

УДК 577.2+616-006.66

ПРОСТРАНСТВЕННАЯ ОРГАНИЗАЦИЯ ХРОМАТИНА ПРОМОТОРНОЙ ОБЛАСТИ ГЕНА *KLF5* В КЛЕТКАХ ПРОТОКОВОЙ АДЕНОКАРЦИНОМЫ ПОДЖЕЛУДОЧНОЙ ЖЕЛЕЗЫ

© 2024 г. М. В. Зиновьева^а, Л. Г. Николаев^{а, *}^аИнститут биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова
Российской академии наук, Москва, 117997 Россия

*e-mails: lev@ibch.ru, kinvel@gmail.com

Поступила в редакцию 22.01.2024 г.

После доработки 05.03.2024 г.

Принята к публикации 05.03.2024 г.

Протоковая аденокарцинома поджелудочной железы (Pancreatic Ductal AdenoCarcinoma, PDAC) характеризуется неблагоприятным прогнозом и плохо поддается современным методам лечения. Моделью развития PDAC может служить ряд клеточных культур, отражающих различные степени дифференцировки и злокачественности опухоли. Высокодифференцированные низкоклеточные клетки отличаются повышенной экспрессией гена *KLF5*. Белок *KLF5* – яркий представитель многофункциональных факторов транскрипции, участвует во множестве клеточных процессов, в частности, связанных с различными видами рака. Методом циркулярной идентификации конформации хромосомы (4C-seq) нами изучена пространственная организация хроматина, включающего регуляторные области гена *KLF5*, в высокодифференцированных клетках PDAC Capan2 с высоким уровнем экспрессии *KLF5* и в низкодифференцированных клетках PDAC MIA PaCa2 с низким уровнем экспрессии этого гена. Выявлены существенные различия в числе и распределении контактов регуляторной области *KLF5* с другими областями хроматина в этих типах клеток: в клетках Capan2 число контактов значительно выше. Существует корреляция между уровнем экспрессии соседствующих с *KLF5* генов и интенсивностью их контактов с регуляторной областью *KLF5*, что указывает на согласованность их экспрессии, возможно, в рамках транскрипционной фабрики. Для Capan2 характерен высокий уровень контактов регуляторной области *KLF5* со свободным от генов участком генома, содержащим кластер ассоциированных с PDAC полиморфных сайтов (SNP/InDel). Таким образом, при повышении степени злокачественности клеток наблюдается снижение общего числа контактов промоторной области гена *KLF5* и уровня экспрессии большинства окружающих *KLF5* генов.

Ключевые слова: протоковая аденокарцинома поджелудочной железы, *KLF5*, хроматин, пространственная структура, регуляция транскрипции

DOI: 10.31857/S0026898424050069, EDN: HUOYYD

ВВЕДЕНИЕ

Протоковая аденокарцинома поджелудочной железы (Pancreatic Ductal AdenoCarcinoma, PDAC) занимает четвертое место среди причин смертности от рака [1], характеризуется неблагоприятным

прогнозом (5-летняя выживаемость менее 10%) и плохо поддается существующим методам лечения [2, 3].

Для понимания механизмов развития онкопатологий важно изучать взаимодействие различных генов и регуляцию их активности. Современные методы молекулярного картирования позволяют выявить в геноме регуляторные области целевых генов и оценить влияние их окружения. Эпигенетические механизмы, регулирующие такие сети, например, ремоделирование хроматина путем модификации гистонов [4] и изменения пространственной организации хроматина [5], также привлекают

Сокращения: PDAC – протоковая аденокарцинома поджелудочной железы; 4C-seq – циркулярная идентификация конформации хромосомы; SNP – однонуклеотидный полиморфизм; FDR – доля ложных заключений (false discovery rate); TAD – топологически ассоциированный домен; FPKM – число фрагментов, картирующихся в 1000 п.н., на миллион прочтений (Fragments Per Kilobase Million).

значительное внимание. В данной работе нами изучено влияние геномного окружения гена транскрипционного фактора *KLF5* на прогрессию PDAC человека.

Развитие PDAC сопровождается дерегулированием согласованной активности транскрипционных факторов, в частности, нарушением регуляторных взаимодействий в транскрипционных сетях, необходимых для правильной дифференцировки и развития клеток поджелудочной железы, что направляет их развитие к агрессивному злокачественному фенотипу [6].

Моделью развития PDAC может служить ряд клеточных культур, отражающих различные степени дифференцировки и злокачественности (grade) опухоли. Показано, что в ряду линий клеток CFPAC1, Capan2, HPAF2, Capan1, Bx-PC3, AsPC1, PANC1, MIA PaCa2 повышается степень злокачественности и понижается степень дифференцировки опухолевых клеток, при этом происходит согласованное повышение или понижение уровней транскрипции ряда генов транскрипционных факторов [7]. В частности, высокодифференцированные низкоккачественные клетки характеризуются повышенной транскрипцией генов *ELF3*, *FOXA1*, *KLF5*, *HNFI1B* и др., тогда как в низкодифференцированных и высокозлокачественных клетках на высоком уровне транскрибируются, в том числе, гены *ZEB1*, *GATA2*, *ETV5* и *TCF12*. Эти данные подтверждены нами ранее для этих и некоторых других факторов транскрипции как на уровне транскрипции, так и на уровне синтеза соответствующих белков [8]. Мы предположили, что клетки линий MIA PaCa2 и Capan2 могут наилучшим образом моделировать соответственно мезенхимальный (низкодифференцированный, высокозлокачественный) и эпителиальный (высокодифференцированный, низкоккачественный) этапы развития PDAC.

Один из ключевых факторов транскрипции, определяющих прогрессию PDAC [9], – Kruppel-подобный фактор 5 (Kruppel-Like Factor 5, *KLF5*), принадлежит к семейству транскрипционных факторов KLF, состоящему более чем из 20 членов [10]. Ген *KLF5* располагается вблизи одного из локусов предрасположенности к PDAC, содержащих несколько SNP, ассоциированных с PDAC [11–13]. Фактор транскрипции *KLF5* является ярким представителем многофункциональных белков, он вовлечен во множество клеточных процессов, определяющих, в частности, патологические изменения при различных видах рака [9, 14], сердечно-сосудистых заболеваниях [15], болезнях почек [16] и др. Ген *KLF5* принимает

непосредственное участие и в прогрессии PDAC у мышей [17], он экспрессируется на высоком уровне в низкоккачественных клетках PDAC в культуре [7, 8].

В настоящей работе методом циркулярной идентификации конформации хромосомы (Circular Chromosome Conformation Capture, 4C-seq) нами изучена пространственная организация хроматина, включающего регуляторные области гена *KLF5*, в культурах высокодифференцированных клеток PDAC (клетки Capan2) с высоким уровнем экспрессии *KLF5* и низкодифференцированных клеток PDAC MIA PaCa2 с низким уровнем экспрессии этого гена. Такой подход позволил нам выявить изменения в пространственной организации хроматина вблизи промоторной области гена *KLF5* при переходе клеток от низкоккачественного к высококкачественному состоянию и, соответственно, при переходе экспрессии гена *KLF5* от высокого уровня к низкому.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Культивирование клеток. Клеточную линию рака поджелудочной железы человека MIA PaCa2 (ATCC CRL-1420) культивировали в среде DMEM/F12, содержащей 10% эмбриональной сыворотки крупного рогатого скота (FBS), 100 ед./мл пенициллина, 100 мкг/мл стрептомицина и 0.25 мг/мл амфотерицина (“Gibco”, “Thermo Fisher Scientific”, США). Клеточную линию PDAC человека Capan2 (ATCC HTB-80) культивировали в среде RPMI 1640 с добавлением 12.5% FBS и тех же антибиотиков. Клетки выращивали при 5% CO₂, 37°C и влажности 90%.

Получение 4C библиотек. Для получения 4C библиотек использовали в основном протокол van de Werken H.J. и соавт. [18] с модификациями. Клетки (~10⁷) обрабатывали 2.25%-ным формальдегидом в течение 10 мин, собирали центрифугированием и ресуспендировали в 5 мл 50 мМ Трис-HCl pH 7.5, 150 мМ NaCl, 5 мМ EDTA, 0.5% NP-40, 1% Тритона X-100 с добавлением 25 мкл коктейля ингибиторов протеаз (“Sigma-Aldrich” P8340, США). Далее клетки обрабатывали эндонуклеазой Hind III (FastDigest, “Thermo Scientific”, США) в соответствии с протоколом [18] полноту расщепления контролировали электрофорезом в агарозном геле. Фермент инактивировали (20 мин, 80°C), ДНК лигировали при высоком разведении с использованием лигазы (5 ед./мкл) и буфера для лигирования (“Thermo Scientific”). Далее ДНК очищали согласно протоколу, расщепляли Dpn II (“New England Biolabs” R0543T, США) и повторно лигировали. Для каждого типа клеток получено

по два независимых образца сшитой библиотеки, а также дополнительно по контрольной библиотеке, полученной по той же схеме, но без добавления формальдегида (несшитый контроль).

В качестве якорных точек (viewpoints, baits) были выбраны два фрагмента Hind III–Dpn II генома человека (K1 и K3), максимально сближенных с точкой начала транскрипции гена *KLF5* и с последовательностью [19], необходимой для активности гена (см. дополнительный рис. S1 см. Дополнительные материалы в электронном виде по DOI статьи и на сайте http://www.molecbio.ru/downloads/2024/5/supp_Zinovyeva_rus.zip).

Библиотеки последовательностей, лигирующихся с этими якорями, далее получали при помощи двустадийной обращенной ПЦР. Структуры праймеров для амплификации методом обращенной ПЦР приведены в табл. 1. Там же приведены структуры вспомогательных последовательностей, необходимых для масштабного секвенирования на секвенаторах Illumina. Окончательная структура библиотек после двух стадий ПЦР приведена на рис. 1а.

