

WASF4 – НОВЫЙ ЧЛЕН СЕМЕЙСТВА ГЕНОВ WAS

© 2008 г. Е. А. Климов*, Э. Р. Рахманалиев, Г. Е. Сулимова

Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991

Поступила в редакцию 29.12.2007 г.

Принята к печати 29.02.2008 г.

Определена экзон-интронная структура гена *WASF4* человека. Проведен компьютерный анализ промоторной области гена, в которой выявлены участки связывания факторов транскрипции. Наибольшее сходство белок *WASF4* имеет с белком *WASF2* человека. В геномах шимпанзе и макаки обнаружен гомолог гена *WASF4*, в геномах других позвоночных *WASF4*-подобные нуклеотидные последовательности не найдены. Экспрессия гена *WASF4* в тканях человека не обнаружена.

Ключевые слова: *WASF4*, экзон-интронная структура гена, анализ *in silico*, человек.

A NEW MEMBER OF THE WAS GENES FAMILY (*WASF4*), by E. A. Klimov, E. R. Rakhmanaliev, G. E. Sulimova (Vavilov Institute of General Genetics, Russian Academy of Science, Moscow, 119991 Russia; *e-mail: klimov_eugene@mail.ru). The exon-intron structure of the human *WASF4* gene has been determined. The *in silico* analysis of the gene promoter region was performed and the presence of transcription factor binding sites was shown. The highest similarity between the *WASF4* protein and the human *WASF2* protein was revealed. The *WASF4* gene homolog was found in chimpanzee and macaque genomes; *WASF4* like nucleotide sequences were not found in other vertebrate genomes. The *WASF4* gene expression in human tissues was not detected.

Key words: *WASF4*, exon-intron structure, *in silico* analysis, human.

Белки семейства WAS (Wiskott-Aldrich syndrome, синдром Вискотта–Олдрича) играют важную роль в полимеризации актина и реорганизации цитоскелета. Эти два процесса необходимы для подвижности клеток, роста нейронов и других функций клеток [1]. В настоящее время известно пять генов человека, входящих в семейство WAS: *WASF1* (6q21–q22), *WASF2* (1p36.11–p34.3), *WASF3* (13q12), *WAS* (Xp11.4–p11.21) и *WASL* (7q31.3). Все белки, кодируемые этими генами, имеют сходную доменную структуру. С-концевой район содержит небольшой (21 аминокислотный остаток) домен WH2, обогащенный кислыми аминокислотами, который встречается только в белках WAS и может связываться с актином, а также кофилин-подобный домен, способный взаимодействовать с полимерным и мономерным актином. С помощью этих доменов белки WAS активируют полимеризацию актина [2]. Пролин-богатый домен PRD (P21-Rho-binding domain, около 200 аминокислотных остатков) консервативен у всех белков WAS, он служит для связывания с Cdc42p и/или Rho-подобными малыми GTPазами. Показано, что домен PRD служит посредником между WAS-белками и сигнал-усиливающими SH3-содержащими белками [3–5]. N-концевой район

WAS-белков содержит домен EVH1 (115 аминокислотных остатков), который может взаимодействовать с 25 аминокислотными остатками пролин-богатого мотива белка WIP, взаимодействующего с WASP (WASP interactive protein). Присутствие нескольких функциональных мотивов свидетельствует о существовании различных способов регуляции активности белков WAS и о возможности их связывания со многими белками [3, 6–10].

Ген WAS человека идентифицировали благодаря мутации, которая приводит к развитию синдрома Вискотта–Олдрича [11, 12] – наследственной болезни, проявляющейся тромбоцитопенией, иммунодефицитом и экземой [13]. Впоследствии идентифицировали гены *WASF1*, *WASF2* и *WASF3* [14, 15]. Изучение транскрипционной активности выявили тканевую специфичность экспрессии генов WAS [11, 12, 15–17]. WAS экспрессируется в основном в крови, а WASL – в крови и нервных клетках [11]. *WASF1* наиболее активен в головном мозге и легких, его мРНК обнаружена также в плаценте, печени и поджелудочной железе. Ген *WASF2* экспрессируется в плаценте и легких, другие ткани содержат следовые количества его мРНК. Высокий уровень экспрессии *WASF3* выявлен в головном мозге, хотя его мРНК обнаружена в легких, семенниках, почках и скелетной мускулатуре [15, 16].

