одной хромосомы, а также определена взаимная ориентация “head-to-head”, “tail-to-tail”, “head-to-tail” или “tail-to-head”.

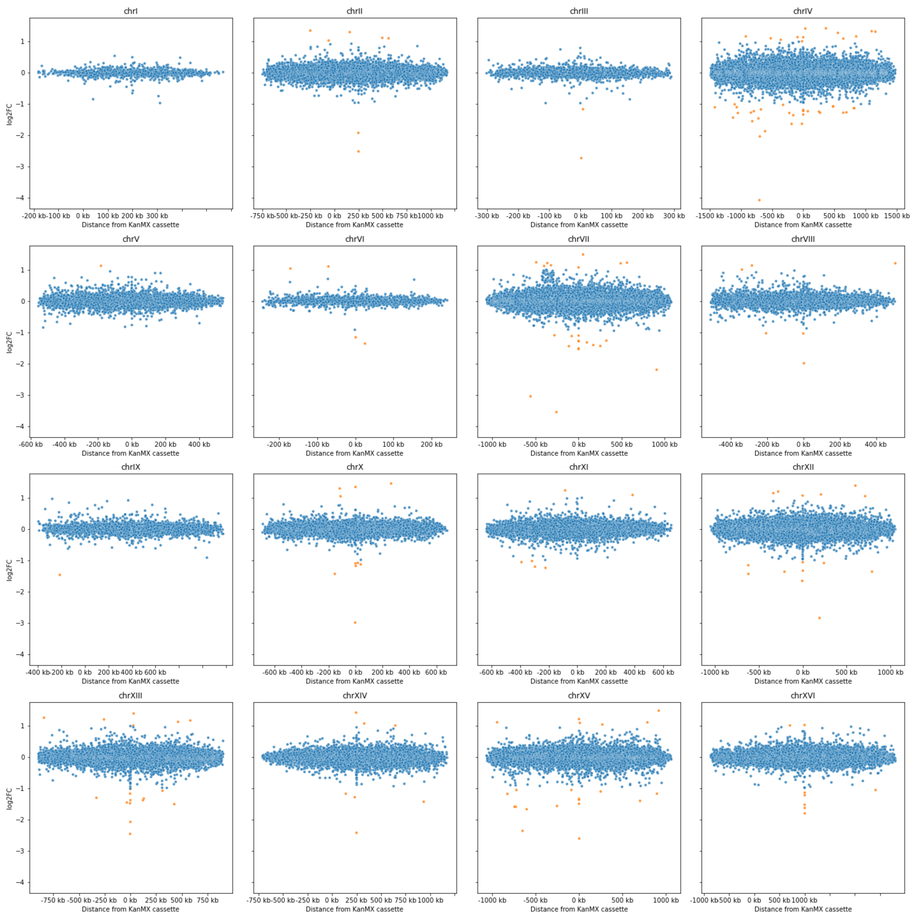


Рис. 4: Похромосомная демонстрация эффекта вставки касеты KanMX на экспрессию окружающих генов. По оси x отмечены расстояния между стартом рассматриваемого гена и касетой. Таким образом, гены расположенные выше, имеют отрицательную дистанцию. По оси y отмечен двоичный логарифм изменения экспрессии генов. Заметно снижение эффекта при увелении расстояния от вставки. Оранжевым выделены точки, имеющие |log2(FC)| > 1. Примечательно преимущественное скопление таких точек в окрестностях касеты. Причины локального и дальнего влияния вставки на экспрессиюю генов остаются неясными.

Дальнейшие планы по проекту

Далее мы планируем изучить локальную кластеризацию генов в регуляторных доменах и сопоставить их с доменами структурными. Кроме того, остается интересным изучение механизма изменения экспрессии генов, соседних с касетой KanMX. Мы планируем проанализировать взаимосвязь изменения экспрессии генов и их положения в одном структурном и\или регуляторном домене. В дальнейшем также планируется изучить взаимосвзять с эпигенетическими метками и уровнем открытости хроматина. Интересным для нас остается вопрос активности транскрипционных факторов в стуктурных и регуляторных доменах.