ПЦР библиотек проводили в два этапа — на первом этапе к последовательностям библиотеки добавляли индекс и праймеры P7 и read-

ing primer (Illumina, США), на втором этапе (step-out) к 5'-концу добавляли последовательность, содержащую праймер P5 (Illumina, США). Оба этапа ПЦР проводили при двух температурах отжига — два первых цикла при температуре отжига геномной части праймера (52–55°C), а затем 30–32 цикла при более высокой температуре отжига полного праймера (60–65°C). Реакцию первого этапа проводили в объеме 800 мкл (16 × 50 мкл), второго этапа — в объеме 100 мкл (4 × 25 мкл) с использованием полимеразы Encyclo (“Евроген”, Россия) и амплификатора PTC-100 (“MJ Research”, США). Продукт очищали согласно [18]. Условия амплификации каждой из библиотек приведены в табл. S1 (см. Дополнительные материалы на сайте http://www.molecbio.ru/downloads/2024/5/supp_Zinovyeva_rus.zip), а картина электрофореза амплифицированных библиотек в агарозном геле — на рис. 1б.

Смесь равных количеств библиотек (суммарно ~400 нг) передана в компанию “Евроген” для секвенирования.

Секвенирование и картирование. Смесь библиотек секвенировали при помощи системы Illumina NovaSeq 6000 в формате односторонних прочтений 1 × 100 п.н. В результате полу-

Таблица 1. Структура праймеров для ПЦР

Библиотека	Праймер
<i>Праймеры Hind III (H-primers) для первого этапа ПЦР*</i>	
Библиотека K1H-3а	5'ACACGACGCTCTTCCGATCT/ATCACG/CATGCTTTCATTTACAAAAGCTT
K1H-4а	5'ACACGACGCTCTTCCGATCT/CGATGT/CATGCTTTCATTTACAAAAGCTT
K1H-5	5'ACACGACGCTCTTCCGATCT/TTAGGC/CATGCTTTCATTTACAAAAGCTT
K1H-6	5'ACACGACGCTCTTCCGATCT/CAGATC/CATGCTTTCATTTACAAAAGCTT
K1H-7	5'ACACGACGCTCTTCCGATCT/ACTTGA/CATGCTTTCATTTACAAAAGCTT
K1H-8	5'ACACGACGCTCTTCCGATCT/GATCAG/CATGCTTTCATTTACAAAAGCTT
K3H-3а	5'ACACGACGCTCTTCCGATCT/ATCACG/TTTTGTGCAATGAAAAAGCTT
K3H-4а	5'ACACGACGCTCTTCCGATCT/CGATGT/TTTTGTGCAATGAAAAAGCTT
K3H-5	5'ACACGACGCTCTTCCGATCT/TTAGGC/TTTTGTGCAATGAAAAAGCTT
K3H-6	5'ACACGACGCTCTTCCGATCT/CAGATC/TTTTGTGCAATGAAAAAGCTT
K3H-7	5'ACACGACGCTCTTCCGATCT/ACTTGA/TTTTGTGCAATGAAAAAGCTT
K3H-8	5'ACACGACGCTCTTCCGATCT/GATCAG/TTTTGTGCAATGAAAAAGCTT
<i>Праймеры Dpn II (D-primers) для первого этапа ПЦР**</i>	
Праймер K1D2N	5'CAAGCAGAAGACGGCATACGAGAT/TCATCAGCCATTCTCTGAAGG
Праймер K3D3N	5'CAAGCAGAAGACGGCATACGAGAT/TCCAGGTTCTACATTGGCTTG
<i>Праймеры для второго (step-out) этапа ПЦР***</i>	
Праймер SO2	5'AATGATACGGCGACCAACCGAGATCT/ACACTCTTCCCTACACGACGCTCTTCCGATCT

* Слева направо (части праймера разделены косой чертой): Reading primer (Illumina, частично), индекс, уникальная часть со стороны Hind III.
** Слева направо: праймер P7 (Illumina), уникальная часть со стороны Dpn II.
*** Слева направо: праймер P5 (Illumina), Reading primer (Illumina). Со стороны Dpn II использовали праймеры K1D2N и K3D3N.

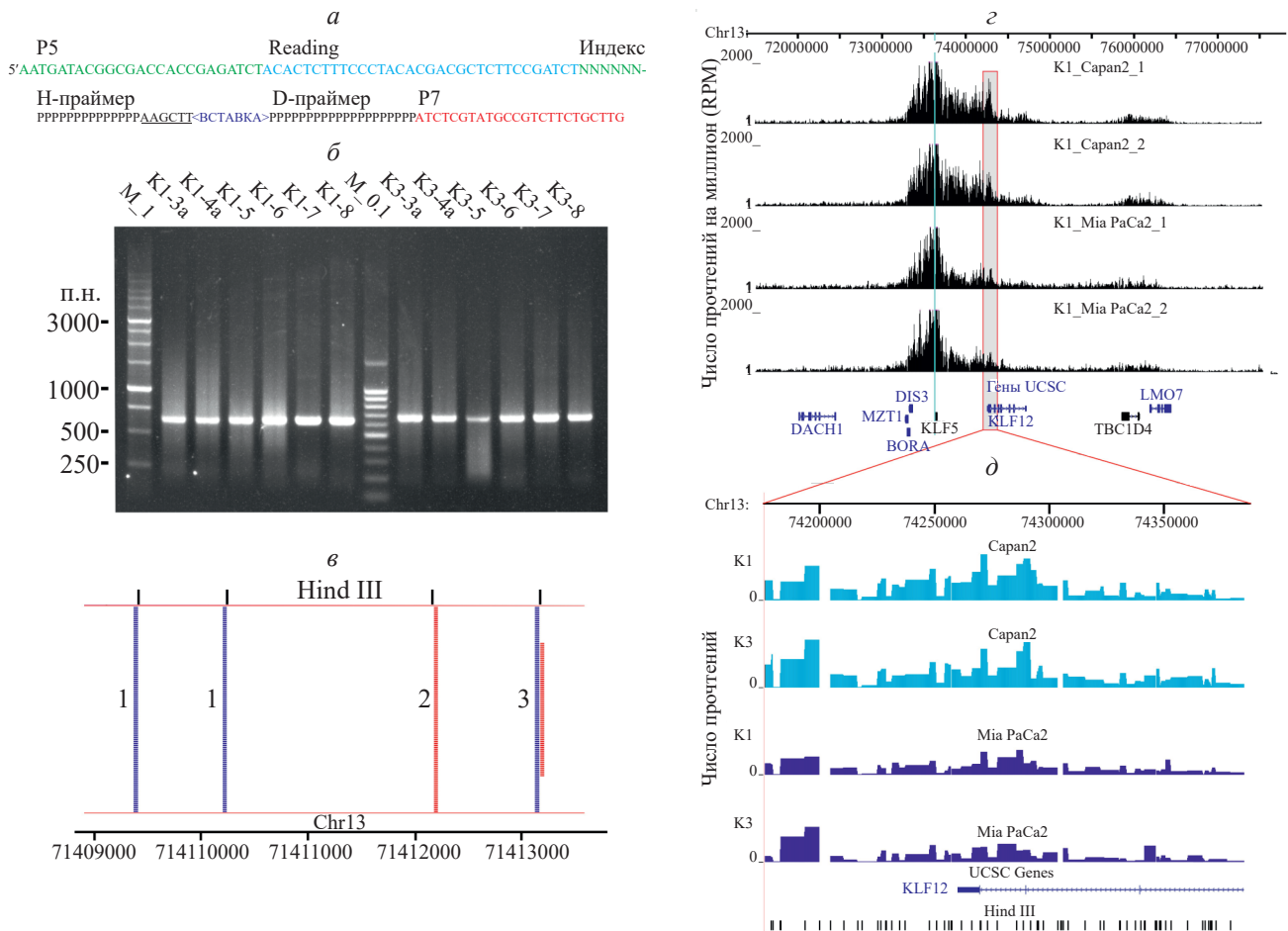


Рис. 1. Структура и секвенирование библиотек. *a* – Структура библиотек. P5, P7, Reading – последовательности для секвенирования Illumina; H- и D-праймеры – уникальные праймеры для обращенной ПЦР со стороны сайтов Hind III и Dpn II соответственно; индекс – индексные последовательности; вставка – амплифицированная геномная последовательность, контактирующая с якорем. *б* – Электрофорез в агарозном геле 4С библиотек после второго этапа ПЦР-амплификации. Номер библиотеки соответствует номеру праймера Hind III (см. табл. 1). M_1 и M_0.1 – ДНК-маркеры 1 и 0.1 т.п.н. соответственно. *в* – Типичное расположение прочтений относительно сайтов расщепления Hind III в участке хромосомы 13. Представлены три варианта лигирования: (1) по одной из цепей; (2) по противоположной цепи; (3) по обеим цепям с разной эффективностью. Рисунок подготовлен при помощи пакета SeqMonk (<https://www.bioinformatics.babraham.ac.uk/projects/seqmonk>). *г* – Расположение контактных участков в области хромосомы 13 длиной 6 млн.п.н. вокруг гена *KLF5*. По оси ординат отложено число прочтений на миллион (RPM) в соответствующей библиотеке, синяя вертикальная линия показывает положение якорных последовательностей. В двух верхних дорожках показано распределение контактов якорной последовательности K1 и клеток Capan2 в двух независимых экспериментах; в следующих двух дорожках – те же данные для клеток MIA PaCa2. *д* – Расположение контактных участков в небольшой (200 т.п.н.) области хромосомы 13 вблизи гена *KLF12*. По оси ординат отложено число прочтений в соответствующей библиотеке. Две верхние дорожки позволяют сравнить распределение контактов якорных последовательностей K1 и K3 в клетках Capan2, следующие две дорожки – те же для клеток MIA PaCa2. Внизу показано расположение сайтов расщепления Hind III.

