

ПОИСК РЕГУЛЯТОРНЫХ SNPs, СВЯЗАННЫХ С РАЗВИТИЕМ РАКА ТОЛСТОЙ КИШКИ, В ГЕНАХ *APC* И *MLH1*

Е.В. Антонцева¹, Л.О. Брызгалов¹, М.Ю. Матвеева¹, Е.В. Кашина¹,
Н.В. Чердынцева², Т.И. Меркулова¹

¹ Учреждение Российской академии наук Институт цитологии и генетики
Сибирского отделения РАН, Новосибирск, Россия, e-mail: schly@mail.ru;

² НИИ онкологии СО РАМН, Томск, Россия

С использованием экспериментальных данных проекта ENCODE по распределению в геноме человека областей связывания различных транскрипционных факторов (ChIP-seq) был проведен отбор полиморфизмов, попадающих в регуляторные области генов *APC* и *MLH1*. Регуляторную значимость таких полиморфизмов тестировали в экспериментах по задержке ДНК-зонда в геле белками ядерного экстракта клеточных линий человека. Оказалось, что более половины полиморфизмов, попадающих в область перекрывания более 8 пиков ChIP-seq, являются регуляторно значимыми.

Ключевые слова: однонуклеотидный полиморфизм, иммунопреципитация хроматина, регуляция экспрессии генов, рак толстой кишки.

Введение

Колоректальный рак занимает четвертое место в структуре онкологической заболеваемости в мире и третье место в России (Тимофеев, 2004). Основная проблема при терапии этого заболевания – обращение пациентов уже на поздних стадиях, когда выживаемость после оперативного вмешательства составляет 5%, в то время как своевременная постановка диагноза на ранних этапах увеличивает этот показатель до 80%. В этой связи нельзя переоценить значение работ по поиску молекулярных маркеров предрасположенности к этому заболеванию, что позволит выделить группу риска среди населения для более частого обследования и проводить раннюю диагностику. Одним из генетических факторов предрасположенности к развитию рака является наличие однонуклеотидных полиморфизмов (Single nucleotide polymorphisms, SNPs) в генах, ассоциированных с тем или иным онкологическим заболеванием.

Большинство известных на сегодняшний день SNPs было обнаружено благодаря секвенированию генома человека с последующим сравнением гомологичных последовательностей ДНК среди людей разных популяций. В

результате такого подхода успешно выявлялись полиморфизмы с частотой встречаемости более 1–5%. В среднем один SNP приходится на 300 нуклеотидов (Seeb *et al.*, 2011). В последние годы в связи с развитием высокопроизводительных методов секвенирования ДНК резко возрос объем информации о насыщенности геномов однонуклеотидными полиморфизмами. Так, если в 2005 г. в базе NCBI dbSNP было зарегистрировано чуть больше 10 млн SNPs в человеческих геномах (Miller *et al.*, 2005), то сейчас их насчитывается более 48 млн. Стремительно увеличивается и объем информации о клинически ассоциированных SNPs, к настоящему времени в базе OMIM их насчитывается уже более 25 тыс. (<http://www.ncbi.nlm.nih.gov/SNP/>). Использование новых технологий генотипирования совместно с базой данных гаплотипов человека (International HapMap Project) (Frazer, 2007; Altshuler *et al.*, 2010) позволяет анализировать значительно большее количество выборок «случай»–«контроль» без какой-либо предварительной гипотезы о связи аллеля с предрасположенностью к заболеванию. В связи с этим активно развиваются и геном-ассоциированные исследования (genome-wide association studies, GWAS), выявляющие на

основе встречаемости SNP-маркеров регионы генома, полиморфные варианты которых ответственны за ассоциацию с фенотипическим проявлением (Casto, Feldman, 2011; Chung, Chanock, 2011).

Большинство SNPs являются нейтральными и не оказывают заметного влияния на фенотип, и только малая доля замен характеризуется фенотипическим проявлением посредством изменения функции гена или уровня его экспрессии (Chung, Chanock, 2011; Matsuzaki, 2011; Zhao, Bracken, 2011). В современной литературе выделяют три функционально значимые группы полиморфизмов: SNPs в кодирующих последовательностях (cSNP, изменяют аминокислотную последовательность белка, влияют на процессинг РНК и белка); SNPs в регуляторных последовательностях (rSNP, влияют на регуляцию экспрессии гена на уровне как транскрипции, так и постраскрипционном) и SNPs в сайтах сплайсинга (sSNP) (Ponomarenko *et al.*, 2002; Wray, 2007). cSNP изучаются особенно интенсивно в связи с относительной простотой их интерпретации, в особенности в тех случаях, когда они приводят к заменам аминокислот (несинонимичные замены). Несинонимичные замены могут изменять функцию белка, либо совершенно разрушая его пространственную организацию, либо тонко корректируя его функциональные домены (Balasubramanian *et al.*, 2005). SNPs, модифицирующие сайты сплайсинга (sSNP), занимают второе по степени изученности место. Большая часть этих замен выявлена в экзонах и существенно меньшая – в интронах (Fairbrother *et al.*, 2004). Все известные случаи sSNP отражены в базе данных ASD, посвященной альтернативному сплайсингу (Stamm *et al.*, 2006).

