

КРАТКИЕ
СООБЩЕНИЯ

УДК 575.22

СИГНАЛЫ НАПРАВЛЕННОГО ОТБОРА В ПОПУЛЯЦИЯХ
КОРЕННОГО НАСЕЛЕНИЯ СИБИРИ

© 2022 г. Н. А. Колесников¹, *. **, В. Н. Харьков¹, А. А. Зарубин¹, М. И. Воевода², М. А. Губина³,
О. В. Штыгашева⁴, Н. Р. Максимова⁵, А. Л. Сухомясова⁵, В. А. Степанов¹

¹Научно-исследовательский институт медицинской генетики, Томский национальный исследовательский
медицинский центр Российской академии наук, Томск, 634050 Россия

²Федеральный исследовательский центр фундаментальной и трансляционной медицины, Новосибирск, 630060 Россия

³Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения
Российской академии наук, Новосибирск, 630090 Россия

⁴Хакасский государственный университет им. Н.Ф. Катанова, Абакан, 655000 Россия

⁵Северо-Восточный федеральный университет им. М.К. Аммосова, Якутск, 677000 Россия

*e-mail: nikita.kolesnikov@medgenetics.ru

**e-mail: nik.fleming@mail.ru

Поступила в редакцию 12.08.2021 г.

После доработки 14.10.2021 г.

Принята к публикации 21.10.2021 г.

Популяции коренных этносов Сибири представляют значительный интерес для популяционной геномики по причине специфичности их генофондов, развивавшихся в различных генетико-демографических условиях. Данные о сигналах направленного отбора являются важным дополнением к существующим данным о эволюции генофондов и механизмах генетической адаптации населения Евразии. Мы использовали массив генотипов по 1779819 SNP в выборке из 477 человек, включающей 20 популяций коренного населения Сибири, для поиска сигналов направленного отбора с помощью теста на протяженную гомозиготность гаплотипов (nSL). Для популяций Сибири обнаружено, что все исследованные популяции сильно отличаются друг от друга по составу генов, которые демонстрируют влияние отбора. Наибольшее число значимых сигналов естественного отбора выявлено в популяциях хантов, коряков и чукчей. Среди локусов генома, несущих наиболее выраженные сигналы направленного отбора, в северных популяциях выделяются гены *ADGRB3*, *ANO3*, *CDH13*, *CUEDC1*, *PCDH15*.

Ключевые слова: популяционная генетика, Сибирь, естественный отбор, адаптация, nSL.

DOI: 10.31857/S0016675822040075

Генетическая структура коренного населения Сибири и Крайнего Севера сложилась под существенным влиянием естественного отбора, в наиболее неблагоприятных условиях среды обитания в арктическом климате Крайнего Севера и резко континентальном умеренном климате Сибири и Дальнего Востока. Длительная эволюция генофонда различных популяций Сибири, в том числе приспособительная, могла привести, с одной стороны, к формированию устойчивых адаптивных комплексов, а с другой – к специфическому спектру заболеваний как моногенной, так и мультифакторной природы. Недавние исследования показывают, что генетические системы, связанные с устойчивостью к температурным шокам, жировым метаболизмом, тонусом сосудистого русла, изменением инфекционной нагрузки [1–4], могут играть существенную роль в адаптации к неблаго-

приятным или экстремальным климато-географическим условиям. Переход от охоты к животноводству, смена типа питания на фоне резкого роста численности популяций могли привести к изменению структуры генофонда популяций по различным генам, продукты которых участвуют в липидном обмене и усвоении различных продуктов.

При анализе древней ДНК в популяциях Европы было показано по каким генам и в каком направлении за последние 8 тыс. лет действовал естественный отбор. Культурные и социальные изменения в неолите сопровождалось изменениями частоты аллелей, связанных с различными признаками. Под отбором находились аллели толерантности к лактозе, пигментации кожи и глаз, метаболизма, а также роста и веса [5]. Население Сибири все еще остается очень слабо изученным

в отношении генетического разнообразия на популяционном уровне. Мы предполагаем, что генетическая структура коренного населения Сибири сложилась под существенным влиянием генетического дрейфа на фоне резкого роста численности популяций, с одной стороны, и естественного отбора, сформировавшего устойчивые адаптивные генетические комплексы, с другой стороны. В совокупности с изоляцией расстоянием это привело к значительной генетической подразделенности населения Сибири.

