

УДК 577.21

В.А. ХАЙЛЕНКО, Ш.А. АТАМБАЕВА, А.Т. ИВАЩЕНКО

## ИЗМЕНЕНИЯ ДЛИНЫ ЭКЗОНОВ И ИНТРОНОВ В ГЕНАХ 7-12 ХРОМОСОМ ЧЕЛОВЕКА

(Казахский национальный университет им. аль-Фараби)

В генах хромосом человека 7–12 происходят изменения длин экзонов, интронов и генов в зависимости от плотности расположения генов в участках ДНК и от числа интронов в генах. Выявлена линейная корреляция между изменениями длин этих структурных элементов генов.

После секвенирования генома человека [1] изучение структурно-функциональной организации его генов позволило выявить много свойств, способствующих пониманию регуляции и механизмов реализации генетической информации [2–5]. Полученные данные активно используются фармацевтическими и биотехнологическими фирмами для получения лекарственных средств и развития молекулярной медицины (<http://www.ornl.gov/hgmis/medicine/medicine.shtml>). Однако сложность структурно-функциональной организации генома человека велика, и предстоит еще огромная работа по ее изучению.

В генах человека длина интронов изменяется от нескольких десятков до десятков тысяч нуклеотидов, и в среднем экзоны в 22 раза короче интронов [1]. Это обстоятельство сильно затрудняет изучение как интронов, так и белок-кодирующей части генов. Многие способы регуляции экспрессии генов связаны с интронами. Например, альтернативный сплайсинг пре-мРНК позволяет увеличивать в несколько раз разнообразие продуктов экспрессии генов [6–8]. Вместе с тем нарушение этого процесса вызывает различные заболевания, ввиду чего важно знать механизмы этого явления. Биологическая роль интронов остается невыясненной, поэтому требуется дальнейшее исследование экзон-интронной организации генов человека.

На модельных организмах установлено, что между длиной экзонов и интронов существует взаимосвязь [9,10], и представляется важным проверить ее наличие в генах человека. При изучении генов 1-6 хромосом человека показано, что гетерогенность распределения генов на нитях хромосом существенно сказывается на характеристиках генов [11]. Установлено, что длина экзонов и интронов существенно зависит от числа

интронов в генах [12]. Хромосомы человека отличаются друг от друга не только по размеру, но и по плотности генов, GC-содержанию, распределению повторяющихся последовательностей и т.д. [1,13–16]. В связи с этим необходимо исследование экзон-интронной организации генов всех хромосом в целях выявления особенностей этой характеристики генов.

### МАТЕРИАЛЫ И МЕТОДЫ

Нуклеотидные последовательности генов и ДНК 7-12 хромосом *Homo sapiens* получены из GenBank (<http://www.ncbi.nlm.nih.gov>). ДНК каждой хромосомы последовательно делили на участки длиной 1 Мбр и в каждом находили число генов. Участки по числу генов в них распределяли в группы с различной плотностью: 1-11, 12-20, 21 генов/Мбр и более. В каждой группе гены распределяли в выборки с 1-2, 3-5, 6-9, 10-14, 15 и более интронами в гене. Определяли длину экзонов ( $l_{ex}$ ), интронов ( $l_{in}$ ), сумму длин экзонов ( $L_{ex}$ ) в гене, длину гена ( $L_{gn}$ ) и долю длины экзонов ( $L_{ex}/L_{gn}$ ) в гене.  $N_{in}$  – число интронов в гене,  $N_{gn}$  – число генов в выборке. Анализировали количество интронов и экзонов с длиной в интервалах 1-20, 21-40, 41-60 п и т. д. до 400 п, а также с длиной более 400 п.

### РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Хромосомы 7-12 генома человека имеют общую длину около 800 Мбр и содержат более 7 тысяч генов [1]. В каждой из хромосом проявляется большая гетерогенность плотности распределения генов. Например, число генов на 1 Мбр ДНК 7 хромосомы равнялось нулю в 23 участках и в 11 участках было выше 21 гена. В 8 хромосоме гены отсутствовали в 28 участках и только 5 участков имели плотность выше 21 гена/Мбр. В хромосомах 11 и 12 генов не было

соответственно в 16 и 9 участках, а в 20 и 16 участках было больше 21 гена.

