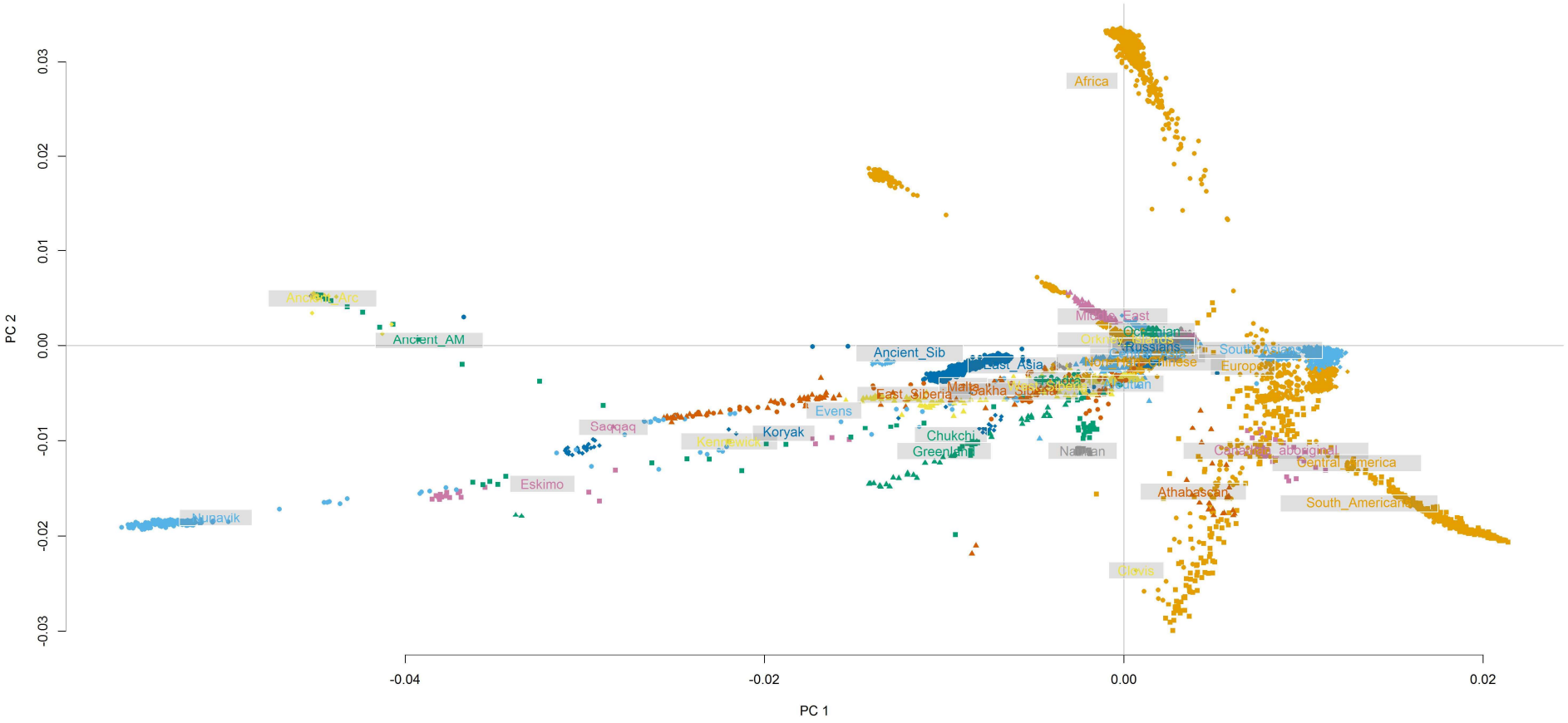


Figure S1: Principle component analysis of Nunavik Inuit with present day worldwide populations. Including 5,422 individuals from 197 sub-populations. Nunavik Inuit are displayed in blue at the left bottom.

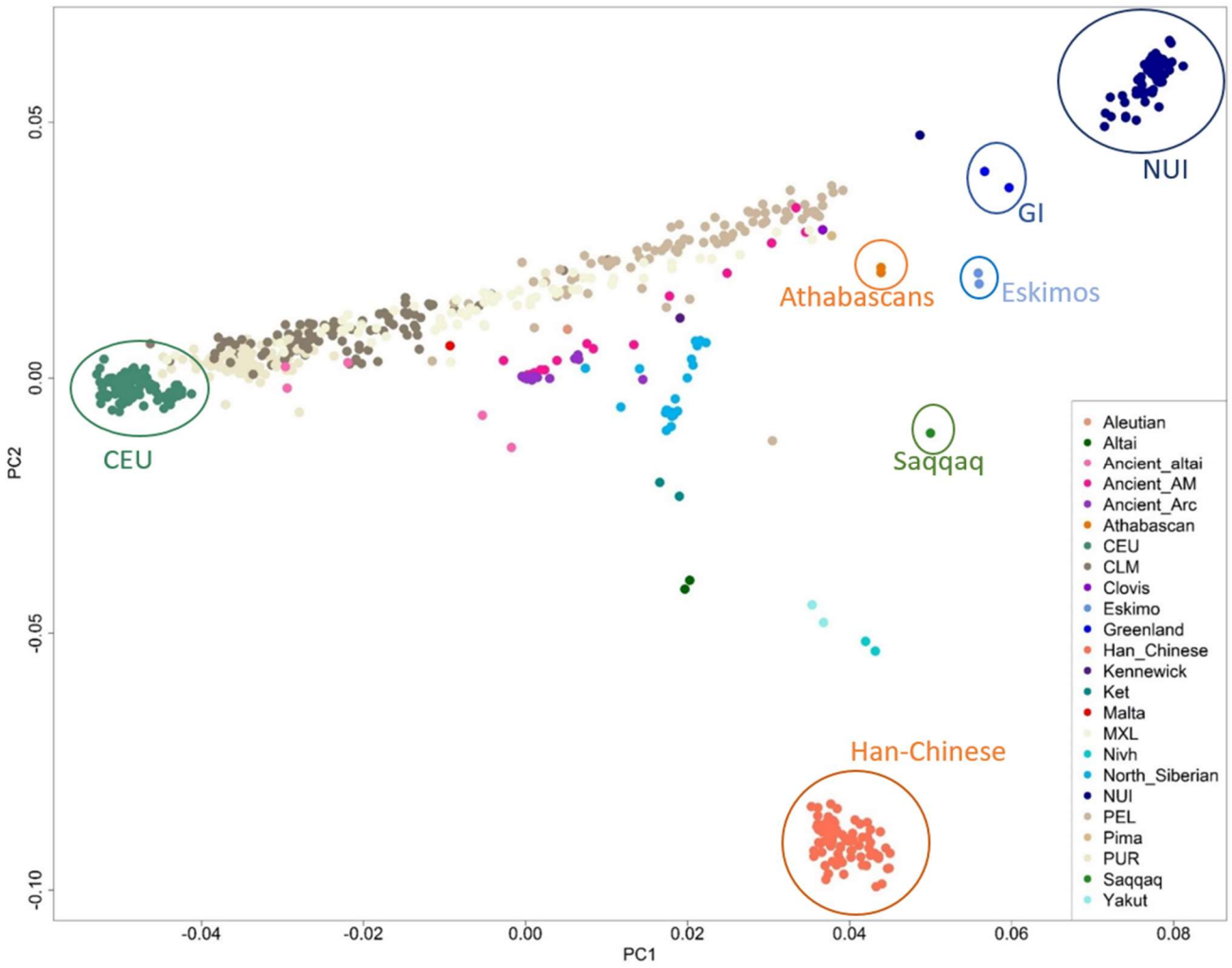


Figure S2: Principle component analysis using high-confidence (HC) exonic regions of Nunavik Inuit, present-day indigenous and ancient populations. 146,668 SNPs from exome HC regions of the Nunavik Inuit (dark blue cluster) and other 649 individuals were used to calculate the PCs, including 4 Native American populations from 1000 genome project, indigenous populations and ancient individuals described in **Table S1.**

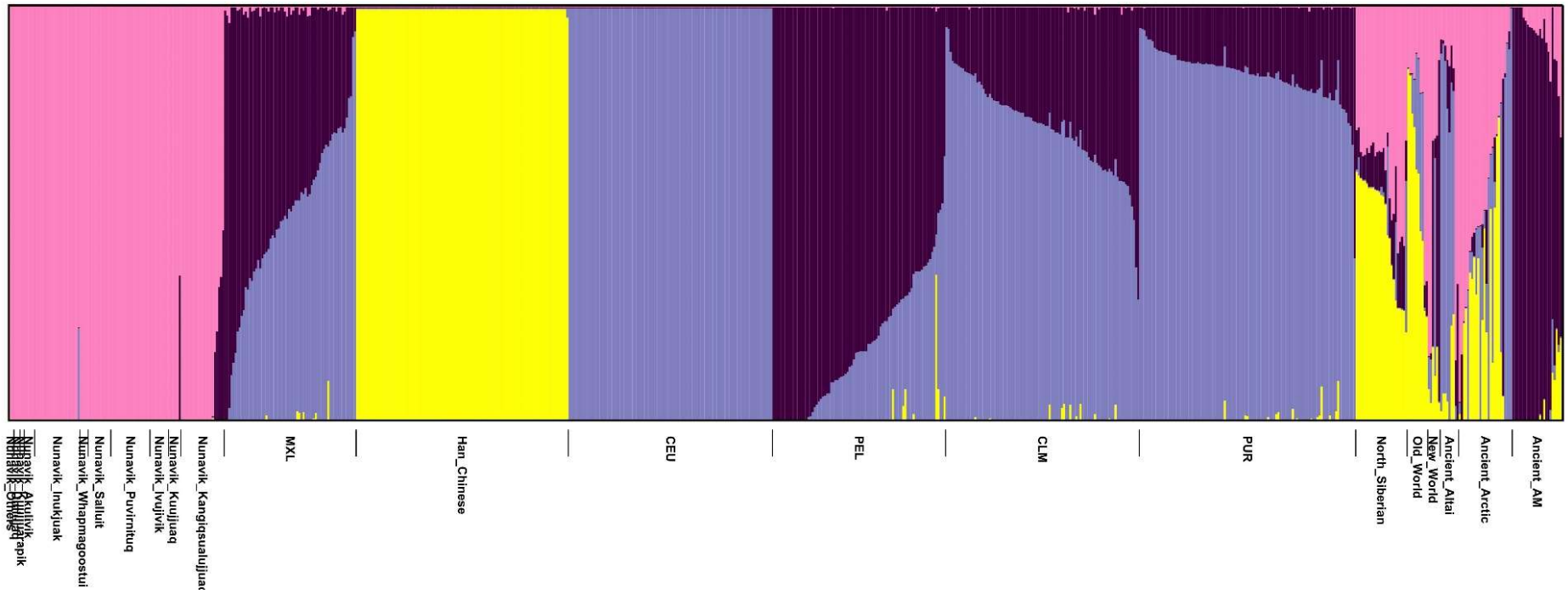


Figure S3. Admixture of Nunavik Inuit with present-day and ancient populations. Using whole exome data. Assuming the scenario of four estimated ancestries. Include Nunavik Inuit (split by villages) with 1KGP Native American populations and ancient populations (K=4, Han, Native-American, European and Inuit).

