



Genetic architecture and adaptations of Nunavik Inuit

Sirui Zhou^{a,b,c}, Pingxing Xie (谢萍星)^d, Amélie Quoibion^{a,e}, Amirthagowri Ambalavanan^{a,e}, Alexandre Dionne-Laporte^{a,f}, Dan Spiegelman^{a,f}, Cynthia V. Bourassa^{a,f}, Lan Xiong^{a,b,f}, Patrick A. Dion^{a,f}, and Guy A. Rouleau^{a,d,f,1}

^aMontreal Neurological Institute and Hospital, McGill University, Montréal, Québec, Canada H3A 2B4; ^bDepartment of Medicine, Faculty of Medicine, Université de Montréal, Montréal, Québec, Canada H3T 1J4; ^cEpidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada H3A 1A2; ^dFaculty of Medicine, McGill University, Montréal, Québec, Canada H3G 1Y6; ^eDepartment of Human Genetics, McGill University, Montréal, Québec, Canada H3A 0C7; and ^fDepartment of Neurology and Neurosurgery, McGill University, Montréal, Québec, Canada H3A 2B4

Edited by Rasmus Nielsen, University of California, Berkeley, and accepted by Editorial Board Member Mary-Claire King March 18, 2019 (received for review June 21, 2018)

The Canadian Inuit have a distinct population background that may entail particular implications for the health of its individuals. However, the number of genetic studies examining this Inuit population is limited, and much remains to be discovered in regard to its genetic characteristics. In this study, we generated whole-exome sequences and genomewide genotypes for 170 Nunavik Inuit, a small and isolated founder population of Canadian Arctic indigenous people. Our study revealed the genetic background of Nunavik Inuit to be distinct from any known present-day population. The majority of Nunavik Inuit show little evidence of gene flow from European or present-day Native American peoples, and Inuit living around Hudson Bay are genetically distinct from those around Ungava Bay. We also inferred that Nunavik Inuit have a small effective population size of 3,000 and likely split from Greenlandic Inuit ~10.5 kya. Nunavik Inuit went through a bottleneck at approximately the same time and might have admixed with a population related to the Paleo-Eskimos. Our study highlights population-specific genomic signatures in coding regions that show adaptations unique to Nunavik Inuit, particularly in pathways involving fatty acid metabolism and cellular adhesion (*CPNE7*, *ICAM5*, *STAT2*, and *RAF1*). Subsequent analyses in selection footprints and the risk of intracranial aneurysms (IAs) in Nunavik Inuit revealed an exonic variant under weak negative selection to be significantly associated with IA (*rs77470587*; $P = 4.6 \times 10^{-8}$).

Nunavik Inuit | genetic architecture | demographic history | natural selection | intracranial aneurysm

The settlement of the Arctic took place over a period of ~6,000 y (1), and the present-day Nunavik Inuit of Canada are the descendants of genetically distinct waves of early settlers. Many believed the occupation of this territory started with the early Paleo-Eskimos, who subsequently were replaced by the Dorset culture and the Thule people (2), the latter becoming the ancestors of all modern Inuit. Nunavik Inuit have adapted to the Arctic environment; however, in recent decades, changes of their lifestyle are believed to have increased their risk of several complex disorders (3), particularly cardiovascular and cerebrovascular diseases (4).

Recently, several studies have genotyped Siberians and Greenlandic Inuit by using high-density SNP arrays to define their genomic profiles. A selection scan of the Siberian populations highlighted genes involved in lipid metabolism and vascular smooth muscle contraction (5), with a strong selective sweep of a *CPT1A* deleterious variant described in a follow-up study (6). Studies of the Greenlandic population revealed substantial recent European gene flow into the Greenlandic Inuit (7), with a strong selection signal reported in a cluster of fatty acid desaturase genes (*FADS1–3*) that determine polyunsaturated fatty acid levels (8). A *TBC1D4* founder mutation was also reported that is associated with increased prevalence of type 2 diabetes in Greenlandic Inuit (9).

Despite these recent advances, there are various aspects of the genetic profile of Inuit populations that remain to be explored. Although the Nunavik and Greenlandic Inuit are descendants of common ancestors, significant genetic drift has accumulated between them. The underlying genetic factors potentially associated with higher prevalence of cardiovascular and cerebrovascular

disorders in Arctic indigenous peoples are still poorly understood. Through a combined analysis of high-density SNP-chip genotyping and whole-exome sequencing (WES) of an Inuit cohort that includes more than 1% of the population of Nunavik, our study provides clues that will help to clarify the nature of the underlying genetic factors.

Results

Population Structure of Nunavik Inuit. Principal component analysis (PCA) of 5,422 present-day individuals and ancient genomes (*SI Appendix, Table S1*) revealed that Nunavik Inuit form a cluster relative to other worldwide populations (*SI Appendix, Fig. S1*). PCA using genotype and WES data on a selection of Asian and American indigenous populations suggested that the closest relatives to the Nunavik Inuit are Siberian Eskimos and Greenlandic Inuit (*Fig. 1A* and *SI Appendix, Fig. S2*). The Saqqaq individual (posited to belong to a Paleo-Eskimo group) (10) was genetically closest to Siberian Koryaks and did not cluster with Inuit from Nunavik or Greenland. The Nunavik Inuit form two clusters: the two Ungava Bay villages (Kuujuuaq and Kangiqsualujjuaq) were distinct from the rest of the villages located near Hudson Bay (*Fig. 1B*).

Analysis with ADMIXTURE software (11) revealed that individuals from Ungava Bay villages might be admixed with Native Americans (*SI Appendix, Fig. S3*, dark purple). Further analysis inferred an average of 16% of admixture in Kangiqsualujjuaq Inuit. Interestingly, this admixed ancestry was observed mainly

Significance

We present a genetic study to highlight the genomewide architecture of Nunavik Inuit with emphasis on selection in gene coding regions. We discovered that the majority of Nunavik Inuit have negligible admixture with present-day populations and small effective population size, and show evidence of genetic relatedness from ancient genomes. We identified genetic differentiations in Nunavik Inuit villages that correlate with their migration route and placed Nunavik Inuit in a population tree in relation to Siberian and Native Americans. Nunavik Inuit also had genetic footprints that reflect high levels of natural selection in functionally relevant genes, from which may arise the genetic risk responsible for their predisposition toward diseases such as intracranial aneurysm.

