

ПРИЛОЖЕНИЕ А

Таблица А1. Результаты иерархического кластерного анализа экзонов первых пар 1-5 модельных генов

			1			2			3			4			5		
№	Расстояние	Метод	κ	χ^2	Q	κ	χ^2	Q	κ	χ^2	Q	κ	χ^2	Q	κ	χ^2	Q
1	euclidean	ward	0.611	93.985	0.837	0.380	104.661	0.668	0.488	37.146	1.590	0.410	60.530	1.031	0.530	42.444	1.443
2	euclidean	single	0.767	2251.606	0.406	0.679	2000.279	0.364	0.581	1984.610	0.316	0.743	1777.242	0.400	0.692	1078.444	0.392
3	euclidean	complete	0.739	222.088	0.595	0.495	263.147	0.438	0.633	98.553	0.824	0.561	177.589	0.562	0.692	205.667	0.589
4	euclidean	average	0.843	716.372	0.491	0.776	872.749	0.445	0.770	497.854	0.485	0.841	1269.233	0.460	0.858	741.111	0.496
5	euclidean	mcquitty	0.759	681.073	0.453	0.672	276.406	0.517	0.651	229.081	0.544	0.644	363.388	0.460	0.755	270.222	0.563
6	euclidean	median	0.683	2670.168	0.360	0.594	2374.016	0.318	0.573	2223.789	0.309	0.648	1906.767	0.350	0.656	2083.444	0.352
7	euclidean	centroid	0.748	2646.825	0.393	0.692	2351.020	0.367	0.560	2291.854	0.302	0.800	1863.909	0.427	0.755	2172.778	0.401
8	maximum	ward	0.722	98.350	0.869	0.446	63.641	1.009	0.539	38.943	1.553	0.476	33.105	1.748	0.609	37.667	1.632
9	maximum	single	0.580	524.409	0.385	0.414	2084.805	0.231	0.452	1457.528	0.260	0.556	1590.612	0.309	0.604	1376.222	0.338
10	maximum	complete	0.696	85.920	0.930	0.482	42.096	1.429	0.507	96.016	0.774	0.557	49.845	1.282	0.518	47.222	1.318
11	maximum	average	0.777	265.263	0.577	0.639	294.327	0.489	0.593	125.398	0.695	0.699	142.447	0.701	0.745	394.667	0.499
12	maximum	mcquitty	0.739	153.292	0.696	0.526	166.606	0.563	0.657	193.358	0.587	0.474	179.251	0.516	0.725	246.333	0.565
13	maximum	median	0.600	814.489	0.361	0.511	1517.155	0.288	0.468	864.390	0.292	0.522	845.279	0.320	0.447	1391.000	0.259
14	maximum	centroid	0.697	544.431	0.440	0.634	1667.873	0.347	0.611	1606.553	0.337	0.633	1319.571	0.354	0.705	1154.222	0.396
15	manhattan	ward	0.350	80.796	0.794	0.254	72.446	0.817	0.313	65.894	0.915	0.330	32.511	1.703	0.329	26.333	2.063
16	manhattan	single	0.846	2646.920	0.442	0.778	2351.020	0.410	0.720	2223.789	0.382	0.838	1995.571	0.444	0.771	2104.778	0.409
17	manhattan	complete	0.599	637.139	0.378	0.504	364.351	0.389	0.566	253.073	0.481	0.659	524.018	0.425	0.666	597.333	0.417
18	manhattan	average	0.879	2486.270	0.460	0.835	2171.402	0.441	0.813	2026.886	0.431	0.885	1862.247	0.469	0.867	1929.333	0.459
19	manhattan	mcquitty	0.790	902.642	0.450	0.728	1681.649	0.394	0.610	608.407	0.387	0.759	887.662	0.436	0.803	1032.778	0.450
20	manhattan	median	0.833	2670.168	0.435	0.773	2420.319	0.407	0.699	2337.829	0.371	0.822	1995.571	0.436	0.770	2195.667	0.408
21	manhattan	centroid	0.871	2693.606	0.454	0.825	2420.319	0.433	0.768	2360.976	0.405	0.866	2041.041	0.457	0.833	2218.667	0.439
22	minkowski	ward	0.710	92.467	0.896	0.485	48.622	1.271	0.538	35.878	1.663	0.480	46.164	1.323	0.601	35.222	1.720
23	minkowski	single	0.662	978.650	0.382	0.485	2020.685	0.267	0.350	1684.130	0.205	0.618	1630.977	0.340	0.671	1151.556	0.379
24	minkowski	complete	0.732	159.839	0.679	0.474	125.689	0.635	0.627	114.089	0.752	0.596	132.950	0.674	0.584	137.444	0.656