*Эл. почта: klimov_eugene@mail.ru

Таблица 1. Праймеры к экзонам гена *WASF4* человека

Праймер	Нуклеотидная последовательность 5'–3'	Экзон	$T_{\text{отжига}}, ^\circ\text{C}$	Длина фрагмента, п.н.
1Fn	TATTGTCTGGAGCATTATGGTGCACATC	2	58	381
1Rn	CCAGCTGAGTGACTTTAACTTGTAGTCCA	3		
2F	CCAAAGAAAAGCTGGGGCCTTT	3	49	около 1250*
2R (dT)	TTTTGTACAAGCTT ₍₃₀₎	полиА		

* Сигнал полиаденилирования предсказан, а не определен экспериментально, поэтому не представляется возможным определить точную длину предполагаемого продукта ПЦР.

Геномная организация и тканевая специфичность экспрессии генов *WASF1*, *WASF2* и *WASF3* человека и мыши сходны, как и их геномное окружение. Считается, что ген *WASF2* наиболее древний, так как у человека и мыши эти гены имеют наименьшее сходство [17].

Ранее [18] мы картировали в области Xp11.2 нуклеотидную последовательность (Асс. No. AJ431612), обладающую сходством (96%) с последовательностью экзона 8 и 9 гена *WASF2* (локализованного в области 1p36.11–p34.3). В представленной работе приведены результаты анализа картированной нами последовательности.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Компьютерные программы и базы данных. Поиск гомологий проводили с использованием программы BLASTN (<http://www.ncbi.nlm.nih.gov/BLAST/>). Экзон-интронные структуры строили также с помощью программы GENSCAN (<http://genes.mit.edu/GENSCAN.html>). Результаты анализа экзон-интронной структуры гена *WASF4* корректировали вручную.

Поиск промоторных областей проводили с использованием программы PromoterInspector (<http://www.genomatix.de/cgi-bin/eldorado/main.pl>) и Promoter Prediction (http://www.fruitfly.org/seq_tools/promoter.html). Поиск участков связывания факторов транскрипции – с помощью программы MatInspector (<http://www.genomatix.de>). Аминокислотную последовательность, кодируемую новым геном, транслировали *in silico* с помощью сервиса DNA=>Protein (<http://cn.expasy.org/tools/dna.html>).

Анализ экспрессии гена *WASF4*. Матрицей для анализа транскрипционной активности гена *WASF4* человека служила кДНК из молочной железы, почки, легкого, яичника, шейки матки и крови человека.

Препараты мРНК выделяли с использованием набора Trizol RNA Prep 100 (“IsoGen”, Россия), кДНК синтезировали с помощью набора GenePakTM RT Core (“IsoGen”) согласно рекомендациям изготовителя.

Праймеры подбирали с помощью программы OligoCalculator (автор Кубаренко А.В.). Нуклеотидные последовательности праймеров представлены в табл. 1.

Полимеразную цепную реакцию (ПЦР) проводили с использованием набора GenePakTM PCR Core (“IsoGen”) на термоциклере PTC-100TM (“MJ Research Inc.”) по схеме: 95°C – 1 мин; 35 циклов – 94°C – 30 с, 58°C или 49°C – 20 с, 72°C – 30–90 с; 72°C – 5 мин. Температура отжига всех пар праймеров приведена в табл. 1. Продукты ПЦР разделяли путем электрофореза в 2%-ном агарозном геле с бромистым этидием, выделяли из геля с использованием набора Gel-Wizard (“Promega”). Нуклеотидную последовательность продуктов ПЦР определяли на приборе ABI-310 (“Applied Biosystems”) согласно рекомендациям изготовителя.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Анализ геномного окружения картированной последовательности выявил в нем открытую рамку считывания. С помощью программы GENSCAN в изучаемой нуклеотидной последовательности обнаружены три экзона. Нуклеотидная последовательность предполагаемого гена, названного нами *WASF4*, зарегистрирована в базе данных EMBL (Асс. No. AJ509090, далее вся нумерация нуклеотидов приведена согласно этой последовательности). С использованием программы PromoterInspector найден промоторный участок (769–1012 п.н.), содержащий участки связывания основных факторов транскрипции: TATA, OCT1, GATA1, GATA2, GATA3, AP1. С помощью программы Promoter Prediction предсказано присутствие в 5'-области гена двух участков инициации транскрипции (41 и 1105 п.н.). Первый участок находится на расстоянии –728 п.н. от начала промотора, что несвойственно генам, транскрибируемым РНК-полимеразой II. Поэтому наиболее вероятно, что участок инициации транскрипции расположен сразу после промотора, в позиции 1105 п.н. Его положение приходится на начало первого интрона, предсказанного программой GENSCAN. Однако анализ полученной *in silico* аминокислотной последовательности указывает на отсутствие гомологии N-концевого участка данного белка с другими белками семейства WAS, а аминокислотная последовательность, имеющая такую гомологию, может транслироваться, начиная со стартового кодона ATG в позиции 1854–1856 п.н., а не 1013–1015 п.н., как предсказано программой GENSCAN. Найден также стоп-кодон (TAA – 7874–7876 п.н.) и два сигнала полиаденилирования (AATAAT – 8260–8265 п.н. и