чено более 150 млн прочтений. Прочтения были демультиплицированы при помощи программы grep пакета SeqKit [20] и картированы на геном человека (сборка GRCh37/hg19) при помощи программы bowtie ver. 1 [21] (опции – best, -n 2). Данные по эффективности картирования приведены в дополнительной табл. S2.

В результате получены и секвенированы 12 библиотек. По четыре библиотеки содержали (в двух повторностях) потенциальные контакты регуляторной области гена *KLF5* с отдаленными областями генома в двух линиях клеток

(MIA PaCa2 и Capan2) и двух соседних якорных точках (viewpoints). Четыре контрольные библиотеки (те же клетки и якорные точки без повторностей) получены без фиксации клеток формальдегидом. Результаты секвенирования занесены в базу данных NCBI SRA (<https://www.ncbi.nlm.nih.gov/sra/>) и доступны под номером PRJNA1016108.

Анализ контактов и обработка результатов. Файлы картированных библиотек преобразовывали в формат bam при помощи пакета Samtools [22] и наиболее вероятные

места контактов выявляли с использованием специализированных программных пакетов R3Cseq [23] FourCSeq [24] и peakC [25] в программной среде R (<https://www.r-project.org/about.html>). Дифференциальные контакты выявляли с использованием R-пакета 4C-ker [26].

Дополнительные ресурсы. Использовали также R-пакеты xlsx (<https://cran.r-project.org/web/packages/xlsx/>), openxlsx (<https://cran.r-project.org/web/packages/openxlsx/index.html>), broman (<https://cran.r-project.org/web/packages/broman/index.html>), REDseq (<https://bioconductor.org/packages/release/bioc/html/REDseq.html>).

В геномном анализе использовали UCSC Genome Browser (<http://genome.ucsc.edu>, [27, 28]), 3D Genome Browser (<http://3dgenome.fsm.northwestern.edu/index.html>), а также базы данных проекта ENCODE (<https://www.encodeproject.org>, [29]), NCBI GEO (<https://www.ncbi.nlm.nih.gov/geo/>, [30]) и GeneCards (<https://www.genecards.org>, [31]).

Для первичной визуализации и контроля результатов секвенирования и картирования применяли программу анализа данных секвенирования SeqMonk (<https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Выбор клеток, конструирование и картирование библиотек. Для исследования пространственной структуры хроматина в области промотора гена *KLF5* из имеющегося спектра клеток, происходящих из тканей на разных стадиях PDAC, мы выбрали две линии, соответствующие крайним стадиям развития опухоли. Клетки линии MIA PaCa2 конститутивно экспрессируют мезенхимальные гены и моделируют высокозлокачественную PDAC, тогда как линия Saran2 относится к эпителиальному типу и по спектру экспрессируемых белков может служить моделью низкоккачественной опухоли [7, 32]. Ранее показали, что ген *KLF5* и соответствующий белок экспрессируются на высоком уровне в низкоккачественных клетках PDAC, включая Saran2, и на два порядка слабее в высокозлокачественных MIA PaCa2 [7, 8].

О регуляции экспрессии самого гена *KLF5* известно довольно мало. Промотор этого гена в геноме человека находится внутри области $-751...+232$ и содержит последовательность CGCCGCGC, связывающую фактор транскрипции EGR1 в гладкомышечных клетках сосудов [33]. Промотор *KLF5* не имеет ТАТА-последовательности и содержит сайт связывания фактора SP1 с координатами $-239...-219$ [34].

С использованием репортерного гена выявлена минимально необходимая для активности промотора область длиной 186 п.н. ($-68...+118$), содержащая несколько GC-богатых сайтов. Связывание с одним из них фактора SP3 также необходимо для базальной транскрипции *KLF5* [19]. Один из предполагаемых энхансеров гена *KLF5*, активный в клетках HT29 колоректальной аденокарциномы человека, располагается в области примерно $+371000$ п.н. и имеет длину 277 п.н. [35] (рис. S1, см. электронное приложение).

Белок *KLF5* взаимодействует с суперэнхансерами [36] и вместе с рядом других факторов транскрипции выполняет роль мастер-регулятора при раке пищевода [37, 38]. В свою очередь, сам ген *KLF5* может находиться под контролем суперэнхансера [39].

Мы исследовали пространственную организацию хроматина вблизи промоторной области гена *KLF5* в высокодифференцированной низкоккачественной линии клеток PDAC (клетки Saran2) с высоким уровнем экспрессии *KLF5* и в низкодифференцированной высокозлокачественной линии MIA PaCa2 с пониженным уровнем экспрессии этого гена. Использовали метод 4C-seq и расщепление ферментами Hind III и Dpn II. Для повышения надежности результатов выбрали две якорные последовательности, прилегающие к ближайшему к точке начала транскрипции *KLF5* сайту расщепления Hind III (см. рис. S1). Получены, секвенированы и картированы на геном человека (сборка GRCh37/hg19) 12 библиотек (по две биологических повторности для каждого из якорей и двух типов клеток, плюс по одной повторности для контрольных библиотек без обработки формальдегидом, см. электронное приложение, табл. S2), в среднем около 10^7 прочтений на библиотеку. Типы расположения прочтений относительно сайтов расщепления Hind III приведены на рис. 1в.

Воспроизводимость результатов. Прочтения полученных библиотек в формате bedGraph визуализовали в геномном браузере UCSC, после чего сравнивали биологические повторы для якоря K1 и клеток MIA PaCa2 и Saran2 (рис. 1з), а также результаты для якорных последовательностей K1 и K3 (рис. 1д). Видно, что общая картина распределения числа прочтений хорошо воспроизводится в независимых повторах, а также сходна при использовании близких по расположению якорных последовательностей. Коэффициент корреляции Пирсона в случае биологических повторов составил 0.98–0.99. Для дальнейшего анализа объединили результаты, полученные для якорей K1 и K3.

Идентификация контактов. В данной работе мы ограничили анализ контактов относительно небольшой областью хромосомы 13 человека, включающей примерно по 3 млн п.н. с 5'- и 3'-стороны от якорной точки (точнее, с 70671801 по 76628882 п.н.). По нашим данным, эта область содержит основную часть обнаруженных контактов. Более удаленные от *KLF5* контактирующие области будут предметом отдельного анализа.

Для анализа данных 4C-seq и выявления наиболее достоверных участков контакта мы применили три наиболее активно используемых с этой целью программных пакета: R3Cseq [23] FourCSeq [24] и peakC [25]. Так как в этих программах используются разные алгоритмы определения базовой линии, а также разные форматы представления и визуализации результатов, прямое сравнение данных, полученных с их помощью, затруднительно. Поэтому для сравнения результатов мы использовали общий для этих программ параметр — *P*-значение, с поправкой на долю ложных заключений (FDR, False Discovery Rate), определяющее в данном случае значимость отличия величины наблюдаемого сигнала от расчетной базовой линии. Для удобства этот параметр представлен в виде отрицательного логарифма $-\lg(P\text{-value})$, а его распределение в обеих линиях клеток и трех использованных программах можно видеть на рис. 2. Область длиной 15 т.п.н. по обе стороны от якоря не рассматривали в связи с высокой вероятностью неспецифического лигирования [23].

Контакты считали значимыми в тех случаях, когда наблюдаемое *P*-значение было ниже порогового (0.01 для R3Cseq и FourCSeq, 0.05 для peakC, рис. 2) одновременно во всех трех пакетах программ. Чтобы получить представление о степени достоверности контакта каждой точки с учетом результатов всех трех программ, мы рассчитали объединенное *P*-значение. Так как *P*-значения, генерируемые тремя программами с разными алгоритмами и методами определения базовой линии в одной и той же точке контакта, могут отличаться на порядок и более (см. рис. 2), мы нормировали эти значения для каждой отобранной точки контакта (представлены в виде $-\lg(P\text{-value})$) на сумму значений всех отобранных контактов в исследованной области генома. Коэффициент нормирования был принят равным единице для FourCseq. После этого суммировали нормированные значения $-\lg(P\text{-value})$ в каждой точке контакта. Такая операция позволяет уравнивать вклад каждой из программ в суммарное значение объединенного параметра $-\lg(P\text{-value})$, а величина параметра будет указывать на степень значимости данного контакта. Результаты приведены в дополнительной табл. S3.

Обращает на себя внимание тот факт, что число идентифицированных контактов промоторной области гена *KLF5* с другими областями исследованного района генома в клетках Saran2, активно экспрессирующих этот ген, значительно (108/69) превышает количество контактов в клетках MIA PaCa2.

Дифференциальные контакты. Распределение различий в контактах в исследуемой области клеток Saran2 и MIA PaCa2 представлено на рис. 2 в виде распределения значений $-\lg(P\text{-value})$. Различия считали значимыми в тех случаях, когда наблюдаемое значение *P* было ниже 0.01.

Все отобранные контактирующие области были разделены на три группы — специфичные для клеток Saran2 (с высоким уровнем экспрессии *KLF5*), специфичные для клеток MIA PaCa2 (с низким уровнем экспрессии *KLF5*) и общие для обоих типов клеток (дополнительная табл. S4). Специфичными для клеток Saran2 оказались 66 контактов, для клеток MIA PaCa2 — 27, общими в обоих типах клеток были 42 контакта.

Расположение контактов относительно TAD. Как хорошо видно из рис. 2, кластеры контактирующих с якорем *KLF5* участков располагаются в соответствии с TAD (см., например, обзор [5]). Согласно данным, полученным с использованием всех программных пакетов, основная масса контактов, как и ожидалось, располагается внутри TAD4, в котором находится ген *KLF5*. В то же время наблюдается заметная область контактов в TAD7, общая для обоих типов клеток, и две области контактов в TAD5 и TAD8, заметно выраженные только в клетках Saran2. Оставшиеся четыре TAD не содержат последовательностей, контактирующих с *KLF5*. Контактирующие с регуляторной областью *KLF5* участки генома будут рассмотрены более подробно. Таким образом, границы TAD не являются непреодолимыми при организации контактов якорной области с удаленными участками генома, хотя TAD играют определенную роль в ограничении этих контактов.