25	<i>minkowski</i>	<i>average</i>	0.791	314.036	0.555	0.715	435.721	0.472	0.709	157.634	0.672	0.759	250.247	0.579	0.803	397.000	0.527
26	<i>minkowski</i>	<i>mcquitty</i>	0.756	198.839	0.629	0.573	306.239	0.450	0.644	188.496	0.587	0.633	255.708	0.512	0.723	296.333	0.530
27	<i>minkowski</i>	<i>median</i>	0.613	1098.876	0.352	0.553	2087.291	0.300	0.308	1688.992	0.184	0.627	1798.374	0.341	0.586	1887.778	0.319
28	<i>minkowski</i>	<i>centroid</i>	0.699	634.482	0.428	0.690	1808.127	0.373	0.566	1645.764	0.313	0.702	1755.753	0.379	0.664	1381.444	0.368

Таблица А2. Результаты иерархического кластерного анализа экзонов первых пар 6-10 модельных генов

			6			7			8			9			10		
№	Расстояние	Метод	κ	χ^2	Q	κ	χ^2	Q	κ	χ^2	Q	κ	χ^2	Q	κ	χ^2	Q
1	<i>euclidean</i>	<i>ward</i>	0.368	100.982	0.679	0.421	106.007	0.682	0.561	48.652	1.308	0.580	100.901	0.786	0.427	25.813	2.150
2	<i>euclidean</i>	<i>single</i>	0.734	1895.809	0.393	0.736	4850.067	0.378	0.581	1879.330	0.317	0.782	2388.071	0.412	0.715	2376.335	0.379
3	<i>euclidean</i>	<i>complete</i>	0.538	466.282	0.376	0.572	623.692	0.366	0.681	128.161	0.731	0.705	166.454	0.653	0.632	229.969	0.533
4	<i>euclidean</i>	<i>average</i>	0.843	1315.891	0.459	0.819	3530.917	0.424	0.796	613.107	0.480	0.838	882.929	0.476	0.811	1099.907	0.451
5	<i>euclidean</i>	<i>mcquitty</i>	0.712	874.245	0.413	0.717	1531.811	0.391	0.750	477.536	0.480	0.754	492.652	0.478	0.651	600.848	0.409
6	<i>euclidean</i>	<i>median</i>	0.707	1962.345	0.379	0.668	5281.117	0.343	0.454	2009.563	0.252	0.696	2695.369	0.367	0.614	2399.097	0.328
7	<i>euclidean</i>	<i>centroid</i>	0.749	1984.800	0.400	0.773	5375.469	0.396	0.615	1856.232	0.334	0.728	2718.603	0.382	0.680	2399.198	0.361
8	<i>maximum</i>	<i>ward</i>	0.522	52.882	1.207	0.457	102.920	0.714	0.590	42.732	1.465	0.655	114.730	0.763	0.539	29.253	1.979
9	<i>maximum</i>	<i>single</i>	0.519	1585.227	0.291	0.520	4433.584	0.271	0.524	1238.848	0.302	0.505	1773.291	0.281	0.446	1696.895	0.252
10	<i>maximum</i>	<i>complete</i>	0.588	74.036	0.969	0.458	131.477	0.609	0.652	76.973	0.976	0.679	135.752	0.708	0.595	72.957	0.983
11	<i>maximum</i>	<i>average</i>	0.711	411.682	0.477	0.650	517.907	0.422	0.685	248.991	0.543	0.732	237.355	0.577	0.642	410.451	0.443
12	<i>maximum</i>	<i>mcquitty</i>	0.645	196.355	0.577	0.517	417.959	0.378	0.675	290.196	0.510	0.699	299.404	0.516	0.530	201.946	0.513
13	<i>maximum</i>	<i>median</i>	0.577	1194.518	0.330	0.466	4148.887	0.245	0.350	1362.696	0.212	0.352	1717.787	0.205	0.399	1695.984	0.229
14	<i>maximum</i>	<i>centroid</i>	0.643	1373.327	0.358	0.595	4140.108	0.310	0.689	756.339	0.411	0.627	517.085	0.410	0.608	1177.401	0.346
15	<i>manhattan</i>	<i>ward</i>	0.216	90.818	0.659	0.316	120.382	0.573	0.319	36.464	1.531	0.347	93.709	0.707	0.264	26.218	2.039
16	<i>manhattan</i>	<i>single</i>	0.794	1940.009	0.423	0.803	5518.059	0.411	0.754	2077.348	0.401	0.873	2718.511	0.455	0.791	2422.062	0.416
17	<i>manhattan</i>	<i>complete</i>	0.488	584.936	0.329	0.623	1975.354	0.337	0.529	181.205	0.540	0.663	380.631	0.463	0.684	483.191	0.445
18	<i>manhattan</i>	<i>average</i>	0.843	1788.145	0.449	0.862	5257.963	0.441	0.820	1791.348	0.438	0.897	2580.766	0.468	0.839	2219.222	0.442
19	<i>manhattan</i>	<i>mcquitty</i>	0.759	1104.936	0.425	0.611	2294.879	0.327	0.549	622.277	0.355	0.795	1930.213	0.423	0.749	1124.591	0.419
20	<i>manhattan</i>	<i>median</i>	0.733	2030.064	0.391	0.788	5590.271	0.403	0.738	2100.214	0.393	0.833	2788.766	0.434	0.749	2491.564	0.395
21	<i>manhattan</i>	<i>centroid</i>	0.813	2052.873	0.431	0.840	5614.438	0.429	0.803	2100.214	0.425	0.870	2788.766	0.453	0.818	2491.564	0.429