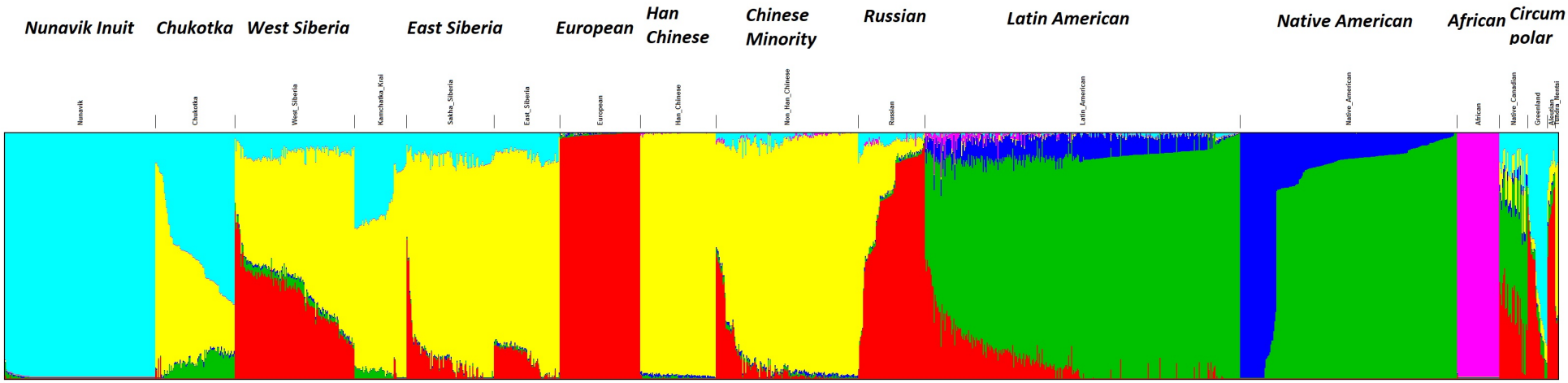


Figure S4. Admixture of Nunavik Inuit without recent European admixture and present-day worldwide populations. Using genotype data. Assuming K=6 ancestral populations. Colors representing admixed proportions from each ancestral component (Inuit=Light Blue; Han Chinese=Yellow; European=Red; African=Purple; First Native American=Green, Second Native American=Dark Blue)

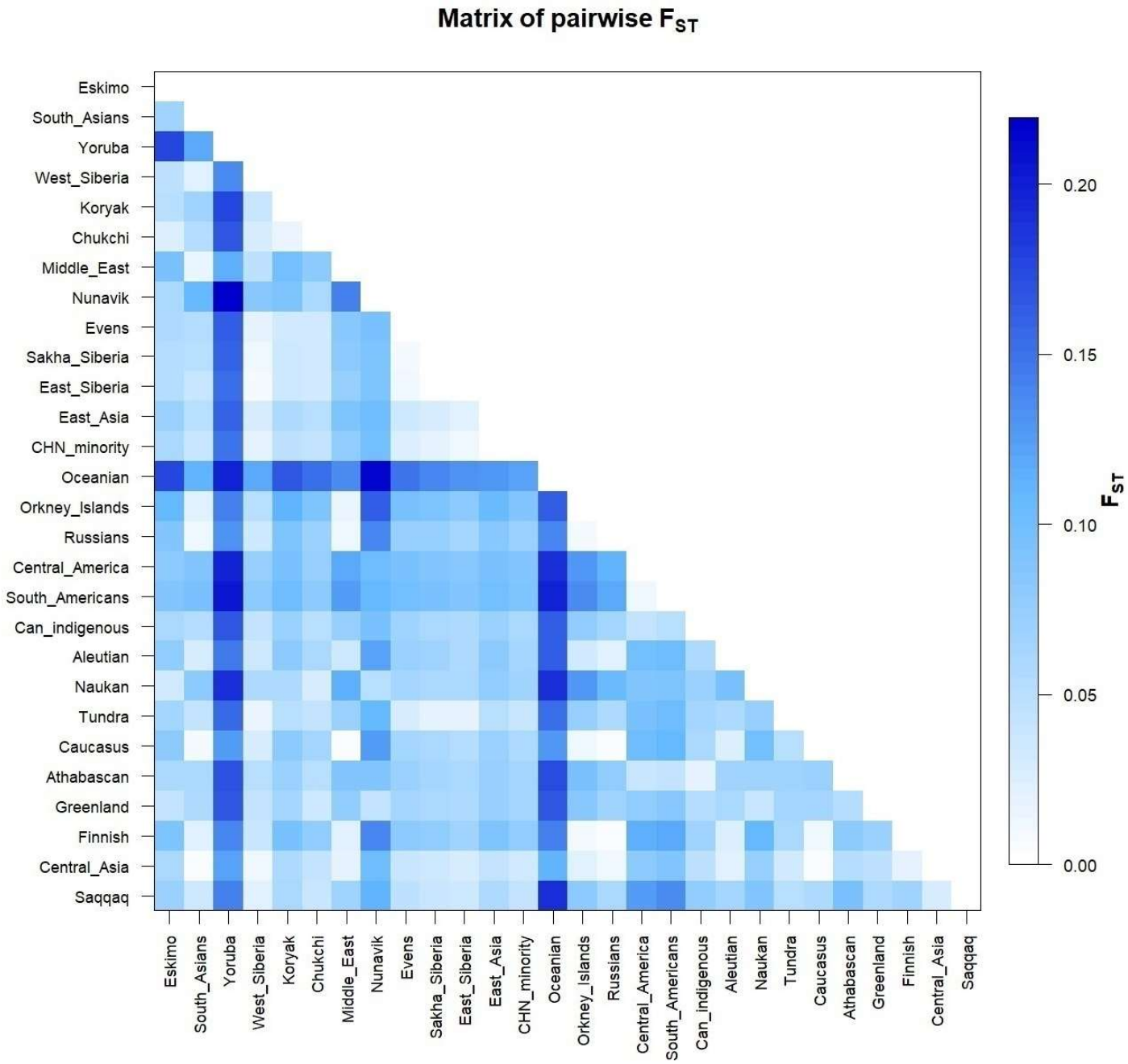
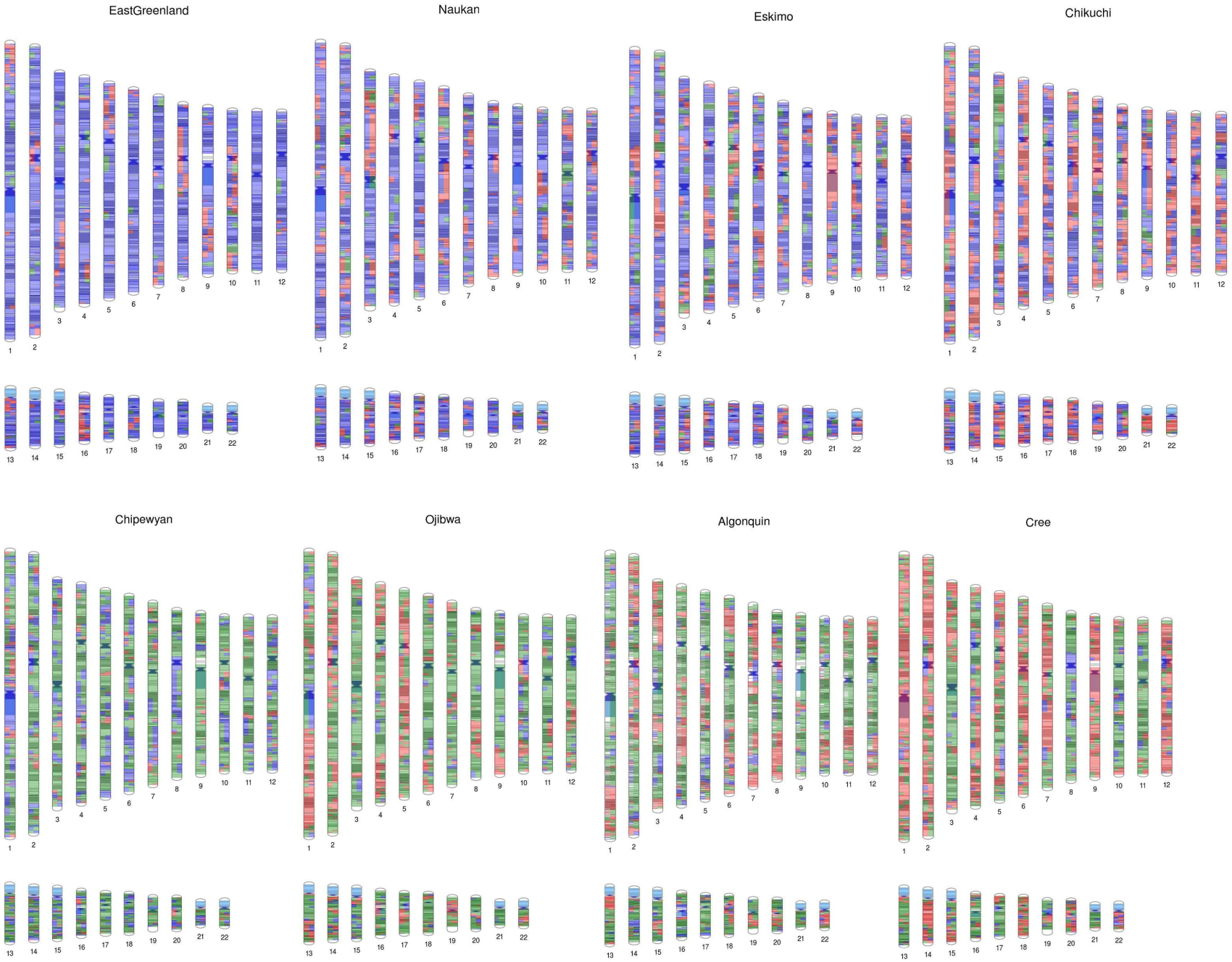


Figure S5. Pairwise FST of Nunavik Inuit and Asian-New World populations. Populations that were genetically more distant from Nunavik Inuit (estimated from ADMIXTURE) were displayed in regional groups. CHN_minority: Minority groups from China; Can_indigenous: indigenous people from Canada.


EastGreenland
Naukan
Eskimo
Chikuchi
Chipewyan
Ojibwa
Algonquin
Cree

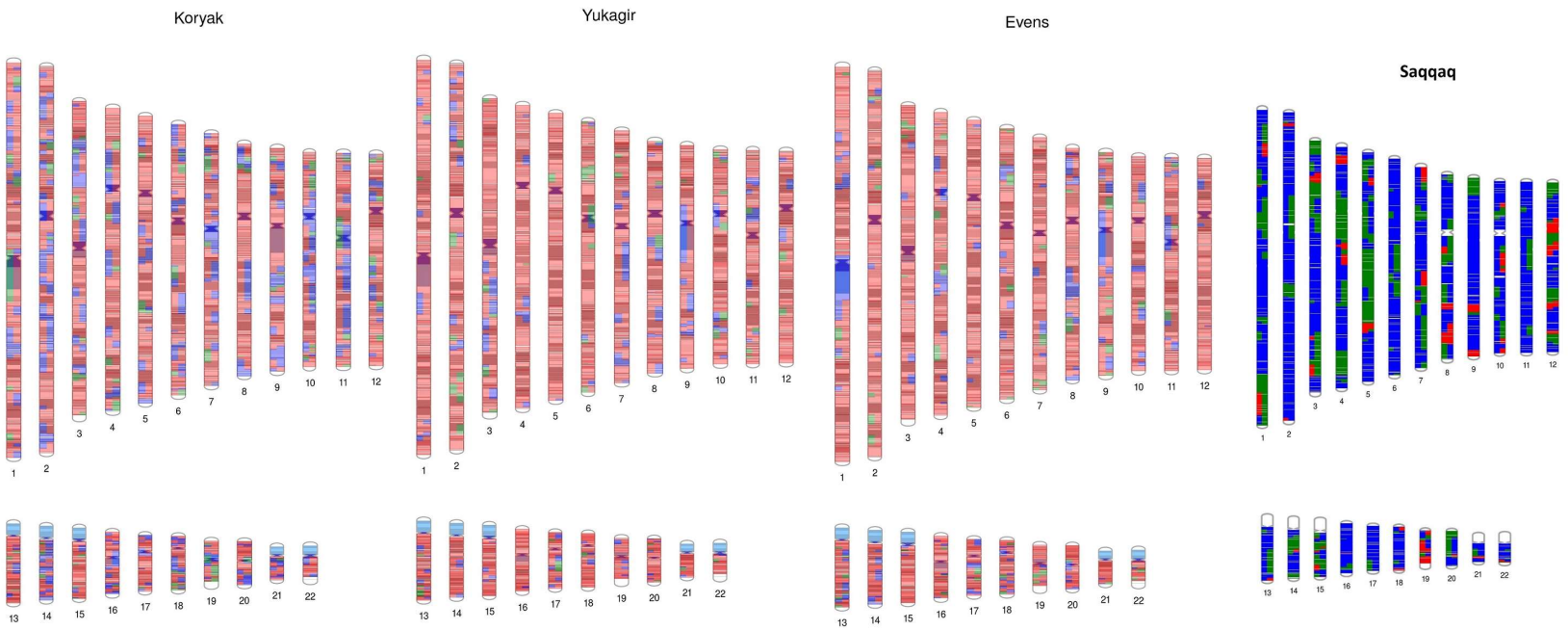


Figure S6: Genome-wide local ancestry of Arctic indigenous populations. For the Siberian populations (Naukan, Chikuchi, Siberian Eskimo, Yukagir, Koryak and Evens), the ancestries were inferred from Nunavik Inuit (NUI, blue), Han Chinese (CHB, red) and Native American (NAT, green); for the Canadian-Arctic populations (Chipewyan, Cree, Ojibwa, Algonquin and Greenland Inuit), the ancestries were inferred from Nunavik Inuit (NUI, blue), European (CEU, red) and Native American (NAT, green). Greenland Inuit shared their most ancestry with Nunavik Inuit (83.4%), followed by Naukan (72.7%), Siberian Eskimo (64.4%), Chikuchi (46.8%), Chipewyan (26.3%), Koryak (24.3%), Algonquin (12.3%), Evens (11.7%), Yukagir (9.75%), Cree (9.5%) and Ojibwa (9.1%). Saqqaq was inferred from Nunavik Inuit (blue), Native American (green) and Han Chinese (red).