Данные были получены с помощью биочипов “Infinium Multi-Ethnic Global-8 Kit”, содержащих 1779819 SNP. Популяции коренного населения Сибири ($N = 477$) представлены: алтайцами (Б – с. Бешпельтир Чемальского муниципального района, $N = 24$ и К – с. Кулада Онгудайского района, $N = 25$), бурятами (А – п. Агинское Агинского района, $N = 23$ и К – с. Курумкан Курумканского района, $N = 28$), калмыками ($N = 29$) (калмыки были собраны в Республике Калмыкия, но были включены в данную выборку, так как имеют центрально-азиатское происхождение и представляют южную часть этого региона), кетами ($N = 15$), коряками ($N = 20$), нивхами ($N = 13$), татарами (Т – г. Томск, $N = 20$), тувинцами ($N = 28$), удэгейцами ($N = 15$), хантами (К – с. Казым Белоярского района, $N = 30$ и Р – д. Русскинская Сургутского района, $N = 26$), хакасами (Т – сагайцами Таштыпского р-на, $N = 29$ и Ш – качинцами Ширинского р-на, $N = 26$), чукчами ($N = 25$), чулымцами ($N = 22$), эвенками (З – забайкальские (п. Чара Каларского района, с. Моклан и с. Тупик Тунгиро-Олёкминского района) $N = 25$ и Я – якутские, $N = 28$) и якутами ($N = 26$).

При биоинформатической обработке полученных данных из анализа были исключены SNP с более чем 10% отсутствующих генотипов. Были исключены все позиции делеций и инсерций. Данные были предварительно отфильтрованы по минимальной частоте редкого аллеля (MAF, minor allele frequency >0.01). В итоге после фильтрации исходных данных по 1779819 маркерам в финальный массив данных вошло 886889 аутосомных SNP. Кроме этого, были исключены образцы, имеющие более 5% пропущенных SNP, а также метисы первого, второго и третьего поколений.

Поиск сигналов направленного отбора в популяциях Сибири на основе анализа данных по высокоплотному набору SNP проводили в пакете SelScan с помощью геномного теста на протяженную гомозиготность гаплотипов (nSL – number of Segregating sites by Length) [6]. Ожидаемая доля ложных отклонений FDR (False discovery rate) была больше или равна 0.05. Обогащенность биологических путей (по базе данных KEGG) генами под отбором была оценена с помощью ресурса webgestalt (WEB-based Gene Set AnaLysis Toolkit).

Результаты поиска сигналов, сформированных под действием направленного отбора в изучаемых популяциях, с помощью теста nSL приведены в табл. 1. Всего было найдено 375 SNP ($fdg \geq 0.05$), попадающих в функциональные участки 287 генов, и 271 SNP – на межгенные участки. Для каждой из популяций обнаружено от одного до 83 генов. Наибольшее число значимых сигналов естественного отбора выявлено в популяциях хантов, коряков и чукчей. Пять генов (*ADGRB3*, *ANO3*, *CDH13*, *CUEDC1*, *PCDH15*), имеющих сигналы направленного отбора, повторяются не более чем у трех популяций, 28 генов – у двух, а остальные являются уникальными для отдельных популяций.

Из списка 287 генов, которые находятся в регионах генома, выявленных как мишени для направленного отбора по результатам тестов nSL, 84 по данным webgestalt оказались аннотированными по базе KEGG. 10 KEGG-путей были статистически значимы с учетом поправки Бенджамини–Хокберга (табл. 2), обогащены генами, находящимися под отбором. Основными молекулярными путями были пути, отвечающие за синаптическую пластичность: долговременная потенция (LTD) и долговременная депрессия (LTP) – отвечающие за усиление или ослабление синаптической передачи между двумя нейронами, сохраняющиеся на протяжении длительного времени после воздействия на синаптический проводящий путь, что обеспечивает нервную систему живого организма возможностью адаптироваться к изменяющимся условиям внешней среды. Также были отмечены пути дофаминергического и глутаматергического синапсов, передача сигналов ретроградного эндоканнабиноида, циркадный захват (отвечающий за синхронизацию эндокринных и поведенческих ритмов организма с сигналами окружающей среды), межклеточный щелевой контакт, обеспечивающий прямой перенос ионов и небольших молекул между соседними клетками, секреция слюны, регуляция липолиза в адипоцитах, сокращение гладкомышечных клеток сосудов.