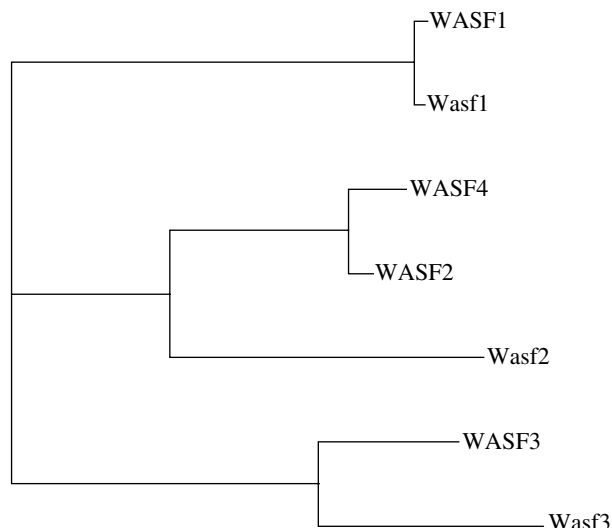
АСАААG – 9181–9186 п.н.), предсказанных ранее [19]. Таким образом, анализируемая последовательность содержит три экзона, из которых только первый не соответствует предсказанию программы GENSCAN. Первый экзон состоит из 18 п.н. (от предсказанной точки инициации транскрипции 1105 п.н. до участка сплайсинга 1122 п.н.), второй – из 304 п.н. (участок 1681–1984 п.н., стартовый кодон в позиции 1854–1856 п.н.), третий – из 1884 п. н. (от участка сплайсинга в позиции 6376 до предсказанного программой первого участка полиаденилирования в позиции 8260). Все экзоны фланкированы каноническими участками сплайсинга (AG...GT).

Предполагаемый продукт изучаемого гена (543 аминокислотных остатка) имеет высокий уровень сходства (90,2%) с белком WASF2 человека (498 аминокислотных остатков) и такую же доменную структуру, что позволяет считать эти белки гомологами, принадлежащими к одному семейству. Первые 46 аминокислотных остатков WASF4 не имеют гомологии с другими членами семейства WAS.

Гомология мРНК генов *WASF2* и *WASF4* составляет 96–98%. Третий, белкокодирующий, экзон гена *WASF4* гомологичен белкокодирующим экзонам гена *WASF2* (экзоны 2–8, всего девять экзонов). Нуклеотидные последовательности экзонам 1 и 2, а также интронам гена *WASF4*, в геноме человека не обнаружено. Нуклеотидная последовательность 5'-области гена *WASF4* (промоторная область и часть первого экзона) также уникальна в геноме человека.

Выравнивание аминокислотных последовательностей *WASF4* и трех белков WASF человека и мыши показывает, что *WASF4* ближе всего к *WASF2* человека (рисунок).

WASF2 найден в геномах грызунов, приматов, человека. *WASF4*-подобные нуклеотидные последовательности обнаружены только в геномах макаки и шимпанзе, в геномах других позвоночных (рыб, грызунов, собаки и др.) таких последовательностей нет. В геноме макаки – это нуклеотидные последовательности, гомологичные экзонам 2 (96% гомологии) и 3 (95%). В геноме шимпанзе найдена последовательность, гомологичная (98%) последователь-



Филогенетическое дерево белков WASF (программа Clustal W). Для сравнения использованы опубликованные аминокислотные последовательности белков WASF1, WASF2, WASF3 человека и Wasf1, Wasf2, Wasf3 мыши [17] и полученная *in silico* аминокислотная последовательность WASF4 человека.