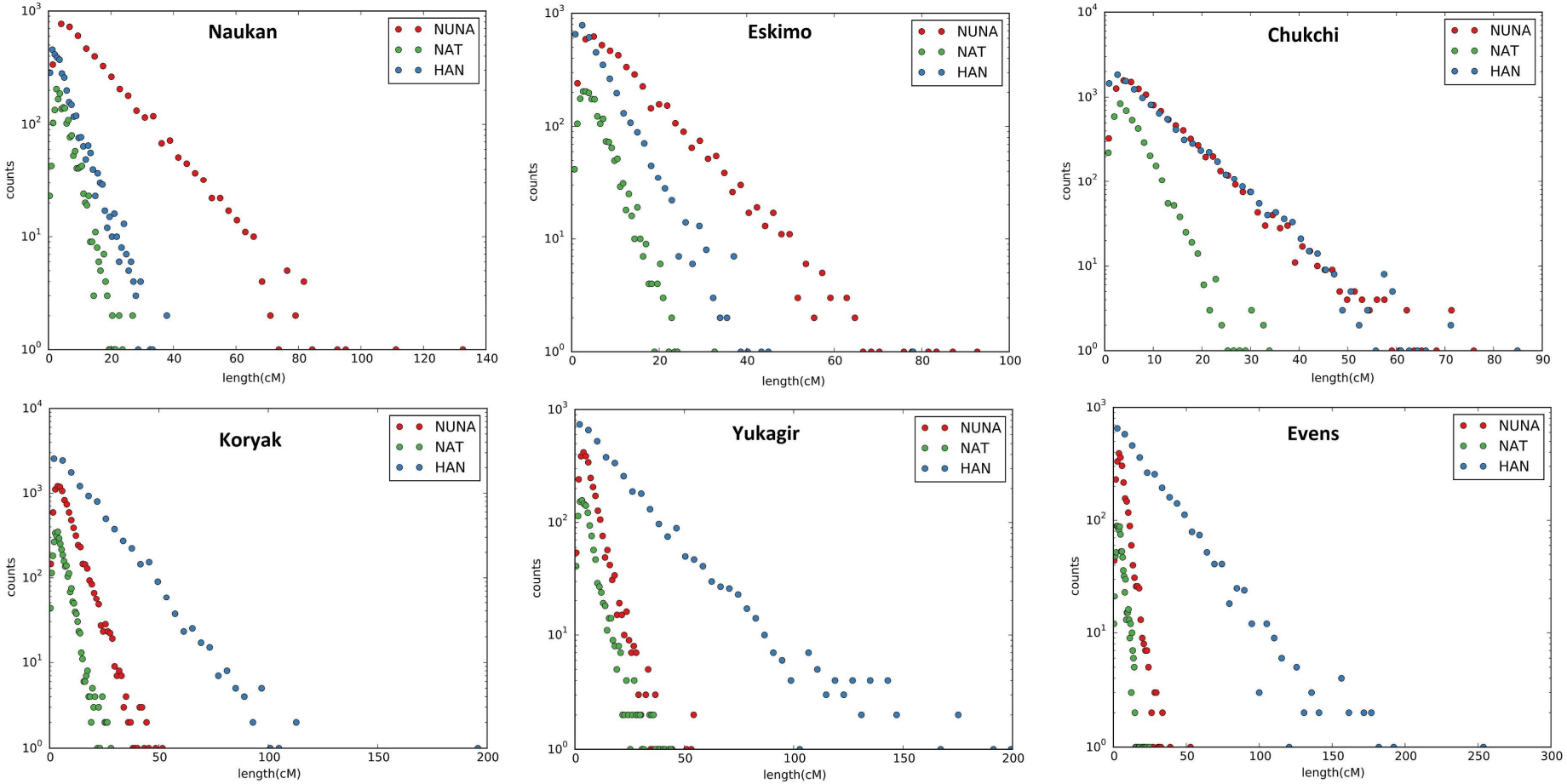


Figure S7. Tracts decay of six Siberian populations. Distribution of continuous ancestry tract lengths of three ancestral components inferred from Nunavik Inuit (NUNA), Han Chinese (HAN) and Native Americans (NAT) in six Siberian individuals representing their respective populations.

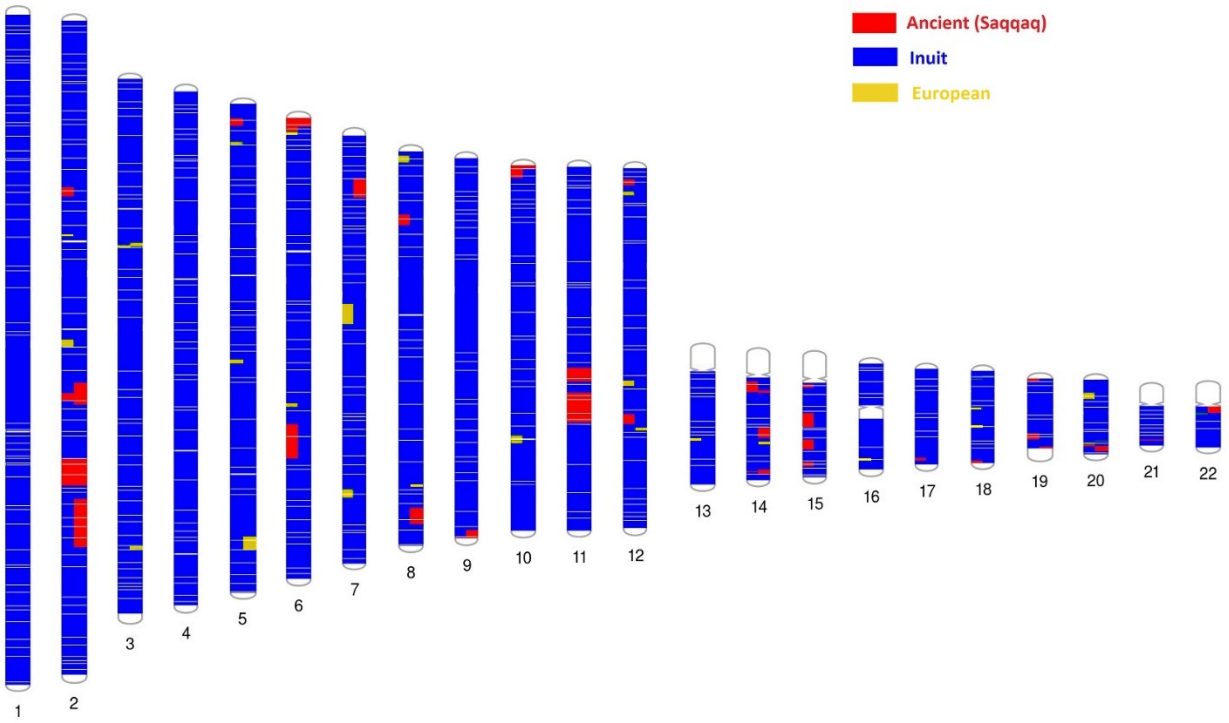


Figure S8. Local ancestry inference of an Inuit individual from the village Kangiqsualujjuaq. Inferred by PCAdmix (3 ancestral components), portion of the genome inferred as Paleo-Eskimo ancestry were depicted in red (3.8%).

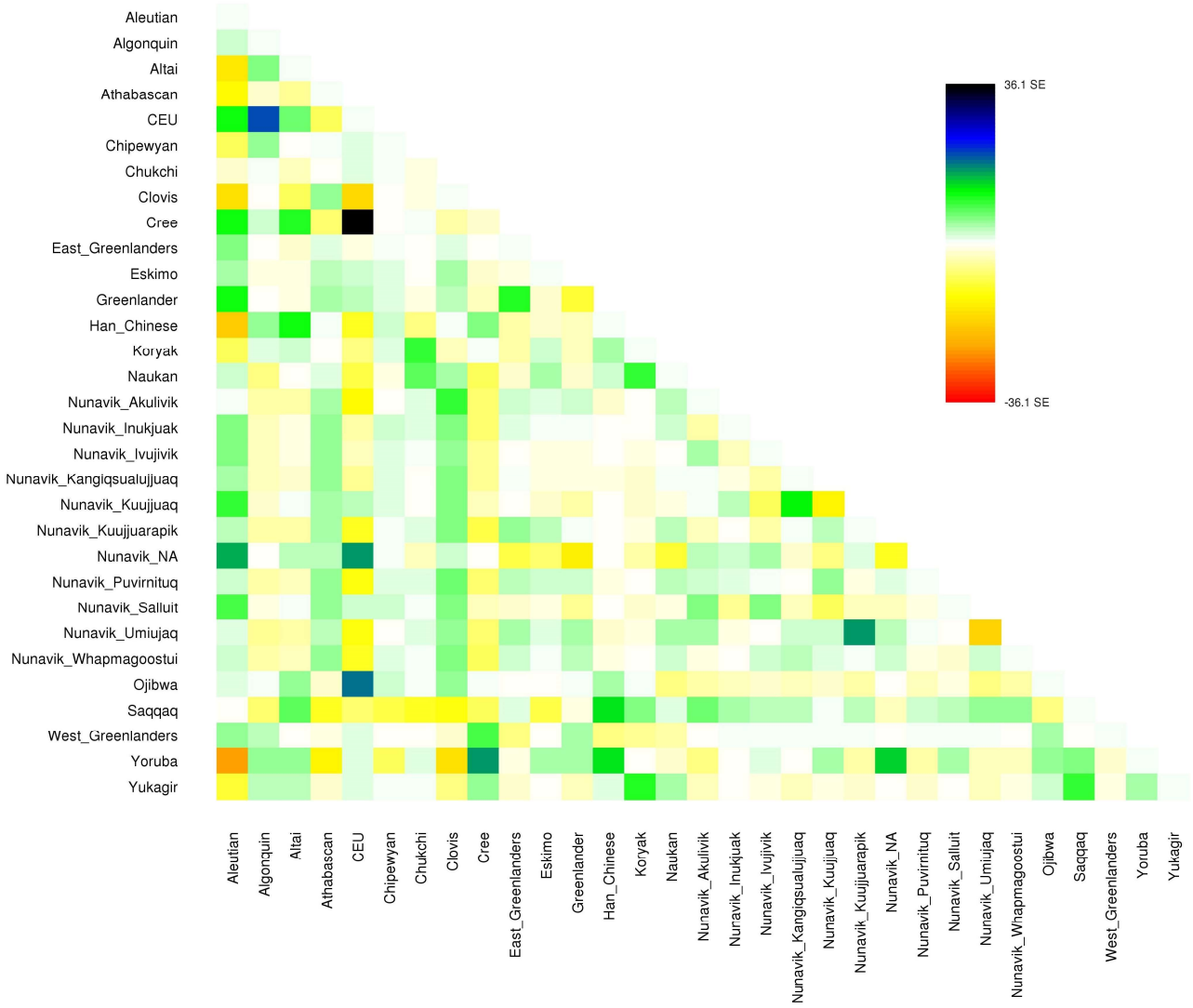


Figure S9. Residual matrix of treemix populations that fit the tree. Tree depicted in Fig 1D with three migration events.

