5 Supplementary Information

5.1 Supplementary Text

5.1.1 The major spliceosome complex is strongly conserved between archaic hominins and modern humans.

Alternative splicing occurs across nearly all eukaryotes and its molecular mechanisms are deeply conserved [41]. Thus, we anticipated that the sequence patterns learned by SpliceAI in humans would generalize to archaic hominins. To test for any significant differences, we first compared the sequences of 147 genes associated with the major spliceosome complex [42] between archaic hominins and modern humans. Analyzing all archaic variants absent in the 1000 Genomes Project (1KG) [43], we identified 19 non-synonymous archaic-specific variants, only 5 possibly or probably damaging by PolyPhen 2 and 8 of which were predicted to be deleterious by SIFT; [45], (**Supplementary Data 1**). Additionally, only two variants were predicted to disrupt protein function by both PolyPhen and SIFT (**Supplementary Data 1**). Furthermore, both deleterious variants were unique to the Altai Neanderthal.

We also considered the extent to which modern humans harbor deleterious variants in spliceosome associated genes. We subset variants that fell within spliceosome genes in each of the randomly sampled 1KG individuals and repeated the above procedure. We also repeated this for the archaics and included all spliceosome variants, regardless of whether or not they were archaic-specific. Modern humans exhibited between 0 and 3 deleterious variants, whereas the Chagyrskaya Neanderthal, Denisovan, and Vindija Neanderthal had 0 and the Altai Neanderthal had the two aforementioned variants (**Supplementary Table 14**).

Therefore, the major spliceosome complex is nearly identical between archaic hominins and modern humans, and there is no evidence that the sequence determinants of splicing have diverged.

5.1.2 Physical characteristics of a gene are associated with the number of SAVs.

Alternative splicing requires genes have at least one intron and the extent of splice altering variants may further be related to gene length. We considered the relationship between the the number of splicing variants (n = 0–11) in each gene and three gene traits: 1) number of exons, 2) length of gene body, and 3) length of coding sequence. These characteristics were positively associated at both Δ thresholds (**Supplementary Fig. 6**) (**Supplementary Table 6**). We also considered the relationship between the number of known isoforms per gene and the number of SAVs. We found a significant, positive association at both thresholds (**Supplementary Table 6**). However, this relationship is likely driven by the number of exons. When we conducted a partial correlation, controlling for the number of exons, both associations were non-significant with minimal effect size ($\Delta \geq 0.2$: partial Spearman, ρ = -0.003, P = 0.702, $\Delta \geq 0.5$: partial Spearman, ρ = 0.0008, P = 0.9174).

5.1.3 Sprime identifies more introgressed SAVs present in all five 1KG superpopulations.

Comparing our results using two different sets of introgressed variants [38, 39] revealed very similar patterns. For example, most ancient SAVs occurred at $\geq$ 0.05 frequency in all five 1KG superpopulations (**Supplementary Fig. 11**). The next most frequent were SAVs that were present in all but east Asians. SAVs shared among non-Africans and those present in all but Europeans were also common.

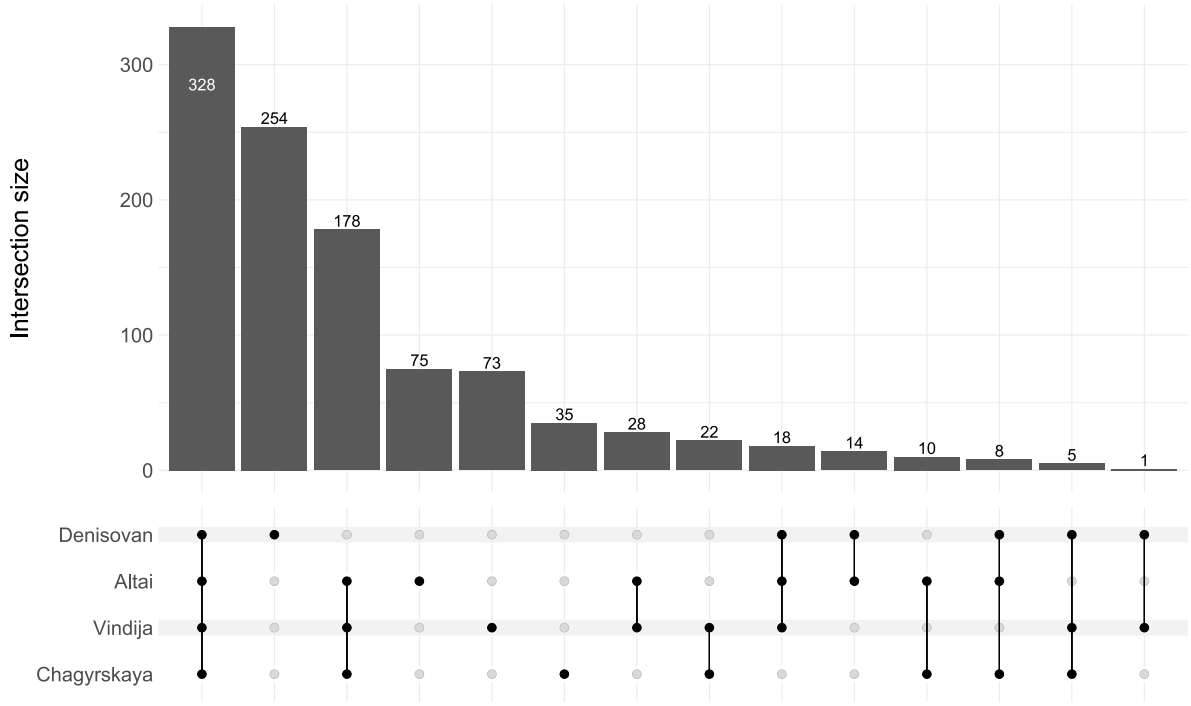
However, there were a few differences between the datasets in the distribution of introgressed SAVs across populations. We found that among those introgressed SAVs classified per [38], many were shared among non-Africans (**Fig. 5D**). However, introgressed SAVs classified per [39] were most commonly present in both Africans and non-Africans, followed by a smaller set of non-African SAVs (**Supplementary Fig. 14**). This difference likely reflects low frequency introgressed variants that are allowed to occur in the reference population in Sprime. While introgression between modern humans and archaics occurred outside of Africa, many Africans have low levels of Neanderthal ancestry due to backflow from Eurasians into Africans [94]. Among non-African subsets, both datasets were largely similar in their distribution (**Supplementary Figs. 14**, **15**).

5.1.4 Ancient and introgressed sQTL SAVs have similar tissue activity patterns in GTEx.

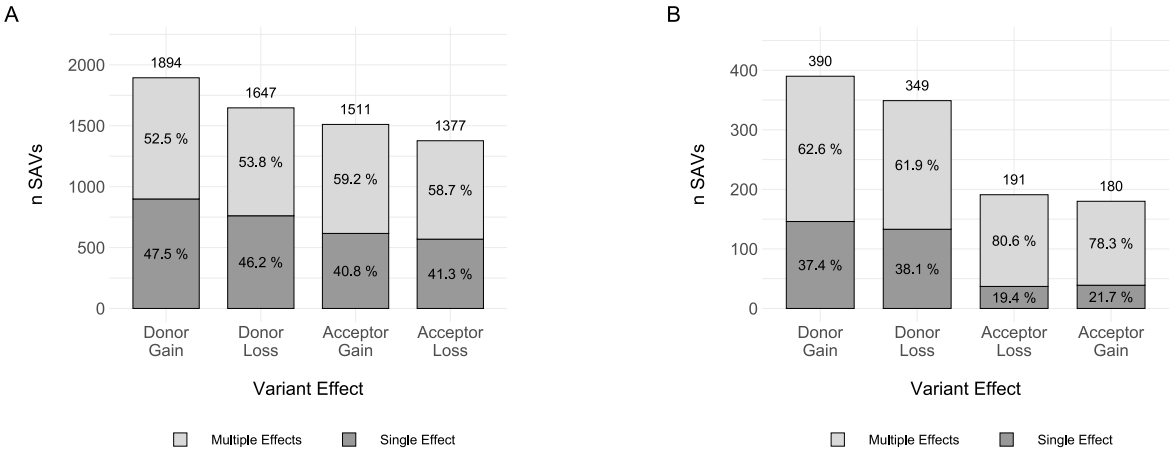
To test for differences in the tissues of activity for SAVs of different origin, we compared the distribution of ancient and introgressed SAVs identified in GTEx (**Supplementary Table 15**). The number of sQTLs in each tissue was positively correlated between ancient and introgressed variants, and this held for both sets of introgressed variants (Browning et al. 2018: Spearman, ρ = 0.94, P = 3.8×10^{-24}; Vernot et al. 2016: Spearman, ρ = 0.92, P = 1.09×10^{-20}) (**Supplementary Fig. 22**). However, the ability to detect sQTL is strongly influenced by sample size, and we found that the number of individuals sampled for each tissue is positively correlated with the number of sQTL SAVs in a tissue for both ancient variants (Browning et al. 2018: Spearman, ρ = 0.91, P = 5.73×10^{-20}; Vernot et al. 2016: Spearman, ρ = 0.91, P = 8.43×10^{-20}) and introgressed variants (Browning et al. 2018: Spearman, ρ = 0.89, P = 2.47×10^{-17}; Vernot et al. 2016: Spearman, ρ = 0.85, P = 1.34×10^{-14}) (**Supplementary Figs. 23**, **24**). We also note that the proportion of sQTL SAVs compared to all significant GTEx sQTLs detected per tissue was largely similar for most tissues (**Supplementary Fig. 25**). These results suggest that tissues of activity between SAVs of different allele origins are similar; however, these analyses are limited by the coverage and power of GTEx sQTL data.

We also considered tissue-level gene expression differences by comparing the number of genes expressed (TPM > 1) in each tissue with at least one SAV of a given origin. We found no significant differences among allele origins for either allele origin classification (**Supplementary Figs. 26**, **27**). As for the sQTL, these patterns were mainly driven by differences in sample size and number of expressed genes in each GTEx tissue.

