

14-я Международная школа молодых ученых «Системная Биология и Биоинформатика»

Анализ регионов высокой гомозиготности в популяциях коренного населения Северной Евразии

Колесников Н.А., Зарубин А.А., Харьков В.Н., Степанов В.А.

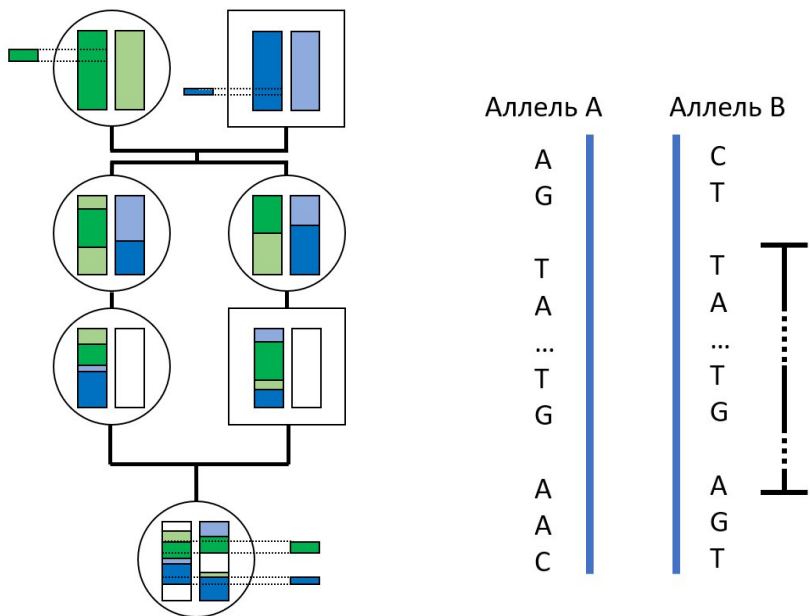
*Федеральное государственное бюджетное научное учреждение
«Томский национальный исследовательский медицинский центр Российской академии наук»
Научно-исследовательский институт медицинской генетики*

Новосибирск 2023

ROH - runs of homozygosity – регионы высокой гомозиготности, непрерывные участки гомозиготных генотипов.

Возникают, если гаплотипы, переданные от матери и отца, идентичны, будучи, в свою очередь, унаследованы от общего предка в какой-то момент в прошлом.

Частоты ROH являются важным ресурсом для понимания структуры генетического разнообразия в популяциях человека, а также ключом для выявления локусов, связанных с заболеваниями, в контексте картирования гомозиготности. Горячие точки ROH представляют собой области, которые могут нести мишени положительного отбора.



Материалы и методы

Микрочип Infinium Multi-Ethnic Global
2100 образцов на 1.7 млн SNP

PLINK v1.9

Исключены:

- образцы с более 5% пропущенных позиций
- SNP с более чем 10% отсутствующими генотипами
 - метисы первого-второго поколения
 - делеций, инсерций, мономорфные SNP
 - MAF > 0.01

1836 образцов 886 889 SNP

Дагестан
23 популяции

Кавказ
11 популяций

Сибирь
20 популяций

Волго-Уральский
географический регион и
популяции Европы
17 популяций

Средняя Азия
5 популяций

Идентификация ROH

Garlic

-- error 0.001 -- winsize 40 -- auto-winsize -- auto-winsize-step 2 -- kde-subsample 0 -- ld-subsample 0 -- resample 40