I. Коэкспрессия генов не связана с доменной огранизацией дрожжей

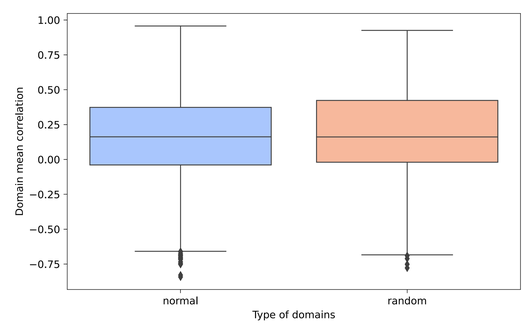


Рис. 1: Распределение средней корреляции Спирмана по структурным доменам. Приведено сравнение для реальных доменов, выделенных методом инсулляции и случайных доменов. Количество случайных доменов на хромосому сохранялось таким же, как и в реальном случае. Длины случайных доменов повторяют распределение настоящих. Согласно тесту Манна-Уитни показатели средней корреляции и в первом, и во втором типе доменов не различаются (p_value = 0.90)

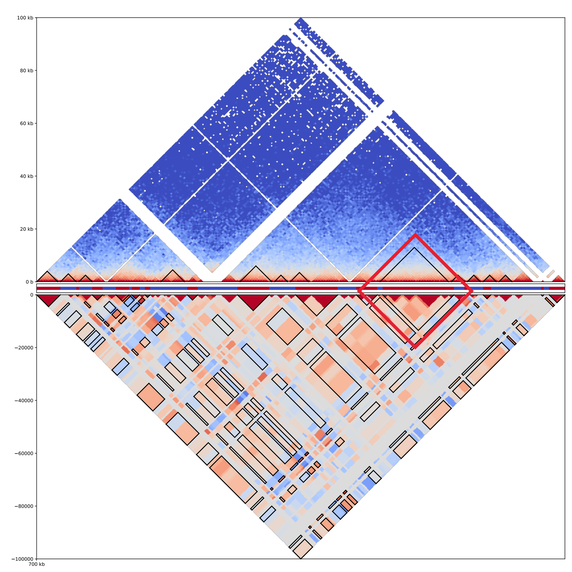


Рис. 2: Фрагмент 12-ой хромосомы. Сверху видна Hi-C карта с разметкой структурных доменов, полученных методом подсчета инсулляции. Ниже представлена теплокарта корреляций генов, расположенных на этом же участке хромосомы. Также выделен глобальный корегуляторный кластер, выделенный методом k-means (cluster number = 2). Примечательно, что даже такая разбивка на кластеры демострирует некоторое сопоставление корегуляторных кластеров и структурных доменов (выделено красным). Полоса между теплокартами демонстрирует разбиение всех генов на 2 кластера: синий и красный. Отрисованы границы красного кластера

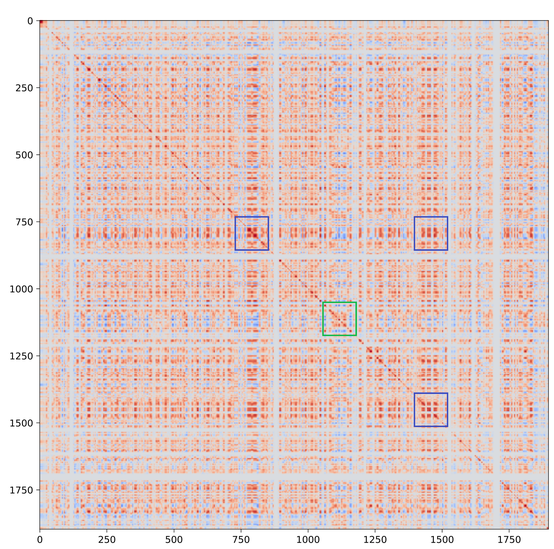
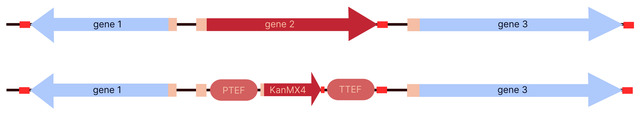


Рис. 3: Теплокарта корреляции экспрессии генов на примере 16 хромосомы. Карта представлена с разрешением в 500 нуклеотидов. Яркая диагональ демонстрирует корреляцию 1 гена с самим собой. Корреляция изменяется в диапазоне от -1 до 1 (от синего к красному). Различимы некоторые локальные корегуляторные кластеры по диагонали и их взаимодействие друг с другом. Примечательно, что некоторые кластеры взаимодействуют друг с другом положительно (выделено синим), а некоторые (выделено зеленым) оказывают противоположный эффект на другие участки хромосомы.

II. Вставка кассеты KanMX влияет на локальную экспрессию генов

Вставка кассеты KanMX производится на месте делетированного гена. Таким образом становится возможным изучение фенотипов дрожжей с делециями генов. Было показано, однако, что вставка кассеты приводит к значимому изменению экспрессии генов-ближайших соседей, что потенциально может приводить к неверной аннотации функции делетированного гена.[9] В нашем проекте мы хотели изучить более широкие локальные изменения, которые вносит вставка.