22	<i>minkowski</i>	<i>ward</i>	0.426	74.627	0.883	0.419	81.937	0.820	0.613	54.455	1.225	0.678	105.787	0.812	0.484	22.171	2.497
23	<i>minkowski</i>	<i>single</i>	0.610	1746.900	0.334	0.610	4652.486	0.316	0.427	1455.438	0.248	0.625	2195.837	0.335	0.597	1489.603	0.332
24	<i>minkowski</i>	<i>complete</i>	0.649	159.718	0.638	0.508	253.566	0.451	0.670	93.339	0.871	0.652	111.319	0.775	0.557	71.136	0.981
25	<i>minkowski</i>	<i>average</i>	0.762	908.755	0.436	0.726	714.282	0.433	0.757	371.795	0.513	0.762	337.851	0.529	0.703	482.179	0.455
26	<i>minkowski</i>	<i>mcquitty</i>	0.657	1020.082	0.378	0.679	564.842	0.428	0.762	320.723	0.537	0.716	280.135	0.536	0.477	380.304	0.370
27	<i>minkowski</i>	<i>median</i>	0.622	1810.955	0.339	0.559	4717.365	0.290	0.458	1571.857	0.261	0.549	2196.021	0.297	0.485	2071.315	0.267
28	<i>minkowski</i>	<i>centroid</i>	0.699	1583.809	0.381	0.666	4629.766	0.344	0.649	1476.679	0.358	0.588	674.468	0.368	0.630	1988.763	0.340

Таблица А3. Точность A иерархического кластерного анализа экспериментально подтвержденных транскриптов 10 пар модельных генов для различных способов вычисления расстояний между транскриптами РНК и методов объединения кластеров

			1	2	3	4	5	6	7	8	9	10
	Расстояние	Метод	A	A	A	A	A	A	A	A	A	A
1	<i>jw</i>	<i>ward</i>	100.0	99.0	100.0	97.6	83.7	98.0	95.5	99.0	99.1	82.8
2	<i>jw</i>	<i>single</i>	100.0	52.6	100.0	59.7	50.0	56.6	66.4	51.4	32.7	100.0
3	<i>jw</i>	<i>complete</i>	69.5	61.9	65.4	61.3	54.1	64.6	68.7	61.0	39.1	57.6
4	<i>osa</i>	<i>ward</i>	68.4	53.6	67.3	77.4	75.5	96.0	65.7	65.7	27.3	59.6
5	<i>osa</i>	<i>single</i>	68.4	57.7	49.0	59.7	50.0	56.6	65.7	51.4	27.3	59.6
6	<i>osa</i>	<i>complete</i>	68.4	50.5	67.3	76.6	68.4	56.6	65.7	72.4	27.3	56.6
7	<i>lv</i>	<i>ward</i>	68.4	53.6	67.3	77.4	75.5	96.0	65.7	65.7	27.3	59.6
8	<i>lv</i>	<i>single</i>	68.4	57.7	49.0	59.7	50.0	56.6	65.7	51.4	27.3	59.6
9	<i>lv</i>	<i>complete</i>	68.4	50.5	67.3	79.0	68.4	56.6	65.7	72.4	27.3	56.6
10	<i>dl</i>	<i>ward</i>	68.4	53.6	67.3	77.4	75.5	96.0	65.7	65.7	27.3	59.6
11	<i>dl</i>	<i>single</i>	68.4	57.7	67.3	59.7	50.0	56.6	65.7	51.4	27.3	59.6
12	<i>dl</i>	<i>complete</i>	68.4	50.5	67.3	76.6	68.4	56.6	65.7	72.4	27.3	56.6
13	<i>lcs</i>	<i>ward</i>	100.0	56.7	100.0	99.2	83.7	100.0	93.3	100.0	100.0	59.6
14	<i>lcs</i>	<i>single</i>	68.4	57.7	71.2	59.7	50.0	56.6	65.7	52.4	27.3	57.6
15	<i>lcs</i>	<i>complete</i>	69.5	71.1	100.0	99.2	72.4	56.6	65.7	95.2	27.3	57.6
16	<i>qgram</i>	<i>ward</i>	100.0	56.7	100.0	99.2	75.5	98.0	94.0	100.0	100.0	56.6
17	<i>qgram</i>	<i>single</i>	68.4	57.7	70.2	59.7	50.0	56.6	65.7	81.0	27.3	59.6
18	<i>qgram</i>	<i>complete</i>	69.5	71.1	100.0	86.3	75.5	56.6	36.6	95.2	27.3	56.6
19	<i>cosine</i>	<i>ward</i>	100.0	97.9	100.0	99.2	100.0	99.0	95.5	100.0	99.1	100.0
20	<i>cosine</i>	<i>single</i>	100.0	50.5	100.0	58.1	100.0	56.6	66.4	100.0	99.1	52.5
21	<i>cosine</i>	<i>complete</i>	69.5	71.1	51.0	64.5	75.5	70.7	78.4	78.1	46.4	66.7
22	<i>jaccard</i>	<i>ward</i>	100.0	99.0	100.0	97.6	100.0	100.0	98.5	100.0	100.0	100.0
23	<i>jaccard</i>	<i>single</i>	100.0	99.0	100.0	58.1	100.0	56.6	66.4	100.0	100.0	50.5
24	<i>jaccard</i>	<i>complete</i>	61.1	72.2	58.7	64.5	75.5	81.8	83.6	54.3	37.3	66.7
25	<i>soundex</i>	<i>ward</i>	53.7	46.4	40.4	63.7	58.2	49.5	76.9	57.1	44.5	63.6
26	<i>soundex</i>	<i>single</i>	61.1	52.6	51.0	59.7	50.0	58.6	68.7	51.4	35.5	58.6
27	<i>soundex</i>	<i>complete</i>	61.1	52.6	51.0	59.7	50.0	58.6	68.7	51.4	35.5	58.6