1 432 099 ROH

A-«короткие» – 1 057 149
B-«средние» – 342 094
C-«длинные» – 32 855

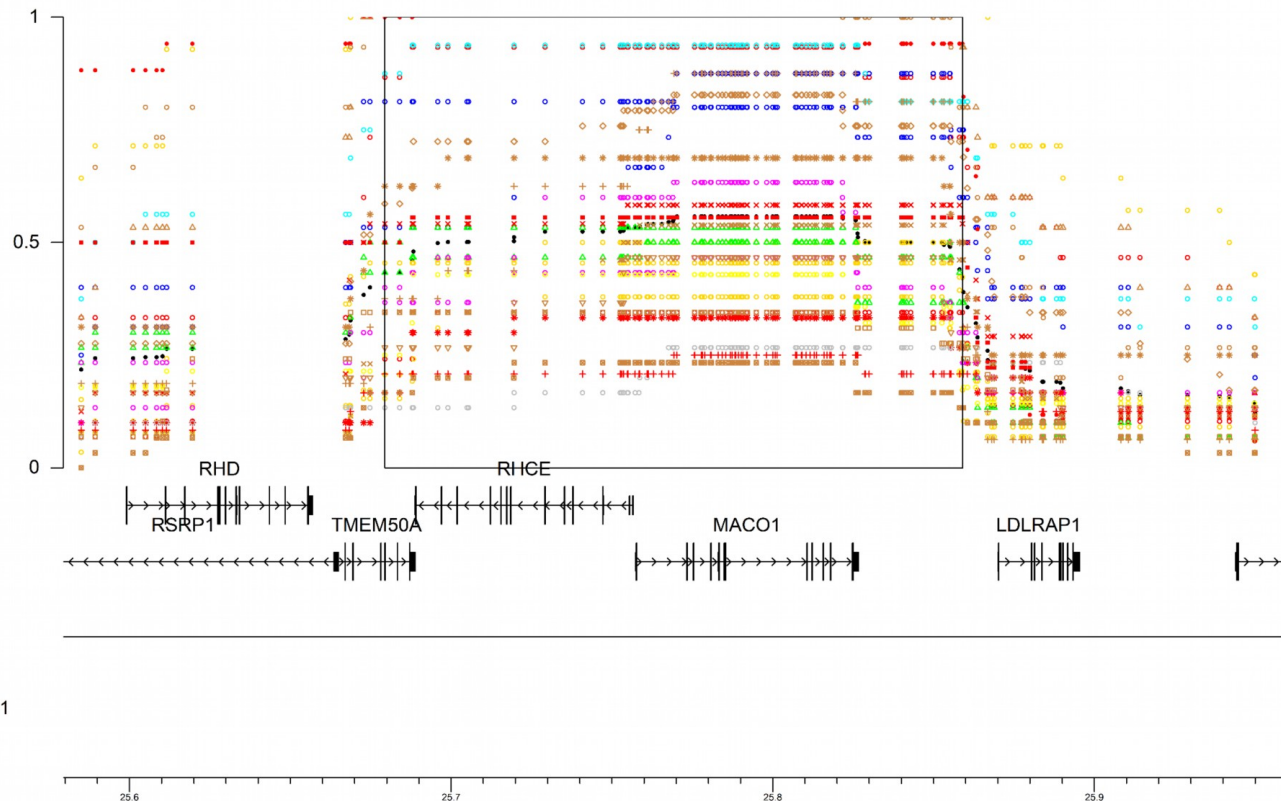
Определение «горячих точек» ROH

R

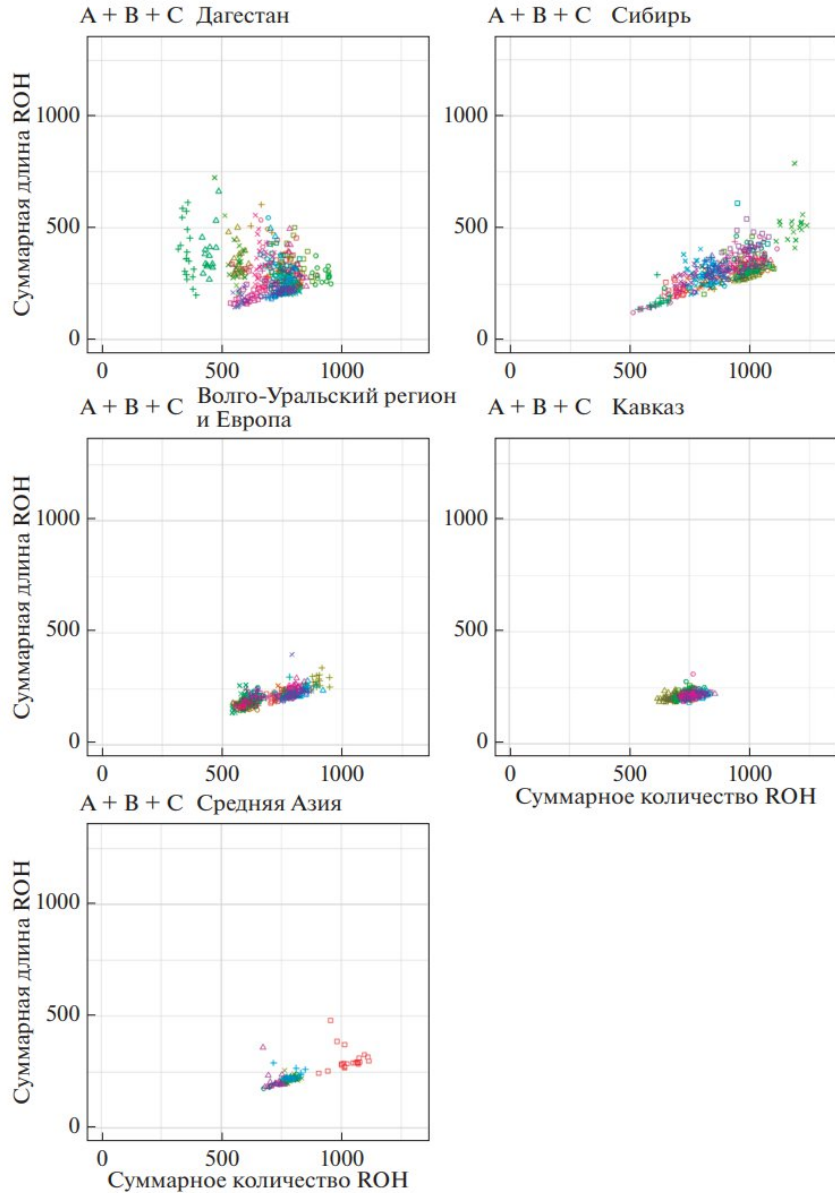
SNP в крайних значениях (> 99,5%) эмпирического распределения общих частот ROH.

если расстояние между SNP было меньше 300 т.п.н.,
регионы были объединены

Частота ROH в исследуемых популяциях



Отношение суммарной длины ROH к суммарному количеству ROH в исследуемых популяциях



- Каждый регион обладает своими особенностями.
- Небольшие популяции имеют большее количество и длину ROH, чем популяции с большей численностью.
- Метисация снижает количество ROH, тогда как эффект бутылочного горлышка увеличивает количество ROH, которые обычно все еще относительно короткие. Инбридинг добавляет небольшое количество очень длинных ROH, таким образом также увеличивая разброс в сумме ROH, видимую как сдвиг вправо на графике NROH по сравнению с SROH.
- Для популяций с эффектом бутылочного горлышка и распространенными кровнородственными браками, будет наблюдаться максимальное количество длинных ROH с тенденцией увеличения длины ROH.