Author contributions: S.Z., L.X., P.A.D., and G.A.R. designed research; S.Z., P.X., and A.Q. performed research; S.Z., A.A., A.D.L., D.S., and C.V.B. analyzed data; S.Z. wrote the paper; and P.A.D. and G.A.R. revised the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. R.N. is a guest editor invited by the Editorial Board.

Published under the PNAS license.

Data deposition: The genotype and WES data reported in this paper have been deposited at Zenodo, <https://zenodo.org/record/3336535>.

¹To whom correspondence may be addressed. Email: guy.rouleau@mcgill.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1810388116/-DCSupplemental.

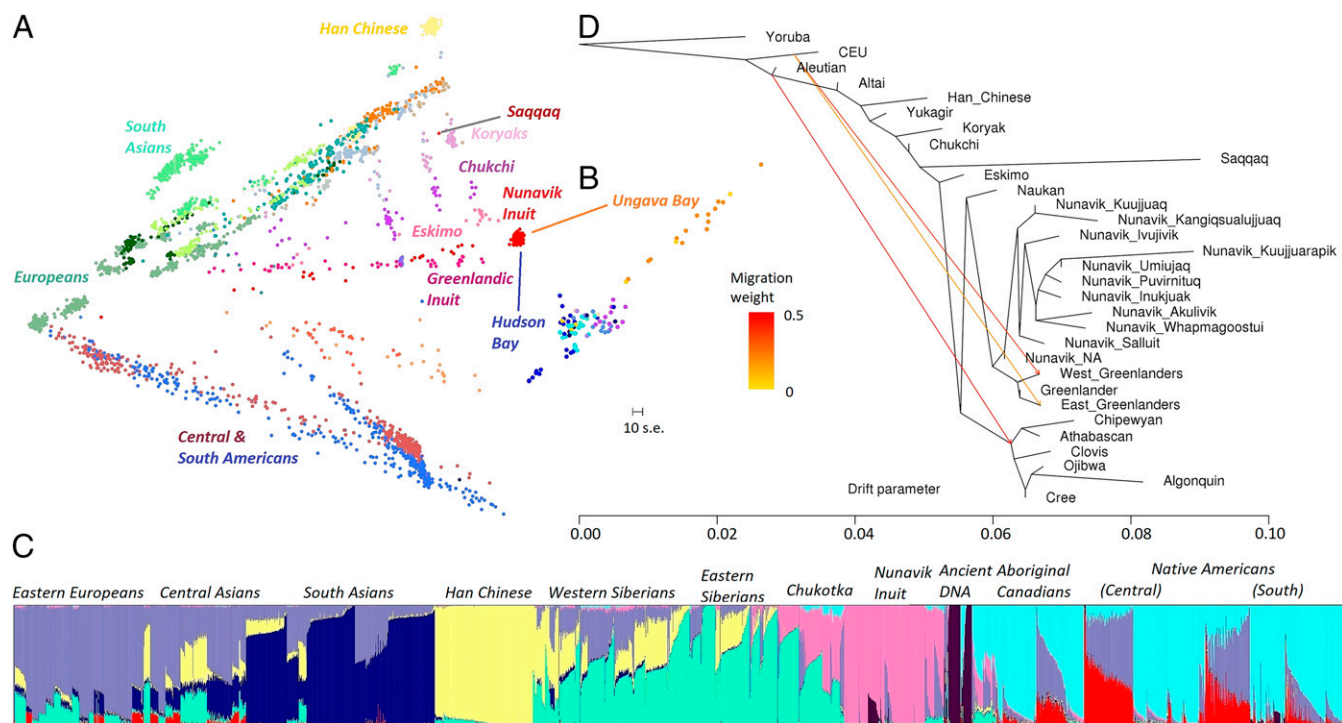


Fig. 1. Genetic structure of Nunavik Inuit and affinities between Nunavik Inuit and worldwide populations. (A) PCA of Nunavik Inuit and out-of-Africa populations. Red cluster represents Nunavik Inuit, with notable populations identified on the figure in their respective colors. (B) Nunavik Inuit by villages per exome data. Orange colors represent Ungava Bay Inuit; blue colors represent Hudson Bay Inuit. (C) Admixture proportions estimated by ADMIXTURE software assuming $K = 8$ ancestral components of Nunavik Inuit and present-day populations. Populations are arranged by their geographical distance with Nunavik Inuit, and colors represent the admixture proportions from each ancestral component. The Native American admixture observed in Nunavik Inuit is displayed in dark purple. (D) Maximum-likelihood genetic-drift tree topology relating people from villages of Nunavik Inuit and other Arctic indigenous populations (estimated by Treemix; three migration events displayed). (Scale bar: 10-unit SE of entries in the covariance matrix.) Yoruba is set as outgroup in rooting the tree. Nunavik_NA, Nunavik Inuit recruitment location unknown.

in the genomes of ancient Americans and Native Americans (Fig. 1C, dark purple). In addition, 21 of 170 Nunavik Inuit showed more than 1% of European admixture, 10 of whom showed more than 25%; this higher percentage could have been introduced by very recent admixing events. Conversely, 87% of the Nunavik Inuit display no evidence of European admixture. Our results also showed that the most distantly related population with a component of Inuit ancestry (Fig. 1C, pink) live around the Altai region, and that such ancestry was increased drastically from Western Siberia to Northeastern Siberia and was significantly reduced in Native American populations. Only groups of indigenous Canadians retained this ancestry.