Определенные функциональные замены в ферментах ADH, которые вызывают различную эффективность превращения этанола в ацетальдегид, могут быть защитными. Goldman и Enoch [7] предположили, что генетические вариации в генах *ALDH* и *ADH* избирательно поддерживаются, и предложили две вероятные селективные силы: микотоксины и инфекционные заболевания. *ADH5* участвует в механизме защиты от эндогенного формальдегида, детоксифицируя его, тем самым защищая клетку от этого генотоксина. В наших данных две популяции характеризуются высокими значениями теста nSL (эвенки (З), тувинцы) для kgp7824500, располагающегося в UTR3 гена *ADH5*, что также может быть связано с защитными функциями. Регион, в котором находится *ADH5*, ранее

Таблица 1. Результаты поиска сигналов направленного отбора с помощью теста nSL

Популяция	Гены с сигналами направленного отбора
Алтайцы (Б)	<i>ADCY1, DHX37, DYNC1H1, FOXP1, LINC00682, LINC02299, LOC101926960, MGLL, PDE10A, PPP2R2B, PTPRG, STK32B, SUCLG2-AS1</i>
Алтайцы (К)	<i>ADGRB3, ARL15, C5orf66, CELF2, LINC01378, LINC01622, RAPGEF2, RBFOX1, RBMS3, SOX5</i>
Буряты (А)	<i>STAG1</i>
Буряты (К)	<i>GLO1, PCDH15, PLSCR1, PRR16, ST8SIA6, TNIK</i>
Чукчи	<i>ACTA2, ADAMTS18, ANKRD60, CACNA2D3, CDCA7L, CDK5R2, CNTN5, FAM135B, FHIT, GABRG3, GMFG, HS1BP3-IT1, HYDIN, ITS2, LOC100507283, MOBP, NALCN-AS1, NCALD, NLRP5, NRXN3, OPCML, PAPP, PCLO, PDLIM5, PGR, PRKACB, SARIA, SCUBE2, TDRP, TENM2, TMEM65, TNIK, TSBP1-AS1, VAV2, VSTM2A</i>
Чулымцы	<i>ANKRD33B, CFAP299, DPP6, GRIA2, HLA-DQA1, LAMA1, LINC00473, LINC01435, LOC100128164, MEIS1, MIR646HG, MYH11, PREP, PTPRN2, RDH10-AS1, RGMB-AS1, SAMD7, SPATS2, STARD4-AS1, TMEM244, TMSB10, TMTC2, TSBP1-AS1, USH2A, XKR4, ZNF423</i>
Эвенки (Я)	<i>ADGRB3, AUTS2, DOCK8, DPP6, GABRB3, GAP43, LINC02294, LINC02743, LOC101928855, MYO15B, NR5A2, NTF3, PTPRN2, TEC, TENT5A</i>
Эвенки (З)	<i>ADH5, ANO3, FAM228B</i>
Калмыки	<i>CUEDC1, ERC1, MGST3, ZNF407</i>
Кеты	<i>LOC101928304, MCF2L2</i>