ПРИЛОЖЕНИЕ Б

Пример анализа данных для гибридного онкогена *RUNX1/RUNX1T1* с использованием разработанного пакета *RNAexploreR*. Для ускорения вычислений и упрощения визуального представления результатов анализа список теоретически предсказанных транскриптов, используемый для демонстрации окон интерфейса пакета, сокращен до 3000 первых транскриптов в списке предсказываемых изоформ РНК.

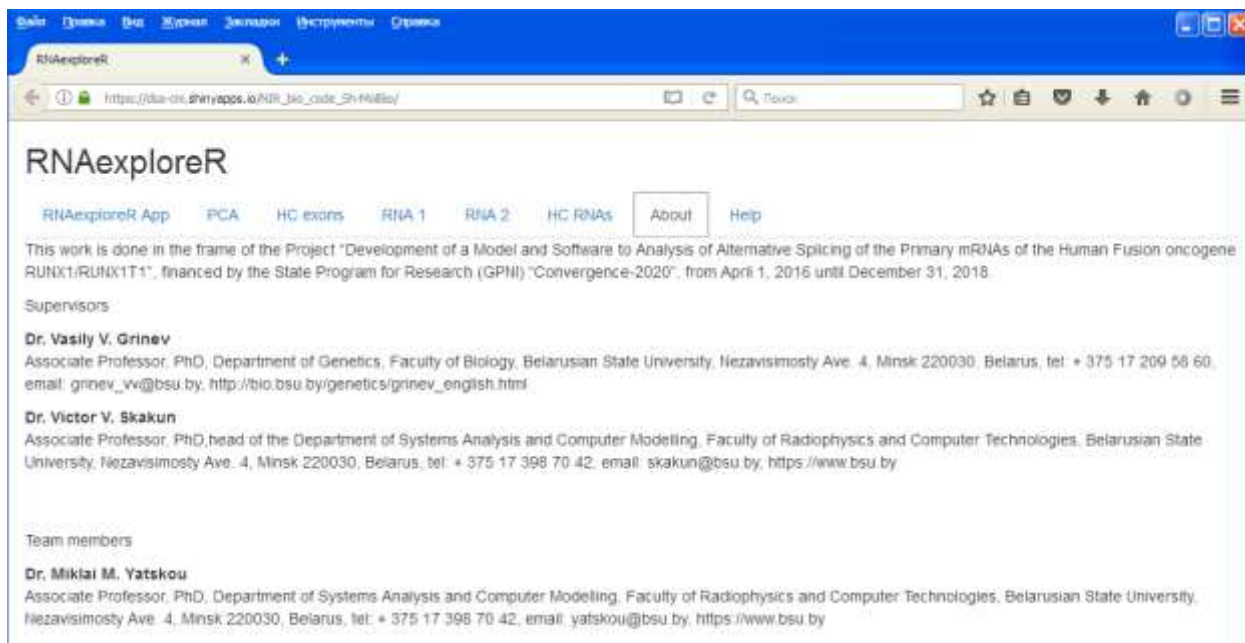


Рисунок Б1 – Вкладка информации об авторах-разработчиках пакета *RNAexploreR*

File

Print

Find

Home

Back

Forward

Stop

RNAexploreR

https://doi.org/10.1093/bioinformatics/btq100

RNAexploreR

RNAexploreR AppPCAHC exonsRNA 1RNA 2HC RNAsAboutHelp

The main window of the package interface consists of eight panels, some of which are corresponding to five stages of analysis. At each stage of the analysis, the user must download the required data file (e.g. - exon features (Figure 1), the transcripts of RNA 1 or the experimentally confirmed transcripts of a gene (Figure 2), the transcripts of RNA 2 or the theoretically predicted transcripts of a gene) and set the system parameters of the data mining algorithms.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1 Exon		Length	MFE	phyloP	phastCons	Sss	3ss	Overall	si	highECI	h	highECI	h	highECI	h
2 chr21:36215920-36216197		278	-0.369	-0.15962	0.102164	-9.44	-9.53	-18.97	0.007491	0.007435	0.01476	0.00738	0.022305	0.011111	
3 chr21:36216364-36216571		208	-0.252	-0.4319	0.008846	-16.3667	-26.49	-42.8567	0.035533	0.020101	0.039801	0.00995	0.01005	0.015	
4 chr21:36217970-36218179		210	-0.31	1.82878	0.666517	-25.0633	-12.84	-37.9033	0.025126	0.024876	0.014778	0.019704	0.034826	0.059406	
5 chr21:36220379-36220725		347	-0.315	-0.2686	0.008823	-11.92	-24.6167	-36.5367	0.029762	0.017751	0.020588	0.014706	0.026627	0.011799	
6 chr21:36221420-36221496		77	-0.196	0.554879	0.303102	-22.4333	-32.6133	-55.0467	0	0.029412	0.014286	0.014286	0	0	
7 chr21:36222592-36222848		257	-0.312	-0.53088	0.002083	-16.2067	-13.7233	-29.93	0.028455	0.012097	0.028	0.02	0.012097	0.012048	
8 chr21:36224756-36224966		211	-0.218	-0.02151	0.050863	-15.1467	-40.4967	-55.6433	0.02	0.014851	0.034314	0.014706	0.009901	0.014778	
9 chr21:36224861-36225523		663	-0.328	-0.28849	0.00959	-25.6533	-11.7967	-37.45	0.023006	0.01682	0.030488	0.022866	0.027523	0.018321	
10 chr21:36225796-36225959		164	-0.174	-0.09081	0.004846	-9.18	-18.76	-27.94	0	0.006452	0.012739	0.012739	0.019355	0.012821	
11 chr21:36226262-36226389		128	-0.298	-0.15771	0.004951	-21.41	-13.5033	-34.9133	0.017094	0.02521	0.016529	0.016529	0.02521	0.025	