Топ 30 «горячих точек» РОН по медиане для всех географических регионов

№	Chr	Gene	p1	p2	kb	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	В предыдущих исследованиях
1	2	-	623216	651380	28.165	0.385	0.607	0.682	0.635	0.687	0.688	
2	15	ADAMTSL3	84557676	84589613	31.938	0.389	0.543	0.564	0.538	0.565	0.565	2015
3	1	F5	169532633	169545587	12.955	0.386	0.509	0.535	0.511	0.537	0.537	
4	8	LPL	19820433	19871113	50.681	0.386	0.414	0.525	0.508	0.580	0.583	2015
5	1	TMEM50A, RHCE, MACO1	25679393	25859024	179.632	0.391	0.487	0.520	0.506	0.540	0.540	
6	1	ANGPTL3, DOCK7	62960250	63194009	233.76	0.401	0.485	0.516	0.504	0.536	0.542	2015
7	6	HLA-DPA1, HLA-DPB1	33031677	33056962	25.286	0.408	0.454	0.487	0.498	0.545	0.586	2015
8	20	CST9, CST3	23578206	23622550	44.345	0.400	0.487	0.495	0.491	0.506	0.507	
9	22	LINC02558, SLC5A1	32381272	32445946	64.675	0.404	0.468	0.496	0.483	0.503	0.508	
10	3	-	86760103	86946458	186.356	0.391	0.457	0.490	0.480	0.508	0.539	
11	6	-	126752884	127036174	283.291	0.388	0.444	0.489	0.477	0.510	0.520	
12	15	LARP6, LRRC49, THAP10	71110305	71187784	77.48	0.387	0.426	0.458	0.471	0.522	0.532	
13	11	OR10AG1	55666553	55737073	70.521	0.388	0.414	0.466	0.467	0.520	0.524	
14	3	LINC01322	165000870	165048175	47.306	0.395	0.446	0.476	0.465	0.487	0.497	
15	18	-	57736270	57905216	168.947	0.391	0.418	0.493	0.465	0.506	0.514	
16	10	-	44745925	44774086	28.162	0.388	0.466	0.478	0.464	0.480	0.480	
17	6	CDKAL1	20649254	20682622	33.369	0.385	0.437	0.480	0.463	0.485	0.486	2015
18	3	-	175673276	175806435	133.16	0.386	0.439	0.460	0.461	0.485	0.533	
19	20	UQCC1, GDF5OS, GDF5	33903354	34022672	119.319	0.386	0.449	0.464	0.456	0.473	0.475	2015
20	4	LCORL	17948668	18028987	80.32	0.388	0.438	0.466	0.453	0.469	0.475	2015
21	1	LRRC7	70303097	70404163	101.067	0.386	0.435	0.455	0.447	0.464	0.467	
22	3	-	147684551	147704038	19.488	0.386	0.432	0.459	0.446	0.464	0.465	
23	4	-	111654124	111714419	60.296	0.385	0.395	0.455	0.441	0.483	0.486	
24	4	CWH43	48950589	49051765	101.177	0.388	0.439	0.451	0.440	0.452	0.459	
25	4	SLC10A4, ZAR1, FRYL	48444367	48618432	174.066	0.385	0.399	0.423	0.440	0.489	0.502	
26	3	-	176057872	176077506	19.635	0.386	0.415	0.447	0.440	0.467	0.478	
27	4	-	156477400	156509595	32.196	0.391	0.422	0.431	0.427	0.434	0.435	
28	1	LEPR	65998467	66159877	161.411	0.385	0.423	0.431	0.426	0.436	0.440	2015
29	6	HLA-DQB2	32686196	32761826	75.631	0.385	0.404	0.411	0.425	0.412	0.530	2015
30	6	OR2B6, OR1F12, ZNF165, ZSCAN12P1, ZSCAN16-AS1, ZSCAN16, ZKSCAN8, ZNF192P1, ZSCAN9, ZKSCAN4, NKAPL, ZSCAN26, PGBD1, ZKSCAN3, ZSCAN31, ZSCAN12, ZSCAN23	27904181	28413491	509.311	0.385	0.413	0.422	0.421	0.426	0.461	2010, 2012, 2015

2010 - Nothnagel M, Lu TT, Kayser M, Krawczak M: Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum Mol Genet* 2010; 19: 2927-2935.

2012 - Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ: Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 2012; 91: 275-292.

