

УДК 577.113.3:57.087.1

## ОЦЕНКА ПО ТЕХНОЛОГИИ rSNP\_Guide SNPs ПРОМОТОРОВ ГЕНОВ *APC* И *MLH1* ЧЕЛОВЕКА, СВЯЗАННЫХ С РАКОМ ТОЛСТОГО КИШЕЧНИКА

© 2013 г. Д.А. Рассказов<sup>1</sup>, Е.В. Антонцева<sup>1</sup>, Л.О. Брызгалов<sup>1</sup>,  
М.Ю. Матвеева<sup>1</sup>, Е.В. Кашина<sup>1</sup>, П.М. Пономаренко<sup>1</sup>, Г.В. Орлова<sup>1</sup>,  
М.П. Пономаренко<sup>1</sup>, Д.А. Афонников<sup>1,2</sup>, Т.И. Меркулова<sup>1,2</sup>

<sup>1</sup> Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики  
Сибирского отделения Российской академии наук, Новосибирск, Россия,  
e-mail: pon@bionet.nsc.ru;

<sup>2</sup> Новосибирский национальный исследовательский государственный университет,  
Новосибирск, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Каждый из 6 регуляторных SNPs (Single nucleotide polymorphisms) генов *APC* и *MLH1* человека (rs75996864, rs76241113, rs78037487, rs80112297, rs80313086 и rs1800734) был оценен по созданной ранее технологии rSNP\_Guide на значимость изменения связывания каждого из 40 факторов транскрипции с соответствующими районами ДНК. В результате для каждого SNP все анализируемые белки были ранжированы по убыванию уровня статистической значимости  $\alpha$  (*t*-тест Стьюдента) изменения их сродства к аллельным вариантам указанной ДНК. Установлено, что самыми вероятными проявлениями SNPs rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC*, а также SNP rs1800734 гена *MLH1* человека являются изменения в связывании именно тех транскрипционных факторов (NF- $\Upsilon$ , NF $\kappa$ B, c-Myb, RAR, YY-1, Sp-1), для которых ранее было показано участие в развитии рака толстого кишечника. Полученные результаты служат новым основанием для исследований ассоциации SNPs rs75996864, rs76241113, rs78037487, rs80112297, rs80313086 гена *APC* с раком толстого кишечника общепринятыми медико-генетическими методами.

**Ключевые слова:** однонуклеотидный полиморфизм, регуляция экспрессии генов, рак толстого кишечника, *APC*, *MLH1*, комплекс ДНК с регуляторным белком, *t*-тест Стьюдента.

### ВВЕДЕНИЕ

Начало III тысячелетия н. э. было ознаменовано эпохальным достижением науки в области молекулярной биологии – расшифровкой генома человека. В 2004 г. было завершено секвенирование так называемого «референсного» (т. е. общепринятого стандарта) генома человека (The International Human Genome Sequencing Consortium, 2004). Это событие считается началом новой постгеномной эры науки о жизни. Для нее характерна быстрая расшифровка индивидуальных геномов пациентов относительно референсного генома человека, что закладывает основы для развития персонализированной ме-

дицины с возможностью диагностики, терапии и мониторинга заболеваний с учетом генетической предрасположенности и индивидуальной чувствительности/устойчивости к лекарственным препаратам.

В этой связи интенсивно ведется статистическое выявление ассоциаций между геномными вариациями и риском заболеваний, NHGRI GWAS catalog (Hindorff *et al.*, 2009), что позволило, в частности, связать наличие ряда однонуклеотидных замен (SNP, Single Nucleotide Polymorphism) с генетической предрасположенностью к раку кожи (Gerstenblith *et al.*, 2010). Экспериментально установленные случаи полиморфизма генов человека докумен-

тируются в базе данных dbSNP (NCBI Resource Coordinators, 2013), ассоциации полиморфизма с патологиями – в базе данных OMIM (Hamosh *et al.*, 2005).

Особые успехи были достигнуты при определении молекулярных механизмов негативного действия SNPs, расположенных в белок-кодирующих районах геномов, вследствие относительной простоты выяснения возможных причин нарушения функции белка: с сохранением пространственной укладки либо с ее нарушением (Sanchez-Ruiz, 2010). Однако молекулярные механизмы влияния SNPs из некодирующих районов генов на возникновение патологий остаются в своем большинстве неясными. В настоящее время выявлено множество регуляторных SNPs, связанных с проявлением различных патологий. В частности, минорный аллель rs1800734 (идентификатор базы данных dbSNP) гена *MLH1* человека, кодирующего белок-гомолог гена *mutL* репарации *E. coli* (Win *et al.*, 2013), был ассоциирован с раком толстого кишечника (Hitchins *et al.*, 2007).

В данной работе мы исследовали аллели rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC* (adenomatous polyposis coli – аденоматозного полипоза толстого кишечника), кодирующего белок-супрессор опухоли (Polakis, 2011). Для этих SNPs ранее (Антонцева и др., 2011; Antontseva *et al.*, 2012) впервые были получены данные по задержке в геле (EMSA, electrophoretic mobility shift assay) белками ядерного экстракта из раковой клеточной линии НСТ-116 (рак толстого кишечника). С помощью технологии rSNP\_Guide (Ponomarenko *et al.*, 2001, 2002) было установлено, что самые статистически достоверные (*t*-тест Стьюдента) аллельные изменения при связывании с белком/белками присущи именно тем участкам ДНК гена *APC*, соответствующим потенциальным сайтам связывания транскрипционных факторов, участие которых в канцерогенезе толстого кишечника уже было экспериментально доказано. Полученные компьютерно-экспериментальные данные являются основанием для проведения дальнейшего стандартного медико-генетического исследования связи SNPs rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC* с раком толстого кишечника человека.