A pairwise fixation index (F_{ST}) examination of population structure comprising 12 single populations (closer geographic distance to Nunavik Inuit) and 16 population groups showed that most populations and groups ($n = 24$) were of high genetic differentiation with Nunavik Inuit ($F_{ST} \geq 0.08$). Greenlandic Inuit and Northeastern Siberians including Naukan, Chukchi, and Siberian Eskimos were moderately differentiated from Nunavik Inuit ($F_{ST} = 0.042, 0.048, 0.060,$ and 0.059 , respectively), which also correlates with ADMIXTURE results (SI Appendix, Fig. S5). A similar trend was observed in the genomewide local ancestries and ancestry tracts of admixed circumpolar individuals [inferred by PCAdmix (12) and tracts (13)], demonstrating that Nunavik Inuit share most of their genetic ancestry with arctic coastal populations such as Greenlandic Inuit (83.4%) and Naukan (72.7%), with less ancestry shared with inland populations (SI Appendix, Fig. S6). Among Siberians, the proportion of long Inuit ancestry tracts also decreases from coastal to inland Siberia regions, with Naukan showing the largest proportion (SI Appendix, Fig. S7). In Nunavik Inuit, individuals from Kangiqsualujuaq were estimated to have an average of 2.0–3.8% of their ancestry from ancient Americans, inferred from the genomes of Saqqaq, Clovis

(14), and Kennewick (15) (SI Appendix, Fig. S8). We estimate that Saqqaq shared 67% of their ancestry with today's Nunavik Inuit (SI Appendix, Fig. S5). However, the ancient ancestry inferred from ancient Arctic genomes (2) accounted for only 6.9% of the Kangiqsualujuaq genome, which could be a result of the low sequencing quality of the ancient genomes.

History of the Populating of Nunavik. We used TreeMix (16) to infer patterns of population splits and mixtures of Nunavik Inuit and the historically related Siberian–Arctic indigenous populations. The result showed the root split between Inuit and other populations, between Nunavik and Greenlandic Inuit, between Nunavik Inuit from Salluit and other villages, and between Ungava Bay and Hudson Bay villages. Such a result suggests that the ancestors of today's Inuit might have entered Nunavik at the northwest point from Nunavut and traveled southward along Ungava Bay and Hudson Bay (Fig. 1D). The split of the Ungava Bay branch was also seen in the F_3 statistics, in which Kangiqsualujuaq and Kuujuaq Inuit have a greater shared genetic shift relative to Han Chinese (CHB; from Beijing) populations than most other Inuit and Siberian–Arctic indigenous populations (Dataset S1). The TreeMix model with three migration events explained 95.6% of the variance of relatedness between populations. After increasing the migration events to 10, the variance of relatedness reached 98.5%. TreeMix models of alternative datasets including additional Siberian and Native American populations yielded similar results (SI Appendix, Fig. S10).

The result of D -statistics (17) (SI Appendix, Fig. S11) also supported the split between Ungava Bay Inuit and Hudson Bay Inuit, with Ungava Bay Inuit being genetically closer to the Greenlandic Inuit and indigenous Canadian branches (SI Appendix, Fig. S12). Nunavik Inuit are genetically closer to indigenous Canadians and West Siberians, whereas Greenlandic Inuit are genetically closer to East Siberians (Fig. 2A), suggesting that Nunavik Inuit might

have arisen from a different migration wave as Greenlandic Inuit. The F_3 statistics also supported a greater shared drift between Nunavik Inuit and Paleo-Eskimo in comparison with other Arctic indigenous peoples (Dataset S1).

Linkage disequilibrium (LD) decay showed that Nunavik Inuit have the highest degree of average LD (measured by r^2) in comparison with other Arctic indigenous peoples and the 1000 Genomes Project (1KGP) populations, with Kangiqsualujuaq Inuit having the highest LD within Nunavik Inuit (SI Appendix, Fig. S13). These observations correlated with runs of homozygosity (ROH) results, in which 27.7% of the ROHs were longer than 10 Mb in the Kangiqsualujuaq Inuit (SI Appendix, Fig. S14). The Inuit from Salluit and Kuujuarapik showed a much smaller percentage of long ROH segments, which might be a result of the inclusion of several recently European-admixed individuals.

Demographic History of Nunavik Inuit. We used SMC++ (18) to infer split times and effective population sizes (N_e values) with respect to time for Nunavik Inuit ($n = 4$) in comparison with Siberians [Altai ($n = 2$), Yakut ($n = 2$), Ket ($n = 2$), Nivkhs ($n = 2$)] and North American indigenous peoples [Athabaskan ($n = 2$), Siberian Eskimos ($n = 2$), Greenlanders ($n = 2$); Fig. 2B]. The N_e of Inuit was estimated to be much lower ($N_e \sim 3,000$) than those of Siberians and North American indigenous populations, with a potential bottleneck that occurred ~ 10 kya. It was also inferred that Nunavik Inuit split from Greenlandic Inuit and Siberian Eskimos approximately 10.5 kya and 11 kya, respectively (SI Appendix, Fig. S15). momi2 (19) was used to fit a subdemography on Nunavik Inuit and Altai (from whom Inuit ancestors were estimated to have originated) and then to iteratively build on this model by adding the ancient Saqqaq genome and inferred migration events. Similar to the results of SMC++, momi2 estimated that the split between Nunavik Inuit and Altai dates approximately 15 kya, the common ancestor of Altai and Nunavik Inuit split from the ancestor of Saqqaq approximately 42 kya, and Nunavik Inuit received a 25% pulse from Saqqaq at ~ 8 kya (Fig. 2C).

Genes Under Natural Selection in Nunavik Inuit. Population branch statistics (PBSs) (20) using pairwise F_{ST} comparisons of three populations were generated to test for genes under natural selection. PBSs were calculated on a total of 387,339 single nucleotide variants (SNVs) in and around coding regions between Nunavik Inuit and CHB populations, with European (Utah residents with European ancestry; CEU) used as outgroup. A total of 38,116 SNVs with PBS > 0.1 were depicted in a Manhattan plot (Fig. 3),

among which there were 9,883 coding SNVs, 784 splicing-site SNVs, and 1,936 SNVs in untranslated regions. Variants with an empirical threshold of the 99.9th percentile and PBS > 0.99 were selected as potentially under natural selection (Dataset S2). The variant with the highest score (PBS = 3.11) was located in *CPT1A* (rs80356779, p.P479L), which validated our previous findings (21). This Arctic-specific variant had a frequency of 68% in the Northeastern Siberian populations (6) and was found in present-day Nivkhs, Athabascans, and Aleutians (22), as well as Saqqaq (4 kya) and a late Dorset individual (1.6–1.4 kya) (10). However, p.P479L was absent in the remaining present-day populations and several ancient individuals including Mal'ta, (24 kya) (23), Clovis (11 kya) (14), and the Kennewick Man (8 kya) (15). This variant may have originated in Neolithic Siberia after the first wave of migration into the Americas, and it reached the highest frequency only around the Arctic sea.