Хакасы (Ш)	<i>ANO3, CCDC138, CDH13, CUEDC1, EDAR, LINC00974, LINC02568, PRKG2, STK32B</i>
Хакасы (Т)	<i>AATK, CYP11B-AS1, FOXP2, GRIK4, HLA-B, KIAA1324, LINC00448, LINC01630, LRP1B, MCPH1, MRC1, TMEM236, NCKAP5L, NCOA7-AS1, NRCAM, NTRK2, PALLD, PLD5, PPM1L, RARB, SDK2, ST6GALNAC1, TRPC6, UBB</i>
Ханты (К)	<i>ARFGEF3, ATP2C1, BTBD9, CCDC33, CD226, CDH13, CSMD2, DGKH, DNAH11, EGFLAM, FAM135B, GAP43, GDI2, GRIN2B, HCG9, ITPR3, KCNAB1, KDM1B, LARGE1, LINC00473, LINC01006, LINC01339, LINC02147, LOC100507283, MAPRE2, PDZD2, PRIMA1, ROR1, RSPO4, SAA3P, SLC6A15, SMARCA2, SMOC1</i>
Ханты (Р)	<i>ADGRA3, AFF3, ALX1, ANK1, ANKRD33B, APBA1, ATP4B, ATXN1, CCND2, CDH13, CDKAL1, CNIH3, CNTN3, CSMD2, CXADR, DCANP1, TIFAB, DDX55, DSCAM, EPHA6, ERC1, FARS2, FAT2, GAREM1, GATAD2A, GDPD5, GLT1D1, GNAL, KIF6, KSR2, LINC00929, LINC01088, LINC01339, LINC02006, LINC02147, LINC02618, LOC101927630, LOC101929595, LOC105370586, LOC105374428, LOC339862, LRFN2, LYRM4-AS1, MTCH1, MYOM1, NAV2, OR5D16, OSBPL6, PBX4, PCDH15, PGR, PIP4K2A, PLEC, PLXNA2, PPP1R16B, PPP2R5C, RASGRP3, RDH13, RFC1, RMND1, RNF7, ROBO2, RYR1, SAA3P, SDK1, SLC14A2, SLC35G2, SLC46A2, SLX4IP, SNAP25, SOX1-OT, ST18, STAG1, SUGP1, TAS1R2, TCERGIL, TCF7L1, TCF7L2, TM6SF2, TMEM163, TSHZ2, TSNA-X-DISC1, WWOX</i>
Коряки	<i>AGBL1, ATP8B1, CACNA2D3, CACNA2D3-AS1, CCDC169-SOHLH2, SOHLH2, CD96, CHI3L1, CLCN1, DOK7, FLJ22447, GACAT2, HDAC9, ITPR1, KCNAB1, LIMS1, LINC02542, LOC100507053, LOC105369911, LRRN2, MMP16, MREG, NLRP11, NRG1, OSTM1-AS1, PCDH15, PRKG1, SLC10A7, SLC16A12, SPIRE2, TMPRSS13, TNIP1, ZBED9</i>
Нивхи	<i>HCG9, LRRC52-AS1, TMEM237</i>
Татары (Т)	<i>ELMO1, LOC101927822</i>
Тувинцы	<i>ADH5, ANO3, BCAR3, CUEDC1, LIMS1, LOC101927822, LYRM4-AS1, RARB, RBFOX1, SULT1C4</i>
Удэгейцы	<i>HPSE, LOC101928764, RBMS3, ST8SIA1</i>
Якуты	<i>ADGRB3, CRTAC1, GLI3, HECW2, LCORL, NTRK3, PRDM6, RHOJ, SMOC1</i>

Таблица 2. Биологические пути (по базе данных KEGG), обогащенные сигналами естественного отбора