В участках с плотностью генов 1-11 их доля составляла 47% от общей нуклеотидной последовательности ДНК. В участках с плотностью 12-20 генов, и 21 генов/Мбр и более гены занимали соответственно 61 и 28% ДНК. Следовательно, в хромосомах 7-12 нет прямой зависимости доли ДНК, занимаемой генами, от их плотности в участках длиной 1 Мбр.

Участки делили на три группы по плотности генов так, чтобы в каждой группе было в среднем примерно равное число генов. Однако вариация числа генов в группах участков 7-12 хромосом была значительной, и пределы чисел генов в группах были таковы: в хромосоме 10 число генов в 1, 2 и 3 группах было 567, 522 и 89 генов, а в хромосоме 11 – соответственно 385, 489 и 794. В целом в каждой из групп выборка генов была репрезентативной для определения достоверности полученных данных.

Результаты определения длины экзонов и интронов в выборках генов из групп участков приведены для всех изученных хромосом в табл. 1 и 2. Полученные данные показывают уменьшение длины экзонов в генах группы участков с плотностью 1-11 генов/Мбр по мере увеличения в них числа интронов. Средняя длина экзонов снижалась в 1,7 раза. В генах из других групп участков уменьшение длины экзонов с ростом числа интронов в генах было близким (см. табл. 1 и 2) с уровнем достоверности  $p < 0,001$ .

Уменьшение средней длины экзонов происходило за счет снижения доли экзонов с длиной более 400 п и образования максимума длин экзонов в интервале 100-140 п (рис. 1). В генах с 1-2 интронами доля экзонов с длиной более 400 п была в несколько раз больше, чем в генах с 15 интронами и более.

В генах первой группы (1-11 генов/Мбр) с 1-2 интронами доля экзонов с длиной более 400 п равнялась 16% и в генах с 15 и более интронами таких экзонов было только 1% (рис. 1). Такой же характер изменения длины экзонов при увеличении числа интронов в гене наблюдался в группе с 21 и более генов/Мбр (см. рис. 1). Соответствующие доли экзонов в этой группе генов составляли 19 и 1%. Следовательно, изменение длины экзонов в гене происходило только с рос-

том числа интронов в гене и не зависело от плотности генов в участках ДНК.

Интроны с длиной менее 400 п входили в состав генов первой группы всех хромосом редко, и доля интронов с длиной более 400 п составляла более 80%. Средняя длина интронов в генах выборок из участков всех хромосом с плотностью 1-11, 12-20, 21 и более генов на Мбр равнялась соответственно 10501, 4121 и 1944 п (по данным табл. 1 и 2), т. е., средняя длина интронов в генах сильно зависела от того, в каких участках эти гены были локализованы. Несмотря на снижение средней длины экзонов в генах с увеличением в них числа интронов, общая длина экзонов увеличивалась в несколько раз (см. табл. 1, 2). Таким образом, интроны делят белок-кодирующую часть гена на части с увеличением ее длины. Средние длины сумм экзонов в генах с 1-2 интронами были близкими в трех группах участков и равнялись соответственно 687, 649 и 678 п. Длины сумм экзонов в генах выборок с 15 и более интронами в расчете на 25 интронов мало отличались между собой и равнялись в трех группах соответственно 3952, 3465 и 3901 п.

Из-за большой разницы средних длин интронов в генах из разных групп участков длина генов в этих участках существенно отличалась для соответствующих выборок генов (см. табл. 1 и 2). В расчете на 25 интронов длина генов для выборок генов с 15 и более интронами равнялась для соответствующих групп 244474, 98300 и 45271 п. В генах с 1-2 интронами их длины в этих группах равнялись в среднем 23160, 7048 и 47201 п. Остается загадкой причина столь разительного отличия длин генов в зависимости от плотности их расположения в хромосомах.