In addition to *CPT1A*, we also confirmed the selection signals previously found in the Greenlandic Inuit (8) within the *MYRF-FADS3* cluster (chr11: 61.5–61.6 Mb), with the highest PBS of 1.17 in *FADS1*. We also discovered several new regions under potential selection, including chr16: 89.6–90.1 Mb (PBS = 1.74; *CPNE7*); chr3: 12.6–12.9 Mb (PBS = 1.76; *TSEN2*); chr8: 143.9–144.0 Mb (PBS = 1.52; *CYP11B1*); chr7 (PBS = 1.65; *IMPDH1*); and chr12 (PBS = 1.53; *STAT2*; Fig. 3 and SI Appendix, Table S2).

According to the previous selection signals based on the WES data of the Greenland Inuit, there was a strong correlation between the Nunavik and the Greenland Inuit populations. Among the 132 SNVs (79 genes) with PBS > 1.0 we found in Nunavik Inuit, 61 SNVs were also found in Greenland Inuit with PBS > 0.3 , 25 of which had PBS > 1.0 ($P = 2.8 \times 10^{-13}$, binomial test).

We also calculated the WES-based PBS between the Nunavik Inuit and the Northeastern Siberians with CHB as outgroup to detect more recent selection events. The top SNV was in *CAND2* (rs180768267; PBS = 1.09), followed by *ATPI0D* (rs16851681; PBS = 1.07), *ICAM5* (rs1056538; PBS = 1.04), *CPT1A* (rs80356779; PBS = 0.84), and *STAT2* (rs2066815; PBS = 0.84). SNVs in *CAND2*, *ICAM5*, *CPT1A*, and *STAT2* were predicted to be highly deleterious (Combined Annotation Dependent Depletion score > 15), suggesting that they may have important function in certain biological processes (SI Appendix, Table S3).

Real-time quantitative PCR (RT-qPCR) was used to test expression levels of genes under selection in Inuit and another founder population that lived in a restricted geographic area for a much shorter period. We used French-Canadians (FCs) from the Saguenay–Lac-Saint-Jean region of Québec as controls for

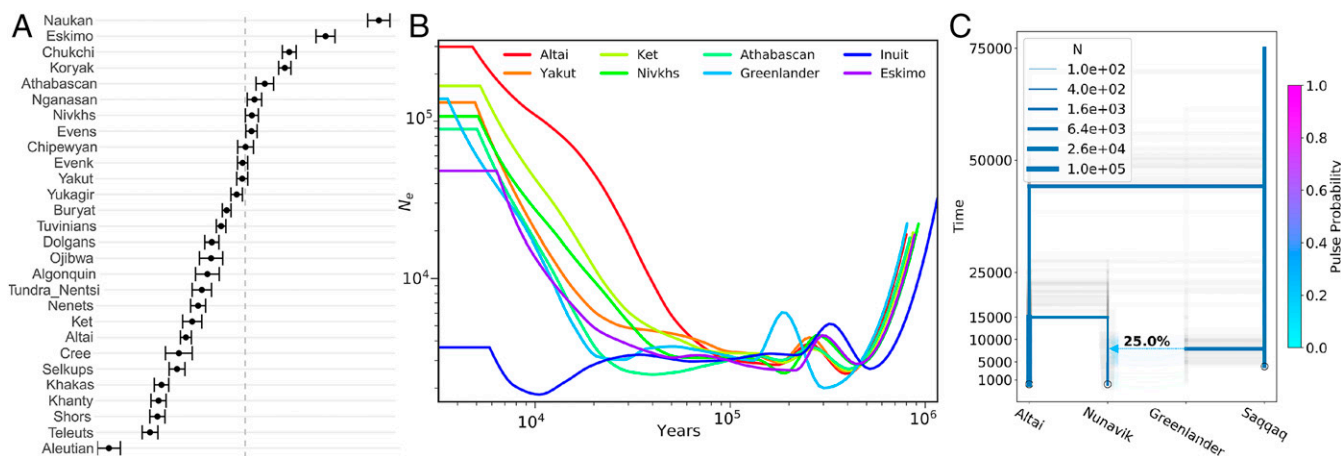


Fig. 2. Demographic history of Nunavik Inuit in the context of arctic indigenous populations. (A) D -statistics in the form of D (Nunavik, Greenlandic Inuit; X, CHB). X represents each arctic indigenous population. CHB is the outgroup. Dashed line represents $D = 0$. Error bars show 95% CIs estimated from Z values. $D < 0$ suggests Nunavik Inuit is closer to population X; $D > 0$ suggests Greenlandic Inuit is closer to population X. (B) SMC++-inferred historical N_e values in respect to times for Nunavik Inuit and other populations. (C) momi2-fitted demographic history of Nunavik Inuit, Altai, and Saqqaq using single pulse model. Estimates with the highest likelihood from 50 bootstraps resampling data are displayed. “Greenlander” represents a ghost population in Greenland.