Явное преобладание информации о cSNP и sSNP над сведениями о rSNP в базах данных показывает, что основные усилия по выявлению и изучению SNPs сосредоточены на вариантах, приводящих к изменению структуры, а следовательно, и функции белка. SNPs, способные хотя бы потенциально влиять на интенсивность экспрессии того же белка на уровне транскрипции, остаются мало исследованными (Chorley *et al.*, 2008; Ponomarenko *et al.*, 2002). Так, на сегодняшний день в базе PubMed NCBI зарегистрировано более 50 000 публикаций по SNPs,

из них по rSNP – менее сотни. Хотя именно эта группа SNPs представляет наибольший интерес как с фундаментальной, так и с прикладной точек зрения. Прежде всего, изучение регуляторных SNPs дает уникальную информацию о регуляторных районах различных генов и функциональной значимости отдельных регуляторных элементов. В последние годы интенсивно развивается также направление, посвященное изучению роли rSNP в адаптивных процессах и процессе эволюции (Wray, 2007; Yokoyama *et al.*, 2011). Регуляторные SNPs могут играть также роль в развитии различных патологических состояний, изменяя уровень экспрессии кандидатных генов. При этом потенциально регуляторными, как правило, считаются SNPs, расположенные в промоторных районах генов (Walitza *et al.*, 2005; Armelin-Correa *et al.*, 2005). Интроны в плане содержания rSNP практически не рассматриваются. Однако хорошо известно, что они также могут содержать регуляторные участки, и к настоящему времени в интронах найдено значительное количество связанных с фенотипическими проявлениями SNPs, не поддающихся интерпретации в качестве повреждающих сайты сплайсинга (Горшкова и др., 2006; Sribudiani *et al.*, 2011). Интересно заметить, что значительная часть таких SNPs располагается в высококонсервативных между видами районах внутри некодирующих последовательностей (Donfack *et al.*, 2005).

До сих пор выявление и изучение rSNP проводилось только в индивидуальных генах. Развитие современных методов массового анализа геномов, накопление информации об однонуклеотидных полиморфизмах и появление методов масштабного поиска мест связывания регуляторных белков поставило задачу разработки подходов к полногеномному варианту поиска регуляторных SNPs с учетом доступных геномных ресурсов и репозитариев данных, таких, как NCBI и проект ENCODE. В данной статье предложен один из вариантов такого подхода для поиска rSNP в генах APC (adenomatous polyposis coli, ген аденоматозного полипоза толстой кишки) (Miyoshi *et al.*, 1992; Fearnhead *et al.*, 2002) и MLH1 (MutL homolog 1, ген неполипозного рака толстой кишки) (Stoffel, 2010), ассоциированных с развитием рака толстой кишки.

Материалы и методы

Базы данных и компьютерная программа

База данных dbSNP NCBI (<http://www.ncbi.nlm.nih.gov/SNP/>) содержит информацию обо всех известных на сегодняшний день полиморфизмах в различных геномах.

В базе данных ChIP-seq проекта ENCODE (<http://genome.ucsc.edu/ENCODE/>) отражено распределение сайтов связывания транскрипционных факторов в геномах различных клеточных линий человека, выявленных с помощью экспериментов по массовой иммунопреципитации хроматина (ChIP-seq, **chromatin immunoprecipitation sequencing**).

Разработанная нами компьютерная программа SNPChIPTools для поиска потенциально регуляторных полиморфизмов в геноме человека на основе данных профилей связывания ChIP-seq реализована на языке Perl и доступна через интернет (<http://lungry.bionet.nsc.ru/cgi-bin/SNPProject/SNPChIPTools.cgi>).

Клеточные линии

Культивирование клеточных линий осуществлялось на базе Центра клеточных технологий ИЦиГ СО РАН. Для проведения экспериментов использовали культуру клеток гепатомы человека HepG2, эпителиальные человеческие клеточные линии аденокарциномы шейки матки HelaS3 и колоректальной карциномы HCT-116, клетки миелоидной лейкемии человека K562. Прикрепленные культуры клеток HepG2, HelaS3, HCT-116 и суспензионную культуру K562 культивировали в пластиковых культуральных флаконах «Greiner» в увлажненной атмосфере 5% CO₂ и 95% воздуха при 37 °С в среде DMEM/F12 «GIBCO», содержащей 10% фетальной бычьей сыворотки «Thermo Scientific HYCLONE», пенициллин (100 ЕД/мл) и стрептомицин (100 мг/мл) производства «Sigma». Жизнеспособность клеток оценивали как после размораживания, так и в процессе культивирования по включению в цитозоль трипанового синего. Она была выше 95% в течение всего процесса культивирования. Смену питательной среды осуществляли 2 раза в неделю.