## МАТЕРИАЛЫ И МЕТОДЫ

Из базы данных dbSNP (NCBI Resource Coordinators, 2013) были извлечены последовательности нуклеотидов  $S = \{s_{-25} \dots s_{-1}(s_0^{WT} / s_0^{minor}) s_1 \dots s_{25}\}$  с аллельной вариацией в центре предковых и минорных аллелей rs1800734, ранее уже ассоциированных с раком толстого кишечника гена *MLH1* человека, и исследуемых 5 SNPs, rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC* человека (табл. 1–3).

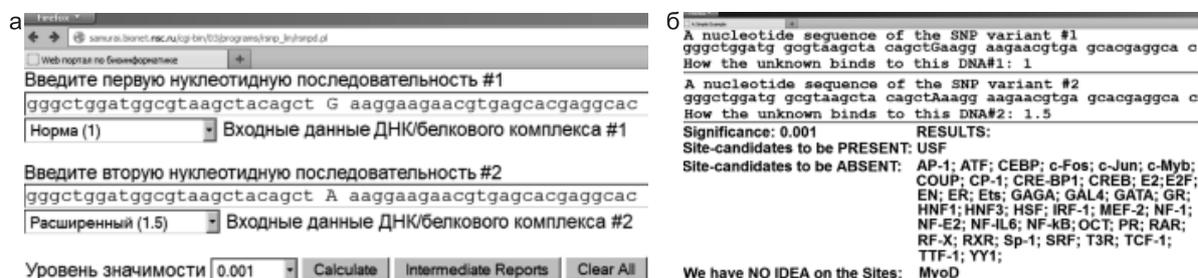
В табл. 1 и 3 представлены взятые из базы данных SNPChIPTools (Антонцева и др., 2011; Antontseva *et al.*, 2012) результаты эксперимента по задержке в геле двуцепочечных олигонуклеотидов, несущих предковые (верхняя дорожка) или минорные (нижняя дорожка) аллели, с белками экстрактов ядер раковых клеток линий НСТ-116 (рак толстого кишечника), HeLaS3 (рак шейки матки), K562 (эритролейкемия) и HepG2 (гепатома).

Эти экспериментальные данные были исследованы по технологии rSNP\_Guide (Ponomarenko *et al.*, 2001, 2002), как это показано на рис. 1 и в табл. 2 на примере аллелей rs1800734, ассоциированных с раком толстого кишечника, гена *MLH1* человека (табл. 1, колонка «НСТ-116»).

После ввода «[http://samurai.bionet.nsc.ru/cgi-bin/03/programs/rsnp\\_lin/rsnpd.pl](http://samurai.bionet.nsc.ru/cgi-bin/03/programs/rsnp_lin/rsnpd.pl)» в Интернет браузер пользователь получает начальную форму (рис. 1, а), которую он должен заполнить исследуемыми им экспериментальными данными. Применяя команду «Calculate», пользователь автоматически получает результат анализа этих данных по технологии rSNP\_Guide (рис. 1, б). Дополнительная команда «Intermediate Report» позволяет получить все данные промежуточных расчетов без исключения (табл. 2). В первой колонке этой таблицы приведены общепринятые обозначения 40 белков-факторов транскрипции, которые представлены в текущей версии rSNP\_Guide. В следующих колонках ii–v даны оценки наибольшего сродства этих белков к анализируемой двунитевой ДНК, полученные на основе анализа контекстных мотивов сайтов связывания этих белков, найденные во «введенной» нити (+) ДНК (рис. 1) и в комплементарной к ней (–) нити ДНК предкового и минорного аллелей «введенного» SNP. В текущей версии rSNP\_Guide (Ponomarenko *et al.*, 2001, 2002)

эти оценки уровней сродства ДНК/белок соответствуют теории статистической механики взаимодействия регуляторных белков с ДНК (Berg, von Hippel, 1987) на основе построения

позиционно-весовых матриц олигонуклеотидов ДНК, как это было описано ранее (Popomarenko *et al.*, 1999). На основе вычисления среднеарифметического  $M_0$  отрицательных



**Рис. 1.** Интерфейс ввода (а) и вывода (б) технологии rSNP\_Guide на примере анализа экспериментальных данных о задержке в геле двуцепочечных олигонуклеотидов, несущих rs1800734, с белками ядерных экстрактов из раковой клеточной линии НСТ-116 (рак толстого кишечника, табл. 1). Комплементарная нить к введенной нити ДНК строится автоматически для их совместного анализа.

**Таблица 1**

Оценка по технологии rSNP\_Guide ассоциированных с раком толстого кишечника аллелей rs1800734 гена *MLH1* человека

Аллель	Последовательность ДНК, ген <i>MLH1</i> , rs1800734 ± 25 п.о.			
Предок, WT минорный	gggctggatggcgtaagctacagct <b>G</b> aaggaagaacgtgagcscgaggcac gggctggatggcgtaagctacagct <b>A</b> aaggaagaacgtgagcscgaggcac			
Клетки Рак	НСТ-116 толстого кишечника	HeLaS3 шейки матки	НepG2 гепатома	K562 эритролейкемия
Задержка в геле				
Комплекс <sup>#</sup>	MINOR больше	WT меньше	WT больше	возник слабый MINOR
Данные для rSNP_Guide	$X_{WT} = 1,0;$ $X_{MINOR} = 1,5$	$X_{WT} = 0,5;$ $X_{MINOR} = 1,0$	$X_{WT} = 1,0;$ $X_{MINOR} = 0,5$	$X_{WT} = 0,0;$ $X_{MINOR} = 0,5$
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I ( $\alpha < 0,001$ )	USF			
II ( $\alpha < 0,0025$ )	USF		USF, MyoD	
III ( $\alpha < 0,005$ )	Ets, GR		GR	
IV ( $\alpha < 0,01$ )	ATF, MyoD, YY1		ATF, Ets, YY1	
V ( $\alpha < 0,025$ )	c-Jun, c-Myb, CREB; GAGA, GATA, NF-IL6, RF-X	c-Fos, c-Myb, CRE-BP1, CREB, GAGA, GATA, MyoD, NF-IL6, RF-X	c-Jun, NF-IL6, CREB, GAGA, RF-X	ATF, c-Fos, c-Jun, CRE-BP1, CREB, GATA, T3R
VI ( $\alpha < 0,05$ )	c-Fos, CEBP, PR, CRE-BP1, RAR, T3R, TCF-1	CEBP, PR, RAR, T3R, TCF-1	CEBP, c-Fos, RAR, c-Myb, PR, CRE-BP1	NF-IL6, RAR, RF-X

<sup>#</sup> Здесь и далее: анализируемый комплекс ДНК/белок (выделен овалом) был охарактеризован референтом базы данных SNPChIPTools (Антонцева и др., 2011; Antontseva *et al.*, 2012).