Контакты и экспрессия генов. Исследуемый нами участок длиной 6 млн п.н. содержит 10 кодирующих белки генов (не считая *KLF5*), образующих два кластера по четыре гена, еще два гена не входят в кластеры (рис. 3а), а также протяженные участки, не содержащие генов (“генные пустыни”). Из рис. 3а видно, что распределение контактов якорной области *KLF5* в целом соответствует распределению генов — плотность контактов высока в областях высокой плотности генов и, наоборот, контакты практически отсутствуют в “генных пустынях”. Важным исключением является область длиной

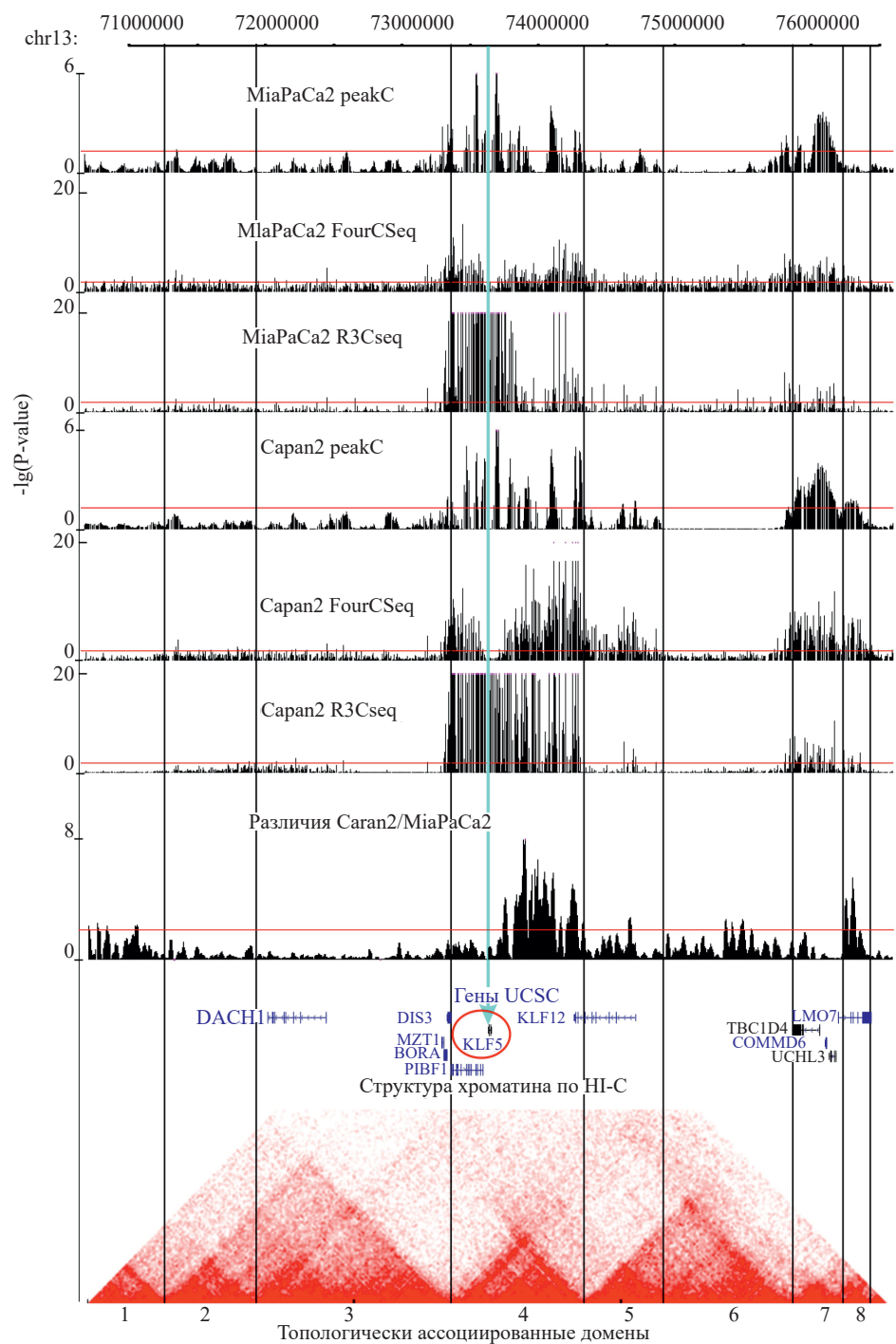


Рис. 2. Распределение контактов якорного участка гена *KLF5* в области хромосомы 13 человека размером 6 млн п.н. (по 3 млн п.н. с 5'- и 3'-сторон от якоря). Приведены результаты анализа клеток MIA PaCa2 и Capan2, полученные при помощи программ peakC, FourCSeq и R3Cseq и представленные в виде $-\lg(P\text{-value})$. Горизонтальная линия соответствует пороговому P -значению (0.05 для peakC и 0.01 для FourCseq и R3Cseq). Для дифференциальных контактов, выявленных при помощи программы 4C-ker, пороговое P -значение составило 0.01. Вертикальная стрелка показывает положение якорной последовательности. Внизу представлено расположение восьми топологически ассоциированных доменов хроматина (Topologically Associated Domains, TAD) согласно данным Hi-C для клеток Capan1 [44].

около 600 т.п.н. между генами *KLF5* и *KLF12*, не содержащая генов, но имеющая большое число контактов, причем преимущественно специфичных для Capan2.

Уровень транскрипции всех генов в клетках Capan2 и MIA PaCa2 (в единицах FPKM по данным RNA-seq [7]) также представлен на рис. 3а, рис. 4 и в табл. 2. Ген *DACH1*, один из

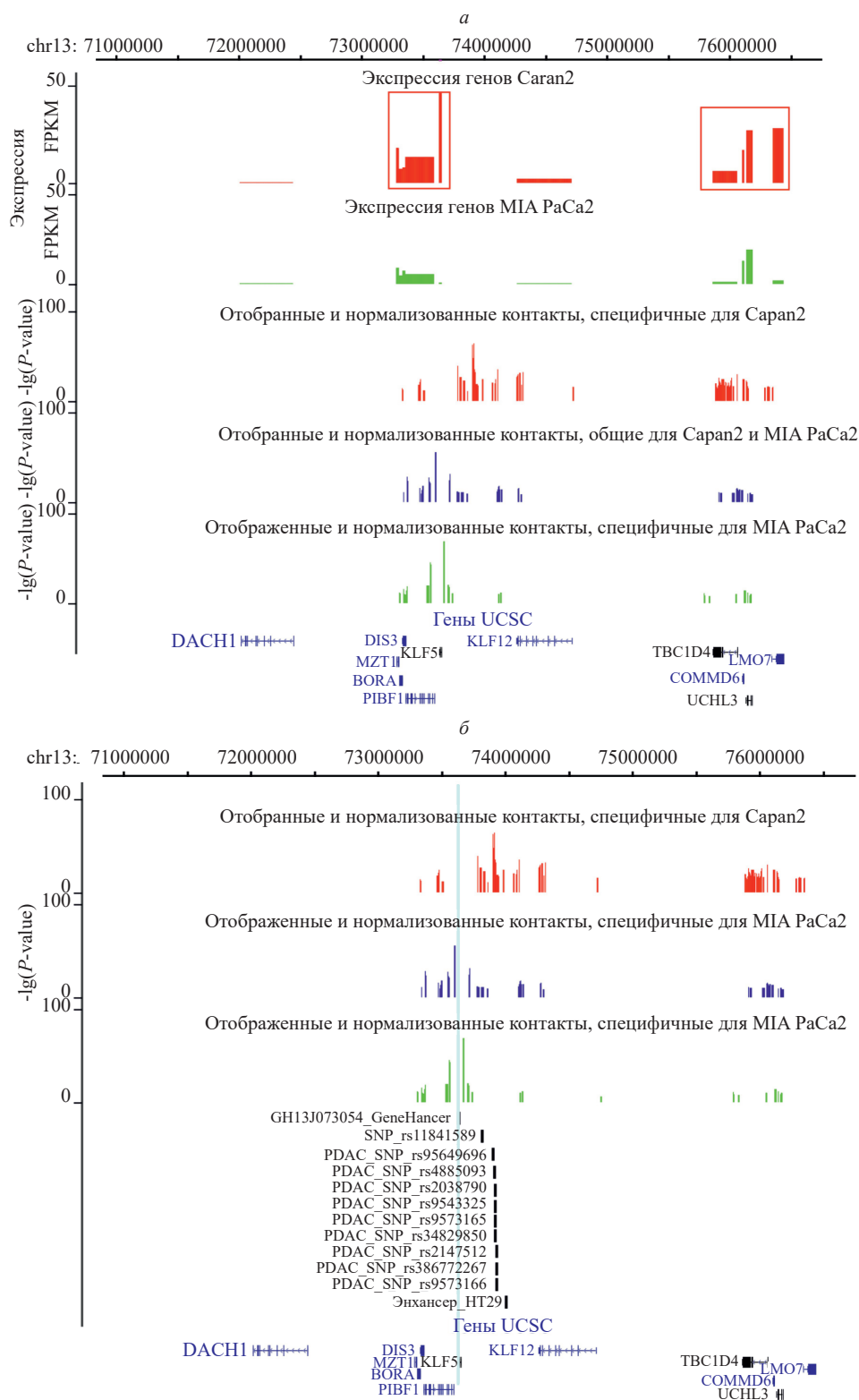


Рис. 3. Расположение отобранных и нормализованных участков контакта якорного участка гена *KLF5* в области хромосомы 13 человека размером 6 млн п.н. относительно некоторых функциональных элементов. Отдельно приведены области контактов, специфичные для клеток *Caran2*, *MIA PaCa2*, и общие для обеих линий клеток. *a* – Тканеспецифичность экспрессии генов в клетках *Caran2* и *MIA PaCa2*, связь уровня экспрессии гена и числа его контактов с промоторной областью *KLF5*. По осям ординат отложены значения FPKM для оценки уровня экспрессии генов и $-\lg(P\text{-value})$ для участков контакта. *б* – Расположение относительно контактов точек однонуклеотидного полиморфизма и делеций/вставок (SNP/indel), ассоциированных с PDAC по данным полногеномного анализа (GWAS, Genome-Wide Association Studies, см. [11, 13]).