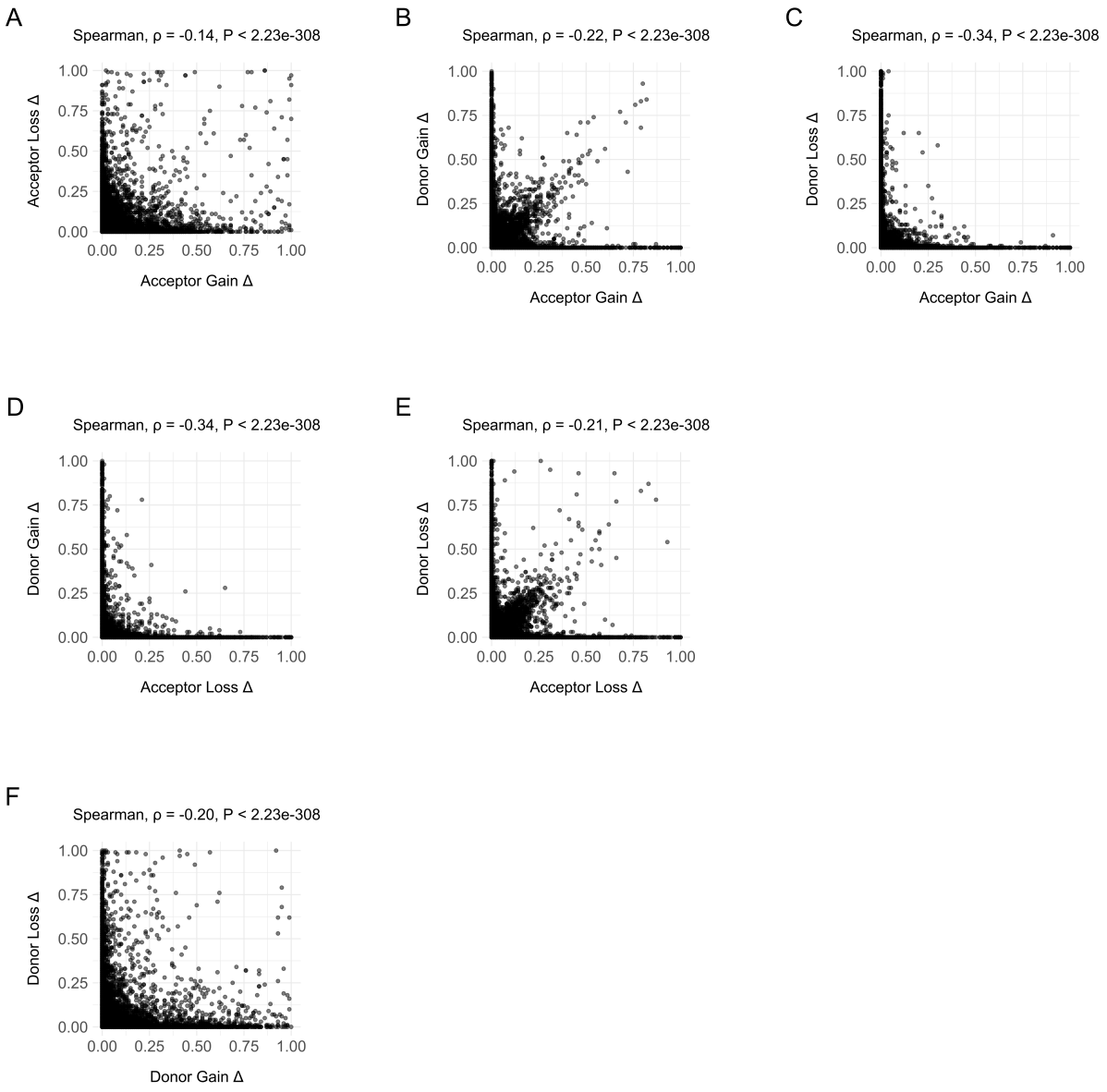
5.2 Supplementary Figures



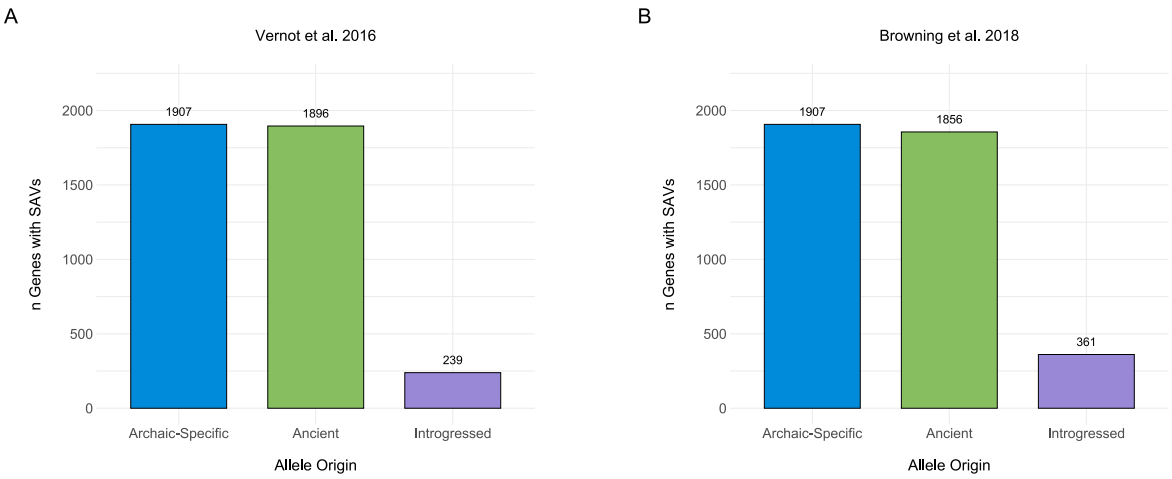
Supplementary Fig. 1 | The distribution of SAVs is consistent with the archaic hominin phylogeny. Unique and shared SAVs at $\Delta \geq 0.5$.



Supplementary Fig. 2 | SAVs most commonly result in a donor gain.
(A) The distribution of variant effects for all variants at the $\Delta \geq 0.2$ threshold. The n above each column indicates the number of variants with that effect above the threshold. Some variants result in multiple effects; therefore, the sum of these classes $\neq$ 5,950. The color in each stacked bar indicates whether the effect is unique (dark grey) or one of many effects (light grey). **(B)** The distribution of variant effects for all variants at the $\Delta \geq 0.5$ threshold. As in **A**, some variants result in multiple effects; therefore, the sum of these classes $\neq$ 1,049. The bar stacks are colored as in **A**.

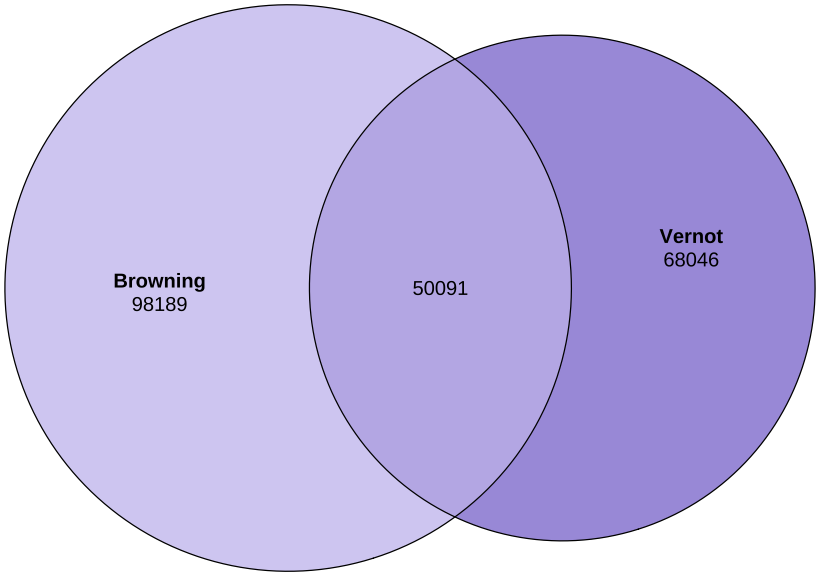


Supplementary Fig. 3 | Delta scores between classes are negatively associated.
(A) Association between acceptor gain Δs and acceptor loss Δs for all variants with at least one Δ > 0. **(B)** Association between acceptor gain Δs and donor gain Δs for all variants with at least one Δ > 0. **(C)** Association between acceptor gain Δs and donor loss Δs for all variants with at least one Δ > 0. **(D)** Association between acceptor loss Δs and donor gain Δs for all variants with at least one Δ > 0. **(E)** Association between acceptor loss Δs and donor loss Δs for all variants with at least one Δ > 0. **(F)** Association between donor gain Δs and donor loss Δs for all variants with at least one Δ > 0.

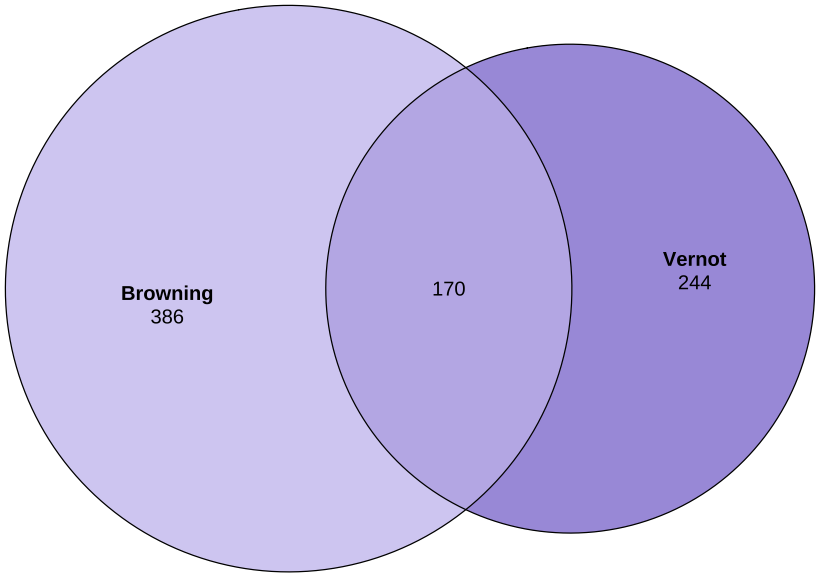


Supplementary Fig. 4 | The number of genes with SAVs by allele origin.
(A) The number of genes encompassed by each SAV allele origin per [39] at $\Delta \geq 0.2$. Some genes may occur in multiple categories. **(B)** The number of genes encompassed by each SAV allele origin per [38] at $\Delta \geq 0.2$. Some genes may occur in multiple categories.