Таблица 2

Пример обоснования полученной по технологии rSNP\_Guide оценки ассоциированного с раком толстого кишечника полиморфизма rs1800734 промотора гена *MLH1* человека (транскрипционный фактор-кандидат USF, для MyoD мало данных)

Фактор транскрипции, ТФ	Сродство ТФ/ДНК, нити (+)/(-)				Сходство Декарта между ТФ и моделью SNP				<i>t</i> -тест Стьюдента ( $\alpha < 0,001$ ) соответствия ТФ модели SNP			
	Предок, WT		Минорный аллель		X <sub>++</sub>		X <sub>+-</sub>		X <sub>+-</sub>		X <sub>--</sub>	
	(+)	(-)	(+)	(-)	X <sub>++</sub>	X <sub>+-</sub>	X <sub>+-</sub>	X <sub>--</sub>	X <sub>++</sub>	X <sub>+-</sub>	X <sub>+-</sub>	X <sub>--</sub>
i	ii	iii	iv	v	vi	vii	viii	ix	x	xi	xii	xiii
AP-1	-0,09	0,01	-0,09	0,01	2,62	1,98	1,80	0,52	-1,19	-1,11	-1,06	0,94
ATF	0,39	0,33	0,39	0,33	1,83	1,54	1,66	1,32	-0,40	-0,68	-0,91	0,14
CEBP	-0,51	0,38	-0,37	0,23	2,79	2,56	1,42	0,89	-1,35	-1,70	-0,68	0,57
c-Fos	0,25	0,20	0,25	0,20	2,10	1,61	1,71	1,05	-0,67	-0,75	-0,96	0,41
c-Jun	0,36	0,30	0,36	0,30	1,90	1,55	1,67	1,26	-0,46	-0,69	-0,92	0,20
c-Myb	-0,00	0,35	0,26	0,35	2,06	1,83	1,46	1,11	-0,62	-0,97	-0,71	0,35
COUP	-0,32	0,10	-0,32	0,08	2,80	2,32	1,67	0,56	-1,37	-1,46	-0,92	0,91
CP-1	-0,34	-0,12	-0,15	-0,12	2,90	2,14	1,97	0,29	-1,46	-1,28	-1,23	1,17
CRE-BP1	0,32	0,12	0,32	0,12	2,12	1,47	1,85	1,06	-0,68	-0,61	-1,11	0,40
CREB	0,19	0,29	0,19	0,28	2,08	1,74	1,56	1,08	-0,64	-0,88	-0,82	0,38
E2	-0,37	-0,12	-0,35	-0,13	3,05	2,32	1,99	0,25	-1,61	-1,46	-1,24	1,21
E2F	-0,16	-0,44	-0,16	-0,44	3,16	2,04	2,43	0,27	-1,72	-1,18	-1,68	1,19
EN	-0,20	-0,53	-0,22	-0,31	3,17	2,11	2,37	0,26	-1,73	-1,25	-1,63	1,20
ER	-0,21	-0,30	-0,29	-0,30	3,10	2,17	2,22	0,07	-1,67	-1,30	-1,48	1,39
Ets	-0,39	0,57	-0,46	0,62	2,60	2,72	0,99	1,28	-1,16	-1,86	-0,25	0,18
GAGA	0,46	-0,60	0,46	-0,60	2,89	1,24	2,86	1,16	-1,46	-0,38	-2,12	0,30
GATA	0,11	-0,01	0,57	-0,01	2,22	1,34	2,05	1,04	-0,78	-0,48	-1,31	0,42
GR	0,56	-0,12	0,56	-0,12	2,22	1,05	2,32	1,24	-0,79	-0,19	-1,57	0,22
HNF1	-0,35	-0,05	-0,28	-0,00	2,89	2,27	1,83	0,38	-1,46	-1,41	-1,09	1,08
HNF3	-0,11	-0,40	-0,07	-0,40	3,05	1,93	2,39	0,32	-1,62	-1,07	-1,64	1,14
HSF	-0,41	-0,37	-0,41	-0,37	3,32	2,38	2,32	0,19	-1,88	-1,51	-1,58	1,27
IRF-1	0,03	-0,47	0,03	-0,47	3,01	1,76	2,50	0,53	-1,58	-0,90	-1,76	0,93
MEF-2	-0,11	-0,37	0,08	-0,37	2,93	1,80	2,35	0,43	-1,49	-0,94	-1,61	1,03
<b>MyoD</b>	<b>0,57</b>	<b>0,80</b>	<b>0,12</b>	<b>0,40</b>	<b>1,81</b>	<b>1,94</b>	<b>1,47</b>	<b>1,63</b>	<b>-0,37</b>	<b>-1,08</b>	<b>-0,72</b>	<b>-0,16</b>
NF-1	-0,33	-0,41	-0,37	-0,04	3,11	2,32	2,09	0,29	-1,67	-1,45	-1,35	1,18
NF-E2	0,08	-0,20	-0,23	-0,20	2,86	1,96	2,13	0,41	-1,43	-1,10	-1,38	1,06
NF-IL6	0,15	0,48	0,15	0,48	1,95	1,93	1,30	1,27	-0,51	-1,07	-0,56	0,19
NF-kB	-0,22	-0,47	-0,22	-0,47	3,24	2,12	2,46	0,26	-1,80	-1,26	-1,71	1,20
OCT	-0,37	-0,31	-0,23	-0,33	3,16	2,21	2,26	0,10	-1,72	-1,35	-1,51	1,36
PR	-0,37	0,38	-0,37	0,26	2,70	2,49	1,38	0,89	-1,26	-1,62	-0,63	0,57
RAR	0,32	0,07	0,21	0,07	2,23	1,54	1,88	0,96	-0,79	-0,67	-1,14	0,50
RF-X	0,41	-0,10	0,33	-0,18	2,40	1,32	2,22	0,97	-0,96	-0,46	-1,48	0,49
RXR	-0,14	0,16	-0,14	0,11	2,56	2,09	1,63	0,66	-1,13	-1,23	-0,88	0,80
Sp-1	-0,31	-0,84	-0,31	-0,80	3,71	2,36	2,95	0,74	-2,27	-1,50	-2,21	0,72
SRF	-0,25	-0,56	-0,02	-0,51	3,22	2,00	2,56	0,44	-1,78	-1,13	-1,82	1,02
T3R	0,24	0,04	0,24	0,04	2,26	1,53	1,90	0,91	-0,83	-0,67	-1,15	0,55
TCF-1	0,16	-0,54	0,28	-0,54	2,95	1,51	2,67	0,82	-1,51	-0,64	-1,92	0,64
TTF-1	0,03	-0,26	0,03	-0,18	2,74	1,76	2,15	0,48	-1,30	-0,90	-1,41	0,99
<b>USF</b>	<b>0,77</b>	<b>0,53</b>	<b>0,77</b>	<b>0,53</b>	<b>1,31</b>	<b>1,39</b>	<b>1,86</b>	<b>1,92</b>	<b>0,12</b>	<b>-0,53</b>	<b>-1,12</b>	<b>0,45</b>
YY1	0,50	-0,24	0,50	-0,37	2,51	1,12	2,52	1,13	-1,08	-0,25	-1,77	0,33