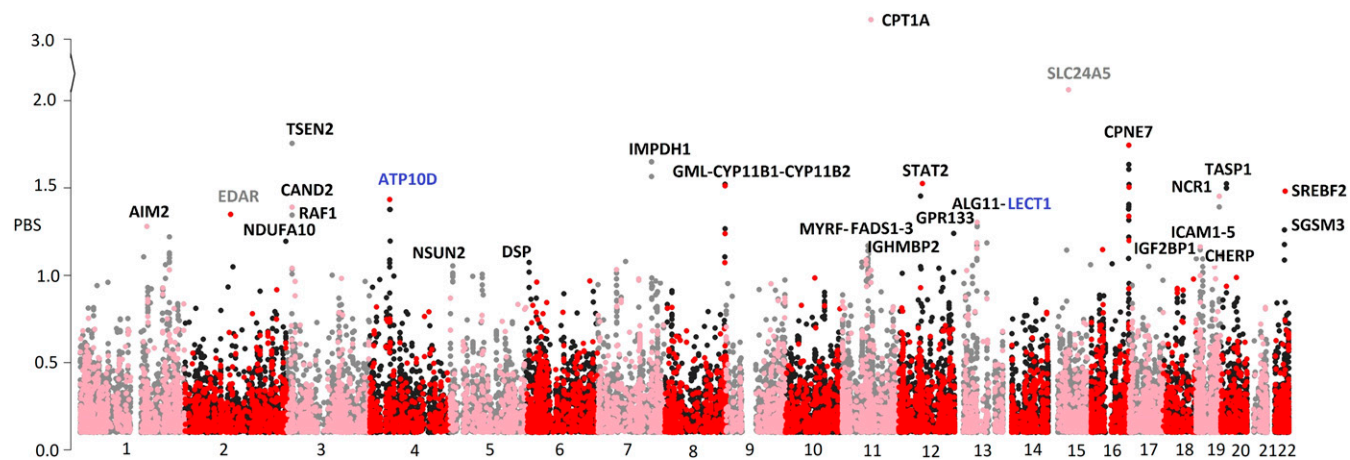


Fig. 3. Manhattan plot depicting PBS of Nunavik Inuit. Variants with PBS > 0.1 in Nunavik Inuit using CHB as a sister population and CEU as an outgroup. Red and pink dots represent exonic variants. Genes included in the subsequent expression analysis are labeled. The y axis is folded at PBS > 2.

this experiment. Thirty-two genes (*SI Appendix, Table S2*) were selected from the SNVs with top PBSs and met the selection criteria (*SI Appendix*). Among the genes tested, *GML*, *CYP11B1*, *CYP11B2*, *IMPDH1*, *NCR1*, *DSP*, *LECT1*, and *IGHMBP2* were excluded from the subsequent analyses because of their limited expression levels in lymphoblastoid cell lines (LCLs). Two genes, *EDAR* and *SLC24A5*, were also not measured here because they were previously reported to be under selection in worldwide populations (24). From the remaining 22 genes, 5 were of special interest because they showed the same direction of differential expression across 3 independent tests (*CPNE7*, $P = 0.045$; *ICAM5*, $P = 0.041$; *STAT2*, $P = 0.050$; *RAF1*, $P = 0.053$; *FADS1*, $P = 0.051$; *SI Appendix, Fig. S16*). However, their P values did not survive correction for multiple testing.

Intracranial Aneurysms and Genes in Selection Footprint Regions of Nunavik Inuit. Intracranial aneurysm (IA) is a complex cerebrovascular disorder characterized by weakness of the intracranial artery walls and may be a consequence of disrupted lipid metabolism. Nunavik Inuit were previously reported to be genetically predisposed to IA; this might be the result of recent Western lifestyle changes introducing detrimental effects linked to previously beneficial variants, or of genetic drift related to their small population size. We tested this hypothesis by looking for genetic associations among 8,291 coding region-enriched variants with PBS > 0.1 and IA and discovered a missense variant in *OR4C3* to be significant on a genome-wide basis (rs77470587; $P = 4.60 \times 10^{-8}$; PBS = 0.10). *SHANK3* is the second most significant locus associated with IA (not genome-wide significant) (rs116959666; $P = 5.71 \times 10^{-5}$; PBS = 0.15; *SI Appendix, Fig. S17*).

Discussion

Genetic Structure and Demographic History of Nunavik Inuit. Our findings showed that the Nunavik Inuit are a very homogeneous population (*SI Appendix, Fig. S18*) with small N_e . Aside from a few individuals with recent admixture from Europeans, the Nunavik Inuit have nearly no ancestry from other present-day populations and are distinct from other Arctic indigenous populations including Greenlandic Inuit, making the present data a valuable addition for the construction of a genomic reference panel for Arctic indigenous peoples. Within the Nunavik Inuit, there was genetic separation between the Hudson Bay villagers and the Ungava Bay villagers, whereby part of the ancestry of Ungava Bay village Inuit is found in genomes from ancient North America (4–11 kya). This evidence potentially supports one of two historical scenarios: (i) that Nunavik Inuit at Ungava Bay had once admixed with late precontact Innu of Labrador, whose ancestors were Paleo-Indians (2); or (ii) there had been admixture

between Dorset and Thule Inuit in Nunavik, which is consistent with a previous result (25), even though the study in Greenland Inuit (7) suggested otherwise. Nevertheless, there was evidence that the Dorset people remained in part of eastern Arctic Canada until much later (1350 CE), whereas the Thule people already had occupied Greenland by 1100 CE (26). This scenario was further supported by a possible gene-flow event inferred from Paleo-Eskimos at ~8 kya, which occurred after the split of Greenlandic and Nunavik Inuit at ~10.5 kya. On the contrary, only 11 Nunavik Inuit had recent European admixture from 2–6 generations ago (admixture between 1% and 50%), which is corroborated by documented history of European encounters starting in the 19th century.

We found evidence suggesting that Thule Inuit came into Nunavik through the Sugluk Inlet in Salluit, followed by coastal migrations toward the east and west of Nunavik. The Greenlandic Inuit split off from the rest of Nunavik Inuit, correlating with the Thule migration from the Canadian high Arctic in 1300 CE into Greenland and Eastern Canada (2). We also discovered a recent split between West Greenland and Ungava Bay Inuit, which could be explained by the fact that a small group of Inuit from Baffin Island went to northwest Greenland in 1864 CE (27).

Similar to Greenlandic Inuit, Nunavik Inuit had high LD, which increased from the West to the East of Nunavik. There was also a higher percentage of large ROHs (>10 Mb) in Nunavik Inuit, especially those from Ungava Bay villages. This evidence agreed with the historically small N_e of Nunavik Inuit ($N_e \sim 3,000$) and population bottleneck. These bottlenecks may translate to an increase in the prevalence of recessive disorders, and, over longer time scales, a relative increase in the frequency of additive deleterious variants and a decrease in the frequency of recessive deleterious variants.