ности всего гена *WASF4* человека и также локализованная на хромосоме X.

Участки сплайсинга экзонов 2 и 3 в генах *WASF4* человека и шимпанзе совпадают, существенные различия имеются только в первом экзоне (табл. 2). Промоторная область гена *WASF4* шимпанзе, определенная с помощью программы PromoterInspector, состоит из 601 п.н. Внутри промоторной области расположен участок инициации транскрипции, выявленный с использованием программы Promoter Prediction. Открытая рамка считывания *WASF4* шимпанзе не соответствует рамке считывания *WASF4* человека. В начале третьего экзона (через 19 п.н. от начала экзона) гена *WASF4* шимпанзе имеется делеция 14 п.н., в результате открытая рамка считывания прерывается несколькими стоп-кодонами. Ген *WASF4* шимпанзе кодирует только белок из 60 аминокислотных остатков, гомологичный фрагменту 1–50 белка *WASF4* человека (гомология 98%).

Таблица 2. Размер экзонов и интронов гена *WASF4* (п.н.) человека и шимпанзе и уровень их гомологии (%)

Организм	Экзон 1	Интрон 1	Экзон 2	Интрон 2	Экзон 3*
Человек	18	558	304	4391	1501
Шимпанзе	18	510	304	4378	1525
Гомология, %	25.9	68.1 (фрагмент 314 п.н. – 99)	99.1	75.8**	95.5

* До стоп-кодона включительно (ТАА).

** В интроне шимпанзе имеется участок неизвестной последовательности (N₅₀₇).

Примечание. 3'-нетранслируемые области не сравнивали, так как у шимпанзе в этой области гена имеется участок неизвестной последовательности (N₁₂₃₉).

Присутствие последовательностей, сходных с *WASF4* человека, только в геноме приматов позволяет предположить, что ген *WASF4* появился недавно и его предшественником, по-видимому, был ген *WASF2*. Можно предположить, что возникновение *WASF4* связано с интеграцией части процессированной кДНК гена *WASF2* в геном гипотетического общего предка шимпанзе, макаки и человека. После встраивания кДНК в предковый геном экзоны 2–8 гена *WASF2* оказываются слитыми, они образуют один экзон в гене *WASF4* (экзон 3).

Невыясненным остается вопрос о транскрипционной активности гена *WASF4* человека. Это связано с высокой гомологией мРНК генов *WASF4* и *WASF2* (96–98%), что затрудняет использование большинства молекулярно-генетических методов. Следует отметить, что в базе данных EST представлены только две кДНК, приписываемые локусу *WASF4* человека – XM_929457 и XM_933679. Причем XM_929457 получена *in silico* и гомологична всей предсказанной нами мРНК гена *WASF4* (включая часть, предсказанную как экзон программой GENSCAN), а XM_933679 – только той части, которая сходна с *WASF2* человека.

Экспрессию гена *WASF4* человека анализировали методом ПЦР с использованием двух пар праймеров, специфичных к экзонам *WASF4*, и кДНК из молочной железы, почки, легкого, яичника, шейки матки, крови человека в качестве матрицы. С помощью пары праймеров 3F + 3R (экзоны 2–3) продуктов ПЦР получить не удалось. С парой праймеров 2F + 2R(dT) (экзон 3–полиА) продукта ПЦР ожидаемой длины (около 1250 п.н.) получить также не удалось. Полученный фрагмент выделили из геля и определили его нуклеотидную последовательность. Показано, что это фрагмент 3'-области гена *WASF2*. Таким образом, нам не удалось обнаружить экспрессию описанного нами гена *WASF4* человека в исследуемых органах. Возможно, она ограничена узким временным отрезком на определенных стадиях онтогенеза. Нельзя исключить также, что ген *WASF4* экспрессируется в других органах. Выяснить это помогут дальнейшие исследования.

Работа выполнена при поддержке Фонда “Содействие отечественной науке” и Программы Президента РФ “Поддержка научных школ” (НШ-3914.2008.4).