оценок конструируются все четыре возможные модели влияния заданного SNP на комплексы ДНК/ белок: модель X<sub>--</sub> оценивает неспецифическое связывание ДНК/белок для обеих нитей ДНК каждого из аллелей (колонки (ix и xiii)); модель X<sub>++</sub> – специфическое связывание ДНК/белок введенными в rSNP\_Guide (рис. 1) M<sub>WT</sub> и M<sub>MINOR</sub>; модели X<sub>+-</sub> и X<sub>-+</sub> – специфическое связывание одной из двух нитей ДНК. В колонках vi–ix для каждой из этих четырех моделей влияния «введенного» SNP на связывание ДНК/белок представлены оценки их сходства с изменениями сродства каждого из рассматриваемых 40 белков к регуляторной ДНК с этим SNP, полученные с помощью наиболее часто используемой Декартовой меры сходства.

Наконец, в колонках (x–xiii) для каждой модели представлены отклонения ее значений от 95 %-й доверительной границы *t*-теста Стьюдента для гипотезы «H<sub>0</sub>: заданная SNP неразличимо одинаково меняет связывание ДНК с любым регуляторным белком». На этой основе rSNP\_Guide автоматически генерирует выходные данные (рис. 1).

Положительные оценки указывают на те из 40 анализируемых белков, сродство которых к содержащей SNP последовательности ДНК изменяется в значимом соответствии с моделью патогенного изменения комплекса ДНК/белок во введенных данных эксперимента по задержке в геле для «введенного» SNP (рис. 1). Одновременное несоответствие модели X<sub>--</sub> при соответствии любой другой модели SNP указывает те белки, для которых изменения их связывания с ДНК соответствуют «введенным» данным (рис. 1). При одновременном отсутствии достоверного сходства всех четырех возможных моделей SNP с каким-либо из анализируемых 40 белков (в качестве примера в табл. 2 рассмотрен транскрипционный фактор MyoD) по технологии rSNP\_Guide принималось решение о невозможности оценить влияние «введенной SNP» на связывание этого белка с ДНК. Такое решение принимается и при одновременном достоверном сходстве какого-либо белка как с моделью X<sub>--</sub>, так и с любой из X<sub>++</sub>, X<sub>+-</sub> или X<sub>-+</sub> моделей. В остальных случаях rSNP\_Guide принимает решение о недостоверном влиянии «введенного SNP» на связывание заданного белка с ДНК.

Результатом rSNP\_Guide для ассоциированного с раком толстого кишечника SNP rs1800734 была оценка в качестве самых достоверных аллельных различий гена MLH1 с теми транскрипционными факторами, участие которых в канцерогенезе толстого кишечника было ранее экспериментально доказано (табл. 1): USF (Bruno *et al.*, 2004; Ansoorge *et al.*, 2007; Pare *et al.*, 2008; Belanger *et al.*, 2010; Christensen *et al.*, 2013), Ets (Wai *et al.*, 2006) и GR (например, Вутне *et al.*, 2010). Этот факт послужил отправной точкой для исследования 5 аллелей: rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена APC человека, которые не были еще ассоциированы с раком толстого кишечника и для которых в предыдущей статье (Антонцева и др., 2011; Antontseva *et al.*, 2012) были представлены первые аргументы в пользу такой ассоциации.

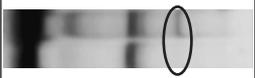
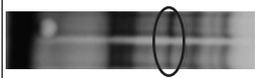
## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Результаты оценки по технологии rSNP\_Guide пяти SNPs: rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена APC человека даны в табл. 3. Для rs75996864 технология rSNP\_Guide предсказала самым вероятным белком-кандидатом, сайт связывания которого с ДНК был изменен, фактор транскрипции RAR. Это согласуется с независимыми экспериментальными данными об участии этого белка в канцерогенезе толстого кишечника (Wei *et al.*, 2003; Groubet *et al.*, 2003; Xu *et al.*, 2004; Mulholland *et al.*, 2005; Kameue *et al.*, 2006). Следующими по значимости кандидатами были транскрипционные факторы YY1 и c-Myb, участие которых в развитии рака толстого кишечника было ранее независимо показано (Chinnappan *et al.*, 2009; Ramsay *et al.*, 2003).