Рисунок Б2 – Вкладка «помощь» пакета *RNAexploreR*

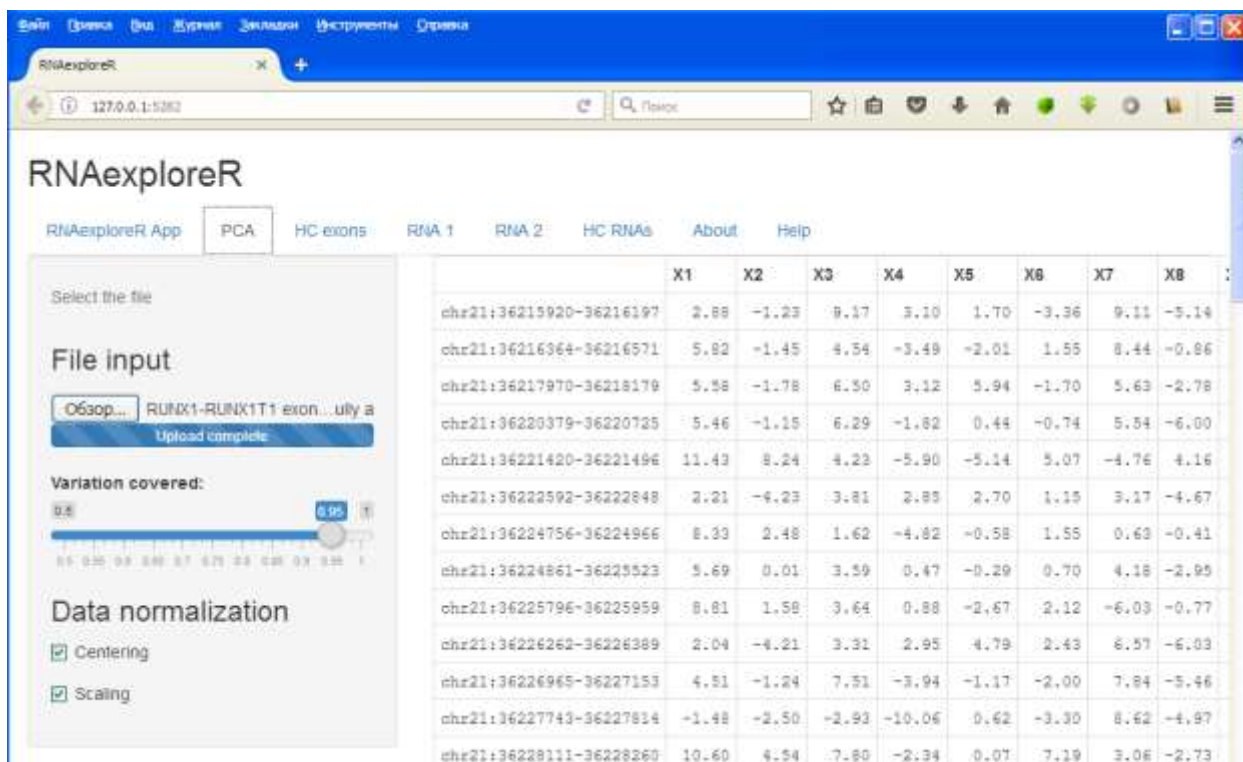


Рисунок Б3 – Представление таблицы экзонов гена в терминах первых главных компонент (относительная доля разброса приходящаяся на первые компоненты – 0.95), полученные