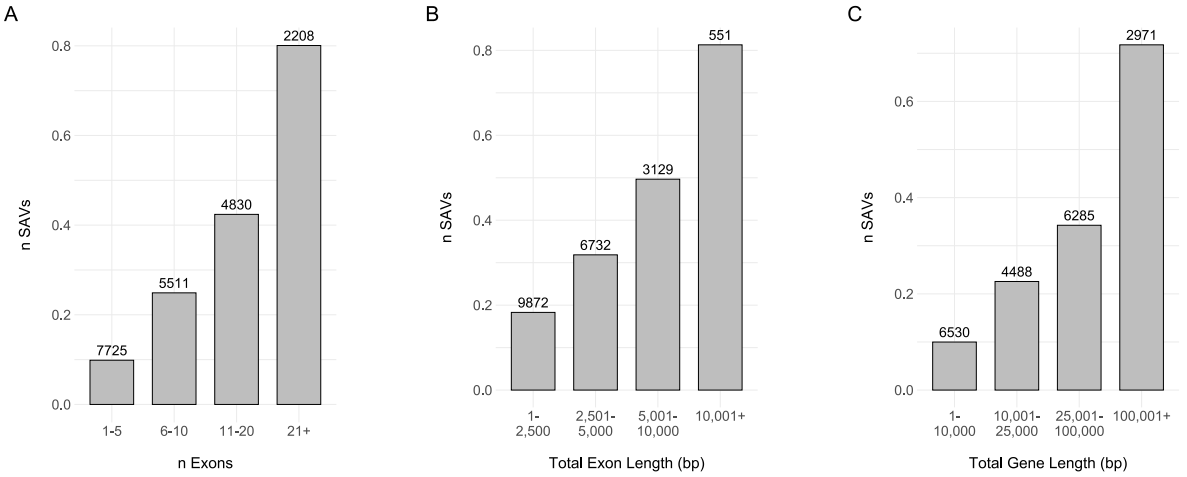
A



B

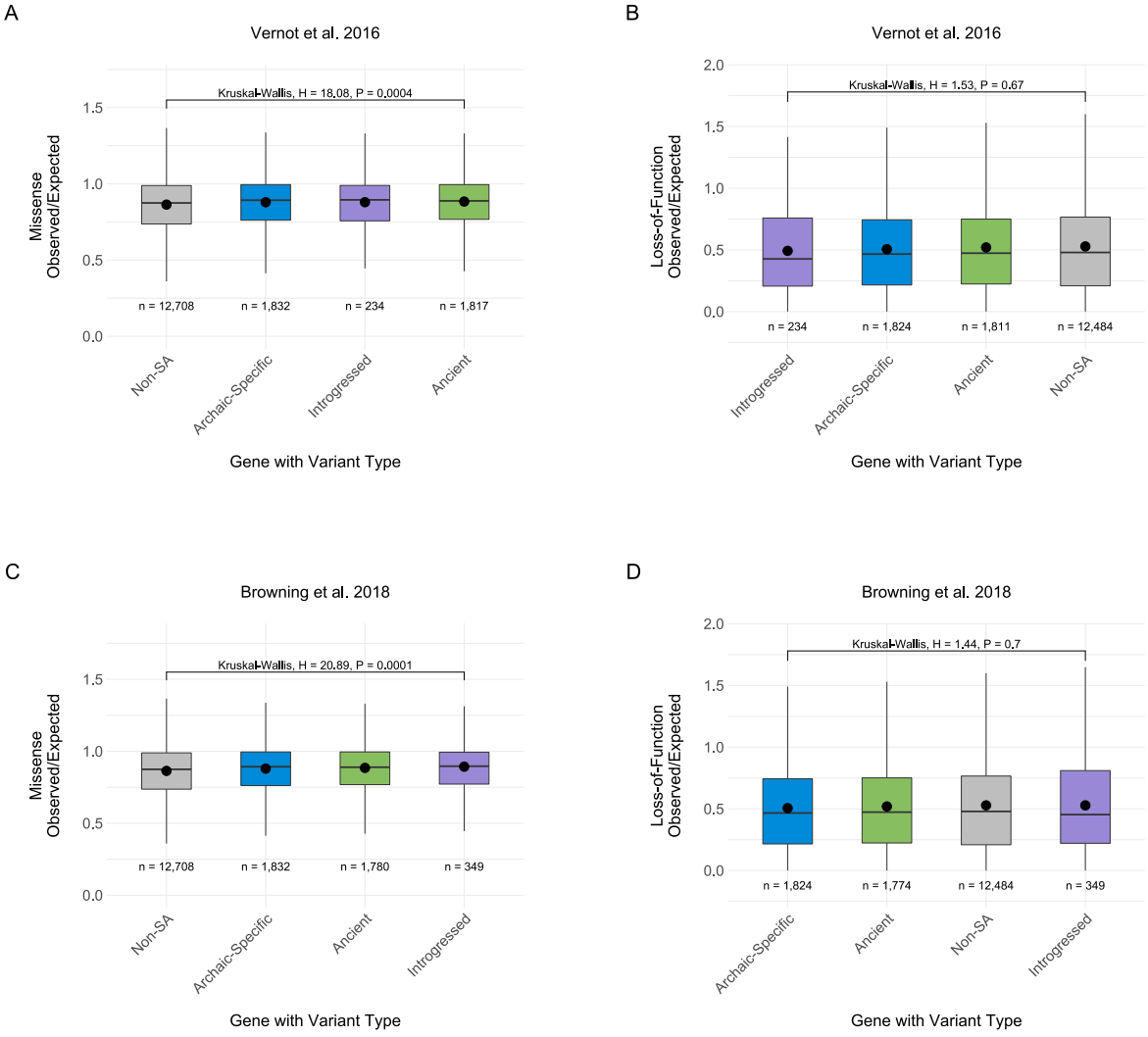


Supplementary Fig. 5 | Two different sets of introgressed variants modestly overlap.
(A) The overlap between introgressed variants from [39] and [38] that match a quality filtered locus in this study and are present in 1KG and/or gnomAD. **(B)** A subset of the overlap for variants with $\Delta \geq 0.2$.



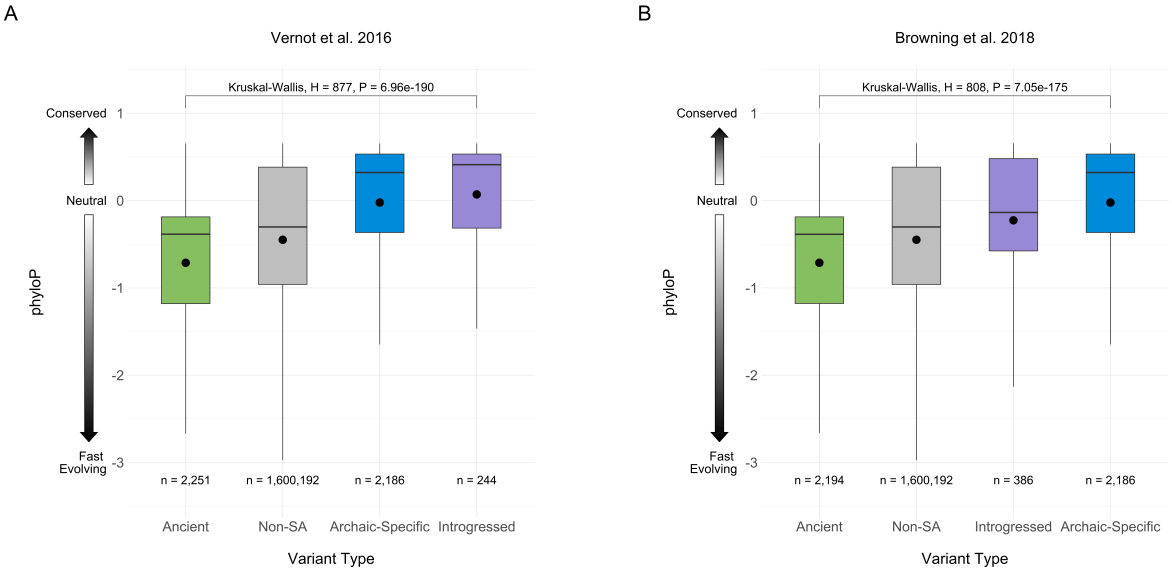
Supplementary Fig. 6 | The number of exons and gene length are associated with more SAVs.

(A) Binned number of exons and the mean number of SAVs at $\Delta \geq 0.2$ per bin. n reflects the number of genes per bin. **(B)** Binned exon length in bp and the mean number of SAVs at $\Delta \geq 0.2$ per bin. n reflects the number of genes per bin. **(C)** Binned gene length in bp and the mean number of SAVs at $\Delta \geq 0.2$ per bin. n reflects the number of genes per bin.



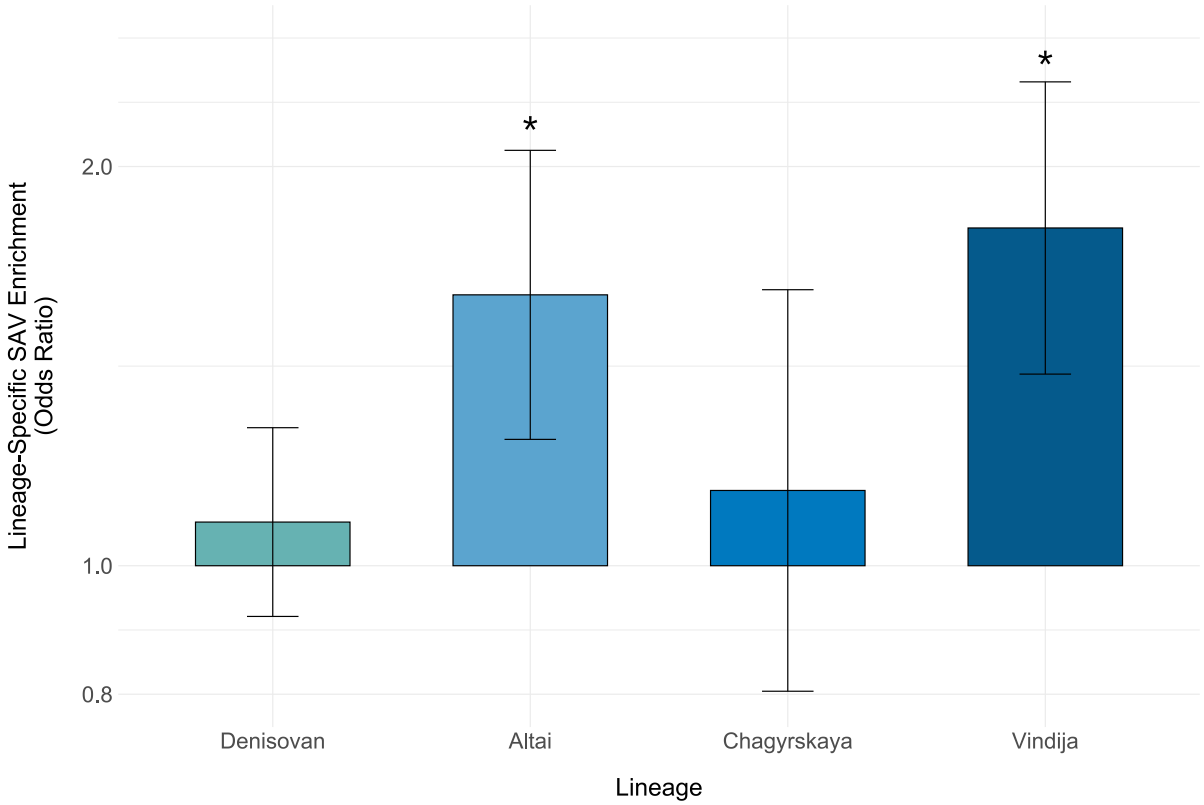
Supplementary Fig. 7 | Gene mutational tolerance does not constrain the evolutionary history of SAVs.