Для четырех других SNPs, rs76241113, rs80112297, rs80313086 и rs78037487, по оценке технологии rSNP\_Guide самыми вероятными оказались вариации потенциальных сайтов связывания транскрипционных факторов, чье участие в развитии рака толстого кишечника было также доказано экспериментально: Sp-1 (см. обзор Allgayer, 2010); MyoD (Arasaradnam *et al.*, 2012), NFκB (Andersen *et al.*, 2010) YY1 (Chinnappan *et al.*, 2009) и CP-1 (синоним NF-Y, см. например, Park *et al.*, 2007).

Таблица 3

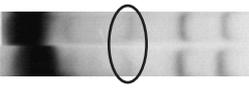
Оценка rSNP\_Guide пяти аллелей rs75996864, rs76241113, rs78037487, rs80112297 и rs80313086 гена *APC* человека для ассоциации с раком толстого кишечника

Аллель	Последовательность ДНК, ген <i>APC</i> , rs75996864 ± 25 п.о.			
Предок, WT минорный	ggcgcacgtgaccgacatgtggctg <b>T</b> attggtgcagcccgccagggtgtca ggcgcacgtgaccgacatgtggctg <b>G</b> attggtgcagcccgccagggtgtca			
Клетки Рак	НСТ-116 толстого кишечника	HeLaS3 шейки матки	НepG2 гепатома	K562 эритролейкемия
Задержка в геле				
Комплекс	WT больше	MINOR больше	возник слабый MINOR	без изменений
Данные для rSNP_Guide	$X_{WT} = 1,0;$ $X_{MINOR} = 0,5$	$X_{WT} = 1,0;$ $X_{MINOR} = 1,5$	$X_{WT} = 0,0;$ $X_{MINOR} = 0,5$	$X_{WT} = 1,0;$ $X_{MINOR} = 1,0$
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I ( $\alpha < 0,0005$ )		USF		
II ( $\alpha < 0,0025$ )	RAR	RAR		RAR
III ( $\alpha < 0,005$ )				USF
IV ( $\alpha < 0,01$ )	c-Myb, T3R, YY1	c-Myb		c-Myb, T3R
V ( $\alpha < 0,025$ )	ATF, c-Jun, CP-1, CREB, MyoD	c-Jun, CP-1, CREB, MyoD, T3R, YY1		ATF, c-Jun, CP-1, CREB, MyoD, YY1
VI ( $\alpha < 0,05$ )	AP-1, c-Fos, RXR, CRE-BP1, RXR, Sp-1, USF	AP-1, ATF, c-Fos, CRE-BP1, RXR, Sp-1	ATF, c-Fos, CREB, CRE-BP1, RXR, Sp-1, T3R	AP-1, c-Fos, Sp-1, CRE-BP1, RXR
Аллель	Последовательность ДНК, ген <i>APC</i> , rs76241113 ± 25 п.о.			
Предок, WT минорный	caggcttgctgcggggggagggggg <b>A</b> agggtggttttccctcgcaactgtctt caggcttgctgcggggggagggggg <b>G</b> agggtggttttccctcgcaactgtctt			
Клетки Рак органа	НСТ-116 толстого кишечника	HeLaS3 шейки матки	НepG2 гепатома	K562 эритролейкемия
Задержка в геле				
Комплекс	MINOR больше	MINOR нет	без изменений	WT больше
Данные для rSNP_Guide	$X_{WT} = 1,0;$ $X_{MINOR} = 1,5$	$X_{WT} = 1,0;$ $X_{MINOR} = 0,0$	$X_{WT} = 1,0;$ $X_{MINOR} = 1,0$	$X_{WT} = 1,0;$ $X_{MINOR} = 0,5$
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I ( $\alpha < 0,0005$ )	Sp-1			
II ( $\alpha < 0,001$ )			Sp-1	
III ( $\alpha < 0,0025$ )				
IV ( $\alpha < 0,005$ )	MyoD		MyoD	MyoD
V ( $\alpha < 0,01$ )	c-Myb, NF-kB, Ets		c-Myb, NF-kB, Ets	
VI ( $\alpha < 0,025$ )	GAGA, E2F		E2F, GAGA	c-Myb, E2F, Sp-1, GAGA, NF-kB
VII ( $\alpha < 0,05$ )	GR, PR, T3R	c-Myb, E2F, PR, GAGA, GR, MyoD, NF-kB, RAR, T3R, YY1	GR, PR, T3R	Ets, GR, NF-E2, PR, T3R, YY1