в ходе анализа 99 экзонов и полного набора из 1438 нормированных атрибутов на этапе 1

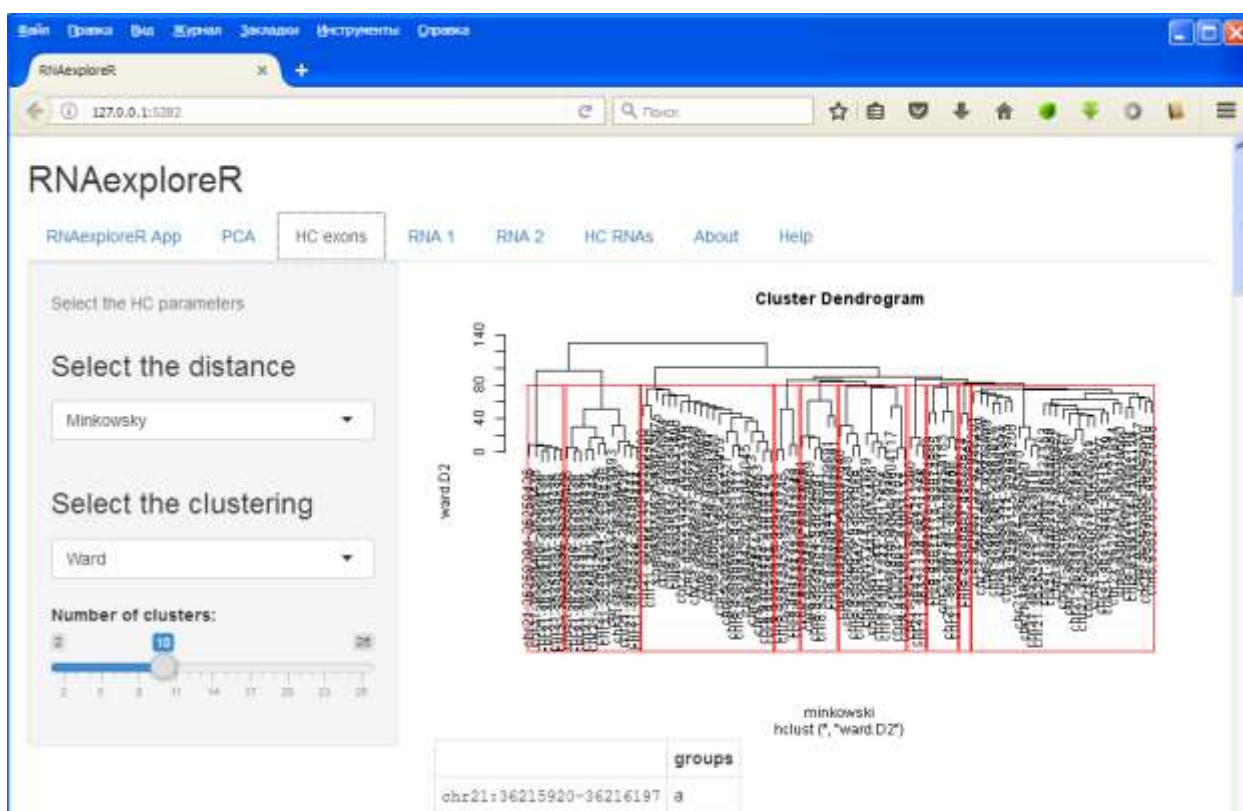


Рисунок Б4 – Результаты работы алгоритма иерархической кластеризации 99 экзонов на основе новых признаков на этапе 2