СПИСОК ЛИТЕРАТУРЫ

- Mitchison T.J., Cramer L.P. 1996. Actin-based cell motility and cell locomotion. *Cell*. **84**, 371–379.
- Takenawa T., Miki H. 2001. WASP and WAVE family proteins: key molecules for rapid rearrangement of cortical actin filaments and cell movement. *J. Cell Sci.* **114**, 1801–1809.
- Fukuoka M., Suetsugu S., Miki H., et al. 2001. A novel neural Wiskott-Aldrich syndrome protein (N-WASP) binding protein, WISH, induces Arp2/3 complex activation independent of Cdc42. *J. Cell Biol.* **152**, 471–482.
- Krugmann S., Jordens I., Gevaert K., et al. 2001. Cdc42 induces filopodia by promoting the formation of an IRSp53:Mena complex. *Curr. Biol.* **11**, 1645–1655.
- Miki H., Takenawa T. 2002. WAVE2 serves a functional partner of IRSp53 by regulating its interaction with Rac. *Biochem. Biophys. Res. Commun.* **293**, 93–99.
- Biesova Z., Piccoli C., Wong W.T. 1997. Isolation and characterization of e3B1, an eps8 binding protein that regulates cell growth. *Oncogene*. **14**, 233–241.
- Carlier M.F., Nioche P., Broutin-L'Hermite I., et al. 2000. GRB2 links signaling to actin assembly by enhancing interaction of neural Wiskott-Aldrich syndrome protein (N-WASP) with actin-related protein (ARP2/3) complex. *J. Biol. Chem.* **275**, 21946–21952.
- Oda K., Shiratsuchi T., Nishimori H., et al. 1999. Identification of BAIAP2 (BAI-associated protein 2), a novel human homologue of hamster IRSp53, whose SH3 domain interacts with the cytoplasmic domain of BAI1. *Cytogenet. Cell Genet.* **84**, 75–82.
- Okamura-Oho Y., Miyashita T., Ohmi K., Yamada M. 1999. Dentatorubral-pallidoluysian atrophy protein interacts through a proline-rich region near polyglutamine with the SH3 domain of an insulin receptor tyrosine kinase substrate. *Hum. Mol. Genet.* **8**, 947–957.
- Rohatgi R., Nollau P., Ho Hy H., et al. 2001. Nck and phosphatidylinositol 4,5-bisphosphate synergistically activate actin polymerization through the N-WASP-Arp2/3 pathway. *J. Biol. Chem.* **276**, 26448–26452.
- Derry J.M.J., Ochs H.D., Francke U. 1994. Isolation of a novel gene mutated in Wiskott-Aldrich syndrome. *Cell*. **78**, 635–644.
- MacCarthy-Morrogh L., Gaspar H.B., Wang Y.C., et al. 1998. Absence of expression of the Wiskott-Aldrich syndrome protein in peripheral blood cells of Wiskott-Aldrich syndrome patients. *Clin. Immunol. Immunopathol.* **88**, 22–27.
- Cooper M.D., Chae H.P., Lowman J.T., et al. 1968. Wiskott-Aldrich syndrome: an immunologic deficiency disease involving the afferent limb of immunity. *Am. J. Med.* **44**, 499–513.
- Miki H., Suetsugu S., Takenawa T. 1998. WAVE, a novel WASP-family protein involved in actin reorganization induced by Rac. *EMBO J.* **17**, 6932–6941.
- Suetsugu S., Miki H., Takenawa T. 1999. Identification of two human WAVE/SCAR homologues as general actin regulatory molecules which associate with the Arp2/3 complex. *Biochem. Biophys. Res. Commun.* **260**, 296–302.
- Sossey-Alaoui K., Su G., Malaj E., et al. 2002. WAVE3, an actin-polymerization gene, is truncated and inactivated as a result of a constitutional t(1;13)(q21;q12) chromosome translocation in a patient with ganglioneuroblastoma. *Oncogene*. **21**, 5967–5974.
- Sossey-Alaoui K., Head K.N., Nowak N., Cowell J.K. 2003. Genomic organization and expression profile of the human and mouse WAVE gene family. *Mamm. Genome*. **14**, 314–322.
- Klimov E., Vinokourova S., Moisjak E., et al. 2002. Human papilloma viruses and cervical tumours: mapping of integration sites and analysis of adjacent cellular sequences. *BMC Cancer*. **2**, 24.
- Климов Е.А. 2008. Потенциальные сигналы терминации транскрипции интегрированной ДНК вируса папилломы человека: анализ *in silico*. *Молекуляр. биология*. **42**, 178–180.