Genetic Signatures Highlight Adaptations of Nunavik Inuit. Two genes under global populationwide natural selection (*SLC24A5* and *EDAR*) were identified in Nunavik Inuit; the rs1426654 A allele of *SLC24A5* and the rs3827760 A allele of *EDAR* were strongly associated with skin- and hair-related traits in South and East Asians (24). Both alleles swept to fixation in Europeans but were nearly absent in Nunavik Inuit. Interestingly, *SLC24A5* rs1426654-A is also absent from certain South Indians, who bear no genetic resemblance to Nunavik Inuit.

Genes involved in lipid metabolism were previously reported to be under positive selection in Arctic indigenous populations, especially in Inuit (6, 8). In addition to confirming the involvement of two genes on chromosome 11 that were previously found to have a high level of selection signal (*CPT1A* and *FADS1*), we also discovered additional loci with unique genetic signatures in Nunavik Inuit by mainly focusing on the protein-coding regions of their genomes.

A strong signal in the Nunavik Inuit was identified on chromosome 16, in which variants in three genes, *CPNE7*, *DPEP1*,

and *CDK10*, showed the highest PBS. A *CPNE7* variant was found to be positively enriched in Greenlandic Inuit (rs139901937; $PBS^{GI} = 1.39$). An alternative splicing variant in *CPNE7* (rs12445560; $PBS = 1.74$) had an extremely high frequency in Nunavik Inuit ($PBS = 0.95$), whereas Northeast Siberians and Native Americans had slightly lower frequencies ($PBS = 0.40$ and 0.74 , respectively). The derived allele frequency of rs12445560 was much lower in Europeans and East Asians (0.1) and much rarer in Africans (0.01). The frequency changes of this splicing variant was interesting and seemed to be in line with the out-of-Africa migration, which has reached near-fixation in Nunavik Inuit. The *CPNE7* protein is a member of the copine family that binds phospholipids and also belongs to the von Willebrand factor A domain-containing proteins (28), which are known to be important in cell adhesion and migration events. Studies on copines suggest that they are associated with coronary artery disease (29) and obesity (30), which indicates that they also function in lipid metabolism pathways. However, whether the high frequency of *CPNE7* variants is caused by neutrality or selection needs further investigation.

Another locus with a significant selection signal was located on chromosome 19, containing the genes *ICAM5* and *ICAM1*. A haplotype of two missense SNVs in *ICAM5* under strong selection was observed at a higher frequency in the indigenous populations (0.60 in Native Americans and 0.56 in Northeast Siberians) compared with outgroups (0.19 in CHB and 0.38 in CEU), and had reached near-fixation in Nunavik Inuit (0.98). *ICAM5* is a neuronal-specific transmembrane glycoprotein involved in adhesion, and *ICAM1* is a critical molecule secreted by endothelium during vascular inflammation and is responsible for formation of atheroma (31). *ICAMs* are known to be involved in fatty acid metabolism, oxidative stress, and inflammatory response (32, 33), making those potentially important biomarkers for cardiovascular and cerebrovascular disease through selective adaptation.

STAT2 is a well studied inflammatory mediator and important factor in antiviral defense, and it is also a prominent modulator of inflammation in vascular cells and in atherosclerosis (34). *RAFI* also encodes an inflammation regulator and is critical in endothelial cell survival during angiogenesis (35); its expression was reported to be altered in patients with arterial hypertension (36).

In addition to fatty acid metabolism, we also identified several genes with high PBS that were enriched in pathways of vascular remodeling and inflammatory response (SI Appendix, Table S4). Differences in allele frequencies caused by genetic drift are to be expected in a small population. However, systematic differences as measured by PBS or allele frequency differences suggest the action of positive selection. These very differentiated loci are prime candidates to explain population differences in genetic risk under a shared Western lifestyle.

Selection Footprints and the Risk of IA in Nunavik Inuit. IA shares pathological similarities with other common cardiovascular diseases, such as the damaged vascular remodeling process and maladapted immunological responses. It was also demonstrated that lipid metabolism could contribute to the pathogenesis of IA (37). We hypothesized that, among the signatures of selection footprints we had observed, some may also be associated with increased risk of IA in Nunavik Inuit. Because of the degree of relatedness of Nunavik Inuit, we incorporated their genetic relationship matrix by using linear mixed models while performing an association test that identified a missense variant in *OR4C3* to be significant at a genomewide threshold. Functional studies of *OR4C3* are limited, but it is known to belong to a family of olfactory receptors that are responsible for the recognition and G protein-mediated transduction of odorant signals. In general, the identification of potential IA risk variants with low PBS scores (~ 0.1) may suggest that the susceptibility of IA in Nunavik Inuit came from only genetic drift or weak selection.

Limitations. Sample size and lack of a usable imputation panel were two major limitations to the power of our association test with IA. The demographic inference algorithms used in this study

may not be robust in modeling recent demographic events. There are also inconsistencies between different demographic models, for which further data are needed for validation. The use of sparse genome data (i.e., exome sequence) and only conserved regions may bias the inference process, leading to slightly decreased N_e , although such bias was shown to be minimal (SI Appendix, Fig. S19). Finally, LCLs may not be the ideal cell type to capture the impact of gene-expression changes brought about by selection.

In summary, we report a systematic population-genetic characterization of the Nunavik Inuit, which demonstrates specific genetic adaptations and susceptibility to IAs as a result of unique genetic background. The combination of WES with genomewide SNP genotyping data represents an effective and unbiased approach to study populations with different genetic profiles. We discovered genetic evidence that had not been reported in previous Greenlandic Inuit studies. The Nunavik Inuit is a unique population with small N_e , and have a substantial ancestry component that had previously been observed primarily in ancient samples. They have distinct genetic signatures in pathways involving lipid metabolisms and cell adhesion, which suggests differential adaptation to extreme environments among Inuit populations. Furthermore, we identified variants in regions under selection that may be associated with risk of IA. All of these have demonstrated the value of genetic and medical research in isolated, indigenous founder populations.