Метод задержки ДНК-зонда в геле белками ядерного экстракта

Получение белкового экстракта ядер из клеточных линий человека проводили следующим методом: 10⁷ клеток инкубировали с 1 мл лизирующего буфера (10 mM Hepes, 10 mM KCl, 1 mM ДТТ, 0,5 mM спермидин, 0,15 mM спермин, 0,1 mM ЭДТА, 0,1 mM ЭГТА, 0,5 mM ПМСФ, коктейль ингибиторов протеаз «Pierce») в течение 15 мин на льду, затем добавляли 62 мкл 10% NP-40 и центрифугировали в течение 5 мин при 4 °С, 400 g. Осадок ядер лизировали в буфере (20 mM Hepes, 420 mM NaCl, 1,5 mM MgCl₂, 0,2 mM ЭДТА, 1 mM ДТТ, 25% глицерин, коктейль ингибиторов протеаз «Pierce») в течении 20 мин на льду, после чего центрифугировали в течение 10 мин при 4 °С, 10000 g. Супернатант, содержащий суммарный белок ядер, отбирали в чистую пробирку, измеряли концентрацию на спектрофотометре по методу Бредфорд, фасовали на аликвоты и хранили при -72 °С.

В качестве ДНК-зондов использовали двуцепочечные олигонуклеотиды, соответствующие различным полиморфным вариантам для каждого SNP (табл. 1, 2). Введение метки в ДНК-зонд осуществляли с помощью достройкой укороченных 3'-концов фрагментом Кленова ДНК-полимеразы I. Реакцию вели 5 мин при комнатной температуре в 10 мкл реакционной смеси, содержащей 0,01 mM олигонуклеотида, 1 мкл 10× буфера для меченя (500 mM Трис-НСl, pH = 8,0, 100 mM NaCl, 100 mM MgCl₂, 1 mM ДТТ, 2 mM dГТФ, 2 mM dТТФ, 2 mM dСТФ), 2 единицы активности фрагмента Кленова, 10 мкКи (α-32P)dАТФ. Для очистки олигонуклеотидов от невключившейся метки реакционную смесь наносили на ионообменную бумагу ДЕАЕ-81 и промывали 0,25 M KH₂PO₄.

Анализ связывания фрагментов ДНК с белками ядерных экстрактов проводили по следующей методике: белковый ядерный экстракт инкубировали с обработанной ультразвуком ДНК спермы лосося (из расчета 1 мкг ДНК на 7 мкг суммарного белка) в течение 10 мин на льду для предотвращения неспецифического связывания белков с ДНК-зондом. После этого добавляли 4 мкг экстракта к пробам, содержащим 50 pM

Таблица 1

Результаты компьютерно-теоретического поиска
и экспериментальной проверки регуляторных полиморфизмов в гене APC