(A) Observed over expected ratio for missense variants per gene from gnomAD among four different sets. The ancient, archaic-specific, and introgressed sets from [39] include any gene that had $\geq$ 1 SAV at the $\Delta \geq 0.2$ threshold. Non-SA genes are all genes that do not occur in any of the other three sets. The boxplots display the median and IQR, with the upper whiskers extending to largest value $\leq$ 1.5 x IQR from the 75th percentile and the lower whiskers extending to the smallest values $\leq$ 1.5 x IQR from the 25th percentile. The mean is noted by the black point. Ns reflect the number of variants per set. **(B)** Observed over expected ratio for loss-of-function variants per gene from gnomAD among four different sets from [39]. The ancient, archaic-specific, and introgressed sets include any gene that had $\geq$ 1 SAV at the $\Delta \geq 0.2$ threshold. Non-SA genes are all genes that do not occur in any of the other three sets. Data are visualized as in **A**. **(C)** Observed over expected ratio for missense variants per gene from gnomAD among four different sets from [38]. Data are visualized as in **A**. **(D)** Observed over expected ratio for loss-of-function variants per gene from gnomAD among four different sets from [38]. Data are visualized as in **A**.

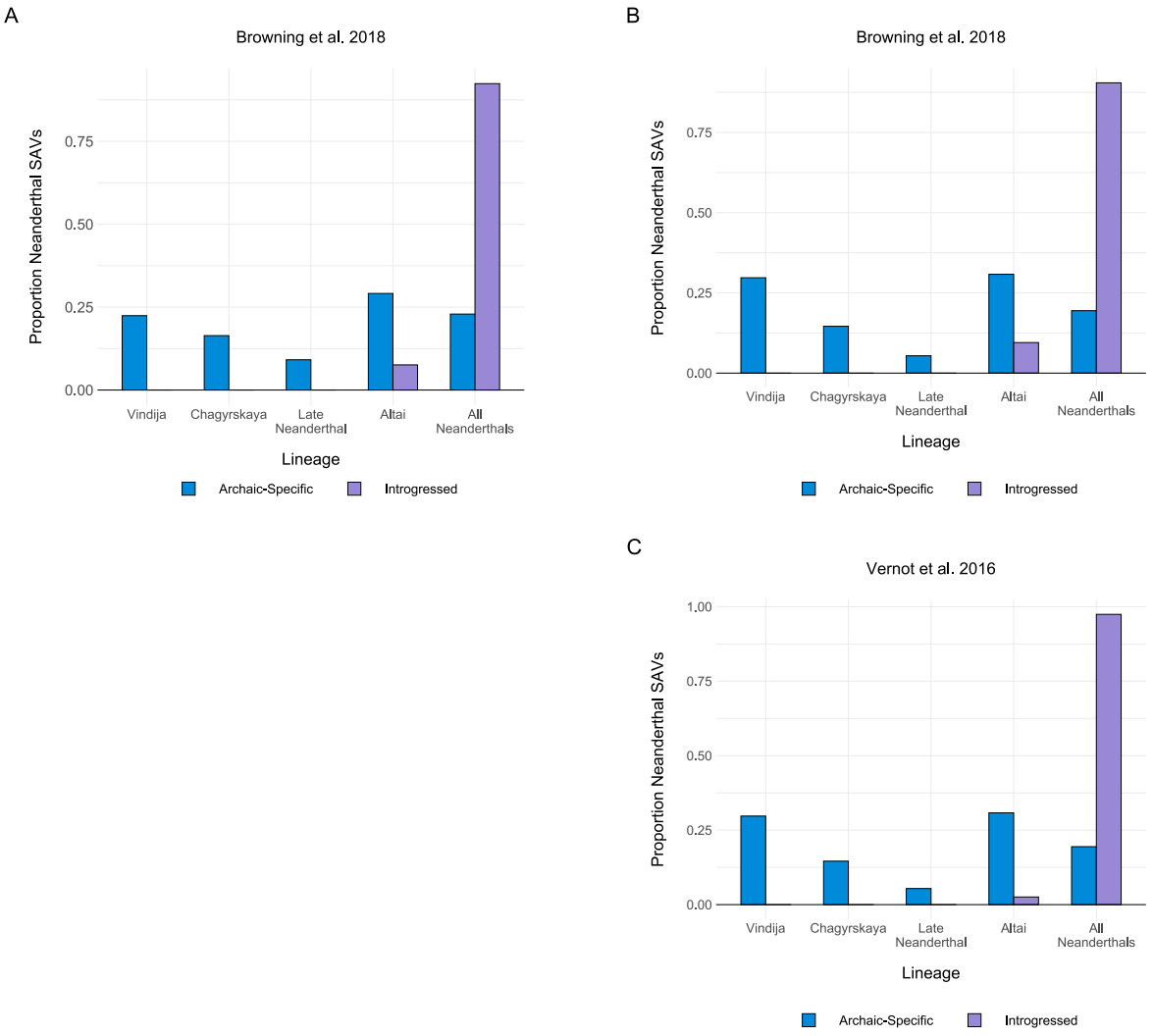


Supplementary Fig. 8 | Site-level evolutionary conservation varies across SAVs with different origins.

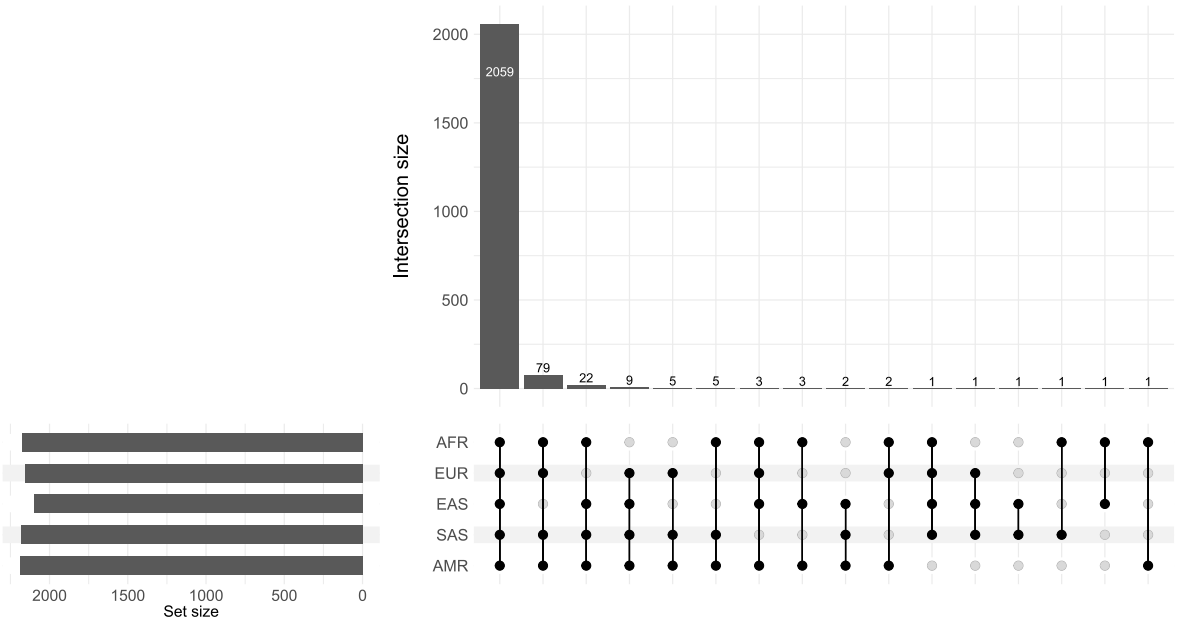
(A) phyloP evolutionary conservation score distributions for archaic SAVs of different origins and non-SAVs. Positive scores indicate substitution rates slower than expected under neutral evolution (conservation), while negative scores indicate higher substitution rates than expected (fast evolution). The boxplots display the median and IQR, with the upper whiskers extending to largest value $\leq$ 1.5 x IQR from the 75th percentile and the lower whiskers extending to the smallest values $\leq$ 1.5 x IQR from the 25th percentile. Ns are the number of variants per set. SAV classifications were based on [38] introgression calls. **(B)** phyloP evolutionary conservation score distributions for archaic SAVs of different origins and non-SAVs based on [39] introgression calls. Data were visualized as in **A**.



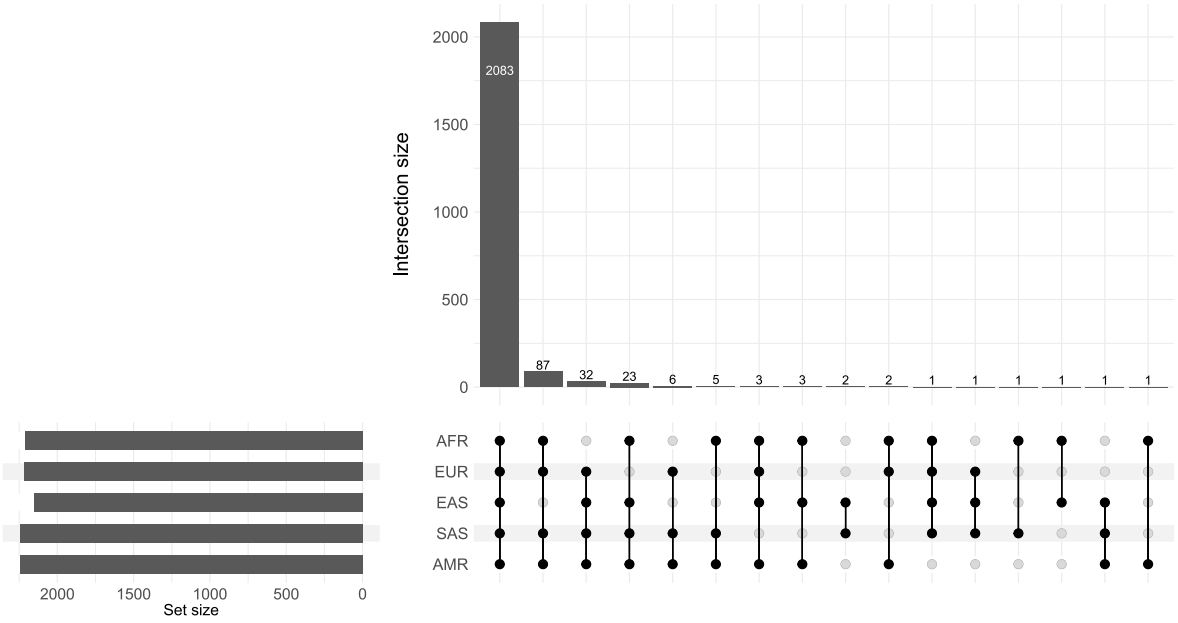
Supplementary Fig. 9 | Lineage-specific high-confidence SAVs are enriched in two Neanderthals. Odds ratios from Fisher's exact test performed for each lineage's unique high confidence SAVs/non-SAVs compared to those shared among all four individuals. Odds ratios were calculated from 81,916 Altai-specific, 53,765 Chagyrskaya-specific, 411,492 Denisovan-specific, 70,999 Vindija-specific, and 573,197 shared variants. The number of lineage-specific and shared SAVs/non-SAVs used in each enrichment test are listed in **Supplementary Table 9**. Bar height indicates the odds ratio. Error bars denote the 95% CI and are centered on the odds ratio. Asterisks reflect significance of Fisher's exact tests using a Bonferroni corrected α (0.0125). Error bars denote the 95% CI. Note the y-axis is $\log_{10}$ transformed.