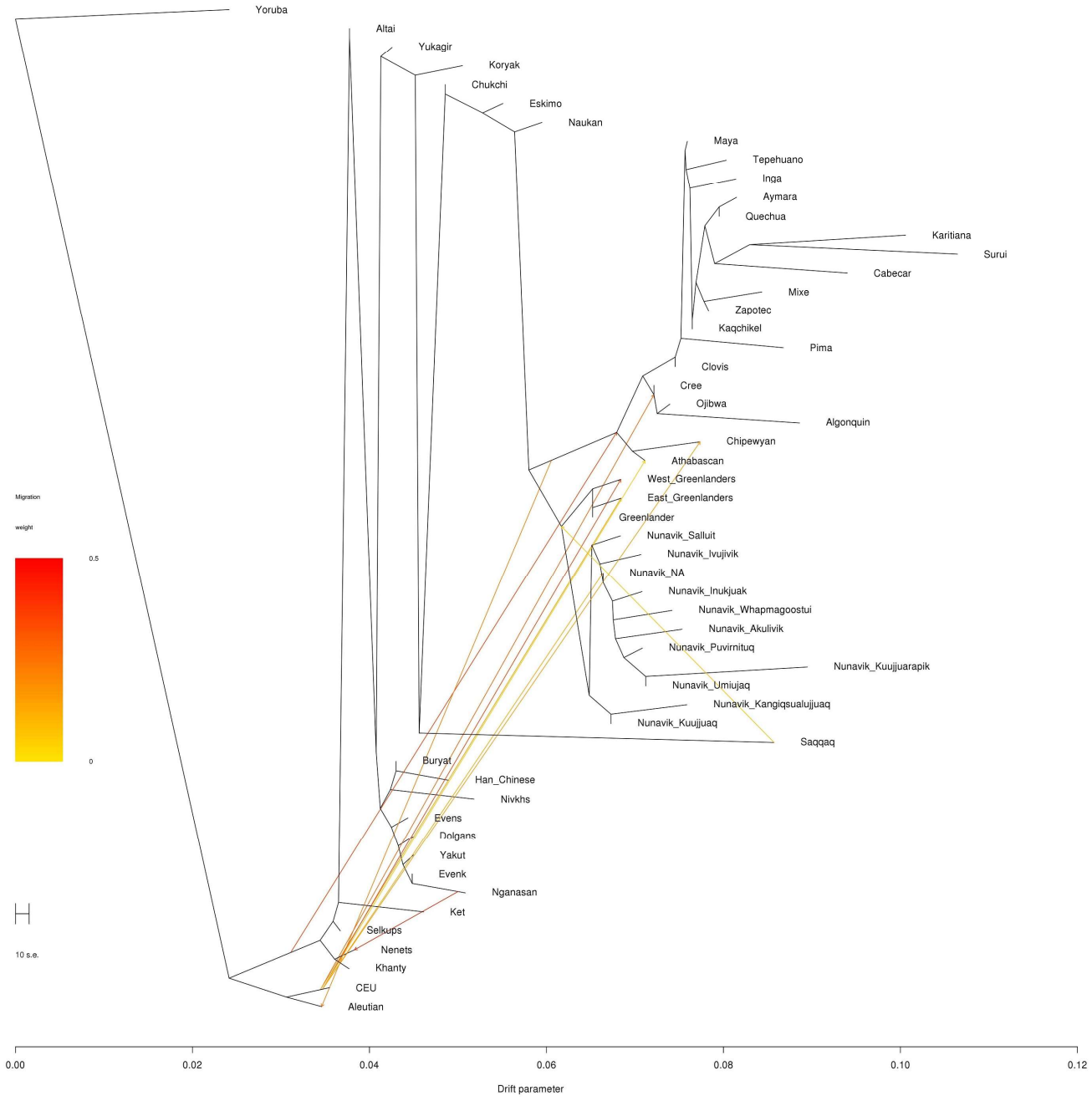


Figure S10. Treemix of Nunavik Inuit with Asian-New World Populations. Using window of SNP=500 for SNP pruning, display 10 migration events. Variance of relatedness explained by this model is estimated to be 97.2%.

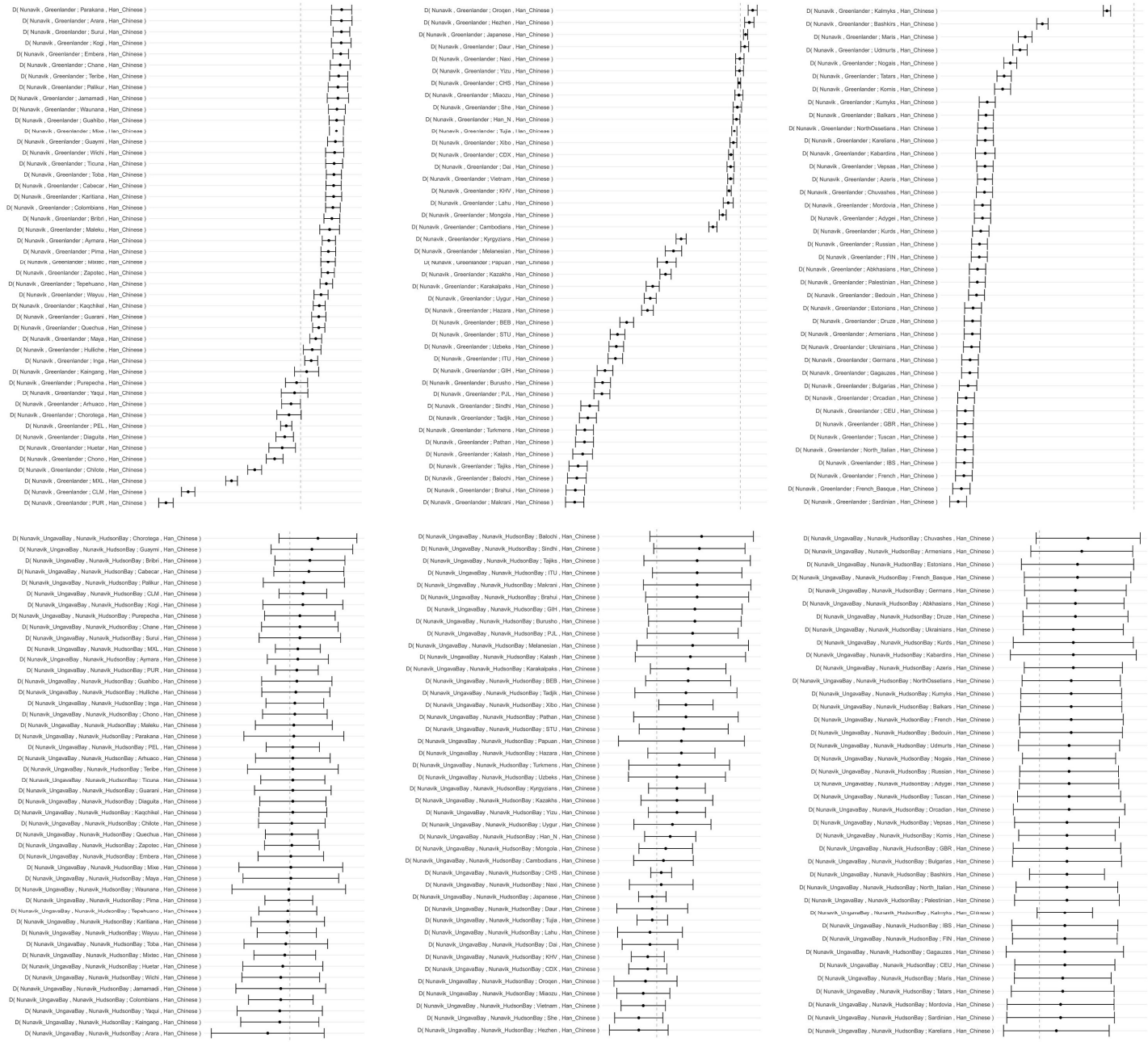


Figure S11. D statistics of Nunavik Inuit and close-by populations

D statistics in the form D (Nunavik Inuit, Greenlandic Inuit; X, Han-Chinese) and D (Ungava Inuit, Hudson Inuit; X, Han-Chinese), where X were Native American populations, Asian and European populations, respectively. Dashed line indicated D=0, error bars indicated confidence interval calculated from D and |Z|. D<0 suggested Nunavik Inuit (or Ungava Inuit) were closer to population X, D>0 suggested Greenlandic Inuit (or Hudson Inuit) were closer to population X.

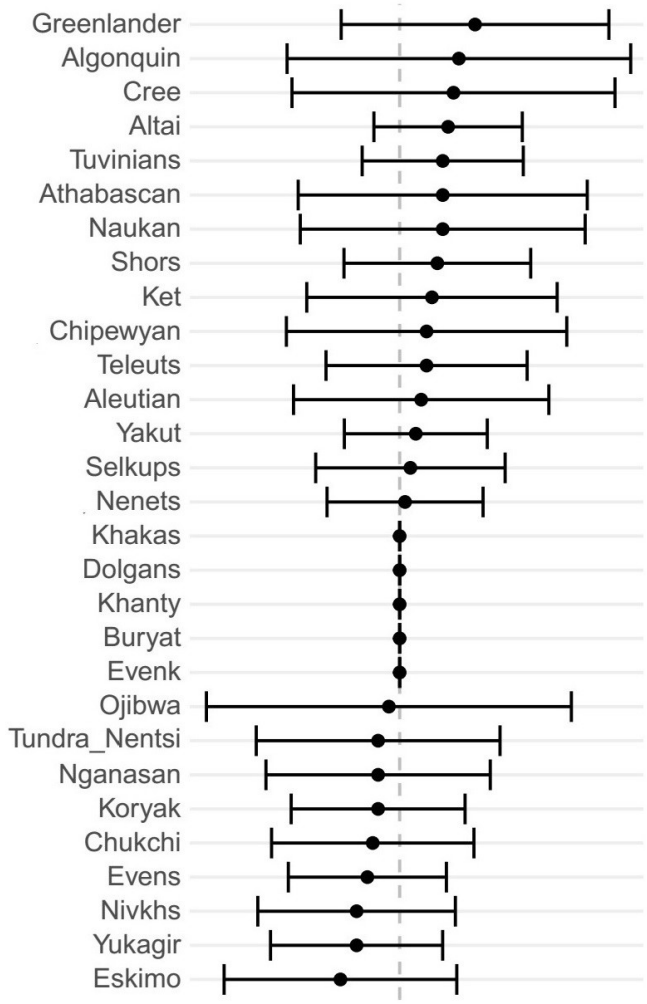


Figure S12. D-statistics of groups within Inuit and arctic indigenous populations. D-statistics in the form of D (Ungava Bay, Hudson Bay; X, Han Chinese). X were the Siberian-arctic indigenous populations. Han Chinese was outgroup. Dashed line represented D=0. Error bars represent the confidence intervals estimated from Z values. D<0 suggested Ungava Inuit are closer to population X, D>0 suggested Hudson Inuit are closer to population X.