## Продолжение таблицы 3

Аллель	Последовательность ДНК, ген <i>APC</i> , rs80112297 ± 25 п.о.			
Предок, WT минорный	ccatggccaggcttgctgcgggggg <b>A</b> ggggggaaggtggttttccctcgca ccatggccaggcttgctgcgggggg <b>G</b> ggggggaaggtggttttccctcgca			
Клетки Рак органа	HCT-116 толстого кишечника	HeLaS3 шейки матки	HepG2 гепатома	K562 эритролейкемия
Задержка в геле				
Комплекс	MINOR больше	MINOR нет	возник MINOR	
Данные для rSNP_Guide	$X_{WT} = 1,0;$ $X_{MINOR} = 1,5$	$X_{WT} = 1,0;$ $X_{MINOR} = 0,0$	$X_{WT} = 0,0;$ $X_{MINOR} = 1,0$	
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I ( $\alpha < 0,0005$ )	Sp-1			
II ( $\alpha < 0,01$ )	Ets, MyoD, NF-kB		Ets	
III ( $\alpha < 0,025$ )	c-Myb, E2F, YY1	c-Myb, Ets, YY1, NF-kB	NF-kB, YY1	
IV ( $\alpha < 0,05$ )	GR, NF-E2, PR, T3R	NF-E2, PR, RAR, T3R	E2F, GR, MyoD, NF-E2, PR	
Аллель	Последовательность ДНК, ген <i>APC</i> , rs80313086 ± 25 п.о.			
Предок, WT минорный (редкий <sup>8</sup> ):	cttgctgcggggggaggggggaagg <b>T</b> ggttttccctcgcaactgtcttaaac cttgctgcggggggaggggggaagg <b>C</b> ggttttccctcgcaactgtcttaaac cttgctgcggggggaggggggaagg <b>G</b> ggttttccctcgcaactgtcttaaac)			
Клетки Рак органа	HCT-116 толстого кишечника	HeLaS3 шейки матки	HepG2 гепатома	K562 эритролейкемия
Задержка в геле				
Комплекс	WT больше	WT меньше	MINOR больше	возник слабый MINOR
Данные для rSNP_Guide	$X_{WT} = 1,0;$ $X_{MINOR} = 0,5$	$X_{WT} = 0,5;$ $X_{MINOR} = 1,0$	$X_{WT} = 1,0;$ $X_{MINOR} = 1,5$	$X_{WT} = 0,0;$ $X_{MINOR} = 0,5$
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I ( $\alpha < 0,0005$ )			Sp-1	
II ( $\alpha < 0,0025$ )				
III ( $\alpha < 0,005$ )			c-Myb, GR	
IV ( $\alpha < 0,01$ )	GR	NF-kB, GR	NF-kB	
V ( $\alpha < 0,025$ )	c-Myb, E2F, Sp-1 MyoD, NF-kB	c-Myb, E2F, Sp-1	E2F	GAGA, NF-E2
VI ( $\alpha < 0,05$ )	CEBP, COUP, Ets, GAGA, MEF-2, NF-E2, PR, T3R, YY1	CEBP, COUP, Ets, GAGA, MEF-2, NF-E2, PR, RAR, T3R	CEBP, COUP, GAGA, Ets, MyoD, MEF-2, NF-E2, PR, T3R	RAR, T3R, TCF-1

## Окончание таблицы 3

Аллель	Последовательность ДНК, ген <i>APC</i> , rs78037487 ± 25 п.о.			
Предок, WT минорный	agggcgctccccattcccgtcgggga <b>G</b> ccccgccgattgggctgggtgtgggcg agggcgctccccattcccgtcgggga <b>C</b> ccccgccgattgggctgggtgtgggcg			
Клетки Рак органа	НСТ-116 толстого кишечника	HeLaS3 шейки матки	НерG2 гепатома	K562 эритролейкемия
Задержка в геле				
Комплекс	без изменений	MINOR нет	возник слабый MINOR	
Данные для rSNP_Guide	$X_{WT} = 1,0;$ $X_{MINOR} = 1,0$	$X_{WT} = 1,0;$ $X_{MINOR} = 0,0$	$X_{WT} = 0,0;$ $X_{MINOR} = 0,5$	
Ранг (значимость)	rSNP_Guide: ранжирование транскрипционных факторов-кандидатов			
I ( $\alpha < 0,0025$ )	c-Myb, CP-1			
II ( $\alpha < 0,01$ )	Ets, Sp-1			
III ( $\alpha < 0,025$ )	NF-E2	c-Myb, GAGA	T3R	
IV ( $\alpha < 0,05$ )	GAGA, RF-X, T3R	Ets, RF-X	NF-E2	

<sup>S</sup> Редкий аллель не был оценен из-за отсутствия других примеров встречаемости такого типа аллелей.

С целью верификации оценок rSNP\_Guide в табл. 1 и 3 приведены результаты для исследуемых аллелей генов *APC* и *MLH1* человека в случае использования раковых клеток линий НСТ-116 (рак толстого кишечника), HeLaS3 (рак шейки матки), K562 (эритролейкемия) и НерG2 (гепатома).

Можно видеть, что как для ранее уже ассоциированного SNP rs1800734 (табл. 1), так и для 5 остальных исследованных нами SNPs (табл. 3) наблюдается одна и та же закономерность: для раковых клеточных линий различных органов человека имеются и элементы сходства оценок rSNP\_Guide, отражающие общие черты канцерогенеза в целом и существенные различия, соответствующие орган-специфическим особенностям патогенеза.

В заключение представляется важным обсудить также ограничения текущей версии технологии rSNP\_Guide (Ponomarenko *et al.*, 2001, 2002) и возможные пути усовершенствования технологии в следующих версиях.

Самым дискуссионным вопросом проведенного анализа, несомненно, является слишком большое количество транскрипционных факто-

ров-кандидатов вблизи общепринятого порога статистической значимости ( $\alpha < 0,05$ ). Это замечание типично для всех биоинформационных методов распознавания сайтов связывания факторов транскрипции и, в первую очередь, обусловлено высоким регуляторным потенциалом ДНК, лишь часть из которого реализуется в каждой конкретной клеточной ситуации (Kolchanov *et al.*, 2007). В этой связи большую актуальность приобретает дополнительная независимая верификация «припороговой» части оценок rSNP\_Guide, например, с помощью данных секвенирования футпринтов для транскрипционных факторов в масштабе генома человека в базе данных экспериментов ChIP-Seq с Энциклопедией ДНК-элементов ENCODE (Auerbach *et al.*, 2013).

Очевидным недостатком текущей версии rSNP\_Guide (Ponomarenko *et al.*, 2001, 2002) является также ее ограниченность 40 регуляторными белками (табл. 2), тогда как, например, система SITECON (Oshchepkov *et al.*, 2004) включает более 100 транскрипционных факторов. Следовательно, требуется обновление rSNP\_Guide за счет существенного расширения количества анализируемых белков.