Materials and Methods

Nunavik Inuit. One hundred seventy Inuit individuals from Nunavik (Québec, Canada) were included in this study; 155 were recruited from 10 villages, and the remaining 15 had unknown village origins (SI Appendix, Fig. S20). Written informed consent was obtained from all participants, and the study was approved by the McGill University Ethics Committee and the Nunavik Nutrition and Health Ethics Committee.

Illumina HumanOmniExpress-12 and -24 were used to genotype 165 Nunavik Inuit, each with 730,525 and 716,503 SNPs, respectively. WES was performed on 114 selected genotyped Nunavik Inuit. Illumina GenomeStudio was used for the SNP-chip genotype calling. An in-house pipeline used for variant generation from the exome sequencing data was described in a previous study (21).

Other Population Panels and Data Handling. We obtained SNP array-based data from the publicly available populations from the Siberians and the Native Americans. Whole-genome sequencing (WGS) data were also obtained from the 1KGP phase III and ancient genomes from Siberia and North America. A total of 5,252 individuals from 187 populations were included in this study (SI Appendix, Table S1). The merging and quality control of data are described in SI Appendix.

PCA and Admixture. PCA implemented in Eigensoft 6.1 (38) was performed on (i) a total of 5,422 individuals from the aforementioned structure panel and (ii) 3,130 individuals selected from Asia-American (including Siberian, Middle Asian, East European, and Native American populations located along the migration path of the ancestors of today's Nunavik Inuit).

Ancestry analysis was also performed by using ADMIXTURE software (11). $K = 8$ was used for ancestry estimation for each dataset based on pruned data used for PCA analysis. An additional PCA and admixture analysis ($K = 6$) were also performed by using the WES data of Nunavik Inuit and WGS controls to minimize population bias induced by the European population-based SNP-chip.

PCAdmix (12) was used for the local ancestry estimation for individual genomes selected from highly admixed Inuit individuals and each of the populations with high-level admixture of the Inuit ancestry. We inferred three ancestral populations (ancient American–Saqqaq, European, and Inuit) based on the ADMIXTURE results when performing PCAdmix analysis on highly admixed Inuit individuals. Another 11 indigenous populations were also selected for comparison with Inuit. The results were illustrated on the chromosomal ideographs. Distribution of continuous ancestry tract lengths for selected admixed Siberian individuals were also calculated (13). It is assumed that historical demography has an effect on LD in isolated populations such as Nunavik Inuit. We calculated the r^2 from SNP-chip data of 14 indigenous and other present-day populations including Nunavik Inuit to compare the patterns of their LD decay. Each population contains 20–50 individuals to minimize inflation. ROHs were used in the estimation of individual genomewide autozygosity of Nunavik Inuit and other indigenous populations. To further characterize population genetic differentiation, a pairwise F_{ST} for selective populations relevant to Nunavik Inuit was also calculated and plotted by using the Arlequin v3.5 package (39).

TreeMix. We used TreeMix (16) to infer population splits and mixtures of Nunavik Inuit and other populations by using genotype data. The algorithm used Gaussian approximation to estimate genetic drift between populations. We focused on different groups of Asian–American populations in the constructing of trees. TreeMix was run in windows of 500 SNPs to account for LD, with migration events added from 0 to 10 and variance in relatedness between populations calculated for each model. F_3 statistics were also calculated for each pair of Asian–American populations with CHB as outgroup.

D -statistics (or ABBA-BABA statistics) were calculated for different topologies of four populations to estimate the direction of gene flow between different Inuit populations (Nunavik Inuit and Greenlandic Inuit and within Nunavik Inuit) and worldwide populations with the outgroups of CHB and Yoruba, respectively. AdmixTools (17) and the R package admixturegraph (40) were used to calculate and present the D -statistic result.

Demographic Inference of Nunavik Inuit. To estimate the demographic history of Nunavik Inuit, we first applied the sequentially Markov coalescent-based algorithm SMC++ (v1.11) (18) to calculate the N_e at different time points of Nunavik Inuit and publicly available Siberian and circumpolar populations with WGS data. The inferred N_e at corresponding time points from SMC++ were then used in the joint sample frequency spectrum-based algorithm momi2 (19) to further infer the demography of Nunavik Inuit by using the truncated Newton conjugate method. The sequences overlapping the highly conserved (HC) region (SI Appendix) and exome targeted regions were used, and the four most distantly related Nunavik Inuit individuals were selected. The N_e of the initial model, generation time, and mutation rates of exome regions used in both models were 1.2×10^4 , 29, and 1.45×10^{-8} , respectively, and the ancestral allele was set to unknown. To construct the pulse model including Saqqaq, 20 random initial values were chosen for the model and the best result (log likelihood) was selected. Fifty bootstraps of resampling the data were also used to find the best pulse model.

Detection of Selection by PBS. Populationwide site variant frequencies were calculated from the Nunavik Inuit, CHB, and CEU exome HC variant dataset (SI Appendix, Table S1). The calculation of PBS is described in SI Appendix. The results were shown in a Manhattan plot, and all variants with $PBS > 0.1$ were plotted against concatenated chromosomal locations. The exome variants with $PBS > 0.1$ are shown in red (Fig. 3). We also calculated the PBS values between Nunavik Inuit and Northeastern Siberians, with CHB as outgroup to identify more recent selections.

Expression Analyses of the Genes Under Natural Selection. Genes under potential natural selection were chosen for RT-qPCR analysis to detect whether their levels of expression are significantly changed among the Nunavik Inuit samples. EBV-transformed LCLs from 14 randomly selected unrelated Nunavik Inuit and 14 unrelated FC individuals as controls were used. Relative quantification (41) for each gene was calculated for each individual and were compared by nonparametric Mann–Whitney U test.

Genetic Risk of IA in Nunavik Inuit. To test the risk of IA in Nunavik Inuit-specific genetic loci, we performed association tests on 8,291 coding region enriched variants identified with $PBS > 0.1$ in 42 unadmixed Nunavik IA cases and 62 unadmixed Nunavik controls by using a linear mixed model that accounted for individual relatedness (42). The diagnosis of IA in Inuit was described in our previous study (43). After Bonferroni correction, genomewide significance was set to $P < 6.03 \times 10^{-6}$.