Набор генов	Описание пути	p Value	FDR	Гены
hsa04713	Циркадный захват	1.09E-06	3.6E-04	<i>ADCY1, GRIA2, GRIN2B, ITPR1, ITPR3, PRKACB, PRKG1, PRKG2, RYR1</i>
hsa04730	Долговременная депрессия	5.21E-05	8.4E-03	<i>GRIA2, ITPR1, ITPR3, PRKG1, PRKG2, RYR1</i>
hsa04720	Долговременная потенция	9.75E-05	8.4E-03	<i>ADCY1, GRIA2, GRIN2B, ITPR1, ITPR3, PRKACB</i>
hsa04728	Дофаминергический синапс	1.03E-04	8.4E-03	<i>GNAL, GRIA2, GRIN2B, ITPR1, ITPR3, PPP2R2B, PPP2R5C, PRKACB</i>
hsa04723	Ретроградный эндоканнабиноидный сигнал	2.40E-04	1.5E-02	<i>ADCY1, GABRB3, GABRG3, GRIA2, ITPR1, ITPR3, MGLL, PRKACB</i>
hsa04724	Глутаматергический синапс	2.76E-04	1.5E-02	<i>ADCY1, GRIA2, GRIK4, GRIN2B, ITPR1, ITPR3, PRKACB</i>
hsa04923	Регуляция липолиза в адипоцитах	3.27E-04	1.5E-02	<i>ADCY1, MGLL, PRKACB, PRKG1, PRKG2</i>
hsa04270	Сокращение гладкомышечных клеток сосудов	3.97E-04	1.6E-02	<i>ACTA2, ADCY1, ITPR1, ITPR3, MYH11, PRKACB, PRKG1</i>
hsa04540	Щелевидное соединение	4.39E-04	1.6E-02	<i>ADCY1, ITPR1, ITPR3, PRKACB, PRKG1, PRKG2</i>
hsa04970	Секреция слюны	4.95E-04	1.6E-02	<i>ADCY1, ITPR1, ITPR3, PRKACB, PRKG1, PRKG2</i>

был отмечен как регион, находящийся под отбором у восточно-азиатских популяций [8].

В восточно-азиатском регионе одним из самых популярных кандидатов-участков является *EDAR*. Он кодирует белок, отвечающий за толщину волос и морфологию резцов, и неоднократно идентифицировался как кандидат на отбор у жителей Восточной Азии [8–10]. В нашем исследовании сигнал направленного отбора в этом гене, а также в близкорасположенном гене *CCDC138* встречается в популяции хакасов-качинцев (Ш). Помимо этого рядом с *EDAR* располагается ген *LIMS1*, попадающий в один регион, находящийся под отбором у восточно-азиатских популяций [8]. Для популяций коряков, тувинцев наблюдаются высокие значения по гену *LIMS1*, белок которого участвует в регуляции выживаемости, пролиферации и дифференцировки клеток.

Ранее для коренного населения Сибири был обнаружен целый ряд генов, предположительно вовлеченных в процесс адаптации к холоду (*THADA*, *ITPR3*, *GNGT1*, *PRKG1*, *RELN*, *CPT1A* и *LRP5*) [3]. Среди них в наших популяциях сигналы направленного отбора встречаются у хантов (Р) – в гене *ITPR3*, играющем ключевую роль в экзокринной секреции, лежащей в основе энергетического метаболизма и роста, коряков – в гене *PRKG1*, отвечающем за сокращение гладкомышечных клеток сосудов.

Таким образом, показано, что все исследованные популяции сильно отличаются друг от друга по составу генов, которые демонстрируют влияние отбора. Наибольшее число значимых сигналов естественного отбора выявлено в популяциях хантов, коряков и чукчей. Среди локусов генома, несущих наиболее выраженные сигналы направленного отбора, в северных популяциях выделяются гены *ADGRB3*, *ANO3*, *CDH13*, *CUEDC1*, *PCDH15*. Различные комплексы генов демонстрируют вовлеченность в разные метаболические пути, отвечающие за синаптическую пластичность, межклеточный щелевой контакт, секрецию слюны, регуляцию липолиза в адипоцитах, сокращение гладкомышечных клеток сосудов.

В итоге было выявлено существенное число сигналов направленного естественного отбора, вероятно повлиявших на генетическую структуру популяции в процессе долговременной адаптации к климато-географическим условиям обитания в резко континентальном или арктическом климате.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-34-90101.

Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального и/или национального комитета по исследовательской этике и Хель-

синкской декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики.