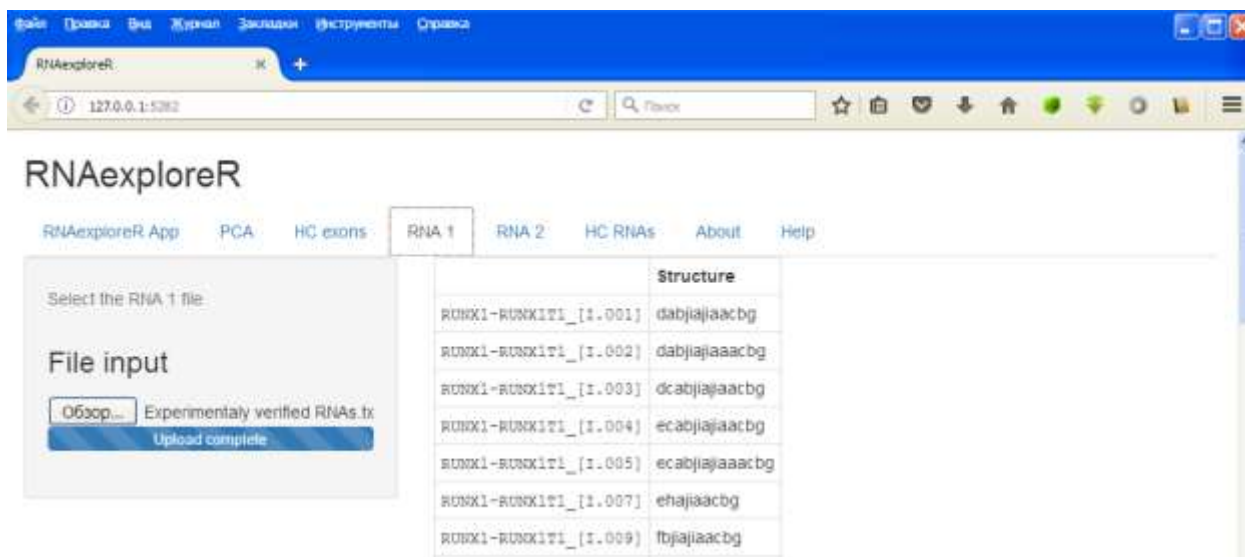


Рисунок Б5 – Преобразование символов последовательностей экспериментально подтвержденных транскриптов РНК к меткам кластеров на этапе 3