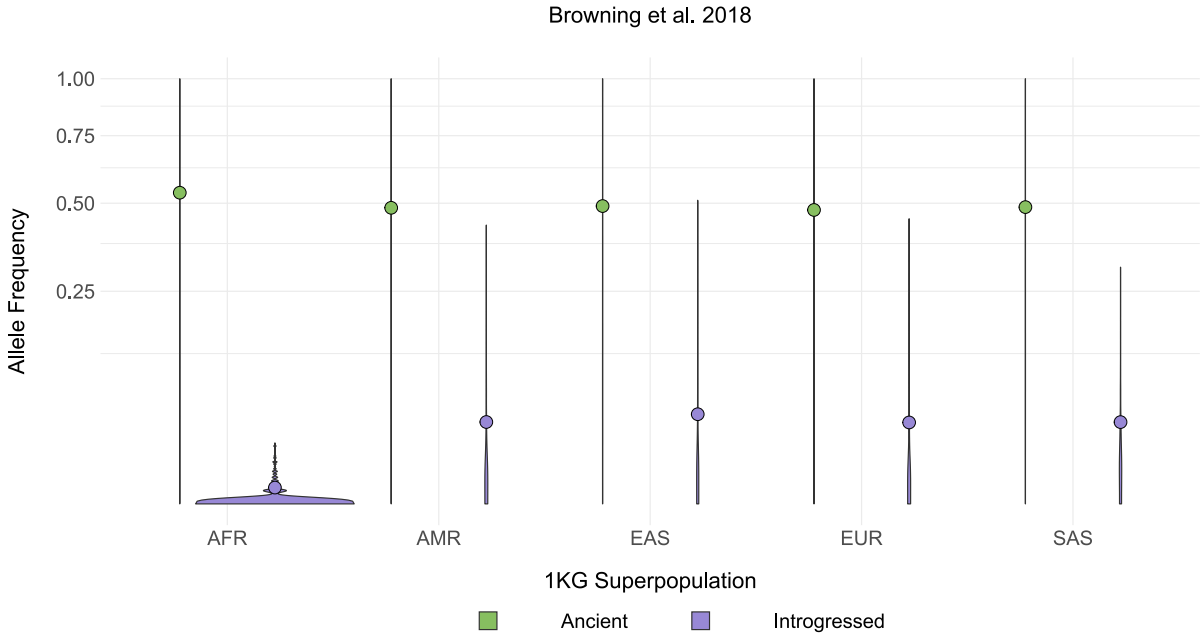
Supplementary Fig. 10 | Introgressed SAVs are largely older variants.
(A) Proportion of SAVs at $\Delta \geq 0.2$ per Neanderthal lineage among archaic-specific SAVs (expected) and introgressed SAVs (observed) from [39]. Proportions were calculated from the sum of all Neanderthal lineages because power to detect introgressed Denisovan SAVs is low. All data are presented in **Supplementary Table 10**. **(B)** Proportion of SAVs at $\Delta \geq 0.5$ using introgressed SAVs from [39]. **(C)** Proportion of SAVs at $\Delta \geq 0.5$ using introgressed SAVs from [38].



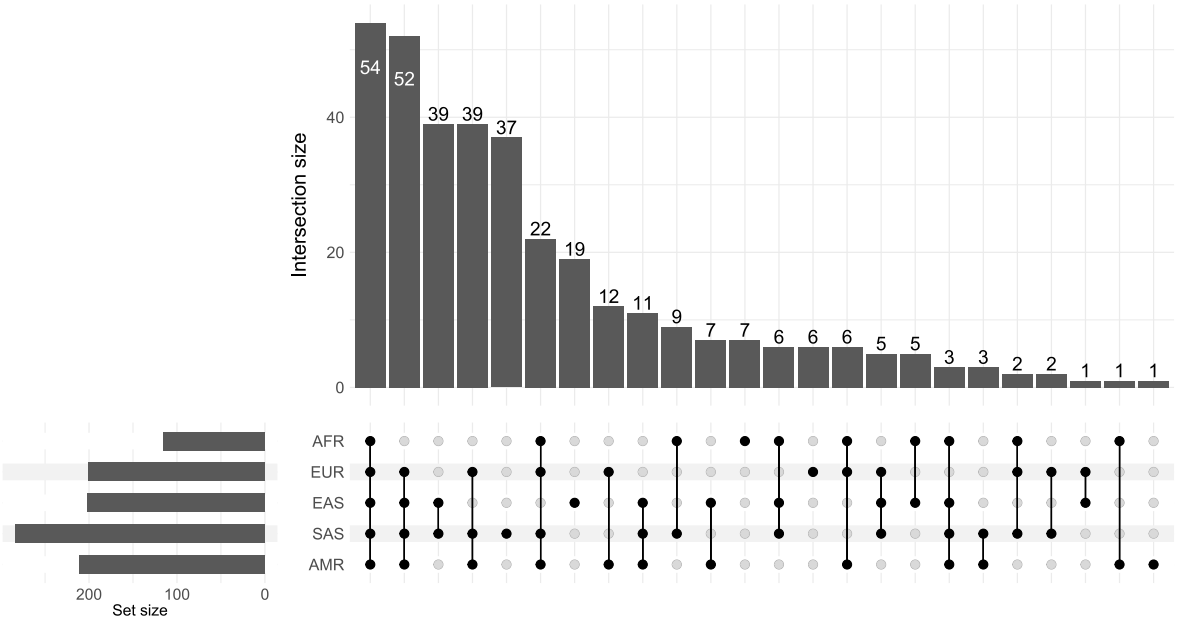
Supplementary Fig. 11 | Most ancient SAVs occur in all 1KG superpopulations. Unique and shared ancient SAVs per [39] at $\Delta \geq 0.2$. Allele frequencies are from 1KG. By definition, each ancient SAV was considered present in a population if the allele frequency was ≥ 0.05 for at least two superpopulations. AFR = African, AMR = American, EAS = East Asian, EUR = European, SAS = South Asian.



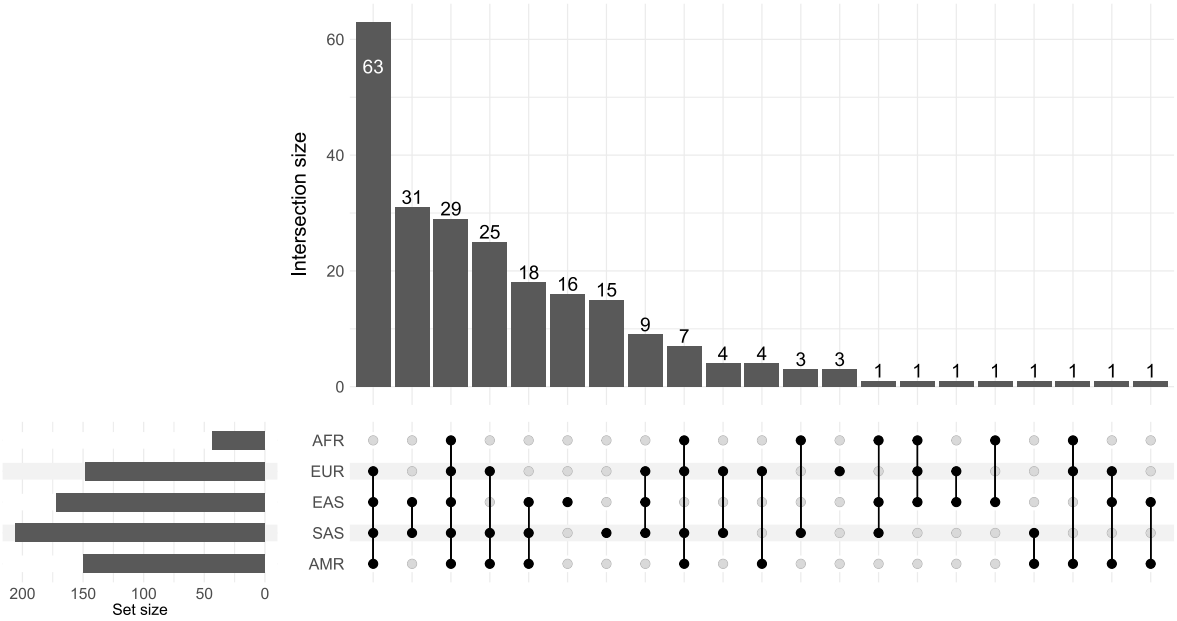
Supplementary Fig. 12 | Most ancient SAVs occur in all 1KG superpopulations. Unique and shared ancient SAVs per [38] at $\Delta \geq 0.2$. Allele frequencies are from 1KG. By definition, each ancient SAV was considered present in a population if the allele frequency was ≥ 0.05 for at least two superpopulations. AFR = African, AMR = American, EAS = East Asian, EUR = European, SAS = South Asian.