Из таблиц 1 и 2 видно, что с повышением числа интронов в генах увеличивалась в несколько раз суммарная длина экзонов. Эта связь для генов из участков с различной плотностью генов хромосом 7-12 описывается зависимостью с высокими коэффициентами корреляции ( $r$ ):  $N_{in} = aL_{ex} + b$ , где  $a$  и  $b$  – параметры линейной регрессии. На рис. 2 изображена эта взаимосвязь суммы длин экзонов и числом интронов для генов хромосомы 11. Соответствующие величины коэффициентов корреляции и параметров линейной регрессии для генов всех изученных хромосом приведены в табл. 3.

Таблица 1. Изменения длин интронов, экзонов, суммарной длины экзонов и генов в генах 7–9 хромосом *H. sapiens*

ген/Mbp	$N_{in}$	$l_{ex}$	$l_{in}$	$L_{ex}$	$L_{gn}$	$L_{ex}/L_{gn}, \%$	$N_{gn}$
Хромосома 7							
5	1,6	238	22564	624	37059	1,70	122
	3,7	176	14310	827	54065	1,50	93
	7,6	179	11399	1540	88383	1,70	76
	11,7	143	8973	1812	106422	1,70	82
	23,7	157	10077	3869	242798	1,60	107
16	1,6	220	3268	568	5729	9,90	88
	3,8	162	4710	785	18889	4,20	128
	7,5	147	5083	1255	39510	3,20	76
	11,6	125	3137	1571	37960	4,10	70
	19,9	137	4600	2874	94617	3,04	70
32	1,5	269	2351	661	4083	16,20	79
	3,9	187	2175	909	9323	9,70	106
	7,3	175	2575	1453	20270	7,20	52
	12,2	137	2136	1807	27792	6,50	42
	22,7	137	2848	3249	68003	4,80	30
Хромосома 8							
5	1,6	272	18628	718	31185	2,30	118
	4,1	165	13356	845	55737	1,52	91
	7,3	171	14141	1409	104044	1,35	97
	11,9	136	10719	1755	129652	1,35	73
	26,3	145	10636	3978	284191	1,40	84
14	1,5	263	4825	658	7896	8,33	44
	4	186	4047	920	16906	5,44	60
	7,3	135	4792	1111	35880	3,10	51
	12	138	3511	1796	43934	4,09	20
	19,2	165	3906	3327	78242	4,30	22
36	1,5	245	2473	606	4248	14,30	55
	3,5	213	1428	957	5956	16,10	38
	7,5	143	1244	1218	10546	11,60	28
	11,8	143	2096	1824	26510	6,90	18
	20,7	137	1219	2984	28253	10,60	19
Хромосома 9							
5	1,5	297	9025	738	14151	5,21	109
	3,7	174	6529	823	25157	3,27	110
	7,3	152	10412	1256	76856	1,63	92
	12	151	6455	1959	79123	2,48	65
	24,2	160	11429	4032	281153	1,43	93
15	1,5	225	4090	567	6767	8,38	95
	4	169	3380	841	14323	5,87	87
	7,3	151	3442	1249	26328	4,74	49
	12,1	150	3171	1962	40202	4,88	49
	26,1	132	3827	3560	103524	3,44	42
32	1,5	274	1800	682	3361	20,30	41
	4	148	2549	741	10937	6,78	65
	7,3	137	2242	1135	17446	6,51	58
	11,3	145	1655	1784	20559	8,68	32
	23,8	146	1538	3615	40296	8,97	45

Таблица 2. Изменения длин интронов, экзонов, суммарной длины экзонов и генов в генах 10 – 12 хромосом *H. sapiens*