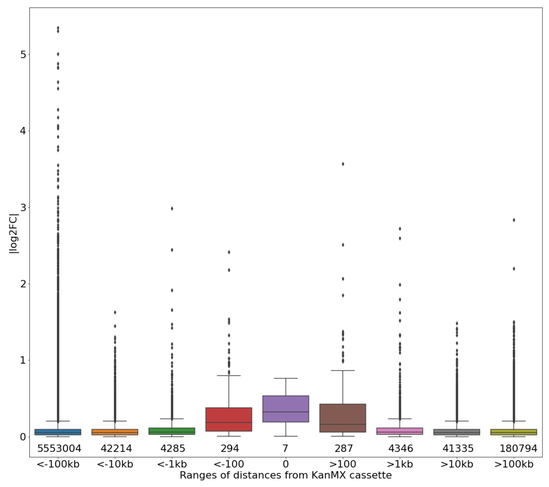


Рис. 5: Демонстарция эффекта вставки кассеты KanMX на экспрессию генов в зависимости от расстояния. По оси x выделены диапазоны расстояний от страта кассеты, по оси y представлен модуль log2(FC). Легко видеть пик влияния вставки на расстоянии [-1kb, 1kb] от старта кассеты. С отдалением от вставки эффект снижается. Однако, интерес также представляют значительные изменения эспрессии генов на большом расстоянии.

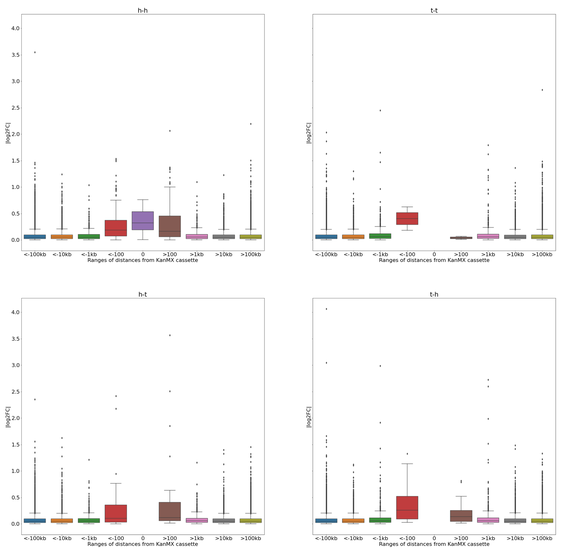


Рис. 6: Эффект от вставки кассеты по категориям взаимного расположения генов: 'head-to-head', 'tail-to-tail', 'head-to-tail', 'tail-to-head'. По оси x отмечены промежутки расстояний от старта кассеты. Ось y демонстрирует модуль от log2(FC).

При изучении эффекта на экспрессию генов вставки кассеты KanMX было замечено следующее. Очень часто наблюдалось сильное изменение экспрессии генов-соседей со вставкой. В целом было замечено, что влияние на экспрессию генов снижается с увеличением расстояния от KanMX кассеты. Возможно, это связано с локальными изменениями структуры хроматина, вызванными экспрессией кассеты, так как она находится под сильным промотором. Наша работа может служить основой для проверки гипотезы о том, что изменение экспрессии генов связано с нахождением в одном структурном и\или регуляторном домене.

Использованные источники

1. Hsieh T-HS, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell. 2015;162: 108–119.
2. Cohen BA, Mitra RD, Hughes JD, Church GM. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet. 2000;26: 183–186.
3. Soler-Oliva ME, Guerrero-Martínez JA, Bachetti V, Reyes JC. Analysis of the relationship between coexpression domains and chromatin 3D organization. PLoS Comput Biol. 2017;13: e1005708.
4. Hsieh T-HS, Fudenberg G, Goloborodko A, Rando OJ. Micro-C XL: assaying chromosome conformation at length scales from the nucleosome to the entire genome. doi:10.1101/071357
5. Venev S, Abdennur N, Goloborodko A, Flyamer I, Fudenberg G, Nuebler J, et al. open2c/cooltools: v0.4.1. 2021 [cited 5 Aug 2022]. doi:10.5281/zenodo.5214125
6. GEO datasets: GSE114077,GSE115556,GSE135568,GSE153609,GSE175794
7. Grisel O, Mueller A, Lars, Gramfort A, Louppe G, Prettenhofer P, et al. scikit-learn/scikit-learn: scikit-learn 1.1.1. 2022 [cited 5 Aug 2022]. doi:10.5281/zenodo.6563718
8. Kemmeren P, Sameith K, van de Pasch LAL, Benschop JJ, Lenstra TL, Margaritis T, et al. Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. Cell. 2014;157: 740–752.
9. Egorov AA, Alexandrov AI, Urakov VN, Makeeva DS, Edakin RO, Kushchenko AS, et al. A standard knockout procedure alters expression of adjacent loci at the translational level. Nucleic Acids Res. 2021;49: 11134–11144.