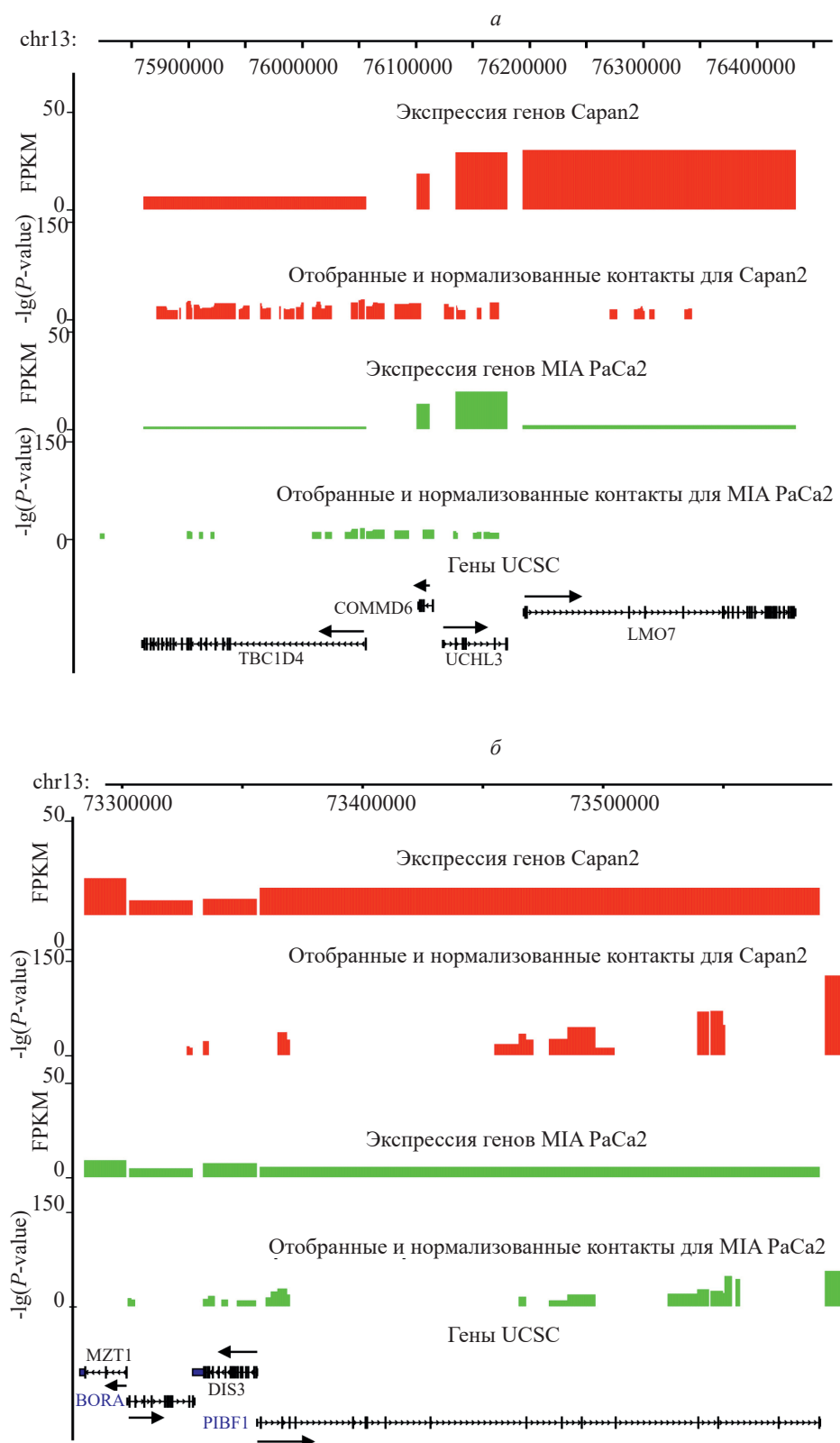


Рис. 4. Связь уровней экспрессии генов и интенсивности их контактов с якорным участком *KLF5* в двух богатых генами участках хромосомы 13 (обведены квадратами на рис. 3а). а – Участок с координатами 75 800 000 – 76 500 000 п.н. б – Участок с координатами 73280000–73600000 п.н. По осям ординат отложены значения FPKM для оценки уровня экспрессии генов и $-\lg(P\text{-value})$ для участков контакта. Стрелками указано направление транскрипции генов. Уровни экспрессии генов приведены по данным [7] (доступны в базе данных NCBI GEO, GSE64558). Показаны все контакты, идентифицированные в каждой линии клеток (как специфичные для данной линии клеток, так и общие).

Таблица 2. Транскрипция генов в исследуемой области

Ген	Координаты на хромосоме 13	Транскрипция в клетках линии Саран2, FPKM*	Транскрипция в клетках линии MIA PaCa2, FPKM*
<i>DACH1</i>	72012097–72441330	0.223	0.186
<i>MZT1</i>	73284460–73301740	19.19	8.89
<i>BORA</i>	73303079–73329354	7.50	4.53
<i>DIS3</i>	73333933–73355970	8.24	7.25
<i>PIBF1</i>	73357608–73590057	14.23	5.33
<i>KLF5</i>	73636011–73650024	98.10	0.57
<i>KLF12</i>	74269627–74708066	2.26	0.036
<i>TBC1D4</i>	75860928–76055903	6.54	0.87
<i>COMMD6</i>	76100725–76111941	18.35	12.82
<i>UCLH3</i>	76134943–76179948	28.93	18.93
<i>LMO7</i>	76378463–76432149	30.12	1.59

* По данным [7].

двух генов, не входящих в кластеры, практически не экспрессируется в обоих типах клеток и не содержит ни одного контакта. Ген *KLF12*, напротив, экспрессируется на заметном уровне в клетках Саран2 и слабо экспрессируется в MIA PaCa2. Соответственно, с этим геном колокализуется значительное число контактов в клетках Саран2 и намного меньшее их число в MIA PaCa2. Возможно, существует прямая корреляция между уровнем экспрессии гена и числом его контактов с промоторной областью *KLF5*, экспрессия которого очень высока в клетках Саран2 и практически отсутствует в MIA PaCa2 (рис. 3а).

Верно ли это предположение для других генов? На рис. 4 представлены в более крупном масштабе два кластера по четыре гена в каждом. На рис. 4а видно, что гены, экспрессия которых в клетках Саран2 заметно выше, чем в MIA PaCa2, характеризуются значительно бóльшим числом колокализованных контактов, специфичных для Саран2. В частности, контакты, специфичные для MIA PaCa2, полностью отсутствуют в области гена *LMO7*. Если же ген экспрессируется в обеих линиях на сходном уровне (*COMMD6*, *UCLH3*), то и число локализованных контактов для них близко. Более сложная картина наблюдается в другом кластере генов (рис. 4б). Если в случае гена *PIBF1* приведенное выше правило выполняется — ген экспрессируется в обеих линиях клеток и содержит близкое число контактов как в Саран2, так и в MIA PaCa2, то подобная картина не наблюдается у трех других генов. В целом, однако, предложенное правило выполняется для семи из 10 генов.

Приведенная выше картина хорошо укладывается в концепцию “транскрипционных фабрик” (transcription factories), то есть внутриядерных частиц размером ~80 нм, содержащих около 10 молекул РНК-полимеразы II и сопутствующие транскрипционные факторы, способных согласованно транскрибировать несколько генов одновременно (см. обзор [40]). Наши данные показывают, что на одной и той же фабрике могут транскрибироваться далеко отстоящие друг от друга в геноме гены, относящиеся к разным TAD, как, например, гены *KLF5* и *LMO7* (TAD 4 и 8, рис. 2).

Таким образом, согласованно экспрессироваться гены могут не только за счет двунаправленного промотора [41], но и находясь на большом удалении друг от друга в геноме в составе одной и той же транскрипционной фабрики. Нельзя также исключить влияния на них одного и того же энхансера *in trans*, как мы предположили ранее [42].

Контакты и модификации гистонов. Регуляция транскрипции. Как можно видеть из рис. 5, расположение в исследуемой области меток активного хроматина (H3K27ac, H3K4me3 и H3K4me1) в обеих линиях клеток хорошо согласуется между собой и с расположением активных генов. Также можно видеть, что области высокой плотности модификаций гистона H3 в целом совпадают по положению с областями контактов регуляторной области *KLF5* (как сказано выше, контакты в области, непосредственно примыкающей к гену *KLF5*, исключены нами из рассмотрения). Обратное неверно — имеются протяженные участки, обогащенные контактами, но не содержащие

Таблица 3. Регуляторные элементы, потенциально находящиеся в контакте с промоторной областью гена *KLF5*

Координаты на хромосоме 13	Функциональные сегменты хроматина*	Модификации гистона H3	Ген	Транскрипция, FPKM (Capan2/MIA PaCa2)
Специфичные для MIA PaCa2				
73302581–73305305	Промотор	+++	<i>BORA</i> , интрон 1	4.53
73347814–73362175	Промотор	+++	<i>DIS3</i> , экзоны 2–7	7.25
76106595–76115846	Промотор	+++	<i>COMMD6</i> , интрон 1	12.8
Специфичные для Capan2				
73909003–73913511	PDAC SNP rs4885093	++	–	–
73918091–73919235	PDAC SNP rs34829850	-	–	–
73924989–73937996	PDAC SNP rs9573166, rs386772267	++	–	–
74277232–74285530	Сайт CTCF	++	<i>KLF12</i> , интрон 7	2.26
74709920–74715862	Промотор/Энхансер	+++	<i>KLF12</i> , 5'-область	-
75989063–75991974	Энхансер	++	<i>TBC1D4</i> , интрон 1	6.54
76125003–76128339	Промотор	+++	<i>UCHL3</i> , интрон 2	28.9
76336300–76342754	Промотор/Энхансер	+++	<i>LMO7</i> , интрон 5	30.1
Общие для Capan2 и MIA PaCa2				
74288571–74291701	Энхансер	++	<i>KLF12</i> , интрон 6–7	2.26/0.04
76056653–76061748	Промотор	+++	<i>TBC1D4</i> , 5'-область	-

* По [43] либо экспериментально найденные (см. текст).