Идентификационный № SNP в базе dbSNP NCBI	Число пиков ChIP-seq	Последовательности олигонуклеотидов, соответствующие разным полиморфным вариантам	Влияние SNPs на связывание с белками
rs113017087	27	GTGTAATCCGCTGGAC <u>CG</u> CGGACCAGGGCGCT GTGTAATCCGCTGGAT <u>TG</u> CGGACCAGGGCGCT	нет
rs115242894	27	GTATTGGTGCAGCCCC <u>CC</u> AGGGTGTCACTGG GTATTGGTGCAGCCCC <u>G</u> CCAGGGTGTCACTGG	нет
rs115658307	23	CAAGATGGCGGAGGG <u>CA</u> AGTAGCAAGGGGGC CAAGATGGCGGAGGG <u>T</u> AAGTAGCAAGGGGGC	нет
rs138386816	21	GCGGGGTGTGGCCGC <u>CG</u> GGAAGCCTAGCCGCT GCGGGGTGTGGCCGC <u>T</u> GGAAGCCTAGCCGCT	нет
rs35417795	16	GTGCGGTTGGGCGGG-CCCTGTGCCCACTG GTGCGGTTGGGCGGG <u>G</u> CCCTGTGCCCACTG	да
rs75996864	26	ACCGACATGTGGCTG <u>G</u> ATTGGTGCAGCCCGC ACCGACATGTGGCTG <u>T</u> ATTGGTGCAGCCCGC	да
rs76241113	20	GCGGGGGGAGGGGGG <u>A</u> AGGTGGTTTTCCCTC GCGGGGGGAGGGGGG <u>G</u> AGGTGGTTTTCCCTC	да
rs77733015	15	GGGCTAGGCAGGCTG <u>G</u> GCGGTTGGGCGGGGC GGGCTAGGCAGGCTG <u>T</u> GCGGTTGGGCGGGGC	нет
rs78037487	26	CCATTCCCGTTCGGGAC <u>CC</u> CGCCGATTGGCTG CCATTCCCGTTCGGGAG <u>C</u> CCCGCCGATTGGCTG	да
rs78429131	17	AGGAAGGTGAAGCAC <u>G</u> CAGTTGCCTTCTCGG AGGAAGGTGAAGCAC <u>T</u> CAGTTGCCTTCTCGG	нет
rs78597499	16	AGGCAGGCTGTGCGG <u>G</u> TGGGCGGGGCCCTGT AGGCAGGCTGTGCGG <u>T</u> TGGGCGGGGCCCTGT	нет
rs79896135	21	AGCCTAGCCGCTGCT <u>C</u> GGGGGGGACCTGCGG AGCCTAGCCGCTGCT <u>G</u> GGGGGGGACCTGCGG	да
rs80112297	19	GCTTGCTGCGGGGGG <u>A</u> GGGGGGAAGGTGGTT GCTTGCTGCGGGGGG <u>G</u> GGGGGGAAGGTGGTT	да
rs80313086	20	GGGGAGGGGGGAAGG <u>C</u> GGTTTTCCCTCGCAC GGGGAGGGGGGAAGG <u>G</u> GGTTTTCCCTCGCAC GGGGAGGGGGGAAGG <u>T</u> GGTTTTCCCTCGCAC	да

ДНК-пробы, меченной (^{32}P), и доводили общий объем реакционной смеси до 15 мкл $0,8\times\text{PBS}$. После инкубации при комнатной температуре в течение 15 мин смесь подвергали электрофорезу в 4,5 % ПААГ в $0,5\times\text{TBE}$ при 4°C , затем гель выдерживали в фиксирующем растворе, высушивали и экспонировали с рентгеновской пленкой.

Выделение ДНК из крови

Формирование коллекции образцов крови больных раком толстой кишки было проведено на базе клиники НИИ онкологии СО РАМН, г. Томск. Для выделения ДНК использовали 2 мл крови, содержащей 0,5 мМ ЭДТА. К 2 мл крови добавляли 12 мл буфера 1 (320 мМ сахараза, 1 %

Таблица 2

Результаты компьютерно-теоретического поиска и экспериментальной проверки регуляторных полиморфизмов в гене *MLH1*

Идентификационный № SNP в базе dbSNP NCBI	Число пиков ChIP-seq	Последовательности олигонуклеотидов, соответствующие разным полиморфным вариантам	Влияние SNP на связывание с белками
rs112765329	15	AGGAAGAGCGGACAGAGATCTCTAACGCGCA AGGAAGAGCGGACAGCGATCTCTAACGCGCA	нет
rs114683294	25	GGCGGAGGCCGCCGGATTCCCTGACGTGCCA GGCGGAGGCCGCCGGGTTCCCTGACGTGCCA	нет
rs34566456	28	GATGAGGCGGCGACACACCAGGCACAGGGCC GATGAGGCGGCGACAGACCAGGCACAGGGCC	нет
rs35032294	16	AACCAATAGGAAGAGCGGACAGCGATCTCTA AACCAATAGGAAGAGGGGACAGCGATCTCTA	да
rs4647205	14	ACCTTTTAAGGGTTGGTTGGAGTGTAAGTGG ACCTTTTAAGGGTTGTTGGAGTGTAAGTGG	нет
rs55927901	25	ACGACTTAACGGGCCCGTCACTCAATGGCG ACGACTTAACGGGCCGCGTCACTCAATGGCG	нет
rs1800734	28	GCGTAAGCTACAGCTAAGGAAGAACGTGA GCGTAAGCTACAGCTGAAGGAAGAACGTGAG	да