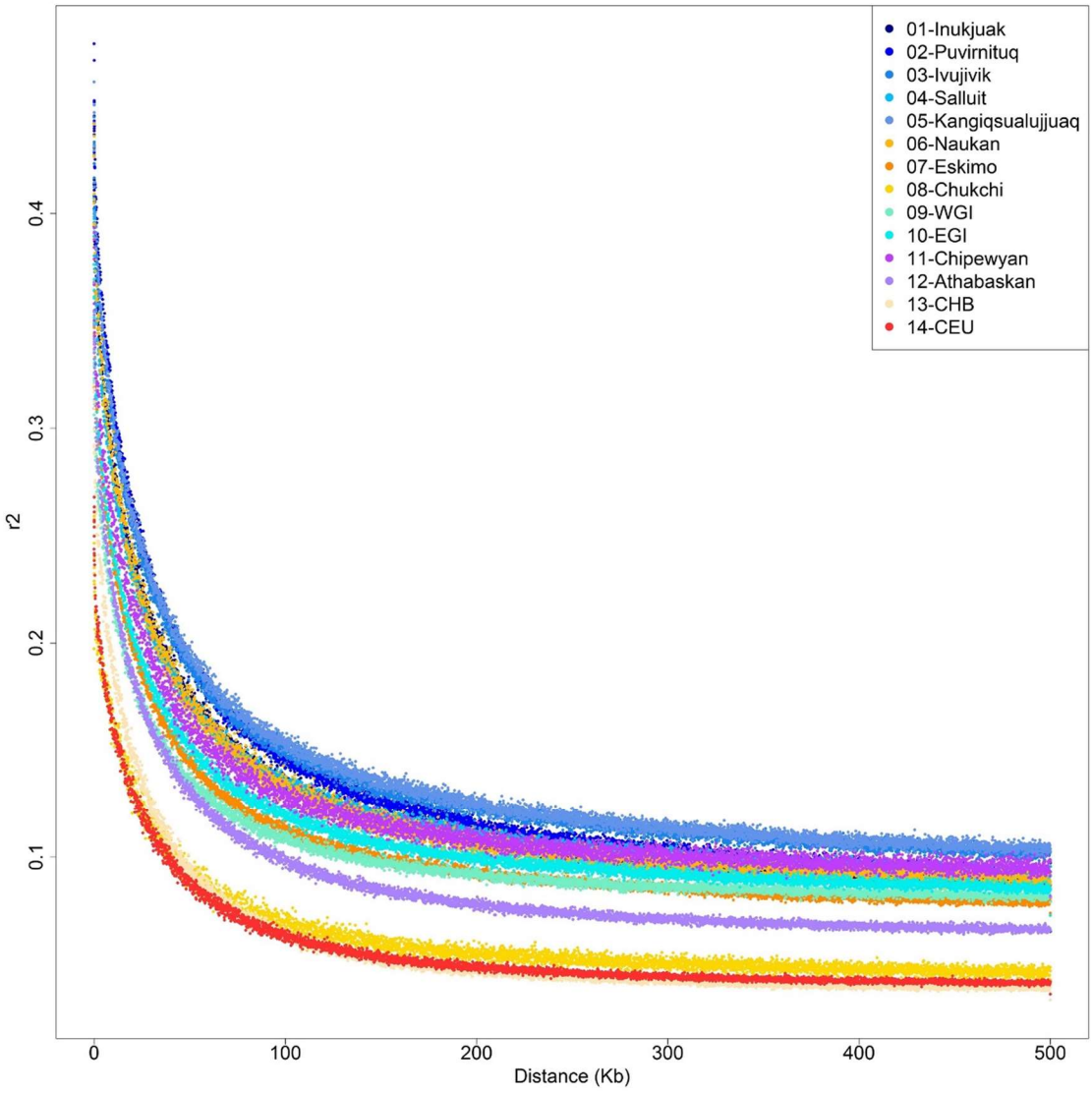


Figure S13. LD decay for different populations. Only populations which have more than 15 individuals were included for accuracy. Nunavik villages: 01-05, Northeastern Siberians: 06-08, Greenlandic Inuit: 09-10, Arctic indigenous peoples: 11-12 and 1KGP reference populations: 13-14. WGI: Western Greenlandic Inuit; EGI: Eastern Greenlandic Inuit; CHB: Han Chinese from Beijing.

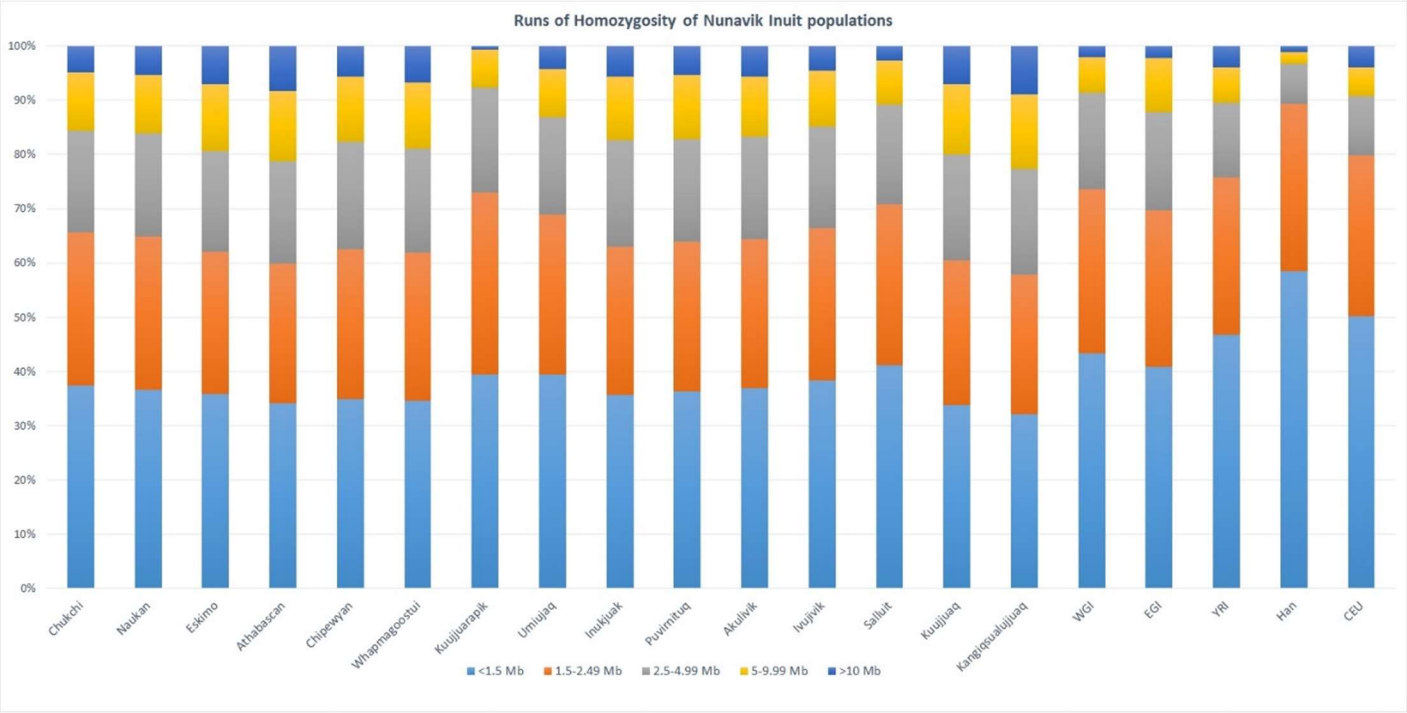


Figure S14. Estimated Runs of Homozygosity (ROH) lengths of different populations.

Including Nunavik Inuit from different villages, Arctic indigenous peoples and 1KGP populations. ROH in windows of different sizes were shown in proportion in each population.

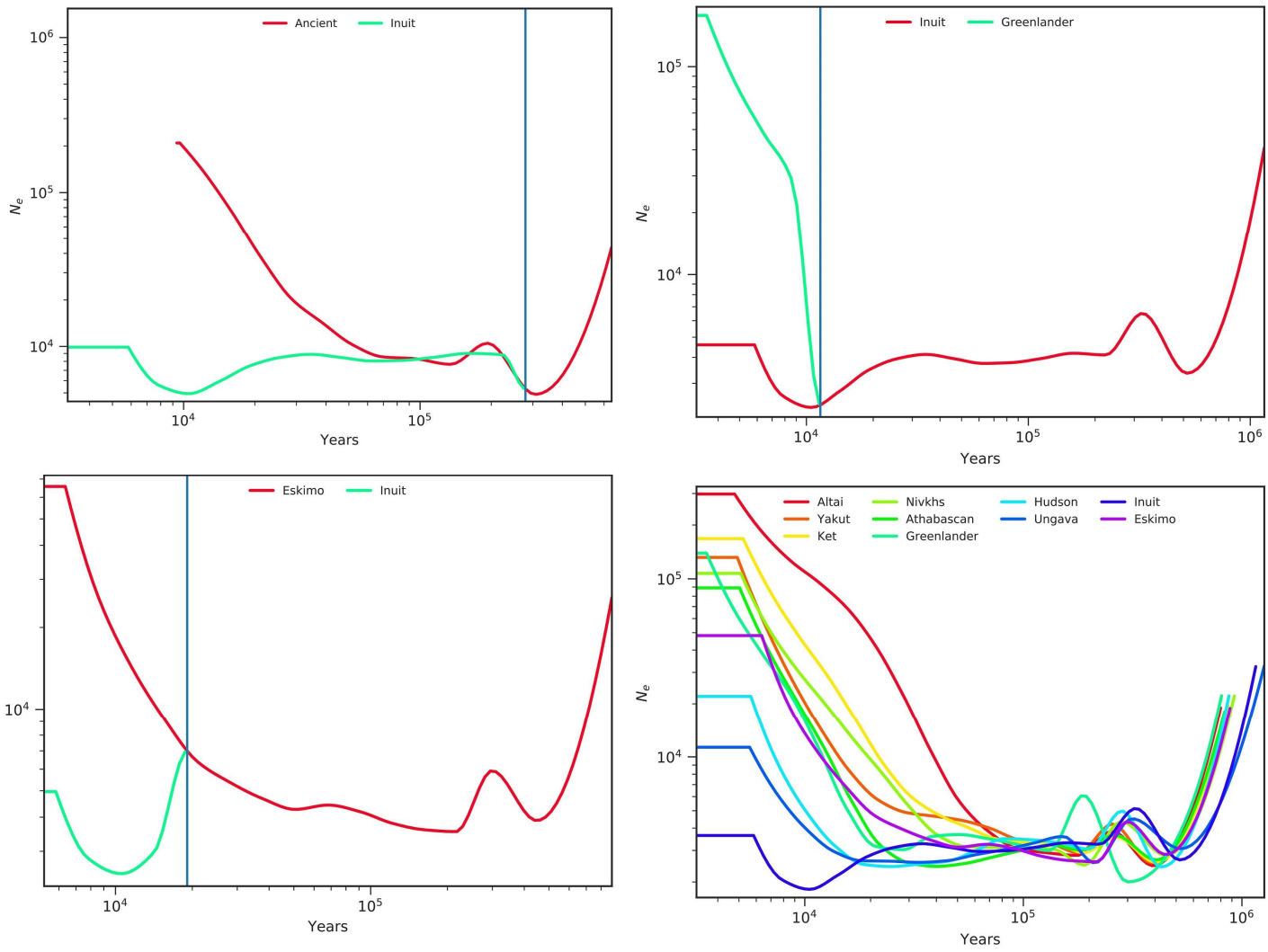


Figure S15. SMC++ estimated split time of Nunavik Inuit

Split time between Nunavik Inuit and Ancient Saqqaq, Siberian Eskimo and Greenlandic Inuit, depicting effective population sizes in respect to times. Demographic inference are also performed in Inuit from Hudson Bay and Ungava Bay along with other arctic indigenous populations.