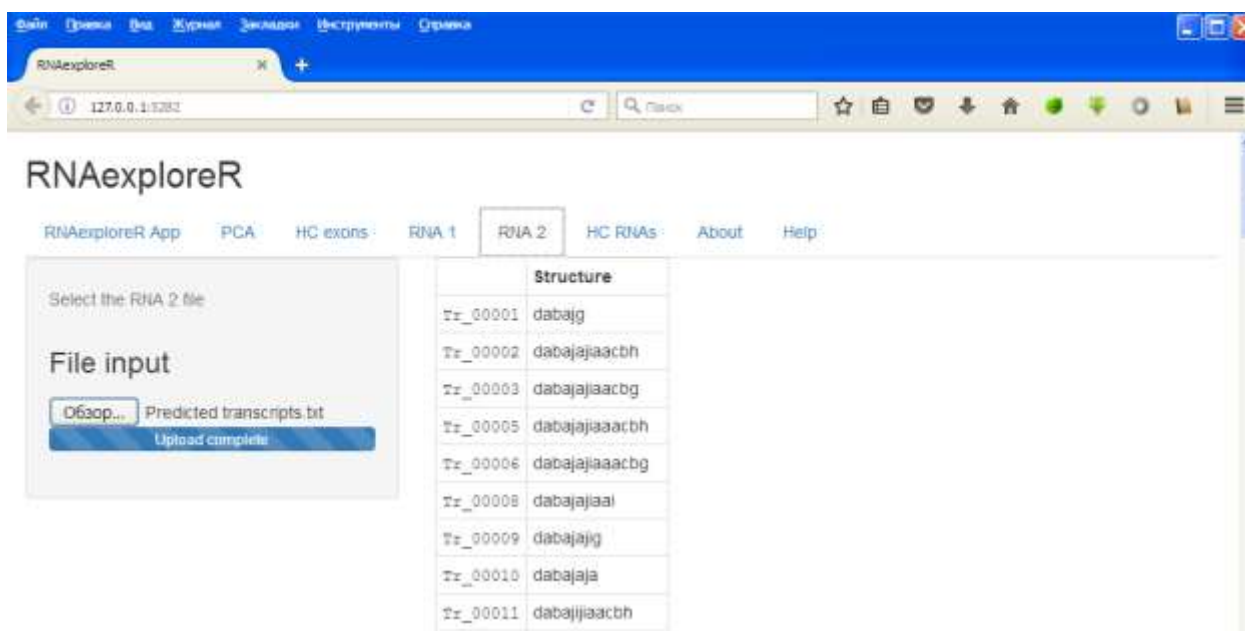


Рисунок Б6 – Преобразование символов последовательностей теоретических транскриптов РНК к меткам кластеров на этапе 4

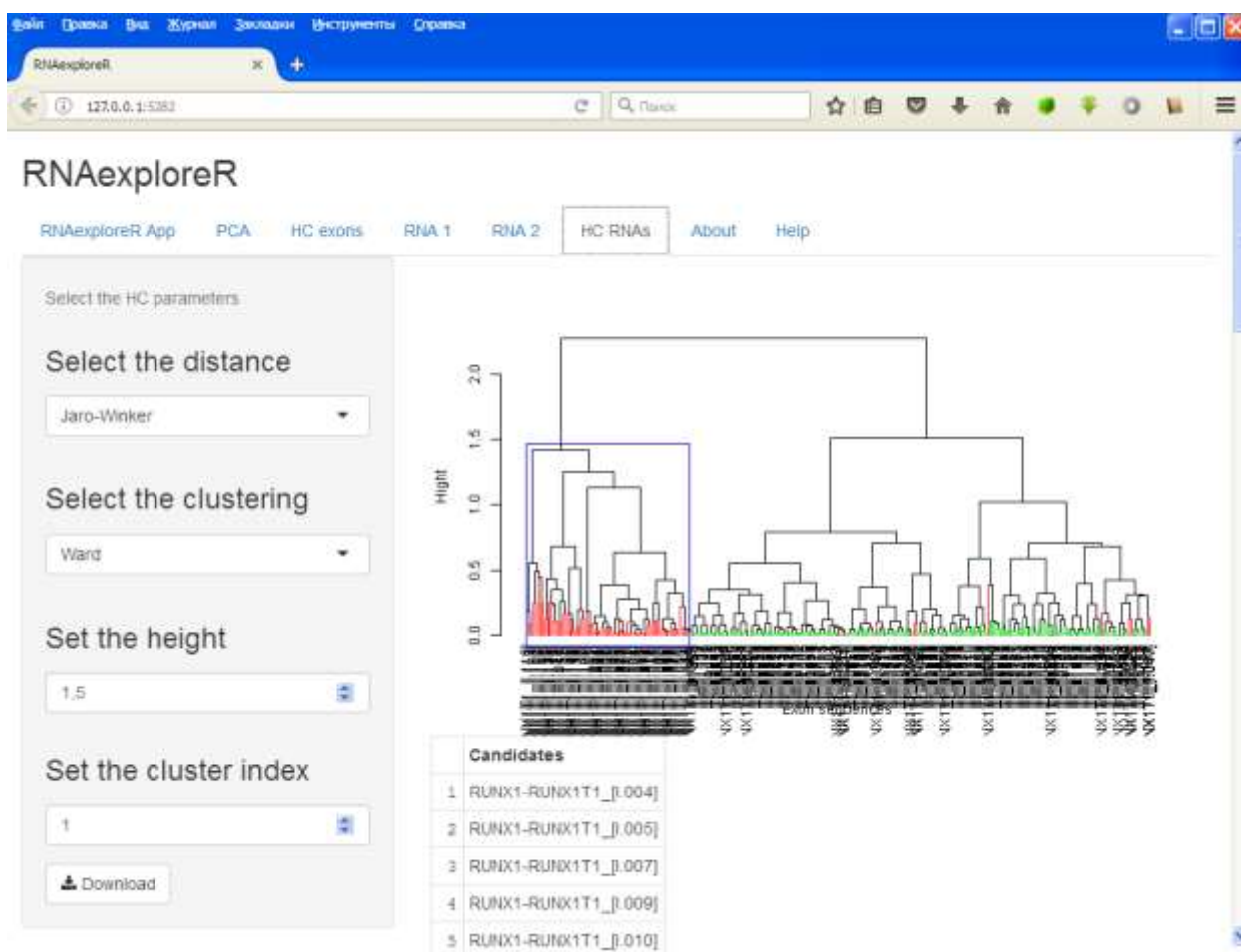


Рисунок Б7 – Результаты работы алгоритма иерархической кластеризации пула уникальных экспериментальных (красный цвет) и теоретических (зеленый) транскриптов на этапе 5