Triton X-100, 5 mM Трис Cl, pH 7,6, 5 mM MgCl₂), инкубировали 2 часа на льду, центрифугировали в течение 15 мин при 4°C, 1500 g. **Осадок ядер** суспендировали в 1,6 мл буфера 2 (25 mM ЭДТА, pH 8, 75 mM NaCl) и затем центрифугировали в течение 4 мин при 4°C, 700 g. **Осадок суспендировали** в 1 мл буфера 2, добавляли ДДС до 1% и протеиназу К до концентрации 50 мкг/мл и инкубировали при комнатной температуре в течение 18 ч при перемешивании. После этого добавляли AcNa, pH 7,5, до концентрации 0,3 M, а затем равный объем фенола, pH 8,0, смесь инкубировали 10 мин при 0°C и центрифугировали в течение 10 мин при 4°C, 15000 g. **Водную фазу отбирали** в новые пробирки и добавляли равный объем смеси фенола и хлороформа в отношении 1 : 1. Перемешивали и инкубировали 5 мин при 0°C, после чего центрифугировали в течение 10 мин при 4°C, 15000 g. **Снова отбирали водную фазу** и добавляли к ней равный объем хлороформа, перемешивали и инкубировали 5 мин при 0°C. После центрифугирования образцов в течение 10 мин при 4°C, 15000 g **водную фазу отбирали** и добавляли к ней равный объем изопропанола. Инкубировали 10 мин при комнатной темпера-

туре и центрифугировали в течение 15 мин при 15000 g. **Осадок промывали 70%-м спиртом** и растворяли в бидистиллированной воде.

Секвенирование ДНК

Реакцию Сэнгера проводили в амплификаторе GeneAmp PCR System 9700 (Applied Biosystems) в 10 мкл реакционной смеси, содержащей 1 мкг ДНК-матрицы, 2 пмоль праймера, 3 мкл BigDye Terminator Cycle Sequencing Kit v. 3.1 (Applied Biosystems), 2 мкл BigDye-буфера (Applied Biosystems). Образцы ДНК после реакции Сэнгера очищали с помощью гель-фильтрации на Sephadex G100. **Очищенные образцы подсушивали** на концентраторе Eppendorf 5301 и сдавали в Межинститутский центр секвенирования ДНК для анализа на секвенаторе ABI 3100 (Applied Biosystems).

Последовательности праймеров для секвенирования:

APC1-U 5'-AGAACAGCGAAGCAGTGCCCG-3',
APC1-R 5'-CGGTGCTCCAGGACCCGAGA-3',
APC2-U 5'-GCTTCAACTCAGGCAACCCAGACT-3',
APC2-R 5'-TCTACCCCTCCCCCTTCTTCTCT-3',

MLH1-U 5'-GAAGGAGGCCACGGGCAAGTC-3',
MLH1-R 5'-CACCTCAGTGCCTCGTGCTCA-3',
MLH2-U 5'-CGTGAGCACGAGGCACTGAGG-3',
MLH2-R 5'-GCGTCCTCCGAAATAAAGGACAGG-3'.

Результаты

Компьютерная программа SNPChIPTools для поиска потенциально регуляторных полиморфизмов в геноме человека на основе данных профилей связывания ChIP-seq

В настоящее время в области анализа данных высокопроизводительного секвенирования сложилась ситуация, когда экспериментальные технологии в своем развитии значительно опережают биоинформационные средства их поддержки, анализа и интерпретации результатов экспериментов. В связи с появлением огромного массива экспериментальных данных по распределению участков связывания транскрипционных факторов в геноме человека (проект ENCODE) появилась возможность предсказывать регуляторные районы генов в масштабе целого генома.

В основу разработанной нами компьютерной программы SNPChIPTools (<http://lungry.bionet.nsc.ru/cgi-bin/SNPProject/SNPChIPTools.cgi>) легли экспериментальные данные проекта ENCODE по распределению областей связывания различных транскрипционных факторов методом ChIP-seq. Эти данные представляют собой сведения, полученные с помощью метода массового параллельного секвенирования, о расположении пиков ChIP-seq, соответствующих местам связывания различных факторов транскрипции, во всем геноме. Нами была выдвинута гипотеза о том, что области, обогащенные пиками ChIP-seq, с большой вероятностью являются регуляторными, и расположенные в них полиморфизмы могут влиять на регуляцию экспрессии генов, разрушая или создавая сайты связывания регуляторных белков.

Данная компьютерная программа осуществляет поиск пиков ChIP-seq в месте расположения конкретного полиморфизма. Поиск осуществляется путем сравнения хромосомных координат участков полиморфизмов и групп пиков ChIP-seq по геномному релизу hg19.