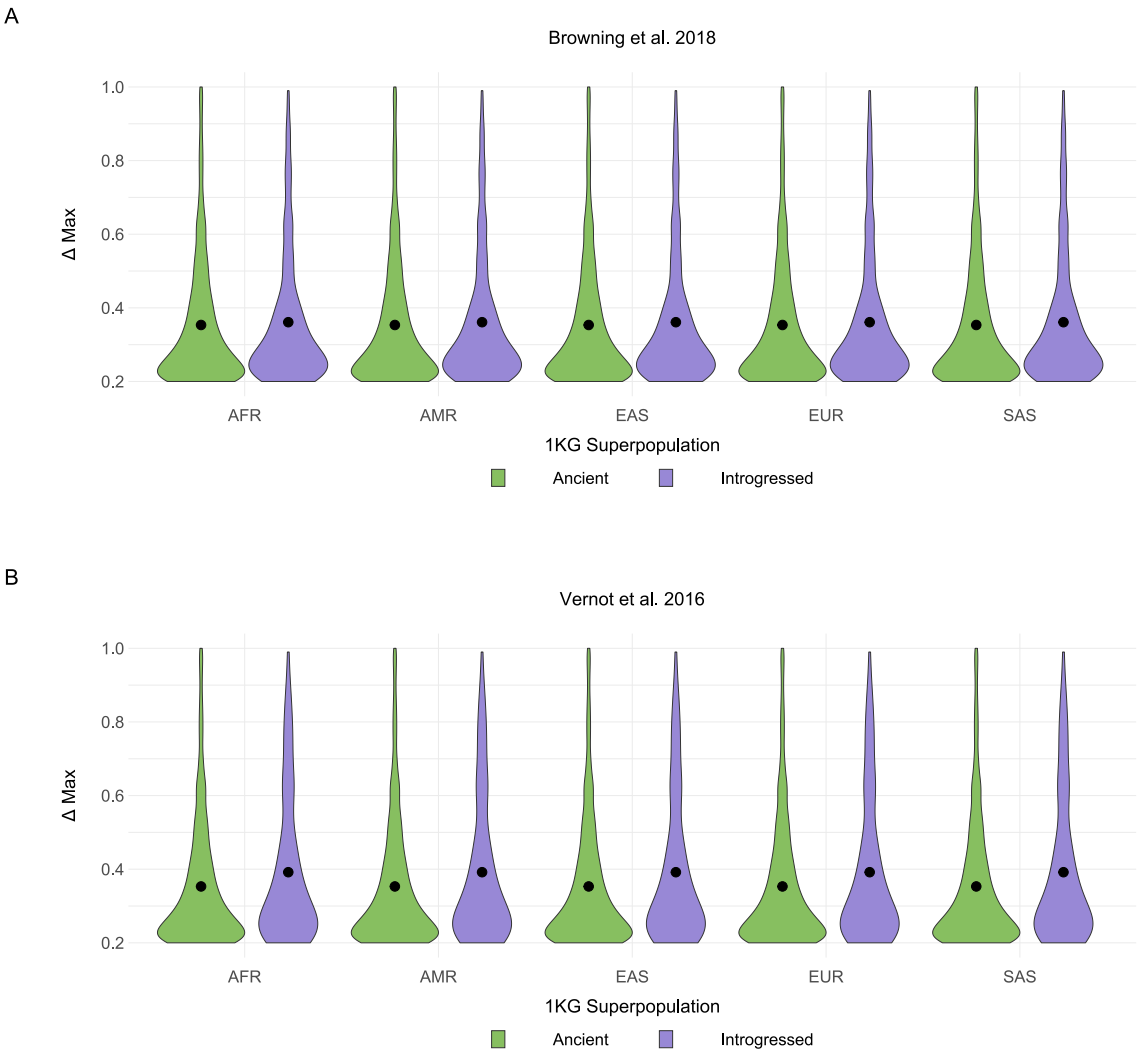
Supplementary Fig. 13 | Ancient SAVs occur a moderate to high frequencies, whereas introgressed SAVs occur at lower frequencies. Allele frequency distributions for SAVs present in both archaic and modern individuals stratified by 1KG superpopulation and origin (ancient vs. introgressed) per [39]. Ancient SAVs (n = 2,195) are colored green and displayed on the left, while introgressed SAVs (n = 377) are colored purple and displayed on the right per superpopulation. If the introgressed allele was the reference allele, we subtracted the 1KG allele frequency from 1. AFR = African, AMR = American, EAS = East Asian, EUR = European, SAS = South Asian. The colored dot represents the mean allele frequency. Note the y-axis is square root transformed.



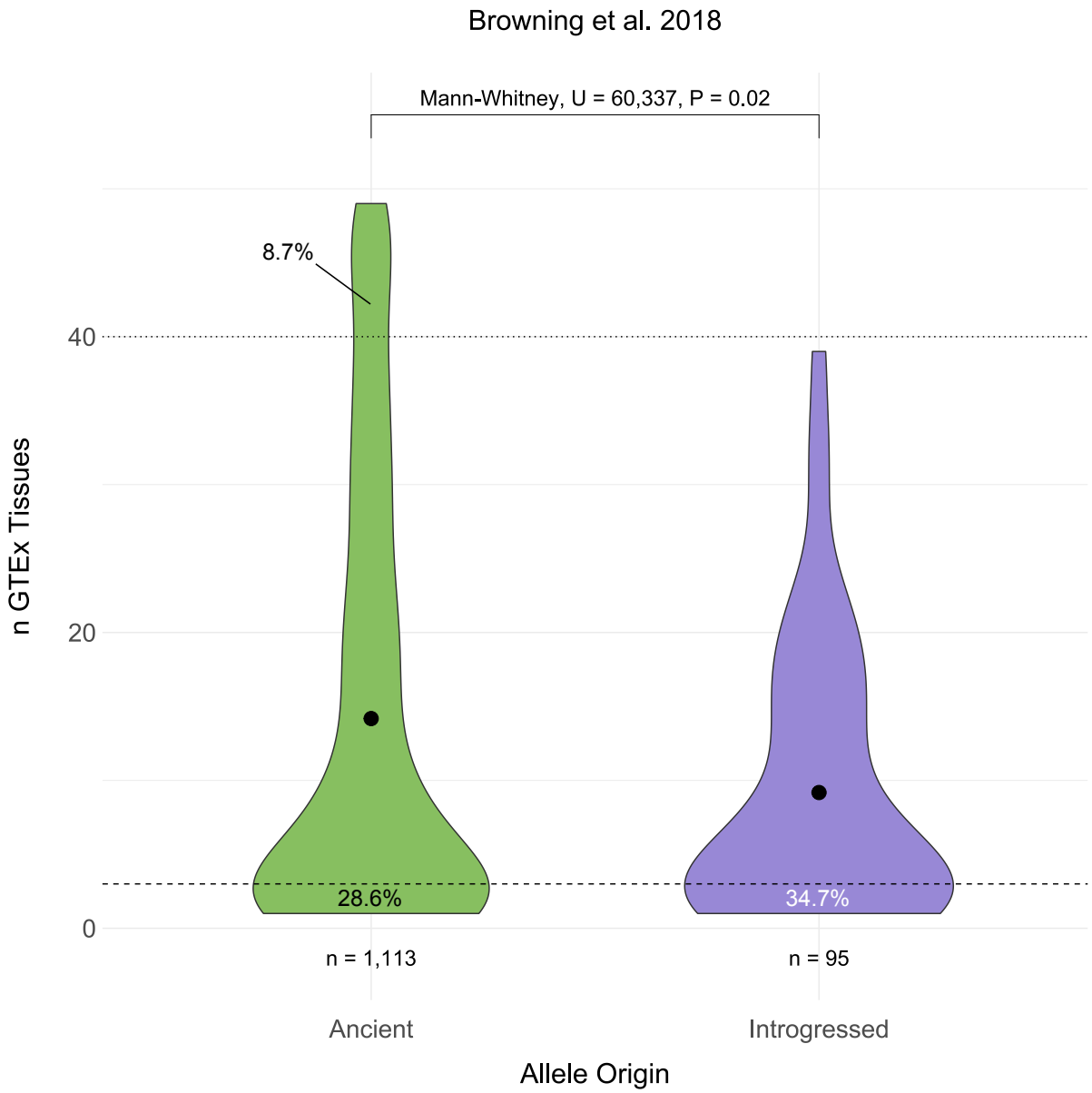
Supplementary Fig. 14 | Many introgressed SAVs occur in multiple 1KG superpopulations. Unique and shared introgressed SAVs per [39] at $\Delta \geq 0.2$. Allele frequencies are from 1KG. If the introgressed allele was the reference allele, we subtracted the 1KG allele frequency from 1. Each SAV was considered present in a population if the allele frequency was > 0.01 (N = 349). The remaining 28 variants occurred at ≤ 0.01 allele frequency. AFR = African, AMR = American, EAS = East Asian, EUR = European, SAS = South Asian.



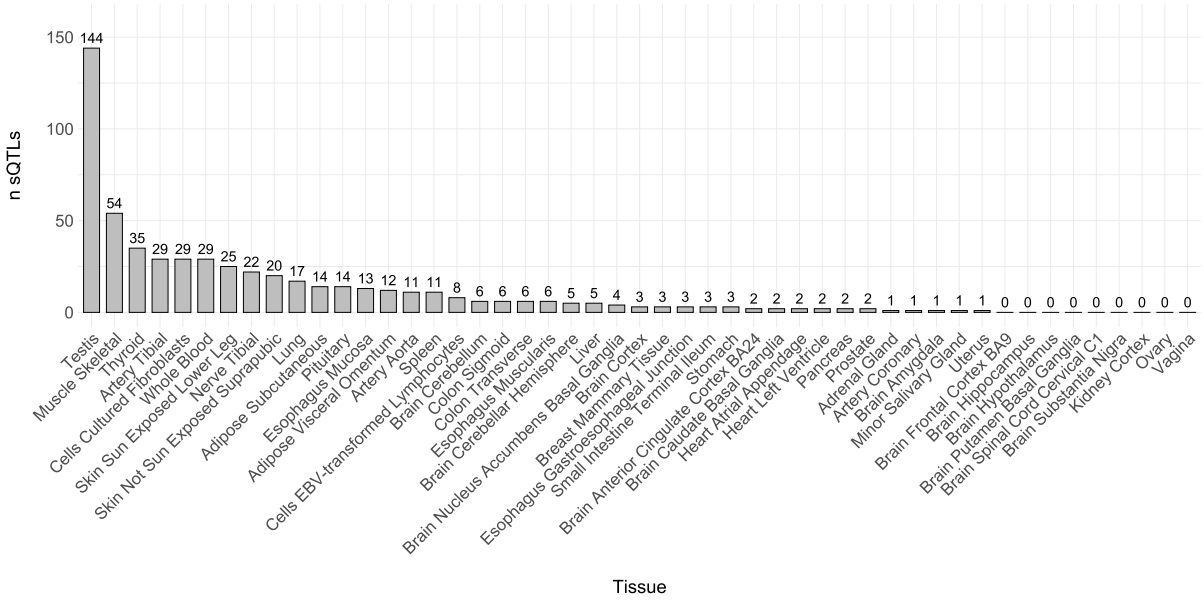
Supplementary Fig. 15 | Many introgressed SAVs occur in multiple 1KG superpopulations. Unique and shared introgressed SAVs per [38] at $\Delta \geq 0.2$. Allele frequencies are from the [38] metadata. Each SAV was considered present in a population if the allele frequency was > 0.01 (N = 203). The remaining 34 variants occurred at ≤ 0.01 allele frequency. AFR = African, AMR = American, EAS = East Asian, EUR = European, SAS = South Asian.