ген/Mbp	N <sub>in</sub>	l <sub>ex</sub>	l <sub>in</sub>	L <sub>ex</sub>	L <sub>gn</sub>	L <sub>ex</sub> /L <sub>gn</sub> , %	N <sub>gn</sub>
Хромосома 10							
5	1,5	284	8031	725	13160	5,51	93
	3,8	172	8692	836	34292	2,44	106
	7,4	149	8711	1245	65289	1,91	105
	11,7	143	8820	1810	104846	1,73	104
	25	144	7889	3760	201261	1,87	111
16	1,6	259	4869	669	8370	7,99	98
	4	185	5737	928	24023	3,86	118
	7,3	150	4934	1244	37369	3,33	112
	11,4	128	3619	1586	42881	3,70	61
28	25,3	113	4229	2966	109751	2,70	75
	1,4	241	1256	586	2380	24,61	14
	4	136	1835	678	8019	8,46	25
	7,2	141	2207	1148	16963	6,77	24
	12,5	154	949	2083	13946	14,94	6
	21	167	2059	3675	46910	7,83	8
Хромосома 11							
5	1,5	275	9282	691	14754	4,69	66
	4,0	193	11774	962	47878	2,01	65
	7,30	168	10489	1391	77974	1,78	83
	11,4	149	7317	1838	85074	2,16	48
	23,8	160	10714	3966	258840	1,53	57
15	1,5	284	3898	709	6528	10,86	71
	3,8	175	3919	842	15786	5,34	107
	7,4	147	5284	1234	40323	3,06	88
	11,7	136	5176	1725	62268	2,77	56
	23,4	141	3279	3433	80039	4,29	101
32	1,6	281	3500	717	6142	11,67	120
	3,9	146	2059	725	8863	8,18	148
	7,4	158	1610	1326	13225	10,03	151
	11,9	136	1583	1742	20510	8,49	109
	21,6	158	1612	3570	38433	9,29	72
Хромосома 12							
5	1,5	247	13489	622	21093	2,95	85
	3,8	189	7196	913	28472	3,21	94
	7,2	152	10206	1246	74813	1,67	96
	11,9	141	5385	1814	65927	2,75	85
	27,1	150	8392	4219	231327	1,82	128
15	1,4	340	4530	827	7327	11,29	46
	3,8	175	4869	849	19574	4,34	52
	7,3	146	3461	1210	26450	4,57	41
	11,9	140	3026	1814	37950	4,78	51
	23,8	134	3019	3324	75291	4,42	44
28	1,5	277	2618	696	4664	14,92	97
	3,8	151	1984	726	8311	8,74	176
	7,5	157	1528	1338	12806	10,45	107
	11,9	136	1832	1742	23471	7,42	64
	24,6	150	1349	3847	37062	10,38	70

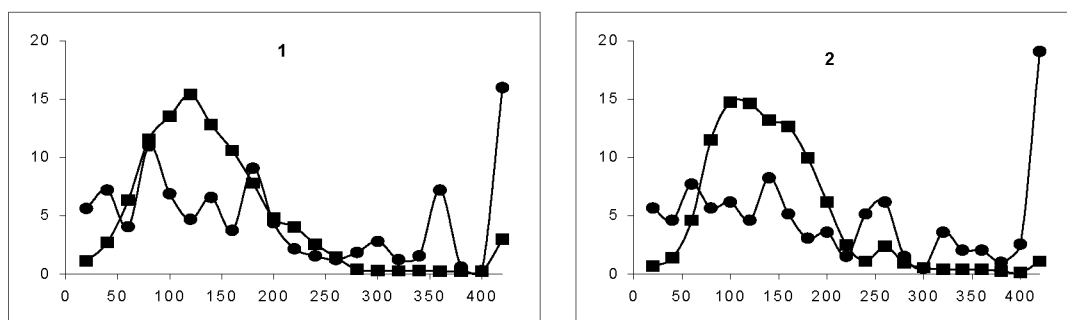


Рис. 1. Длина экзонов в генах с 1-2 (●-), 15 и более (■-) интронами хромосомы 7 из участков с плотностью 1-11 (1) и 21 и более (2) генов/Мбр. Ось абсцисс – длина экзонов, н; ось ординат – доля экзонов соответствующей длины, %