Поиск регуляторных полиморфизмов в генах APC и MLH1

Из базы данных dbSNP NCBI были отобраны полиморфизмы, располагающиеся в генах APC (связанном с развитием семейного полипозного рака толстой кишки) и MLH1 (связанном с неполипозным типом рака толстой кишки), 2366 и 1023 полиморфизмов соответственно. С помощью разработанной нами программы SNPChIPTools эти SNP были проанализированы и выбраны те из них, что попадали в область перекрытия более 8 пиков ChIP-seq. В результате этого были отобраны потенциально регуляторные полиморфизмы – 14 для гена APC (rs113017087, rs115242894, rs115658307, rs138386816, rs35417795, rs75996864, rs76241113, rs77733015, rs78037487, rs78429131, rs78597499, rs79896135, rs80112297, rs80313086) (табл. 1) и 7 для гена MLH1 (rs112765329, rs114683294, rs34566456, rs35032294, rs4647205, rs55927901, rs1800734) (табл. 2). Для функционального анализа этих полиморфизмов были синтезированы олигонуклеотиды (табл. 1, 2), которые использовали в качестве ДНК-зондов в экспериментах по задержке их в геле белками ядерного экстракта.

В результате из 14 потенциально регуляторных полиморфизмов в гене APC (rs113017087, rs115242894, rs115658307, rs138386816, rs35417795, rs75996864, rs76241113, rs77733015, rs78037487, rs78429131, rs78597499, rs79896135, rs80112297, rs80313086) после экспериментальной проверки с помощью метода задержки ДНК-зонда в геле белками ядерного экстракта функционально значимыми были признаны 7 (rs75996864, rs76241113, rs78037487, rs80112297, rs80313086, rs79896135, rs35417795) (табл. 1), поскольку они влияли на формирование комплексов с белками. На рис. 1 представлены типичные примеры результатов таких экспериментов. На основе полученных данных можно предполагать, что эти полиморфизмы способны нарушать регуляторный код гена APC, разрушая сайты связывания одних факторов транскрипции и/или создавая сайты для других, и таким образом, вносить свой вклад в регуляцию экспрессии гена на транскрипционном уровне.

Для гена MLH1 из 7 отобранных по результатам компьютерного поиска полиморфизмов в

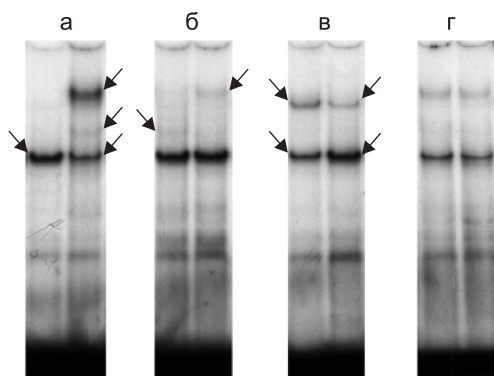


Рис. 1. Влияние полиморфизма на связывание ДНК-зонда белками ядерного экстракта.

Олигонуклеотиды, соответствующие разным вариантам участка ДНК, использовались в эксперименте в качестве ДНК-зондов. а, б – наличие полиморфизма приводит к появлению новых комплексов – полос задержки и ослаблению старых; в – перераспределение интенсивности полос задержки комплексов ДНК–белок; г – наличие полиморфизма не вызывает изменений в картине связывания.

двух случаях (rs35032294, rs1800734) (табл. 2) олигонуклеотиды, соответствующие разным вариантам SNP, формируют разные картины связывания с белками ядерного экстракта. Это свидетельствует о том, что данные полиморфизмы обладают регуляторным потенциалом и способны приводить к изменениям в экспрессии гена *MLH1* на уровне транскрипции.

Генотипирование больных раком толстой кишки

Для проверки связи выявленных нами регуляторных полиморфизмов в генах *MLH1* и *APC* с предрасположенностью к развитию рака толстой кишки были начаты работы по генотипированию больных колоректальным раком. Была создана начальная выборка из 10 образцов ДНК. Методом прямого секвенирования по Сенгеру были проанализированы районы ДНК, содержащие выявленные программой SNPChIPTools полиморфизмы в генах *APC* и *MLH1*.

Так, из анализируемых 14 полиморфизмов в гене *APC* в нашей выборке больных только у одного человека было найдено сразу два (rs79896135, rs78429131), причем один из них (rs79896135) менял спектр связывания транскрипционных факторов в экспериментах по задержке ДНК-зонда в геле.

При поиске предсказанных компьютерной программой регуляторных полиморфизмов в гене *MLH1* была выявлена ранее неаннотированная делеция продолжительностью 26 нуклеотидов, удаляющая два из проанализированных нами полиморфизмов (rs35032294, rs112765329). Причем один из них (rs35032294) значительно влиял на связывание факторов транскрипции с данным районом в экспериментах по задержке ДНК-зонда в геле. Примечательно, что данная делеция обнаружена у всех 10 пациентов, что может являться свидетельством ее причастности к развитию этого заболевания.