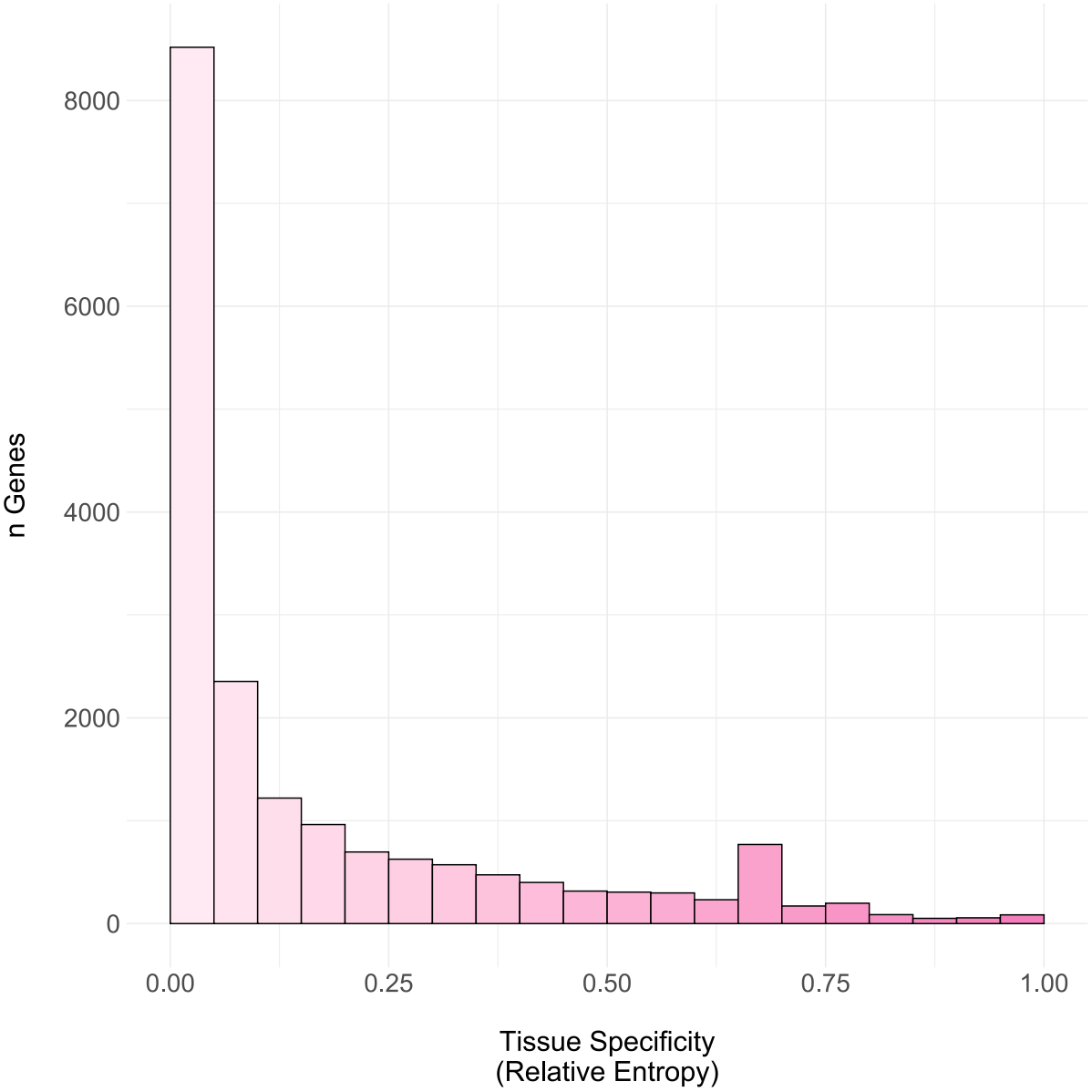
Supplementary Fig. 16 | △ max does not differ between 1KG superpopulations.
(A) △ max for SAVs by 1KG superpopulation and allele origin per [39]. AFR = African, AMR = American, EAS = East Asian, EUR = European, SAS = South Asian. **(B)** △ max for SAVs by 1KG superpopulation and allele origin per [38].



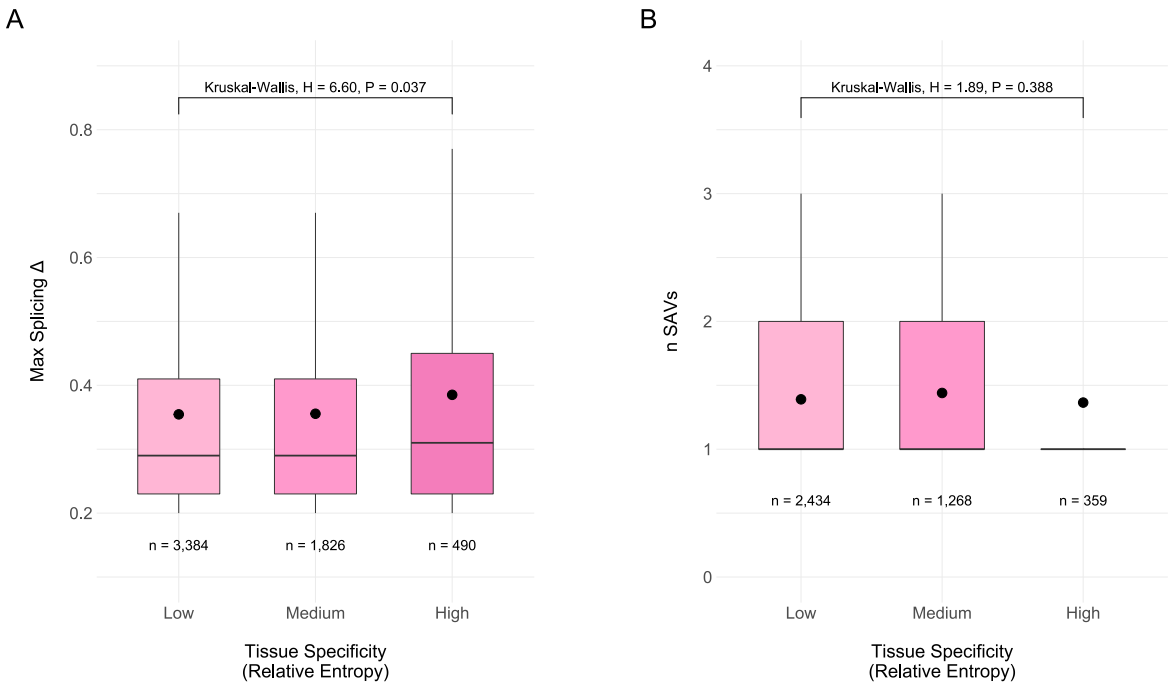
Supplementary Fig. 17 | Introgressed sQTL SAVs are more tissue-specific than ancient variants. The distribution of GTEx tissues in which an ancient or introgressed SAV per [39] was identified as an sQTL. We defined "tissue-specific" variants as those occurring in 1 or 2 tissues and "core" sQTLs as those occurring in > 40 of the 49 tissues. The dashed and dotted lines represent these definitions, respectively. The proportion of SAVs below and above these thresholds for both origins are annotated.



Supplementary Fig. 18 | sQTL SAVs predominantly occur in testis, muscle skeletal, thyroid, tibial artery, fibroblasts, skin, and tibial nerve tissues. The number of tissue-specific (n GTEx tissues = 1 or 2) sQTL SAVs per GTEx tissue.

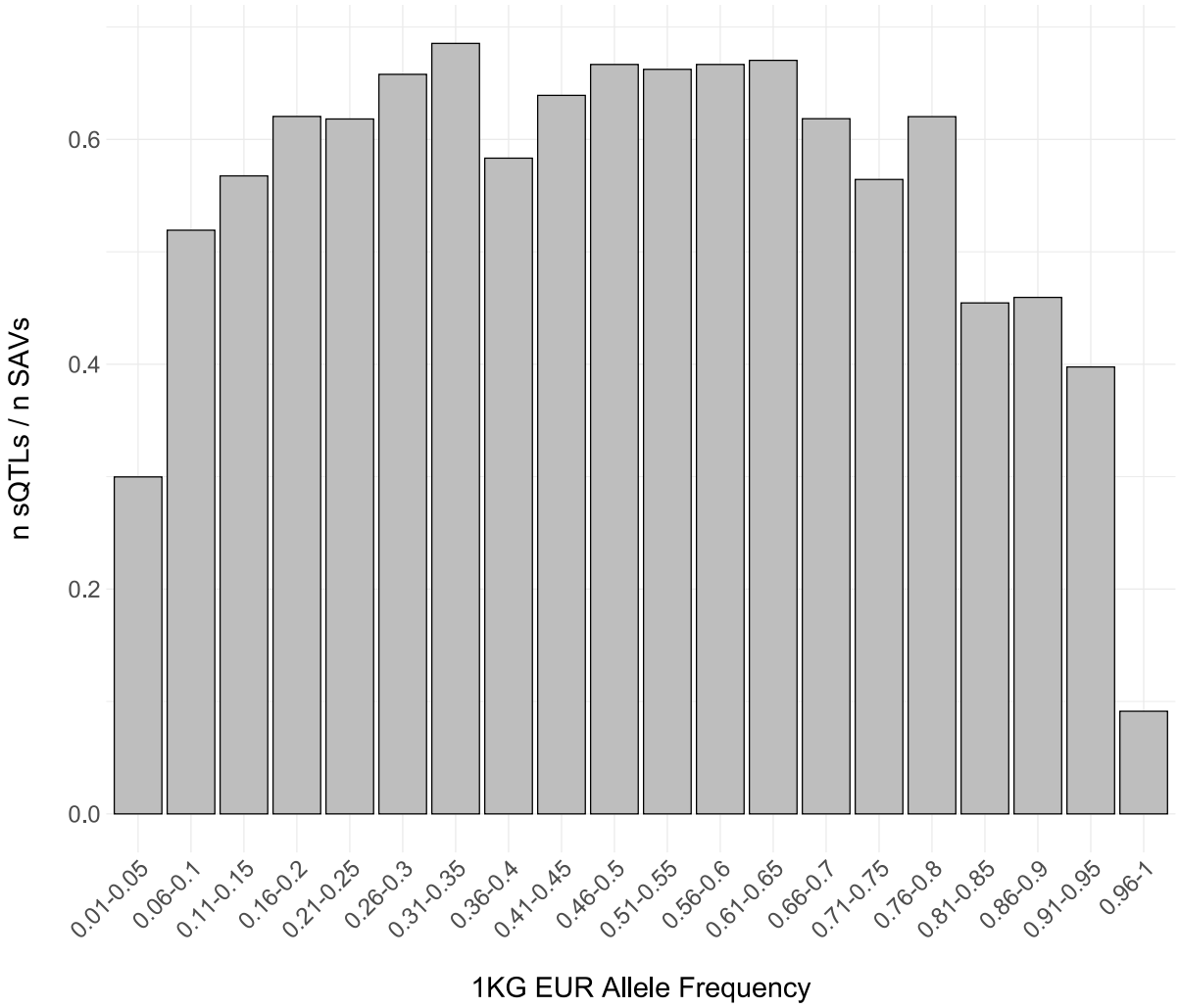


Supplementary Fig. 19 | Most gene expression is not tissue-specific. The distribution of relative entropy in 0.05 bins calculated from GTEx TPM counts across 34 tissues for 18,392 genes.