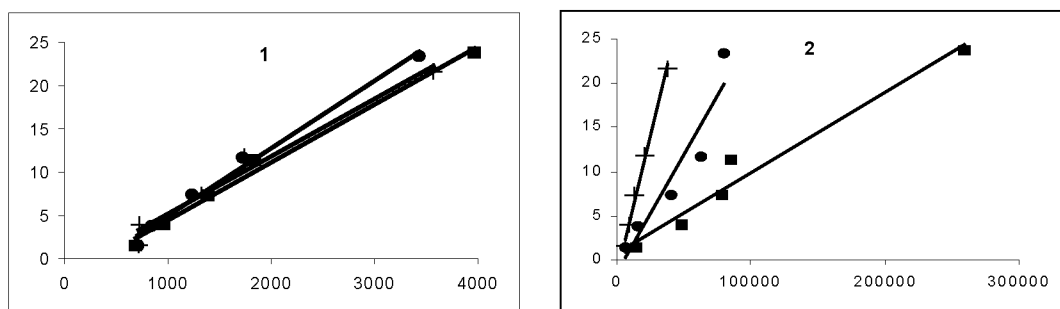


Рис. 2. Связь суммы длин экзонов (1) и генов (2) с числом интронов в генах хромосомы 11. Участки с плотностью 5 (■), 15 (●-) и 32 (+-) гена/Мбр. Ось абсцисс – сумма длин экзонов (1) и генов (2), н; ось ординат – число интронов в генах

Таблица 3. Параметры сопряженного изменения длины генов, суммы длин экзонов от числа интронов в генах 7-12 хромосом *H. sapiens*

гены/Мбр	Параметры линейной регрессии						N <sub>gn</sub>
	a	b	r	c	d	r	
Хромосома 7							
5	0,0067	-2,03	0,995	0,00011	-1,61	0,992	480
15	0,0079	-2,30	0,993	0,00021	0,73	0,971	432
32	0,0082	-3,81	0,996	0,00033	0,97	0,990	309
Хромосома 8							
5	0,0074	-2,58	0,994	0,00010	-1,64	0,997	463
14	0,0064	-1,20	0,978	0,00025	-0,49	0,991	197
36	0,0082	-3,38	0,996	0,00062	-0,33	0,926	158
Хромосома 9							
5	0,0066	-1,91	0,994	0,00008	1,66	0,974	469
15	0,0081	-3,09	0,999	0,00025	0,61	0,996	322
32	0,0072	-1,86	0,994	0,00063	-2,04	0,990	241
Хромосома 10							
5	0,0104	-5,45	0,996	0,00013	-0,59	0,998	519
16	0,0075	-2,61	1,000	0,00024	-0,70	0,992	464
28	0,0060	-0,58	0,992	0,00042	1,80	0,936	77
Хромосома 11							
5	0,0067	-2,25	0,995	0,00009	0,79	0,983	319
15	0,0078	-2,81	0,994	0,00027	-1,34	0,948	423
32	0,0066	-1,47	0,986	0,00061	-1,36	0,996	600
Хромосома 12							
5	0,0070	-2,09	0,996	0,00012	0,50	0,975	488
15	0,0084	-3,91	0,993	0,00034	-1,68	0,997	234
28	0,0070	-1,83	0,992	0,00069	-2,02	0,989	514

От числа интронов зависела общая длина гена. Многократные изменения длины генов хорошо коррелировали с числом интронов в них и описывались уравнением  $N_{in} = cL_{gn} + d$ , где  $c$  и  $d$  – коэффициенты линейной регрессии (табл. 3). В качестве примера на рис.2 приведены соответствующие зависимости для хромосомы 11.

Полученные данные свидетельствуют о том, что длины экзонов, интронов и генов изменяются связано и эта связь зависит от плотности генов в участках хромосом 7–12 человека.