От каждого из включенных в исследование участников было получено информированное добровольное согласие.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Степанов В.А., Харьков В.Н., Вагайцева К.В. и др. Сигналы направленного отбора в популяциях народов Сибири и европейской части России // Генетика. 2019. Т. 55. № 10. С. 1198–1207. <https://doi.org/10.1134/S0016675819100126>
2. Hancock A.M., Alkorta-Aranburu G., Witonky D.B., Di Rienzo A. Adaptations to new environments in humans: The role of subtle allele frequency shifts // *Philosoph. Transactions Royal Soc. B: Biological Sci.* 2010. V. 365. № 1552. P. 2459–2468. <https://doi.org/10.1098/rstb.2010.0032>
3. Cardona A., Pagani L., Antao T. et al. Genome-wide analysis of cold adaptation in indigenous Siberian populations // *PLoS One*. 2014. V. 9. № 5. P. 98076. <https://doi.org/10.1371/JOURNAL.PONE.0098076>
4. Череди́ченко А.А., Трифонова Е.А., Вагайцева К.В. и др. Распространенность аллелей полиморфных вариантов генов, ассоциированных с иммунозависимыми заболеваниями, в популяциях Северной Евразии // *Мол. биология*. 2015. Т. 49. № 6. С. 984–992. <https://doi.org/10.7868/S002689841506004X>
5. Mathieson I., Lazaridis I., Rohland N. et al. Genome-wide patterns of selection in 230 ancient Eurasians // *Nature*. 2015. V. 528. № 7583. P. 499–503. <https://doi.org/10.1038/nature16152>
6. Szpiech Z.A., Hernandez R.D. Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection // *Mol. Biol. Evol.* 2014. V. 31. № 10. P. 2824–2827. <https://doi.org/10.1093/molbev/msu211>
7. Goldman D., Enoch M.A. Genetic epidemiology of ethanol metabolic enzymes: a role for selection // *World Rev. Nutrition Dietetics*. 1990. V. 63. P. 143–160. <https://doi.org/10.1159/000418505>
8. Racimo F. Testing for ancient selection using cross-population allele frequency differentiation // *Genetics*. 2016. V. 202. № 2. P. 733–750. <https://doi.org/10.1534/genetics.115.178095>
9. Sabeti P.C., Varilly P., Fry B. et al. Genome-wide detection and characterization of positive selection in human populations // *Nature*. 2007. V. 449. № 7164. P. 913–918. <https://doi.org/10.1038/nature06250>
10. Grossman S.R., Shlyakhter I., Karlsson E.K. et al. A composite of multiple signals distinguishes causal variants in regions of positive selection // *Science*. 2010. V. 327. № 5967. P. 883–886. <https://doi.org/10.1126/science.1183863>

Signals of Directed Selection in the Indigenous Populations of Siberia

N. A. Kolesnikov^{a, *, **}, V. N. Kharkov^a, A. A. Zarubin^a, M. I. Voevoda^b, M. A. Gubina^c, O. V. Shtygasheva^d, N. R. Maksimova^e, A. L. Sukhomyasova^e, and V. A. Stepanov^a

^aTomsk National Research Medical Center of the Russian Academy of Sciences, Research Institute of Medical Genetics, Tomsk, 634050 Russia

^bFederal State Budget Scientific Institution “Federal Research Center of Fundamental and Translational Medicine”, Novosibirsk, 630060 Russia

^cInstitute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, 630060 Russia

^dFederal State-Funded Educational Institution of Higher Education “Katanov Khakass State University”, Abakan, 655000 Russia

^eAmmosov North-Eastern Federal University, Yakutsk, 677000 Russia

*e-mail: nikita.kolesnikov@medgenetics.ru

**e-mail: nik.fleming@mail.ru

The indigenous populations of Siberia are of significant interest for population genomics, due to the specificity of their gene pools, which developed in various genetic and demographic conditions. Data on directional selection signals is an important addition to the existing data on the evolution of gene pools and the mechanisms of genetic adaptation of the Eurasia population. We used genotype array of 1 779 819 SNPs in a group of 477 unrelated subjects, including 20 indigenous populations of Siberia, to search for directional selection signals using a test for extended homozygosity of haplotypes (nSL). The present study detected that all studied Siberia populations strongly differ from each other in the gene composition that demonstrate the effect of selection. The largest number of significant signals of natural selection was found in the populations of the Khanty, Koryak, and Chukchi. The genes *ADGRB3*, *ANO3*, *CDH13*, *CUEDC1*, *PCDH15* are distinguished among the genomic loci carrying the most pronounced directional selection signals in the northern populations.

Keywords: population genetics, Siberia, natural selection, adaptation, nSL.