2015 - Karafet TM, Bulayeva K B, Bulayev O A, et al. Extensive genome-wide autozygosity in the population isolates of Daghestan. *Eur J Hum Genet.* 2015 Oct;23(10):1405-12.

16 из 61 регионов «горячих точек» ROH популяций Сибири имеют сигналы направленного отбора

N	chr	Gene	p1	p2	kb	Min.	1st_Qu.	Median	Mean	3rd_Qu.	Max.	Тест на отбор*
1	1	DOCK7, ANGPTL3	62960289	63130431	170.143	0.419	0.440	0.459	0.452	0.465	0.468	xp-ehh
2	1	LEPR	65938332	66159877	221.546	0.413	0.451	0.505	0.489	0.526	0.549	xp-ehh
3	1	F5, SELP	169531135	169566827	35.693	0.415	0.471	0.589	0.542	0.589	0.589	PBS
4	1	CD55	207441975	207517696	75.722	0.413	0.428	0.428	0.428	0.430	0.432	xp-ehh
5	2	GCC2, LIMS1, CCDC138, RANBP2	109031721	109412959	381.239	0.415	0.426	0.436	0.435	0.447	0.451	nSL, iHS, xp-ehh
6	2	IRS1, RHBDD1	227647783	227751141	103.359	0.413	0.413	0.413	0.421	0.430	0.440	xp-ehh
7	3	ZBTB38	141096736	141140968	44.233	0.415	0.446	0.449	0.448	0.459	0.459	iHS
8	4	SLC2A9	10024413	10056082	31.67	0.413	0.417	0.426	0.425	0.432	0.438	iHS
9	4	SLAIN2, SLC10A4, ZAR1, FRYL	48408599	48591150	182.552	0.415	0.488	0.518	0.506	0.530	0.558	PBS
10	4	PLRG1, FGB	155452868	155488821	35.954	0.417	0.434	0.440	0.439	0.444	0.444	PBS
11	6	CDKAL1	20646912	20683164	36.253	0.413	0.514	0.556	0.532	0.564	0.564	nSL, iHS
12	6	HLA-DPA1, HLA-DPB1	33032975	33056962	23.988	0.413	0.482	0.507	0.522	0.597	0.614	PBS
13	7	CCDC146	76910774	76927488	16.715	0.413	0.472	0.518	0.512	0.568	0.572	iHS, PBS
14	8	SNTG1	51249185	51296270	47.086	0.413	0.413	0.413	0.414	0.414	0.415	iHS
15	12	ERC1	1377408	1486249	108.842	0.413	0.415	0.417	0.418	0.421	0.426	nSL, iHS
16	20	LINC00261, FOXA2, LNCNEF	22518523	22575792	57.27	0.413	0.510	0.532	0.534	0.574	0.581	iHS

Источники результатов тестов на отбор:

- Cardona A., Pagani L., Antao T. et al. *Genome-wide analysis of cold adaptation in indigenous Siberian populations* // PLoS ONE. – 2014. – Vol. 9. – № 5. – P. 98076.
- Степанов В. А., Харьков В. Н., Вагайцева К. В. и др., *Сигналы направленного отбора в популяциях народов Сибири и европейской части России*. January 2019 *Genetika* 55(10):1198-1207
- Колесников Н.А., Харьков В.Н., Зарубин А.А., и др., *Сигналы направленного отбора в популяциях коренного населения Сибири* // Генетика. 2022. Т. 58. № 4. С. 470-474.

- Данные о частотах ROH в популяциях важны для исследователей в области популяционной, эволюционной и медицинской генетики.
- Данные о разнообразии ROH в различных популяциях человека могут быть использованы для идентификации связанных с признаками ROH в исследованиях по картированию гомозиготности.
- Горячие точки ROH могут содержать мишени положительного отбора, которые испытали общее снижение генетического разнообразия и увеличение гомозиготности вокруг выбранных локусов.