модифицированного гистона H3. Особенно хорошо это видно на примере клеток MIA PaCa2 (три такие области обведены кружками на рис. 5). Таким образом, указанные выше активирующие модификации гистона H3 (открытый хроматин) не являются необходимыми для образования контактов. Тем не менее в ряде случаев при увеличении числа контактов гена с регуляторной областью *KLF5* в его области появляется или существенно возрастает степень всех трех модификаций гистона H3, что отчетливо видно на примере генов *LMO7* и *KLF12*.

Контакты и PDAC-ассоциированные SNP. В ходе полногеномного анализа ассоциаций (GWAS, см. [11–13]) выявлен ряд точек SNP/indel, ассоциированных с PDAC. Эти участки располагаются плотным кластером в небольшой области, расположенной между генами *KLF5* и *KLF12* и характеризующейся также высокой плотностью контактов и модификаций гистона H3, строго специфичных для клеток Capan2. В линии MIA PaCa2 модификации гистона H3 и контакты в этой области полностью отсут-

ствуют (рис. 3б). Можно предположить, что присутствие одного или нескольких SNP мешает образованию контактов, что в свою очередь нарушает регуляцию *KLF5* и корегулируемых генов, что и приводит к прогрессии PDAC.

Функционально значимые области. Из областей, контактирующих с промоторной областью гена *KLF5*, отбирали функционально значимые, удовлетворяющие следующим критериям:

- контактирующая область содержит значительный уровень модифицированного гистона H3 (модификации K27Ac, K4me1 и K4me3, характерные для активного хроматина);
- область содержит также либо экспериментально охарактеризованные регуляторные элементы, либо сегменты хроматина с предсказанной функцией по данным [43].

Некоторые функциональные области представлены в табл. 3, полный список всех контактов и их функциональных характеристик приведен в дополнительных табл. S5, S6 и S7.

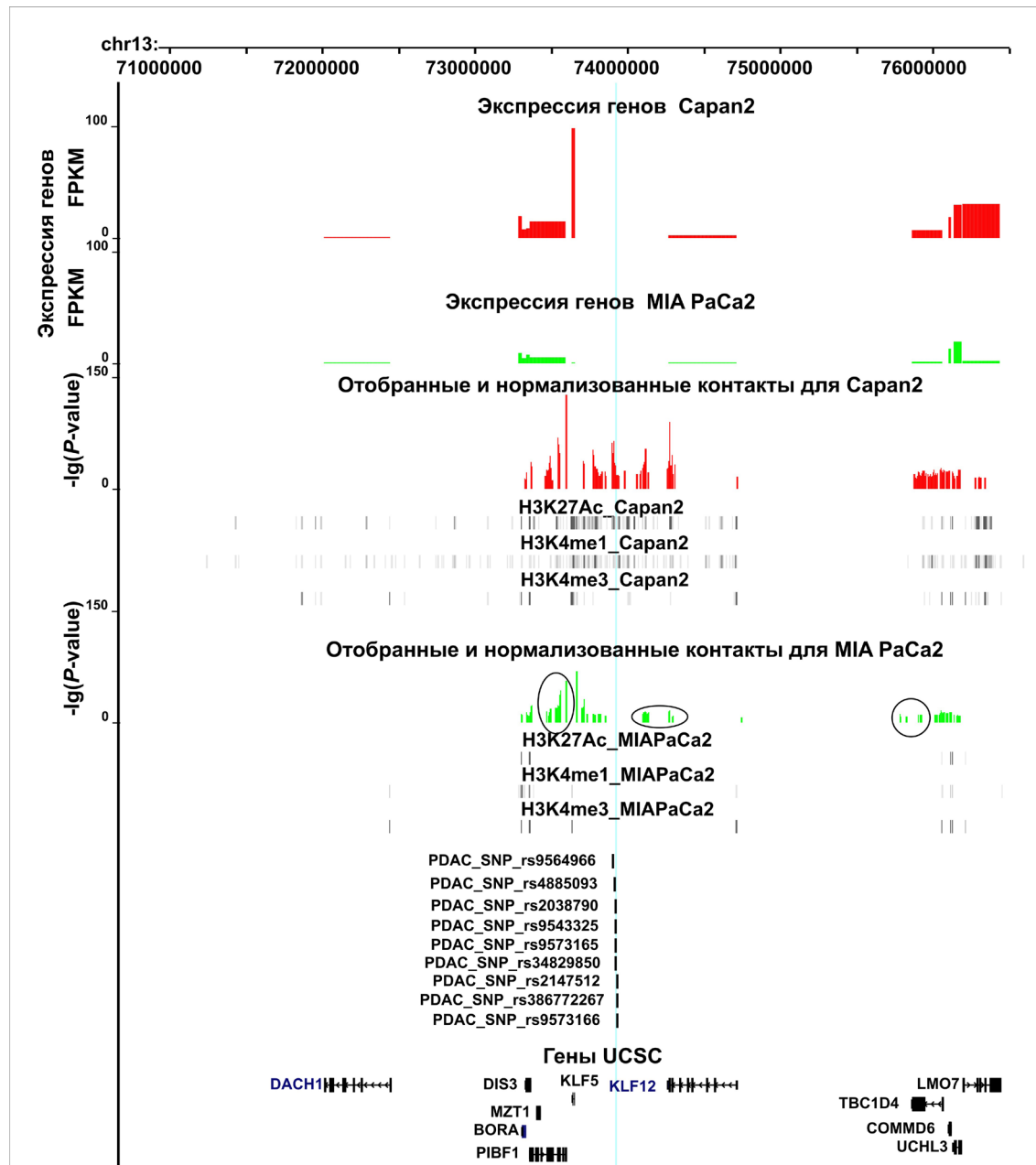


Рис. 5. Распределение геномных контактов якорного участка гена *KLF5* и модификаций гистона H3 в хроматине клеток Capan2 и MIA PaCa2 (доступны в базе данных NCBI GEO, GSE64560) в области хромосомы 13 длиной 6 млн п.н. По осям ординат отложены значения FPKM для оценки уровня экспрессии генов и $-\lg(P\text{-value})$ для участков контакта. Уровни экспрессии генов приведены по данным [7] (доступны в базе данных NCBI GEO, GSE64560). Показаны все контакты, идентифицированные в каждой линии клеток (как специфичные для данной линии клеток, так и общие).

Одна из функциональных областей контакта, специфичная для клеток MIA PaCa2, располагается вблизи точек начала транскрипции генов *MZT1* и *BORA*, транскрибирующихся в противоположных направлениях, и обладает свойствами потенциального промотора. Возможно, она является промотором одного из этих генов, либо общим для них двунаправленным промотором. Оба этих гена экспрессируются как в клетках

Capan2, так и в MIA PaCa2 на достаточно высоком уровне (табл. 3). Два других потенциальных промотора располагаются в 5'-кодирующих областях генов *DIS3* и *COMMD6*.

Специфичная для клеток Capan2 область длиной около 40 т.п.н. содержит группу SNP, ассоциированных с PDAC (рассмотрена выше). Кроме того, для Capan2 специфичны две области контакта со свойствами промотора. Одна из них

(74709920–74715862) расположена вблизи точки начала транскрипции *KLF12*, другая — во втором интроне (и вблизи точки начала транскрипции) *UCHL3*. При этом *KLF5* экспрессируется преимущественно в Саран2, а *UCHL3* — в обеих линиях клеток. Также для Саран2 специфичны сайт связывания фактора транскрипции CTCF (потенциальный инсультатор) в интроне 7 гена *KLF12* и две области контакта со свойствами энхансера, расположенные в интроне 1 гена *TBC1D4* и интроне 5 гена *LMO7*. Эти два элемента можно рассматривать как кандидаты на роль энхансера *KLF5* в этих клетках.

Общим для двух линий клеток оказался контакт со свойствами промотора, расположенный вблизи точки начала транскрипции гена *TBC1D4*; при этом в клетках Саран2 ген экспрессируется намного сильнее, возможно, за счет описанного в предыдущем абзаце энхансера. Область контакта, расположенная вблизи 3'-конца гена *KLF12*, также обладает свойствами потенциального энхансера.

ЗАКЛЮЧЕНИЕ. РАЗЛИЧИЯ МЕЖДУ КЛЕТКАМИ САРАН2 И MIA PACA2

Из полученных нами и уже опубликованных данных можно сделать следующие заключения о различиях в структуре и регуляции гена *KLF5* и окружающего этот ген хроматина в клетках линии Саран2 (низкоккачественные высокодифференцированные, высокая экспрессия *KLF5*) и линии MIA PaCa2 (высококкачественные и низкодифференцированные, низкая экспрессия *KLF5*):

— число и распределение контактов регуляторной области *KLF5* с другими областями хроматина в этих типах клеток существенно различаются (рис. 2) — в линии Саран2 число контактов значительно выше;

— в линии Саран2 интенсивность активирующих хроматин модификаций гистона H3 в исследованной области значительно выше, чем в MIA PaCa2. Этому соответствует и повышенный в клетках Саран2 уровень экспрессии большинства генов этой области;

— существует корреляция между уровнем экспрессии генов и интенсивностью их контактов с регуляторной областью *KLF5*, что указывает на согласованность их экспрессии, возможно, в рамках транскрипционной фабрики;

— только для клеток линии Саран2 характерен высокий уровень контактов регуляторной области *KLF5* со свободным от генов участком генома, содержащим кластер ассоциированных с PDAC SNP.