Наконец, среди транскрипционных факторов-кандидатов, сайты связывания которых изменяются в случае SNPs rs75996864, rs76241113, rs78037487, rs80112297, rs80313086 гена *APC* человека был предсказан фактор ТЗР, участие которого в развитии рака толстого кишечника до сих пор не было описано. Поскольку в более чем 30 экспериментальных статьях описано участие этого белка в развитии других типов рака у человека, например, при патогенезе рака молочной железы (Alvarado-Pisani *et al.*, 1986), представляется целесообразным экспериментально проверить влияние изученных SNPs в гене *APC* на связывание ТЗР.

Представленные нами результаты свидетельствуют о получении по технологии rSNP\_Guide (Ponomarenko *et al.*, 2001, 2002) новых оснований для дальнейшего исследования ассоциации SNPs rs75996864, rs76241113, rs78037487, rs80112297, rs80313086 гена *APC* с раком толстого кишечника общепринятыми медико-генетическими методами.

Работа была поддержана Госконтрактом № 14.512.11.0094 и Соглашением № 8740 Минобрнауки РФ, а также частично – молодежным проектом поддержки ведущих научных школ НШ-5278.2012.4, программами Президиума РАН «Молекулярная и клеточная биология» № 6.6 и «Фундаментальные науки – медицине» № 23.

## ЛИТЕРАТУРА

- Антонцева Е.В., Брызгалов Л.О., Матвеева М.Ю. и др. Поиск регуляторных SNPs, связанных с развитием рака толстой кишки, в генах *APC* и *MLH1* // Вавилов. журн. генет. и селекции. 2011. Т. 15. Вып. 4. С. 644–652.
- Allgayer H. Pcd4, a colon cancer prognostic that is regulated by a microRNA // Crit. Rev. Oncol. Hematol. 2010. V. 73. No. 3. P. 185–191.
- Alvarado-Pisani A.R., Chacon R.S., Betancourt L.J., Lopez-Herrera L. Thyroid hormone receptors in human breast cancer: effect of thyroxine administration // Anticancer Res. 1986. V. 6. No. 6. P. 1347–1351.
- Andersen V., Christensen J., Overvad K. *et al.* Polymorphisms in NFkB, PXR, LXR and risk of colorectal cancer in a prospective study of Danes // BMC Cancer. 2010. V. 10. P. 484.
- Ansoorge N., Juttner S., Cramer T. *et al.* An upstream CRE-Box element is essential for gastrin-dependent activation of the cyclooxygenase-2 gene in human colon cancer cells // Regul. Pep. 2007. V. 144. No. 1/3. P. 25–33.
- Antontseva E.V., Bryzgalov L.O., Matveeva M.Yu. *et al.* Search for regulatory SNPs associated with colon cancer in the *APC* and *MLH1* genes // Russ. J. Genet. Appl. Res. 2012. V. 2. No. 3. P. 222–228.
- Arasaradnam R.P., Quraishi M.N., Commane D. *et al.* MYOD-1 in normal colonic mucosa—role as a putative biomarker? // BMC Res. Notes. 2012. V. 5. P. 240.
- Auerbach R.K., Chen B., Butte A.J. Relating genes to function: identifying enriched transcription factors using the ENCODE ChIP-Seq significance tool // Bioinformatics. 2013. V. 29. No. 15. P. 1922–1924.
- Belanger A.S., Tojcic J., Harvey M., Guillemette C. Regulation of UGT1A1 and HNF1 transcription factor gene expression by DNA methylation in colon cancer cells // BMC Mol. Biol. 2010. V. 11. P. 9.
- Berg O.G., von Hippel P.H. Selection of DNA binding sites by regulatory proteins, Statistical-mechanical theory and application to operators and promoters // J. Mol. Biol. 1987. V. 193. No. 4. P. 723–750.
- Bruno M.E., West R.B., Schneeman T.A. *et al.* Upstream stimulatory factor but not c-Myc enhances transcription of the human polymeric immunoglobulin receptor gene // Mol. Immunol. 2004. V. 40. No. 10. P. 695–708.
- Byrne A.M., Foran E., Sharma R. *et al.* Bile acids modulate the Golgi membrane fission process via a protein kinase Ceta and protein kinase D-dependent pathway in colonic epithelial cells // Carcinogenesis. 2010. V. 31. No. 4. P. 737–744.
- Chinnappan D., Xiao D., Ratnasari A. *et al.* Transcription factor YY1 expression in human gastrointestinal cancer cells // Int. J. Oncol. 2009. V. 34. No. 5. P. 1417–1423.
- Christensen L.L., Tobiasen H., Holm A. *et al.* MiRNA-362-3p induces cell cycle arrest through targeting of E2F1, USF2 and PTPN1 and is associated with recurrence of colorectal cancer // Int. J. Cancer. 2013. V. 133. No. 1. P. 67–78.
- Gerstenblith M.R., Shi J., Landi M.T. Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis // Pigment Cell Melanoma Res. 2010. V. 23. No. 5. P. 587–606.
- Groubet R., Pallet V., Delage B. *et al.* Hyperlipidic diets induce early alterations of the vitamin A signalling pathway in rat colonic mucosa // Endocr. Regul. 2003. V. 37. No. 3. P. 137–144.
- Hamosh A., Scott A.F., Amberger J.S. *et al.* Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders // Nucl. Acids Res. 2005. V. 33. P. D514–D517.
- Hindorff L.A., Sethupathy P., Junkins H.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits // Proc. Natl Acad. Sci. USA. 2009. V. 106. No. 23. P. 9362–9367.
- Hitchins M.P., Wong J.J.L., Suthers G. *et al.* Inheritance of a cancer-associated *MLH1* germ-line epimutation // New Eng. J. Med. 2007. V. 356. No. 7. P. 697–705.
- Kameue C., Tsukahara T., Ushida K. Alteration of gene expression in the colon of colorectal cancer model rat by dietary sodium gluconate // Biosci. Biotechnol. Biochem. 2006. V. 70. No. 3. P. 606–614.
- Kolchanov N.A., Merkulova T.I., Ignatieva E.V. *et al.* Combined experimental and computational approaches to study the regulatory elements in eukaryotic genes // Brief Bioinform. 2007. V. 8. No. 4. P. 266–274.