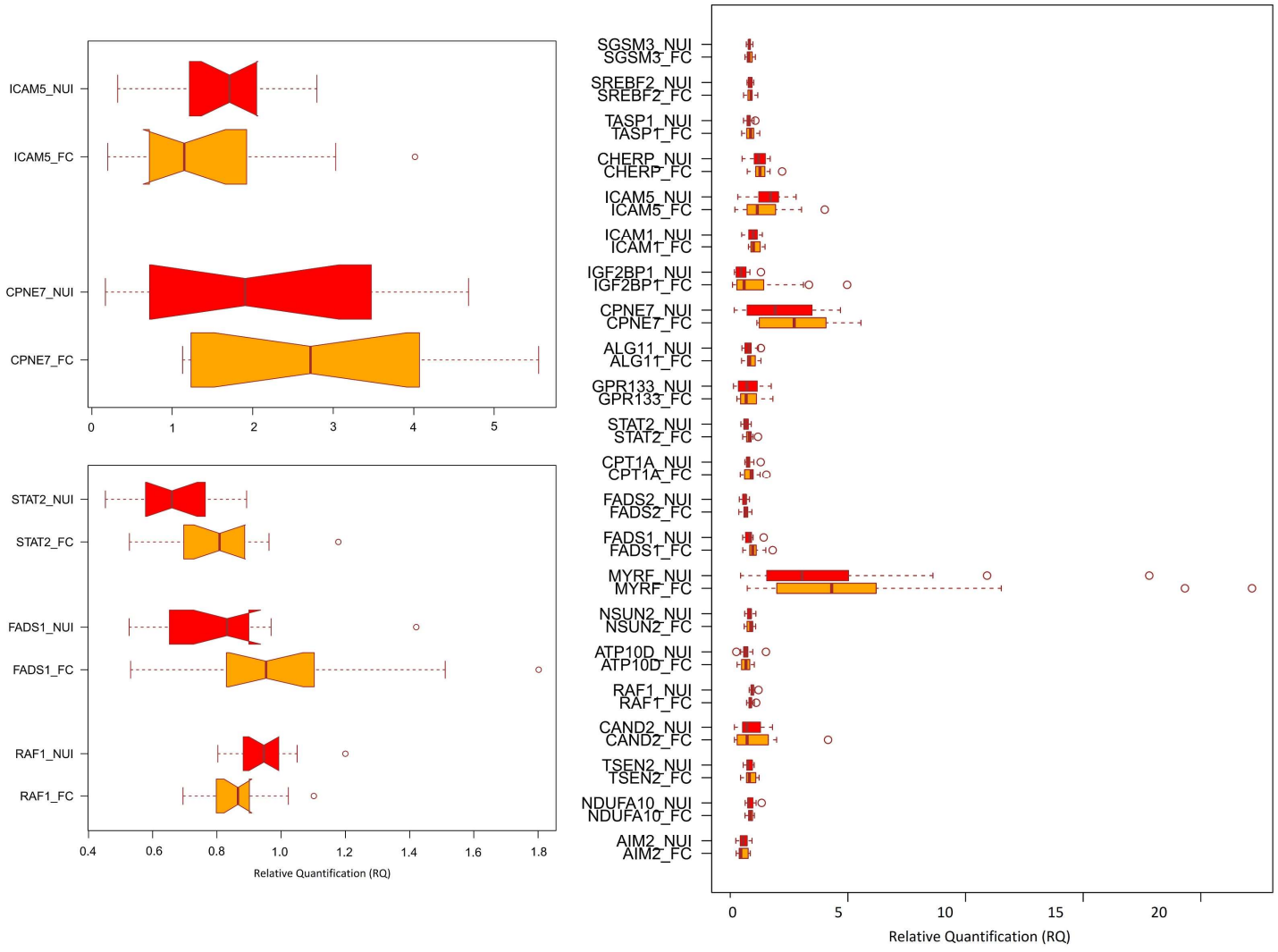


Figure S16: Differential expression of genes in regions of selection footprints of Nunavik Inuit and French-Canadian. Showing relative quantification (RQ) scores of genes (*ICAM5, CPNE7, STAT2, FADS1* and *RAF1*) with highest differential expression level changes in LCL, and all 22 selected genes with highest PBS, between randomly selected Nunavik Inuit and French-Canadians. Results from three independent measurements were analyzed.

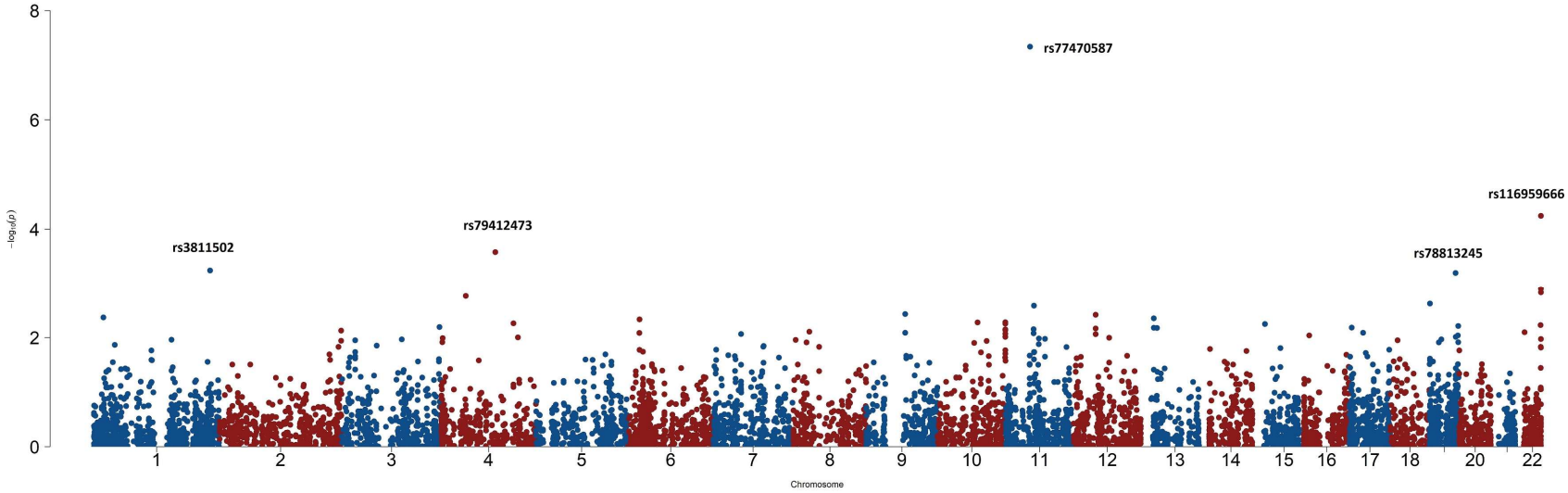


Figure S17: Association of variants within selection footprints of Nunavik Inuit with intracranial aneurysms. Displayed in Manhattan plot. 8,291 exonic variants with PBS>0.1 were included in the association analysis by fastLMM. Display top five variants: *OR4C3* (rs77470587), *SHANK3* (rs116959666), *TBCK* (rs79412473), *TTC13* (rs3811502), and *SIGLEC10* (rs78813245).

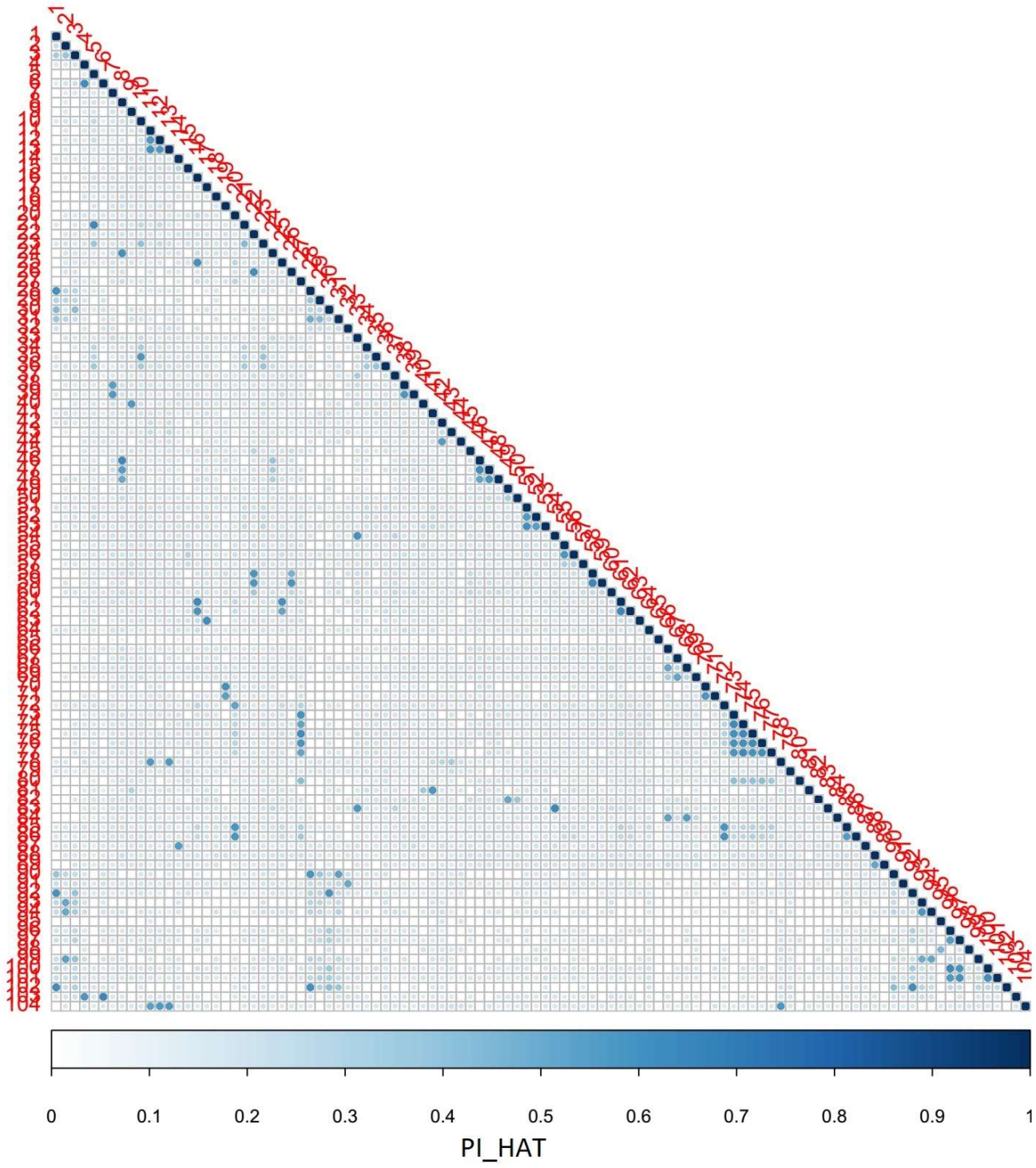


Figure S18. IBD analysis of 104 Nunavik Inuit without recent European admixture. PI_HAT = P(IBD=2) + 0.5*P(IBD=1). PI_HAT > 0.5 indicate first-degree relationship between two individuals; PI_HAT > 0.25 indicate second-degree relationship between two individuals; PI_HAT > 0.125 indicate third-degree relationship between two individuals.