Однако на данном этапе для оценки эффективности использования предложенного нами подхода поиска потенциально регуляторных полиморфизмов как кандидатных для выявления их клинических ассоциаций наша выборка больных колоректальным раком мала. В связи с этим был проведен анализ клинически ассоциированных полиморфизмов, аннотированных в базе данных OMIM. Оказалось, из 14 полиморфизмов, предсказанных нашей программой SNPChIPTools в гене *APC*, 7 уже идентифицированы как ассоциированные с предрасположенностью к развитию колоректального рака, причем в их число не входят SNP, выявленные нами в ходе генотипирования. В гене *MLH1* всего 1 полиморфизм из 7 предсказанных оказался ассоциированным с этим заболеванием, согласно данным базы OMIM.

Заключение

Выяснение молекулярных механизмов генетической предрасположенности к наиболее распространенным заболеваниям, таким, как сердечно-сосудистые, психоневрологические и онкологические, является одной из основных проблем современной медицинской генетики, молекулярной физиологии и патологии. В рамках этой проблемы в настоящее время во всем мире проводятся широкомасштабные исследования, посвященные выявлению кандидатных генов, изучению полиморфизмов в нуклеотидных последовательностях этих генов и установлению связей между тем или иным вариантом последовательности и изучаемой патологией (Hata, Onouchi, 2009; Toyoda, Ishikawa, 2010;

Kingsley, 2011; Chung, Chanock, 2011). Однако объектом большинства таких работ являются некие конкретные полиморфизмы в отдельных генах-кандидатах, что не дает полного представления о генетически обусловленных причинах развития многофакторных заболеваний.

В связи с развитием методов полногеномного секвенирования появились принципиально новые возможности для поиска различных молекулярных маркеров предрасположенности к развитию таких заболеваний (Shiraishi *et al.*, 2010). В частности, результаты экспериментов по массовой иммунопреципитации хроматина позволяют искать в геноме регуляторные районы и, следовательно, говорить о регуляторном потенциале полиморфизмов, попадающих в эти районы.

В нашей работе описан один из таких подходов. Путем сравнения хромосомных координат с помощью разработанной нами программы SNPChIPTools был осуществлен поиск полиморфизмов в потенциально регуляторных районах, т. е. в местах, обогащенных пиками ChIP-seq, в генах *APC* и *MLH1*, ассоциированных с развитием колоректального рака. Экспериментальными методами мы подтвердили эффективность теоретического предсказания регуляторных полиморфизмов.

Работа была поддержана Госконтрактом № 16.512.11.2274 Минобрнауки РФ и частично программой президиума РАН «Фундаментальные науки – медицине» № 23.

Литература

- Горшкова Е.В., Каледин В.И., Кобзев В.Ф., Меркулова Т.И. SNP в начале интрона 2 гена *K-ras* мыши, ассоциированные с развитием рака легких, влияют на связывание с факторами транскрипции // Бюлл. эксп. биол. медицины. 2006. № 6. С. 681–684.
- Тимофеев Ю.М. Колоректальный рак: современные аспекты диагностики и лечения // Рус. мед. журнал. 2004. № 11. С. 653–656.
- Altshuler D.M., Gibbs R.A., Peltonen L. *et al.* Integrating common and rare genetic variation in diverse human populations // Nature. 2010. V. 467 (7311). P. 52–58. doi:10.1038/nature09298.
- Armelin-Correa L.M., Lin C.J., Barbosa A. *et al.* Characterization of human collagen XVIII promoter 2: interaction of Sp1, Sp3 and YY1 with the regulatory region and a SNP that increases transcription in hepatocytes // Matrix Biol. 2005. V. 24. P. 550–559.
- Balasubramanian S., Xia Y., Freinkman E., Gerstein M. Sequence variation in G-protein-coupled receptors: analysis of single nucleotide polymorphisms // Nucl. Acids Res. 2005. V. 33. P. 1710–1721.
- Casto A.M., Feldman M.W. Genome-wide association study SNPs in the human genome diversity project populations: does selection affect unlinked SNPs with shared trait associations? // PLoS Genet. 2011. V. 7(1). e1001266.
- Chorley B.N., Wang X., Campbell M.R. *et al.* Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies // Mutat. Res. 2008. 659(1/2). P. 147–157. Epub. 2008. May 4.
- Chung C.C., Chanock S.J. Current status of genome-wide association studies in cancer // Hum. Genet. 2011. 130(1). P. 59–78. Epub. 2011. Jun 16.
- Donfack J., Schneider D.H., Tan Z. *et al.* Variation in conserved non-coding sequences on chromosome 5q and susceptibility to asthma and atopy // Respir. Res. 2005. V. 6. P. 145.
- Fairbrother W.G., Holste D., Burge C.B., Sharp P.A. Single nucleotide polymorphism-based validation of exonic splicing enhancers // PLoS Biol. 2004. V. 2. P. 1388–1395.
- Fearnhead N.S., Wilding J.L., Bodmer W.F. Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumorigenesis // Br. Med. Bull. 2002. V. 64. P. 27–43.
- Frazer K.A., Ballinger D.G., Cox D.R. *et al.* A second generation human haplotype map of over 3.1 million SNPs // Nature. 2007. V. 449 (7164). P. 851–861. doi:10.1038/nature06258.
- Hata A., Onouchi Y. Susceptibility genes for Kawasaki disease: toward implementation of personalized medicine // J. Hum. Genet. 2009. V. 54(2). P. 67–73.
- Kingsley C.B. Identification of causal sequence variants of disease in the next generation sequencing era // Methods Mol. Biol. 2011. V. 700. P. 37–46.
- Matsuzaki S. DNA microarray analysis in endometriosis for development of more effective targeted therapies // Front. Biosci. (Elite Ed). 2011. V. 3. P. 1139–1153.
- Miller R.D., Phillips M.S., Jo I. *et al.* The SNP Consortium Allele Frequency Project. High-density single-nucleotide polymorphism maps of the human genome // Genomics. 2005. V. 86. P. 117–126.
- Miyoshi Y., Nagase H., Ando H. *et al.* Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene // Hum. Mol. Genet. 1992. V. 1. P. 229–233.
- Ponomarenko J.V., Orlova G.V., Merkulova T.I. *et al.* rSNP_Guide: an integrated database-tools system for studying SNPs and site-directed mutations in