#### ЛИТЕРАТУРА

1. Venter J.C., Adams M.D., Myers E.W. et al. The sequence of the human genome // Science. 2001. V. 291. P. 1304-1351.
2. Миронов А.А., Гельфанд М.С. Предсказание и компьютерный анализ экзон-интронной структуры генов человека // Молекулярная биология. 2004. Т.38, №1. С.82-91.
3. Smith A.D., Sumazin P., Xuan Z., Zhang M.Q. DNA motifs in human and mouse proximal promoters predict tissue-specific expression // Proc.Nat.Acad.Sci.USA. 2006. V.103. P.6275-6280.
4. Rigoutsos I., Huynh T., Miranda K., Tsirigos A., McHardy A., Platt D. Short blocks from the noncoding parts of the human genome have instances within nearly all known genes and relate to biological processes // Proc. Nat. Acad. Sci.USA. 2006. V.103. P.6605-6610.
5. Louie E., Ott J., Majewski J. Nucleotide frequency variation across human genes // Genome Research. 2003. V. 13. P. 2594-2601.
6. Thanaraj T.A., Stamm S., Clark F., Riethoven J.-J., Texier V., Muilu J. ASD: the alternative splicing database // Nucl. Acids Research. 2004. V.32. P.D64-69.
7. Minovitsky S., Gee S.L., Schokrpur S., Dubchak I., Conboy J.G. The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons // Nucl. Acids Research. 2005. V.33. P.714-724.
8. Brudno M., Gelfand M.S., Spengler S., Zorn M., Dubchak I., Conboy J.G. Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing // Nucl. Acids Research. 2001. V. 29. P.2338-2348.
9. Atambaeva S.A., Ivashchenko A.T. The length of exons and introns in genes of *Arabidopsis thaliana* and *Caenorhabditis elegans* // Proceedings of the fifth international conference on bioinformatics of genome regulation and structure, Novosibirsk, 2006. P.35-38.
10. Иващенко А., Атамбаева Ш., Хайленко В. Изменение длины интронов и экзонов в генах 1, 7, 8 и 10 хромосом риса // Биотехнология. Теория и практика. 2006. №1. С.33-38.
11. Атамбаева Ш.А., Хайленко В.А., Иващенко А.Т. Изменения длины экзонов и интронов в генах пятой и шестой хромосом человека // Вестн. КазНУ. Сер. биол. 2006. № 4 (30). С.27 -31.
12. Atambaeva S.A., Ivashchenko A.T., Khailenko V.A., Boldina G.F., Turmagambetova A.S. Length of exons and introns in genes of some human chromosomes // Proceedings of the fifth international conference on bioinformatics of genome regulation and structure. Novosibirsk, 2006. P.31-34.
13. Блинов В.М., Денисов С.И., Сараев Д.В., Швецов Д.В., Уваров Д.Л., Опарина Н.Ю., Сандахчиев Л.С., Киселев Л.Л. Структурная организация генома человека: распределение нуклеотидов, Alu-повторов и экзонов в хромосомах 21 и 22 // Молекулярная биология. 2001. Т.35, №6. С.1032-1038.
14. Хитринская И.Ю., Степанов В.А., Пузырев В.П. Alu-повторы в геноме человека // Молекулярная биология. 2003. Т.37, №3. С.382-391.
15. Chen C., Gentles A.J., Jurka J., Karlin S. Genes, pseudogenes, and Alu sequence organization across human chromosomes 21 and 22 // Proc. Nat. Acad. Sci.USA. 2002. V.99. P.2930-2935.
16. Semon M., Mouchiroud D., Duret L. Relationship between gene expression and GC-content in mammals: statistical significance and biological relevance // Human Molecular Genetics. 2005. V.14. P.421-427.

#### Резюме

7-12 адам хромосомаларындағы гендерде экзондардың, интрондардың және гендердің ұзындықтарында өзгерістер ДНҚ-ның учаскілеріндегі гендердің орналасуының тығыздығына және гендердегі интрондардың санына тәуелді өзгерістер жүреді. Гендердің осы құрылымдық элементтердің ұзындықтарының өзгерістерінің арасында сызықтық корреляция табылды.

#### Summary

There are changes of exons, introns and genes lengths depending on density of disposition of genes in sites of DNA and number of introns in genes of 7-12 human chromosomes. The linear correlation between changes of lengths of these structure elements of genes is revealed.