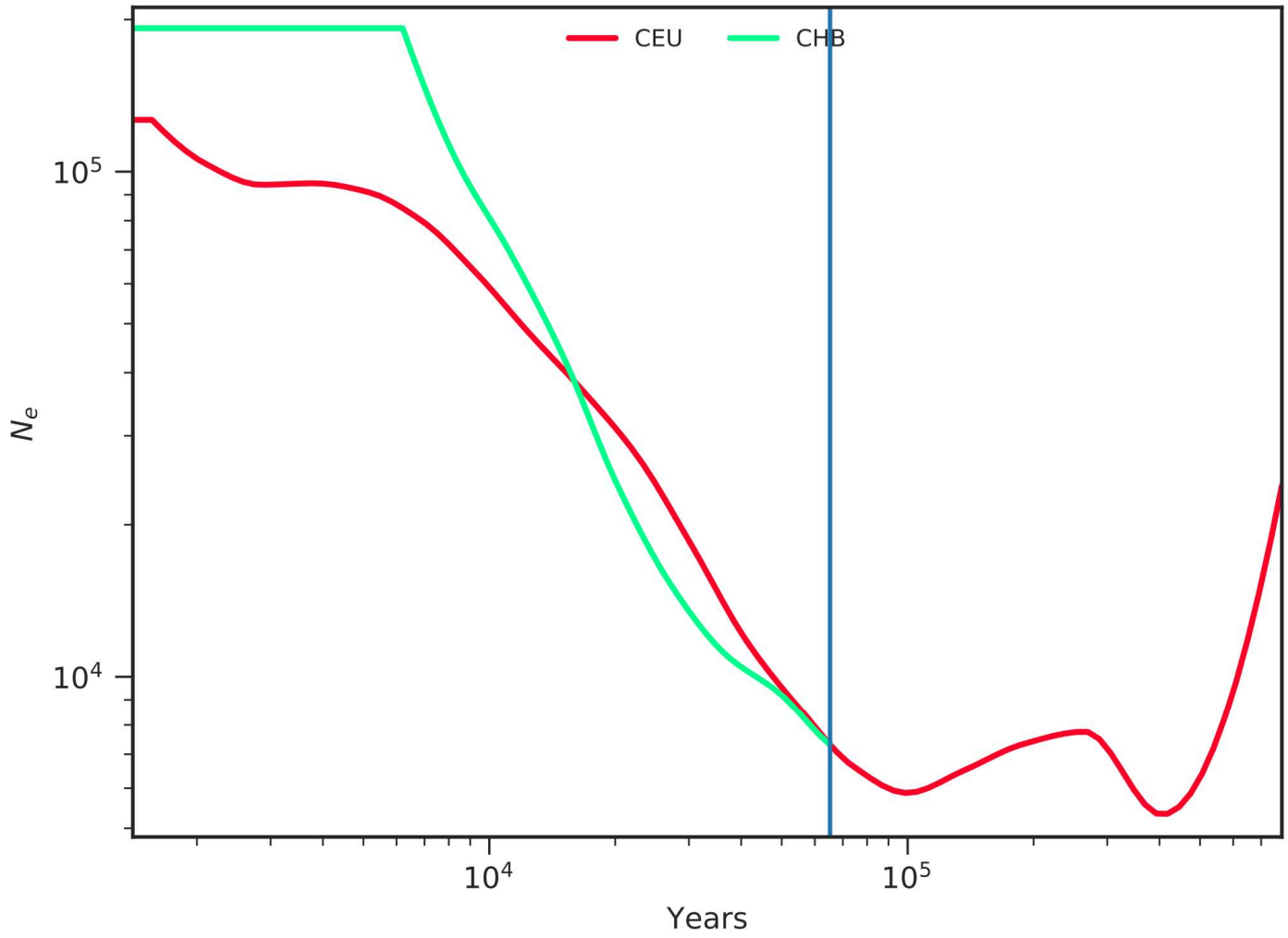


Figure S19. SMC++ estimation of split time between 1000 Genome CEU and CHB using sparse data. The masked region used for demographic inference of Nunavik Inuit and populations in Figure S15 was used for inferring effective population size of CEU and CHB. 100 CEU and 100 CHB individuals were included in the analysis following the descriptions in Terhorst *et al,* 2017. The estimated effect sizes for CEU and CHB at 1 kya are slightly smaller when comparing the use of whole genome data (CEU: Ne=1.04×10^5 vs Ne=1.1×10^5; CHB: Ne=1.09×10^5 vs 1.3×10^5), split time estimated using two dataset remains similar (~47 vs 55 kya).

Figure S20: Nunavik Villages and sample recruitment (figure adapted from http://www.inuitfirstcanadians.com/). 38 individuals were recruited from Eastern Nunavik villages around Ungava Bay: Kuujjuaq (9) and Kangiqsualujjuaq (29); and 117 individuals were recruited from Ivujivik (19), Salluit (23), Puvirnituq (28), Inukjuak (32), Akulivik (5), Kuujjuarapik (2), Umiujaq (3) and Whapmagoostui (5) in Western Nunavik villages around Hudson Bay.

Table S1: **Study populations and controls.**

	SNP-array data[1] (no.populations)	**WGS / WES data[2] (no.populations) [Ancient name]**
Nunavik Inuit (current study)	170 (10)	114 (10)
Siberians (Cardona, 2014; Clemente, 2014)	218 (10)	25 (3)
Native Americans / Siberians (Reich, 2012)	2351 (66)	NA
Sakha-Siberians (Fedorova, 2013)	40 (9)	NA
Caucasus (Yunusbayev, 2012)	204 (13)	NA
Turkic (Yunusbayev, 2015)	322 (32)	NA
Siberia-New World (Rasmussen, 2010)	188 (14)	1 [Saqqaq]
Altai regions (Raghavan, 2014)	85 (9)	1 [Mal'ta]
Siberians (Raghavan, 2015)	20 (5)	9 current + 23 ancient
1000 Genome phase III	NA	2504 (26)
Ancient America (Rasmussen, 2014)	NA	1 [Clovis]
Ancient New World Arctic (Raghavan, 2014)	NA	34 (5)
Ancient Altai (Allentoft, 2015)	NA	5
Ancient America (Rasmussen, 2015)	NA	1 [Kennewick]
TOTAL	5422 (197)	

[1]***SNP-array data****:* Data included 10 Siberian populations (Siberian Eskimo, Altai, Shors, Koryak, Chukchi, Teleuts, Evens, Yakut, Evenk, Buryat) from Cardona *et al.*(1); 114 worldwide populations (including 15 Siberian populations and 51 New World populations) from Reich *et al.*(2); 9 Sakha-Siberian populations from Fedorova *et al.*(3); 13 Caucasus populations from Yunusbayev *et al.*(4); 32 Turkic related populations (Siberians, Middle-Easterners, Central Asians and Caucasus) from Yunusbayev *et al.*(5); 11 Siberians, Greenlanders and Athabascans from Rasmussen *et al.*(6); 9 Altai region populations from Raghavan *et al.*(7) and 5 Siberian populations from Raghavan *et al.*(8)

[2]Whole genome sequencing (WGS) data: Data included 26 populations from 1KGP phase III(9), the Northeastern Siberian populations from Clemente *et al.*(10) and 9 current Old/New World populations(8). Ancient genomes from Siberia and the New World were also included: a 4-kya Saqqaq genome(6); a 24-kya Mal'ta Siberian genome(7); a 11-kya Late Pleistocene human from Clovis(11); 34 ancient Arctic people's genomes (three from PreDorset culture, 14 from Middle Dorset, two from Late Dorset, two from Birnirk culture, one Norton and five Thules)(12); five Bronze age genomes from ancient Altai region(13); the genome of the 8-kya Kennewick Man(14) and 23 ancient Native American genomes(8). ***Whole exome sequencing (WES) data for PBS calculation:*** European (CEU) and Han Chinese (CHB) populations from the 1KGP phase III(9) as well as the Northeast Siberian populations, including Chukchi, Siberian Eskimo, and Koryak(10) were used in combination with the WES data from Nunavik Inuit to calculate the PBS.

Table S2. Genes under strong natural selection in Nunavik Inuit

Gene	Top SNP	Top PBS in NUI-CHB-CEU	Function(CADD)	A1	Freq.A1.NUI	Freq.A1.CHB	Freq.A1.CEU	Top PBS in NUI-NES-CHB	Top SNP (PBS>1) in GI-CEU-CHB	No. SNPs with PBS>1 in NUI
CPT1A	rs80356779	3.11191	missense(18.21)	A	0.9567308	0	0	0.839809	NA	1
SLC24A5	rs1426654	2.06229	missense(19.66)	A	0	0.029126	1	0.126401	NA	1
TSEN2	rs735640	1.75609	3'-UTR(NA)	A	0.8300971	0	0	NA	rs735640 (1.386817257)	1
CPNE7	rs12445560	1.74435	near-splice(9.53)	T	0.9471154	0.126214	0.09596	0.48916	rs139901937 (1.386817257)	4
IMPDH1	rs72624969	1.65097	intronic(NA)	T	0.9759615	0.286408	0.07071	NA	rs4731447 (1.074576366)	3
STAT2	rs2066815	1.526	missense-near-splice(25)	A	0.7932692	0.009709	0	0.839179	NA	1
TASP1	rs11697393	1.52427	intronic(NA)	T	0.8300971	0.048544	0.0202	NA	rs11697393 (1.145574318)	3
CYP11B1	rs4534	1.51184	missense(0.09)	T	0.9615385	0.446602	0.01515	0.495403	rs57589970 (1.42684416116379)	1
SREBF2	rs2228314	1.48112	missense(15.63)	C	0.9951923	0.194175	0.25253	NA	rs2228314 (1.052002479)	1
NCR1	rs2278428	1.45223	missense(5.939)	C	0.9567308	0.320388	0.06566	0.510609	rs2278427(0.801)	3
ATP10D	rs16851681	1.43361	missense(0.09)	A	0.9807692	0.194175	0.23232	1.073339	rs13152689 (1.01525329436256)	5
CAND2	rs180768267	1.38952	missense(16.91)	G	0.7884615	0.021739	0	1.094545	rs181307051 (1.067382254)	2
EDAR	rs3827760	1.3487	missense(21.7)	A	0.008065	0.067961	1	0.096885	rs3827760 (0.039097027)	1
RAF1	rs5746223	1.34431	intronic(NA)	C	0.8267327	0.033981	0.09596	NA	rs5746223(0.72)	3
LECT1	rs62637607	1.28517	intronic(NA)	G	0.7355769	0	0.0101	NA	rs62637607 (1.376472464)	1
AIM2	rs2276405	1.27947	missense(16.34)	T	0.75	0.029126	0	0.634696	rs2276405 (2.536097555)	1
CYP11B2	rs6432	1.26684	intronic(NA)	G	0.9509804	0.519608	0.01515	NA	rs4536 (1.30937734894169)	1
SGSM3	rs55844816	1.25915	intronic(NA)	T	0.8798077	0.169903	0.09596	NA	rs55844816 (1.413484817)	4
GPR133	rs1212936	1.24033	intronic(NA)	C	0.9663462	0.160194	0.32828	NA	rs1195923(0.671033087505734)	2
GML	rs3764795	1.23792	missense(15.17)	C	0.038835	0.504854	0.95455	0.368187	rs3750247(0.238465870959858)	1
NDUFA10	rs77816205	1.19479	intronic(NA)	A	0.7548077	0	0.06633	NA	rs77816205 (1.427053791)	1
ALG11	rs17480245	1.18738	missense(0.017)	G	0.7058824	0	0.0101	0.532059	rs17480245 (1.212805669)	1
FADS1	rs174556	1.16915	intronic(NA)	C	0.0048077	0.664706	0.68966	NA	rs174547(0.763990036106579)	3
ICAM5	rs1056538	1.16354	missense(16.96)	G	0.0192308	0.81068	0.61616	1.038068	rs2228615(0.546499505209769)	3
FADS2	rs174602	1.14054	intronic(NA)	C	0.9326923	0.257282	0.19192	NA	rs174602 (1.14510508)	1
CHERP	rs12460141	1.09348	intronic(NA)	C	0.6923077	0.005952	0.00549	NA	rs12460141 (1.508902039)	2
MYRF	rs174536	1.08532	intronic(NA)	A	0.0096154	0.674757	0.67677	NA	rs108499(0.770196821564124)	4
DSP	rs7741957	1.07435	intronic(NA)	C	0.8543689	0.315534	0	NA	rs7741957 (1.907001643)	2

NSUN2	rs506416	1.05426	intronic(NA)	A	0.0148515	0.650485	0.70202	NA	rs6887702(0.781014952248378)	4
IGF2BP1	rs4265867	1.05156	intronic(NA)	A	0.6634615	0	0	NA	rs4265867 (1.735874019)	1
IGHMBP2	rs560096	1.02829	missense(1.24)	C	0.0480769	0.538835	0.88384	0.15293	rs1249463(0.954137807645166)	1
ICAM1	rs5498	0.99744	missense(0.754)	A	0.0192308	0.737864	0.58586	0.881544	rs5498(0.424053850741787)	1

NUI: Nunavik Inuit; NES: Northeastern Siberians; GI: Greenlandic Inuit; Freq.A1.: A1 allele frequency

Genes are selected based on independent variants with top PBS score in Nunavik Inuit for each region using WES data, PBS>1 was defined as variants under strong selection.