Таким образом, в рамках использованной модели при повышении степени злокачественности клеток в исследованной области уменьшается общее число контактов промоторной области гена *KLF5*, понижается степень модификации гистона H3, характерных для активного хроматина, и уровень экспрессии большинства генов. Для подтверждения этих выводов требуется проведение экспериментов на большем числе клеточных линий.

Настоящая работа финансировалась за счет средств бюджета Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова РАН. Никаких дополнительных грантов на проведение или руководство данным конкретным исследованием получено не было.

Работа выполнена без привлечения людей или животных в качестве объектов исследования.

Авторы заявляют об отсутствии конфликта интересов.

Дополнительные материалы. Дополнительная информация размещена в электронном виде по DOI статьи и на сайте http://www.molecbio.ru/downloads/2024/5/supp_Zinovyeva_rus.zip.

СПИСОК ЛИТЕРАТУРЫ

1. Collisson E.A., Bailey P., Chang D.K., Biankin A.V. (2019) Molecular subtypes of pancreatic cancer. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 207–220.
2. Rawla P., Sunkara T., Gaduputi V. (2019) Epidemiology of pancreatic cancer: global trends, etiology and risk factors. *World J. Oncol.* **10**, 10–27.
3. Kung H.C., Yu J. (2023) Targeted therapy for pancreatic ductal adenocarcinoma: mechanisms and clinical study. *MedComm.* (2020). **4**, e216.
4. Pandey S., Gupta V.K., Lavania S.P. (2023) Role of epigenetics in pancreatic ductal adenocarcinoma. *Epigenomics.* **15**, 89–110.
5. Hafner A., Boettiger A. 2023. The spatial organization of transcriptional control. *Nat. Rev. Genet.* **24**, 53–68.
6. van Roey R., Brabletz T., Stemmler M.P., Armstark I. (2021) Dereglulation of transcription factor networks driving cell plasticity and metastasis in pancreatic cancer. *Front. Cell Dev. Biol.* **9**, 753456.
7. Diaferia G.R., Balestrieri C., Prosperini E., Nicoli P., Spaggiari P., Zerbi A., Natoli G. (2016) Dissection of transcriptional and cis-regulatory control of differentiation in human pancreatic cancer. *EMBO J.* **35**, 595–617.
8. Абжалимов И.Р., Зиновьева М.В., Николаев Л.Г., Копанцева М.Р., Копанцев Е.П., Свердлов Е.Д. (2017) Экспрессия генов транскрипционных

- факторов в линиях клеток, соответствующих разным стадиям прогрессии рака поджелудочной железы. *Докл. РАН.* **475**, 333–335.
9. Luo Y., Chen C. (2021) The roles and regulation of the KLF5 transcription factor in cancers. *Cancer Sci.* **112**, 2097–2117.
 10. Зиновьева М.В., Костина М.Б., Чернов И.П., Кондратьева Л.Г., Свердлов Е.Д. (2016) KLF5, новый игрок и новая мишень в постоянно меняющемся составе молекулярных двигателей рака поджелудочной железы. *Биоорганическая химия.* **42**, 668–674.
 11. Hoskins J.W., Ibrahim A., Emmanuel M.A., Man-miller S.M., Wu Y., O'Neill M., Jia J., Collins I., Zhang M., Thomas J.V., Rost L.M., Das S., Parikh H., Haake J.M., Matters G.L., Kurtz R.C., Bamlet W.R., Klein A., Stolzenberg-Solomon R., Wolpin B.M., Yarden R., Wang Z., Smith J., Olson S.H., Andreson T., Petersen G.M., Amundadottir L.T. (2016) Functional characterization of a chr13q22.1 pancreatic cancer risk locus reveals long-range interaction and allele-specific effects on DIS3 expression. *Hum. Mol. Genet.* **25**, 4726–4738.
 12. Wolpin B.M., Rizzato C., Kraft P., Kooperberg C., Petersen G.M., Wang Z., Arslan A.A., Beane-Freeman L., Bracci P.M., Buring J., Canzian F., Duell E.J., Gallinger S., Giles G.G., Goodman G.E., Goodman P.J., Jacobs E.J., Kamineni A., Klein A.P., Kolonel L.N., Kulke M.H., Li D., Malats N., Olson S.H., Risch H.A., Sesso H.D., Visvanathan K., White E., Zheng W., Abnet C.C., Albanes D., Andreotti G., Austin M.A., Barfield R., Basso D., Berndt S.I., Boutron-Ruault M.C., Brotzman M., Buchler M.W., Bueno-de-Mesquita H.B., Bugert P., Burdette L., Campa D., Caporaso N.E., Capurso G., Chung C., Cotterchio M., Costello E., Elena J., Funel N., Gaziano J.M., Giese N.A., Giovannucci E.L., Goggins M., Gorman M.J., Gross M., Haiman C.A., Hassan M., Helzlsouer K.J., Henderson B.E., Holly E.A., Hu N., Hunter D.J., Innocenti F., Jenab M., Kaaks R., Key T.J., Khaw K.T., Klein E.A., Kogevinas M., Krogh V., Kupcinskas J., Kurtz R.C., LaCroix A., Landi M.T., Landi S., Le Marchand L., Mambrini A., Mannisto S., Milne R.L., Nakamura Y., Oberg A.L., Owzar K., Patel A.V., Peeters P.H., Peters U., Pezzilli R., Piepoli A., Porta M., Real F.X., Riboli E., Rothman N., Scarpa A., Shu X.O., Silverman D.T., Soucek P., Sund M., Talar-Wojnarowska R., Taylor P.R., Theodoropoulos G.E., Thornquist M., Tjonneland A., Tobias G.S., Trichopoulos D., Vodicka P., Wactawski-Wende J., Wentzensen N., Wu C., Yu H., Yu K., Zeleniuch-Jacquotte A., Hoover R., Hartge P., Fuchs C., Chanock S.J., Stolzenberg-Solomon R.S., Amundadottir L.T. (2014) Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat. Genet.* **46**, 994–1000.
 13. Petersen G.M., Amundadottir L., Fuchs C.S., Kraft P., Stolzenberg-Solomon R.Z., Jacobs K.B., Arslan A.A., Bueno-de-Mesquita H.B., Gallinger S., Gross M., Helzlsouer K., Holly E.A., Jacobs E.J., Klein A.P., LaCroix A., Li D., Mandelson M.T., Olson S.H., Risch H.A., Zheng W., Albanes D., Bamlet W.R., Berg C.D., Boutron-Ruault M.C., Buring J.E., Bracci P.M., Canzian F., Clipp S., Cotterchio M., de Andrade M., Duell E.J., Gaziano J.M., Giovannucci E.L., Goggins M., Hallmans G., Hankinson S.E., Hassan M., Howard B., Hunter D.J., Hutchinson A., Jenab M., Kaaks R., Kooperberg C., Krogh V., Kurtz R.C., Lynch S.M., McWilliams R.R., Mendelsohn J.B., Michaud D.S., Parikh H., Patel A.V., Peeters P.H., Rajkovic A., Riboli E., Rodriguez L., Seminara D., Shu X.O., Thomas G., Tjonneland A., Tobias G.S., Trichopoulos D., Van Den Eeden S.K., Virtamo J., Wactawski-Wende J., Wang Z., Wolpin B.M., Yu H., Yu K., Zeleniuch-Jacquotte A., Fraumeni J.F., Jr., Hoover R.N., Hartge P., Chanock S.J. (2010) A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat. Genet.* **42**, 224–228.
 14. Zhang Y., Yao C., Ju Z., Jiao D., Hu D., Qi L., Liu S., Wu X., Zhao C. (2023) Kruppel-like factors in tumors: key regulators and therapeutic avenues. *Front. Oncol.* **13**, 1080720.
 15. Palioura D., Lazou A., Drosatos K. (2022) Kruppel-like factor (KLF)5: an emerging foe of cardiovascular health. *J. Mol. Cell. Cardiol.* **163**, 56–66.
 16. Li J., Liu L., Zhou W.Q., Cai L., Xu Z.G., Rane M.J. (2021) Roles of Kruppel-like factor 5 in kidney disease. *J. Cell. Mol. Med.* **25**, 2342–2355.
 17. He P., Yang J.W., Yang V.W., Bialkowska A.B. (2018) Kruppel-like factor 5, increased in pancreatic ductal adenocarcinoma, promotes proliferation, acinar-to-ductal metaplasia, pancreatic intraepithelial neoplasia, and tumor growth in mice. *Gastroenterology.* **154**, 1494–1508 e1413.
 18. van de Werken H.J., de Vree P.J., Splinter E., Holwerda S.J., Klous P., de Wit E., de Laat W. (2012) 4C technology: protocols and data analysis. *Meth. Enzymol.* **513**, 89–112.
 19. Mihara N., Chiba T., Yamaguchi K., Sudo H., Yagishita H., Imai K. (2017) Minimal essential region for kruppel-like factor 5 expression and the regulation by specificity protein 3-GC box binding. *Gene.* **601**, 36–43.
 20. Shen W., Le S., Li Y., Hu F. (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One.* **11**, e0163962.
 21. Langmead B., Trapnell C., Pop M., Salzberg S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.
 22. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* **25**, 2078–2079.
 23. Thongjuea S., Stadhouders R., Grosveld F.G., Soler E., Lenhard B. (2013) r3Cseq: an R/Bioconductor package for the discovery of long-range genomic