ACKNOWLEDGMENTS. We thank the Nunavik Inuit participants, the Nunavik communities, and the clinicians, as well as Le Module du Nord Québécois and the Nunavik Nutrition and Health Committee, for their support and contributions to this research. We also thank Dr. Jonathan Terhorst (University of Michigan) for help with SMC++; and Jay Ross and Dr. Vince Forgetta (McGill University) for their efforts in editing this manuscript. G.A.R. holds a Canada Research Chair in Genetics of the Nervous System and the Wilder Penfield Chair in Neurosciences.

- Reich D, et al. (2012) Reconstructing Native American population history. *Nature* 488:370–374.
- Raghavan M, et al. (2014) The genetic prehistory of the New World Arctic. *Science* 345:1255832.
- Anand SS, et al. (2001) Risk factors, atherosclerosis, and cardiovascular disease among aboriginal people in Canada: The Study of Health Assessment and Risk Evaluation in Aboriginal Peoples (SHARE-AP). *Lancet* 358:1147–1153.
- Noël M, et al. (2012) Cardiovascular risk factors and subclinical atherosclerosis among Nunavik Inuit. *Atherosclerosis* 221:558–564.
- Cardona A, et al. (2014) Genome-wide analysis of cold adaptation in indigenous Siberian populations. *PLoS One* 9:e98076.
- Clemente FJ, et al. (2014) A selective sweep on a deleterious mutation in CPT1A in Arctic populations. *Am J Hum Genet* 95:584–589.
- Moltke I, et al. (2015) Uncovering the genetic history of the present-day Greenlandic population. *Am J Hum Genet* 96:54–69.
- Fumagalli M, et al. (2015) Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* 349:1343–1347.
- Moltke I, et al. (2014) A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature* 512:190–193.
- Rasmussen M, et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757–762.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664.
- Brisbin A, et al. (2012) PCAmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum Biol* 84:343–364.
- Gravel S (2012) Population genetics models of local ancestry. *Genetics* 191:607–619.
- Rasmussen M, et al. (2014) The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506:225–229.
- Rasmussen M, et al. (2015) The ancestry and affiliations of Kennewick Man. *Nature* 523:455–458.
- Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8:e1002967.
- Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192:1065–1093.
- Terhorst J, Kamm JA, Song YS (2017) Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet* 49:303–309.
- Kamm JA, Terhorst J, Durbin R, Song YS (2018) Efficiently inferring the demographic history of many populations with allele count data. bioRxiv:10.1101/287268. Preprint, posted March 23, 2018.
- Yi X, et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329:75–78.
- Zhou S, et al. (2015) Increased missense mutation burden of Fatty Acid metabolism related genes in nunavik inuit population. *PLoS One* 10:e0128255.
- Raghavan M, et al. (2015) POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349:aab3884.
- Raghavan M, et al. (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505:87–91.
- Duforet-Frebourg N, Luu K, Laval G, Bazin E, Blum MG (2016) Detecting genomic signatures of natural selection with principal component analysis: Application to the 1000 genomes data. *Mol Biol Evol* 33:1082–1093.
- Helgason A, et al. (2006) mtDNA variation in Inuit populations of Greenland and Canada: Migration history and population structure. *Am J Phys Anthropol* 130:123–134.
- Collins HB (1956) Vanished mystery men of Hudson-Bay. *Natl Geogr Mag* 110:669–687.
- Mary-Rousselière G (1980) *Qitlarssuaq, l'Histoire d'une Migration Polaire* (Presses de l'Université de Montreal, Montreal).
- Perestenko PV, et al. (2010) Copines-1, -2, -3, -6 and -7 show different calcium-dependent intracellular membrane translocation and targeting. *FEBS J* 277:5174–5189.
- Tan B, et al. (January 3, 2018) Low CPNE3 expression is associated with risk of acute myocardial infarction: A feasible genetic marker of acute myocardial infarction in patients with stable coronary artery disease. *Cardiol J*, 10.5603/CJ.a2017.0155.
- Wang KS, Zuo L, Pan Y, Xie C, Luo X (2015) Genetic variants in the CPNE5 gene are associated with alcohol dependence and obesity in Caucasian populations. *J Psychiatr Res* 71:1–7.
- Ley K, Miller YI, Hedrick CC (2011) Monocyte and macrophage dynamics during atherosclerosis. *Arterioscler Thromb Vasc Biol* 31:1506–1516.
- Fan XJ, et al. (2015) Role of inflammatory responses in the pathogenesis of human cerebral aneurysm. *Genet Mol Res* 14:9062–9070.
- Fukami K, Yamagishi S, Okuda S (2014) Role of AGEs-RAGE system in cardiovascular disease. *Curr Pharm Des* 20:2395–2402.
- Lagor WR, et al. (2014) Genetic manipulation of the ApoF/Stat2 locus supports an important role for type I interferon signaling in atherosclerosis. *Atherosclerosis* 233:234–241.
- Deng Y, et al. (2013) Endothelial RAF1/ERK activation regulates arterial morphogenesis. *Blood* 121:3988–3996, S1–S9.
- Timofeeva AV, et al. (2006) Altered gene expression pattern in peripheral blood leukocytes from patients with arterial hypertension. *Ann N Y Acad Sci* 1091:319–335.
- Synowiec E, et al. (2016) Expression and variability of lipid metabolism genes in intracranial aneurysm. *Cell Mol Biol* 62:73–82.
- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2:e190.
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567.
- Leppälä K, Nielsen SV, Mailund T (2017) admixturegraph: An R package for admixture graph manipulation and fitting. *Bioinformatics* 33:1738–1740.
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta(\Delta C_T)}$ method. *Methods* 25:402–408.
- Eu-Ahsunthornwattana J, et al.; Wellcome Trust Case Control Consortium 2 (2014) Comparison of methods to account for relatedness in genome-wide association studies with family-based data. *PLoS Genet* 10:e1004445.
- Zhou S, et al. (2016) RNF213 is associated with intracranial aneurysms in the French-Canadian population. *Am J Hum Genet* 99:1072–1085.