- Mulholland D.J., Dedhar S., Coetzee G.A., Nelson C.C. Interaction of nuclear receptors with the Wnt/beta-catenin/Tcf signaling axis: Want you like to know? // *Endocrinol. Rev.* 2005. V. 26. No. 7. P. 898–915.
- NCBI Resource Coordinators, Database resources of the National Center for Biotechnology Information // *Nucl. Acids Res.* 2013. V. 41. P. D8–D20.
- Oshchepkov D.Y., Vityaev E.E., Grigorovich D.A. *et al.* SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition // *Nucl. Acids Res.* 2004. V. 32. Web Server issue. P. W208–W212.
- Pare L., Marcuello E., Altas A. *et al.* Transcription factor-binding sites in the thymidylate synthase gene: predictors of outcome in patients with metastatic colorectal cancer treated with 5-fluorouracil and oxaliplatin? // *Pharmacogenomics J.* 2008. V. 8. No. 5. P. 315–320.
- Park S.H., Yu G.R., Kim W.H. *et al.* NF-Y-dependent cyclin B2 expression in colorectal adenocarcinoma // *Clin. Cancer Res.* 2007. V. 13. No. 3. P. 858–867.
- Polakis P. An Introduction to Wnt Signaling // *Targeting the Wnt Pathway in Cancer*. N.Y.: Springer, 2011. P. 1–18.
- Ponomarenko J.V., Merkulova T.I., Vasiliev G.V. *et al.* rSNP\_Guide, a database system for analysis of transcription factor binding to target sequences: application to SNPs and site-directed mutations // *Nucl. Acids Res.* 2001. V. 29. No. 1. P. 312–316.
- 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. V. 20. No. 4. P. 239–248.
- Ponomarenko M.P., Ponomarenko J.V., Frolov A.S. *et al.* Oligonucleotide frequency matrices addressed to recognizing functional DNA sites // *Bioinformatics.* 1999. V. 15. P. 631–643.
- Ramsay R.G., Ciznadija D., Vanevski M., Mantamadiotis T. Transcriptional regulation of cyclo-oxygenase expression: three pillars of control // *Int. J. Immunopathol. Pharmacol.* 2003. V. 16. No. 2 (Suppl). P. 59–67.
- Sanchez-Ruiz J.M. Protein kinetic stability // *Biophys. Chem.* 2010. V. 148. P. 1–15.
- The International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome // *Nature.* 2004. V. 431. No. 7011. P. 931–945.
- Wai P.Y., Mi Z., Gao C. *et al.* Ets-1 and runx2 regulate transcription of a metastatic gene, osteopontin, in murine colorectal cancer cells // *J. Biol. Chem.* 2006. V. 281. No. 28. P. 18973–18982.
- Wei H.B., Han X.Y., Fan W. *et al.* Effect of retinoic acid on cell proliferation kinetics and retinoic acid receptor expression of colorectal mucosa // *World J. Gastroenterol.* 2003. V. 9. No. 8. P. 1725–1728.
- Win A.K., Hopper J.L., Buchanan D.D. *et al.* Are the common genetic variants associated with colorectal cancer risk for DNA mismatch repair gene mutation carriers? // *Eur. J. Cancer.* 2013. V. 49. No. 7. P. 1578–1587.
- Xu X.L., Yu J., Zhang H.Y. *et al.* Methylation profile of the promoter CpG islands of 31 genes that may contribute to colorectal carcinogenesis // *World J. Gastroenterol.* 2004. V. 10. No. 23. P. 3441–3454.

## rSNP\_Guide-BASED EVALUATION OF SNPs IN PROMOTERS OF THE HUMAN *APC* AND *MLH1* GENES ASSOCIATED WITH COLON CANCER

D.A. Rasskazov<sup>1</sup>, E.V. Antontseva<sup>1</sup>, L.O. Bryzgalov<sup>1</sup>, M.Yu. Matveeva<sup>1</sup>, E.V. Kashina<sup>1</sup>, P.M. Ponomarenko<sup>1</sup>, G.V. Orlova<sup>1</sup>, M.P. Ponomarenko<sup>1</sup>, D.A. Afonnikov<sup>1,2</sup>, T.I. Merkulova<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, e-mail: pon@bionet.nsc.ru;

<sup>2</sup> Novosibirsk National Research State University, Novosibirsk, Russia

### Summary

Six regulatory SNPs (single nucleotide polymorphisms) of two human *APC* and *MLH1* genes (rs75996864, rs76241113, rs78037487, rs80112297, rs80313086 and rs1800734) were evaluated by the previously developed rSNP\_Guide method to compute the significance of the changes for binding of the SNP region with 40 transcription factors. For each SNP, all analyzed proteins were ranged according to the significance of the changes for protein binding to the alleles evaluated by Student's *t*-test. We found that rs75996864, rs76241113, rs78037487, rs80112297, and rs80313086 of *APC*, as well as rs1800734 of *MLH1* in humans concerned to the greatest extent the binding of transcription factors NF-Y, NFkB, c-Myb, RAR, YY-1, and Sp-1, which are known to be involved in colon cancer development. Our results substantiate investigation of associations of rs75996864, rs76241113, rs78037487, rs80112297, and rs80313086 in the *APC* gene with colon cancer by using commonly accepted medical and genetic protocols.

**Key words:** single-nucleotide polymorphism, gene expression regulation, colon cancer, *APC*, *MLH1* genes, DNA–regulatory protein complex, Student's *t*-test.