- transcription factor binding sites // *Hum. Mutat.* 2002. 20(4). P. 239–248.
- Seeb J.E., Carvalho G., Hauser L. *et al.* Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms // *Mol. Ecol. Resour.* 2011. V. 11. Suppl 1. P. 1–8. doi: 10.1111/j.1755-0998.2010.02979.x.
- Shiraishi T., Matsuyama S., Kitano H. Large-scale analysis of network bistability for human cancers // *PLoS Comput. Biol.* 2010. V. 6(7). e1000851.
- Sribudiani Y., Metzger M., Osinga J. *et al.* Variants in RET associated with Hirschsprung's disease affect binding of transcription factors and gene expression // *Gastroenterology.* 2011. V. 20. N 2. P. 572–582.
- Stamm S., Riethoven J.J., Le Texier V. *et al.* ASD: a bioinformatics resource on alternative splicing // *Nucl. Acids Res.* 2006. V. 34. P. D46–55.
- Stoffel E.M. Lynch Syndrome/Hereditary Non-polyposis Colorectal Cancer (HNPCC) // *Minerva Gastroenterol. Dietol.* 2010. V. 56(1). P. 45–53.
- Toyoda Y., Ishikawa T. Pharmacogenomics of human ABC transporter ABCC11 (MRP8): potential risk of breast cancer and chemotherapy failure. *Anticancer Agents // Med. Chem.* 2010. V. 10(8). P. 617–624.
- Walitza S., Renner T.J., Dempfle A. *et al.* Transmission disequilibrium of polymorphic variants in the tryptophan hydroxylase-2 gene in attention-deficit/hyperactivity disorder // *Mol. Psychiatry.* 2005. V. 10. P. 1126–1132.
- Wray G.A. The evolutionary significance of cis-regulatory mutations // *Nat. Rev. Genet.* 2007. V. 8(3). P. 206–216.
- Yokoyama K.D., Thorne J.L., Wray G.A. Coordinated genome-wide modifications within proximal promoter cis-regulatory elements during vertebrate evolution // *Genome Biol. Evol.* 2011. V. 3. P. 66–74.
- Zhao L., Bracken M.B. Association of CD14 -260 (-159) C>T and asthma: a systematic review and meta-analysis // *BMC Med. Genet.* 2011. 12:93. doi:10.1186/1471-2350-12-93.

SEARCH FOR REGULATORY SNPs ASSOCIATED WITH COLON CANCER IN THE *APC* AND *MLH1* GENES

E.V. Antontseva¹, L.O. Bryzgalov¹, M.Yu. Matveeva¹, E.V. Kashina¹,
N.V. Cherdyntseva², T.I. Merkulova¹

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia,
e-mail: Antonseva@bionet.nsc.ru;

² Tomsk Cancer Research Institute, Siberian Branch of the Russian Academy
of Medical Sciences, Tomsk, Russia

Summary

Polymorphisms belonging to the regulatory regions of the *APC* and *MLH1* genes were detected by invoking ChIP Seq data obtained in the ENCODE project. The significance of these polymorphisms for gene regulation was confirmed by gel retardation of DNA probes by nuclear proteins. More than half of the polymorphisms in the overlapping region of more than eight ChIP-seq peaks were found to be significant for regulation.

Key words: single-nucleotide polymorphism, chromatin immunoprecipitation, gene regulation, colon cancer.