- interactions from chromosome conformation capture and next-generation sequencing data. *Nucl. Acids Res.* **41**, e132.
24. Klein F.A., Pakozdi T., Anders S., Ghavi-Helm Y., Furlong E.E., Huber W. (2015) FourCSeq: analysis of 4C sequencing data. *Bioinformatics.* **31**, 3085–3091.
 25. Geeven G., Teunissen H., de Laat W., de Wit E. (2018) peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucl. Acids Res.* **46**, e91.
 26. Raviram R., Rocha P.P., Muller C.L., Miraldi E.R., Badri S., Fu Y., Swanzey E., Proudhou C., Snetkova V., Bonneau R., Skok J.A. (2016) 4C-ker: a method to reproducibly identify genome-wide interactions captured by 4C-Seq experiments. *PLoS Comput. Biol.* **12**, e1004780.
 27. Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M., Haussler D. (2002) The human genome browser at UCSC. *Genome Res.* **12**, 996–1006.
 28. Lee B.T., Barber G.P., Benet-Pages A., Casper J., Clawson H., Diekhans M., Fischer C., Gonzalez J.N., Hinrichs A.S., Lee C.M., Muthuraman P., Nassar L.R., Nguy B., Pereira T., Perez G., Raney B.J., Rosenbloom K.R., Schmelter D., Speir M.L., Wick B.D., Zweig A.S., Haussler D., Kuhn R.M., Haeussler M., Kent W.J. (2022) The UCSC Genome Browser database: 2022 update. *Nucl. Acids Res.* **50**, D1115–D1122.
 29. Moore J.E., Purcaro M.J., Pratt H.E., Epstein C.B., Shores N., Adrian J., Kawli T., Davis C.A., Dobin A., Kaul R., Halow J., Van Nostrand E.L., Freese P., Gorkin D.U., Shen Y., He Y., Mackiewicz M., Pauli-Behn F., Williams B.A., Mortazavi A., Keller C.A., Zhang X.O., Elhajjajy S.I., Huey J., Dickel D.E., Snetkova V., Wei X., Wang X., Rivera-Mulia J.C., Rozowsky J., Zhang J., Chhetri S.B., Victorson A., White K.P., Visel A., Yeo G.W., Burge C.B., Lecuyer E., Gilbert D.M., Dekker J., Rinn J., Mendenhall E.M., Ecker J.R., Kellis M., Klein R.J., Noble W.S., Kundaje A., Guigo R., Farnham P.J., Cherry J.M., Myers R.M., Ren B., Graveley B.R., Gerstein M.B., Pennacchio L.A., Snyder M.P., Bernstein B.E., Wold B., Hardison R.C., Gingeras T.R., Stamatoyannopoulos J.A., Weng Z. (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature.* **583**, 699–710.
 30. Barrett T., Wilhite S.E., Ledoux P., Evangelista C., Kim I.F., Tomashevsky M., Marshall K.A., Phillippy K.H., Sherman P.M., Holko M., Yefanov A., Lee H., Zhang N., Robertson C.L., Serova N., Davis S., Soboleva A. (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucl. Acids Res.* **41**, D991–995.
 31. Stelzer G., Rosen N., Plaschkes I., Zimmerman S., Twik M., Fishilevich S., Stein T.I., Nudel R., Lieder I., Mazor Y., Kaplan S., Dahary D., Warshawsky D., Guan-Golan Y., Kohn A., Rappaport N., Safran M., Lancet D. (2016) The GeneCards Suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics.* **54**, 1.30.1–1.30.33.
 32. Collisson E.A., Sadanandam A., Olson P., Gibb W.J., Truitt M., Gu S., Cooc J., Weinkle J., Kim G.E., Jakkula L., Feiler H.S., Ko A.H., Olshen A.B., Danenberg K.L., Tempero M.A., Spellman P.T., Hanahan D., Gray J.W. (2011) Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med.* **17**, 500–503.
 33. Kawai-Kowase K., Kurabayashi M., Hoshino Y., Ohyama Y., Nagai R. (1999) Transcriptional activation of the zinc finger transcription factor BTEB2 gene by Egr-1 through mitogen-activated protein kinase pathways in vascular smooth muscle cells. *Circ. Res.* **85**, 787–795.
 34. Chen C., Zhou Y., Zhou Z., Sun X., Otto K.B., Uht R.M., Dong J.T. (2004) Regulation of KLF5 involves the Sp1 transcription factor in human epithelial cells. *Gene.* **330**, 133–142.
 35. Takeda T., Yokoyama Y., Takahashi H., Okuzaki D., Asai K., Itakura H., Miyoshi N., Kobayashi S., Uemura M., Fujita T., Ueno H., Mori M., Doki Y., Fujii H., Eguchi H., Yamamoto H. (2022) A stem cell marker KLF5 regulates CCAT1 via three-dimensional genome structure in colorectal cancer cells. *Br. J. Cancer.* **126**, 109–119.
 36. Koutsi M.A., Poulidou M., Champezu L., Vatsellas G., Giannopoulou A.I., Piperi C., Agelopoulos M. (2022) Typical enhancers, super-enhancers, and cancers. *Cancers (Basel).* **14**, 4375.
 37. Chen L., Huang M., Plummer J., Pan J., Jiang Y.Y., Yang Q., Silva T.C., Gull N., Chen S., Ding L.W., An O., Yang H., Cheng Y., Said J.W., Doan N., Dinjens W.N., Waters K.M., Tuli R., Gayther S.A., Klempner S.J., Berman B.P., Meltzer S.J., Lin D.C., Koeffler H.P. (2020) Master transcription factors form interconnected circuitry and orchestrate transcriptional networks in oesophageal adenocarcinoma. *Gut.* **69**, 630–640.
 38. Jiang Y.Y., Jiang Y., Li C.Q., Zhang Y., Dakle P., Kaur H., Deng J.W., Lin R.Y., Han L., Xie J.J., Yan Y., Doan N., Zheng Y., Mayakonda A., Hazawa M., Xu L., Li Y., Aswad L., Jeitany M., Kanojia D., Guan X.Y., Said J.W., Yang W., Fullwood M.J., Lin D.C., Koeffler H.P. (2020) TP63, SOX2, and KLF5 establish a core regulatory circuitry that controls epigenetic and transcription patterns in esophageal squamous cell carcinoma cell lines. *Gastroenterology.* **159**, 1311–1327 e1319.
 39. Chen C.H., Yang N., Zhang Y., Ding J., Zhang W., Liu R., Liu W., Chen C. (2019) Inhibition of super enhancer downregulates the expression of KLF5 in basal-like breast cancers. *Int. J. Biol. Sci.* **15**, 1733–1742.
 40. Kimura H., Sato Y. (2022) Imaging transcription elongation dynamics by new technologies unveils the organization of initiation and elongation in transcription factories. *Curr. Opin. Cell Biol.* **74**, 71–79.

41. Didych D.A., Shamsutdinov M.F., Smirnov N.A., Akopov S.B., Monastyrskaya G.S., Uspenskaya N.Y., Nikolaev L.G., Sverdlov E.D. (2013) Human *PSENN* and *U2AF1L4* genes are concertedly regulated by a genuine bidirectional promoter. *Gene*. **515**, 34–41.
42. Smirnov N.A., Akopov S.B., Didych D.A., Nikolaev L.G. (2017) *In trans* promoter activation by enhancers in transient transfection. *Gene*. **603**, 15–20.
43. Hoffman M.M., Ernst J., Wilder S.P., Kundaje A., Harris R.S., Libbrecht M., Giardine B., Ellenbogen P.M., Bilmes J.A., Birney E., Hardison R.C., Duham I., Kellis M., Noble W.S. (2013) Integrative annotation of chromatin elements from ENCODE data. *Nucl. Acids Res.* **41**, 827–841.
44. Xu Z., Lee D.S., Chandran S., Le V.T., Bump R., Yasis J., Dallarda S., Marcotte S., Clock B., Haghani N., Cho C.Y., Akdemir K.C., Tyndale S., Futreal P.A., McVicker G., Wahl G.M., Dixon J.R. (2022) Structural variants drive context-dependent oncogene activation in cancer. *Nature*. **612**, 564–572.

Spatial Organization of Chromatin of *KLF5* Gene Promoter Region in Pancreatic Ductal Adenocarcinoma Cells

M. V. Zinovyeva¹, L. G. Nikolaev^{1,*}

¹*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997 Russia*

**e-mails: lev@ibch.ru, kinvel@gmail.com*

Pancreatic Ductal Adenocarcinoma (PDAC) is characterized by a poor prognosis and is poorly amenable to modern therapies. A range of cell cultures reflecting different degrees of tumor differentiation and malignancy can serve as a model of PDAC development. Highly differentiated low malignant cells are characterized by increased expression of the *KLF5* gene. The *KLF5* protein is a vivid representative of multifunctional transcription factors, and its involvement in a variety of cellular processes, particularly in the pathology of various cancers, has been demonstrated. We investigated the spatial organization of chromatin of regulatory regions of *KLF5* gene using highly differentiated Capan2 cells PDAC with a high level of *KLF5* expression and poorly differentiated MIA PaCa2 PDAC cells with a low level of this gene expression by circular chromosome conformation capture (4C-seq). It was shown that the number and distribution of contacts of the *KLF5* regulatory region with other chromatin regions are significantly different for these cell types; the number of contacts is significantly higher for Capan2 cells. There is a correlation between the expression level of genes close to *KLF5* and the intensity of their sequence contacts with the *KLF5* regulatory region, indicating that their expression is coordinated, possibly within the transcriptional factory. Only Capan2 is characterized by a high level of contacts of the *KLF5* regulatory region with the gene free region containing a cluster of PDAC-associated single nucleotide polymorphisms (SNP). Thus, the total number of contacts of the promoter region of the *KLF5* gene and the expression level of most of the surrounding *KLF5* genes decrease as the grade of cell malignancy increases.

Keywords: pancreatic ductal adenocarcinoma, *KLF5*, chromatin, spatial organization, transcription regulation