Table S3: Recent selections of genes in selection footprints of Nunavik Inuit.

chrom	pos	rs	PBS (NUI-CHB-CEU)	gene	function	AF (GI)	PBS (GI)	PBS (NUI-NES-CHB)	AF (NES)	CADD
1	1.59E+08	rs2276405	1.279471376	*AIM2*	missense	0.8611	2.536098	0.634696	0.28	16.34
3	12858557	rs180768267	1.389523908	*CAND2*	missense	#N/A	#N/A	1.094545	0.12	16.91
3	12861600	rs181307051	1.039940373	*CAND2*	missense	0.3889	1.067382	0.625435	0.14	25.7
4	47578971	rs16851681	1.433606237	*ATP10D*	missense	#N/A	#N/A	1.073339	0.46	0.09
8	1.44E+08	rs3764795	1.237917802	*GML*	missense	#N/A	#N/A	0.368187	0.28	15.17
8	1.44E+08	rs4534	1.511837689	*CYP11B1*	missense	#N/A	#N/A	0.495403	0.32	0.009
11	61551356	rs174535	1.008393664	*MYRF*	synonymous	1	0.76399	0.284129	0.08	0.924
11	68548130	rs80356779	3.111911226	*CPT1A*	missense	#N/A	#N/A	0.839809	0.32	18.21
11	68678962	rs560096	1.028292975	*IGHMBP2*	missense	0	0.954138	0.15293	0.12	1.24
12	56743044	rs2066815	1.5259979	*STAT2*	missense-near-splice	#N/A	#N/A	0.839179	0.24	25
13	52513266	rs7334118	1.152985299	*ATP7B*	missense	0.4722	1.25654	0.613102	0.2	11.53
13	52598189	rs17480245	1.187378938	*ALG11*	missense	0.4722	1.212806	0.532059	0.22	0.017
13	52715168	rs55969405	1.096288839	*NEK3*	intron	0.4722	1.212806	0.571077	0.22	#N/A
13	53049267	rs34494025	1.303110102	*CKAP2*	synonymous	#N/A	#N/A	0.502239	0.3	4.588
16	89661807	rs12445560	1.744346679	*CPNE7*	synonymous	#N/A	#N/A	0.48916	0.74	9.53
16	89986154	rs885479	1.199063038	*MC1R*	missense	0.8889	0.57	0.375297	0.18	9.495
16	90130139	rs4264393	1.504595068	*PRDM7*	missense	#N/A	#N/A	0.039348	0.62	8.187
19	10395683	rs5498	0.997435183	*ICAM1*	missense	0.8056	0.424054	0.881544	0.4	0.754
19	10402938	rs1056538	1.163543685	*ICAM5*	missense	#N/A	#N/A	1.038068	0.44	16.96
19	10403368	rs2228615	1.163543685	*ICAM5*	missense	0.8056	0.5465	1.038068	0.44	12.92
19	55418054	rs2278428	1.452229608	*NCR1*	missense	#N/A	#N/A	0.510609	0.3	5.939

GI: Greenlandic Inuit; PBS (NUI-NES-CHB): Northeast Siberians (NES) were used as the sister population in calculating PBS; PBS (GI): PBS of Greenlandic Inuit reported by (Moltke et. al, 2015). AF: allele frequency.

Table S4: Result of GO analysis of genes with weak signals of selection.

GO biological process (PANTHER Overrepresentation Test)	Total (20814)	observed	expected	fold Enrichment	P-value
cell adhesion (GO:0007155)	1039	124	78.87	1.57	6.52E-03
biological adhesion (GO:0022610)	1044	124	79.25	1.56	8.25E-03
localization (GO:0051179)	4838	452	367.25	1.23	3.79E-03
cellular response to stimulus (GO:0051716)	6061	560	460.09	1.22	2.60E-04
signaling (GO:0023052)	5052	461	383.5	1.2	3.78E-02
single organism signaling (GO:0044700)	5049	460	383.27	1.2	4.61E-02
response to stimulus (GO:0050896)	7482	677	567.96	1.19	7.46E-05
single-organism cellular process (GO:0044763)	11573	1021	878.51	1.16	1.57E-09
single-organism process (GO:0044699)	12867	1107	976.74	1.13	3.41E-08
cellular process (GO:0009987)	14439	1211	1096.07	1.1	6.16E-07
biological regulation (GO:0065007)	11293	946	857.26	1.1	3.08E-02
biological_process (GO:0008150)	16739	1368	1270.66	1.08	5.53E-07
Unclassified (UNCLASSIFIED)	4075	213	309.34	0.69	0.00E+00

1,596 genes with one or more coding variants with PBS>0.3 were included.

Supplementary Information (SI):

Merging of genotype data: The genotype data of the aforementioned array-based population controls were merged with Nunavik Inuit using Plink 1.9(15) based on GRCh37/hg19 reference and dbSNP 131 database, with strand correlate with the 1KGP. SNPs without chromosomal positions and mitochondrial SNPs were removed. A total of 1,230,877 SNPs were included after the final merging, and were used as intervals for variant calling from the bam files of the ancient whole genome data aforementioned, as well as the 1KGP phase III data. After the removal of sex-chromosomal SNPs, the final merged data comprised of array-based contemporary populations, ancient populations and the 1KGP populations, which include 1,198,992 SNPs and a total of 5,422 individuals from 197 populations.

Quality control of the sequencing data and construct of HC region: Genomic VCF (g.vcf) files for each sample were generated by Genome Analysis Tool Kit (GATK) haplotype caller to ensure the coverage of non-variant sites(16). Variants from the g.VCF files of Nunavik Inuit were combined with the variants of CHB and CEU samples from the 1KGP phase III VCF files. High quality (HC) intervals of both datasets were generated and intersected with each other to provide high confident regions to obtain high quality exome variants. For the Nunavik Inuit WES data, high-confidence regions were considered to be that 80% individuals have 100% of reads in the refseq gene regions above 10X coverage, which contain 25,941,861 base pairs in the autosomal regions. The exome targets of the 1KGP phase 3 were also considered to be the high-confidence regions for this dataset, which contain 44,584,105 bp in the autosomal regions. GATK CombineVariants function was used to merge the datasets between the Nunavik Inuit

(NUI) and the 1KGP phase3 CHB and CEU and the Northeast Siberians; and variant selection and merging were performed within the intervals of HC regions of both datasets. Indels and SNVs, which covered less than 10X in each population, were further removed in the combined dataset, non-polymorphic variants were also excluded.

A round of variant re-calling using GATK were performed on NUI, CHB and CEU BAM files using variants of the NUI-CHB-CEU merged dataset as intervals. This step was used for the calculation of population branch statistics (PBS) to ensure the accurate calling of homozygous reference genotypes in outgroup (CEU) and sister population (CHB).

Variant calling of WGS controls, including 25 North Siberians, 59 ancient genomes and 16 genomes from Old/New World populations, were also performed by GATK within the exome HC region described above. The VCF file was merged with four Native American populations (AMR) from the 1KGP and NUI-CHB-CEU variant recall file, using positions in NUI-CHB-CEU recall dataset as intervals.

PCA: SNPs with the genotype missing rate over 0.6 was removed from the dataset; and LD-pruned to ensure that no pair of SNPs with r2 greater than 0.5 in windows of 50 SNPs. The first 10 eigenvectors were calculated and maximum number of outlier removal iterations was set to 5.

PCAdmix: The data before applying PCAdmix analysis was phased using BEAGLE 4.1(17) with the 1KGP phase III as reference panel. Another 11 indigenous populations were selected, including Naukan, Chikuchi, Eskimo, Yukagir, Koryak, Evens, Chipewyan, Cree, Ojibwa, Algonquin and Greenland Inuit, the Saqqaq individual was also included as a reference. For each these admixed individuals, three ancestral populations were inferred (**Figure S3**).

ROH: 5,000 kb of sliding window and homozygous SNPs spanning 1,000 kb used to define the homozygous segment. The SNP-chip data of Chukchi, Naukan, Eskimo from Siberia, Athabascan and Chipewyan from Alaska and Canada, East and West Greenlandic populations and YRI, CHB and CEU populations from 1KGP were compared with the Nunavik individuals from different villages. The proportion of each population with one or more ROH up to 1.5 Mb, between 1.5-2.49 Mb, 2.5-4.99 Mb, 5.0-9.99 Mb and more than 10 Mb in length was calculated respectively.

PBS Calculation:

Variants with no frequency in one population was assigned a minimum MAF to reduce the bias introduced by a limited population size. Fixation index (F_{ST}) value for each variant was calculated between each of the two populations, and the classical transformation was used as follows,

$$T = -\log (1 - F_{ST})$$

The value T was calculated between the Inuit and CEU populations (T^{IE}), the Inuit and CHB populations (T^{IH}) and CEU and CHB populations (T^{EH}). PBS values, representing the lengths of the branch leading to the Nunavik Inuit population since its divergence with Han Chinese, were calculated as

$$PBS^{I} = (T^{IE} + T^{IH} - T^{EH})/ 2$$

when using Europeans as outgroup(18). The codes were included in a R script described previously(19).

Gene Expression Analysis

Selection criteria for genes tested for expression:

1) 2 or more SNVs have PBS>1 and in the same region or located in the abovementioned selected regions; 2) the gene also showed signs of selection (PBS>0.5) in Greenlandic Inuit; and 3) the gene expression in lymphoblastoid cell lines (LCL) cells is predicted to be measurable by GTEx (www.gtexportal.org).

qPCR expression analysis:

RNA was extracted from each LCL and cDNA was prepared using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). RT-qPCR was performed using the TaqMan Gene Expression Assay on QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems). For each assay, RNA polymerase II polypeptide A (POLR2A) was used as endogenous control to normalize the gene expression level. Three analyses were done with independent RNA extractions and each sample was tested in triplicate. The triplicates were valid if the delta Ct standard deviation was smaller than 0.25, and the mean delta CT value over three replications was used and with delta delta CT calculated using one additionally randomly selected individual as the calibrator. Relative quantification (RQ)(20) for each gene was calculated for 14 Nunavik and 14 FC individuals using the formula of:

$$RQ = 2^{-\Delta\Delta Ct}$$

GO enrichment analysis and results

Gene ontology (GO) enrichment analyses were performed on 1,596 genes with one or more coding SNVs with weak signals of selection (PBS>0.3) between Nunavik Inuit and CHB using Panther (http://geneontology.org/). It revealed significant enrichment of biological

processes in cell adhesion (GO:0007155) (P = 6.52e-03); localization (GO:0051179) (P = 3.79e-03) and response to stimulus (GO:0050896) (P = 7.46e-05) (**Table S4**), from which 124 genes are categorized under cell adhesion process, including the *ICAM* family genes*, ICAM-1 to ICAM-5,* which showed the top selected signals.

References:

1. Cardona A, *et al.* (2014) Genome-wide analysis of cold adaptation in indigenous Siberian populations. *PLoS One* 9(5):e98076.
2. Reich D, *et al.* (2012) Reconstructing Native American population history. *Nature* 488(7411):370-374.
3. Fedorova SA, *et al.* (2013) Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol Biol* 13:127.
4. Yunusbayev B, *et al.* (2012) The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol Biol Evol* 29(1):359-365.
5. Yunusbayev B, *et al.* (2015) The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet* 11(4):e1005068.
6. Rasmussen M, *et al.* (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463(7282):757-762.
7. Raghavan M, *et al.* (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87-91.
8. Raghavan M, *et al.* (2015) POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349(6250):aab3884.
9. Genomes Project C, *et al.* (2015) A global reference for human genetic variation. *Nature* 526(7571):68-74.
10. Clemente FJ, *et al.* (2014) A Selective Sweep on a Deleterious Mutation in CPT1A in Arctic Populations. *Am J Hum Genet* 95(5):584-589.
11. Rasmussen M, *et al.* (2014) The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506(7487):225-229.
12. Raghavan M, *et al.* (2014) The genetic prehistory of the New World Arctic. *Science* 345(6200):1255832.
13. Allentoft ME, *et al.* (2015) Population genomics of Bronze Age Eurasia. *Nature* 522(7555):167-172.
14. Rasmussen M, *et al.* (2015) The ancestry and affiliations of Kennewick Man. *Nature* 523(7561):455-458.
15. Chang CC, *et al.* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.
16. McKenna A, *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297-1303.
17. Browning SR & Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81(5):1084-1097.
18. Yi X, *et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329(5987):75-78.

19. Fumagalli M*, et al.* (2015) Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* 349(6254):1343-1347.
20. 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(4):402-408.