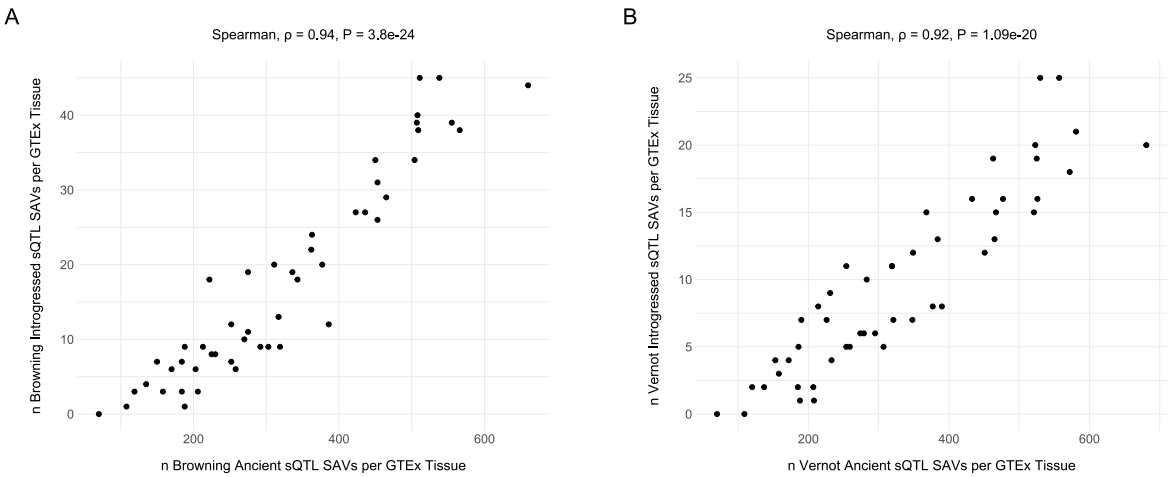


Supplementary Fig. 20 | Maximum splicing probability, but not the number of SAVs, is related to tissue-specific gene expression.

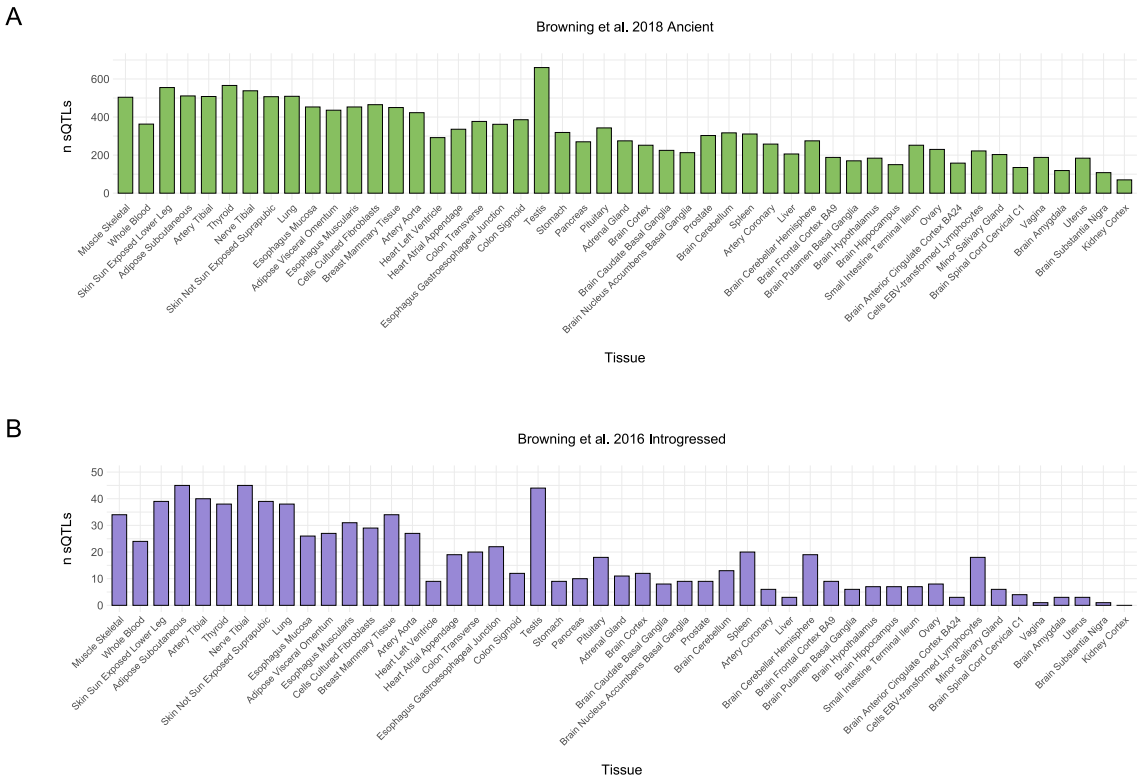
(A) The distributions of maximum splice altering probability (Δ) for SAVs in 4,061 genes binned by tissue-specificity of expression. We quantified the tissue-specificity of each gene as the relative entropy of expression levels across 34 tissues from GTEx compared to genes overall. Low tissue specificity reflects relative entropy ≤ 0.1, medium tissue specificity reflects relative entropy > 0.1 and ≤ 0.5, and high tissue specificity reflects relative entropy > 0.5. (**Supplementary Fig. 19**). The boxplots display the median and IQR, with the upper whiskers extending to largest value $\leq$ 1.5 x IQR from the 75th percentile and the lower whiskers extending to the smallest values $\leq$ 1.5 x IQR from the 25th percentile. **(B)** The number of SAVs by binned gene expression calculated from GTEx TPM counts across 34 tissues. We analyzed a total of 4,061 genes with SAVs. Tissue specificity categories were created and data visualized as in **A**.



Supplementary Fig. 21 | $>$ 50% of SAVs are sQTLs. The proportion of sQTLs also identified as SAVs per binned allele frequencies among 1KG European individuals. We used the precalculated frequencies provided in the 1KG VCFs for this analysis.

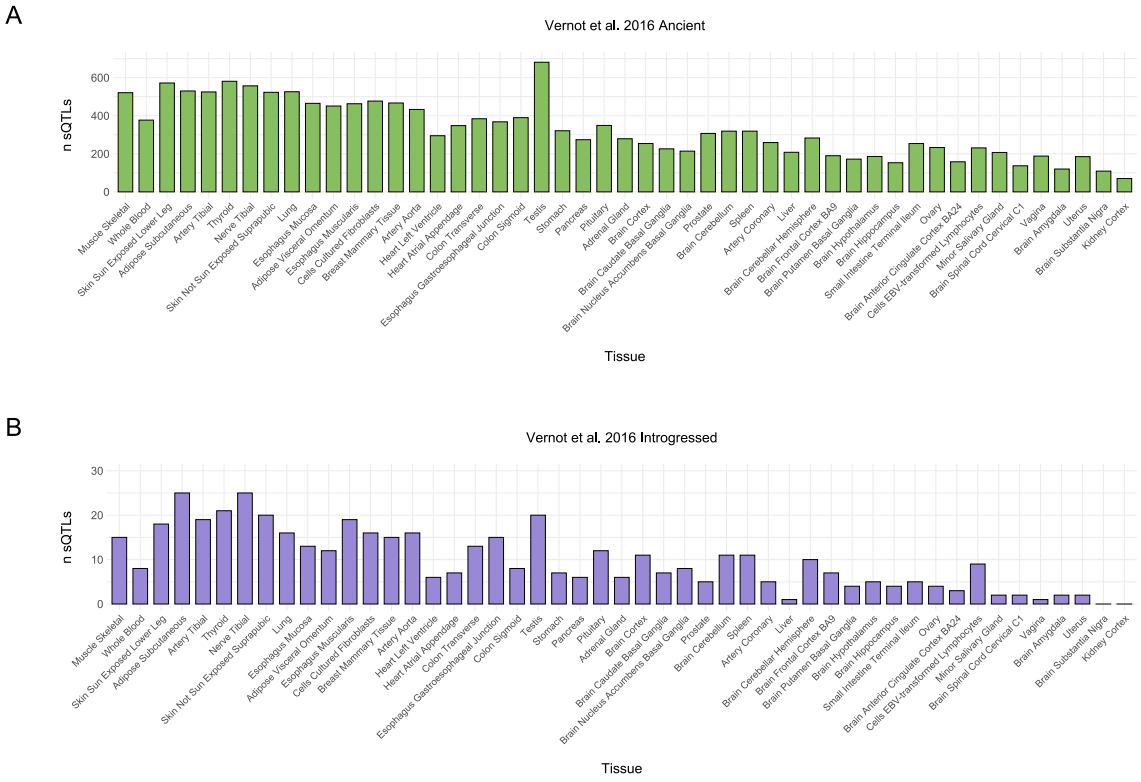


Supplementary Fig. 22 | The number of ancient and introgressed sQTL SAVs are associated among GTEx tissues.
(A) The number of [39] ancient vs. introgressed sQTL SAVs per tissue in GTEx. **(B)** The number of [38] ancient vs. introgressed sQTL SAVs per tissue in GTEx.

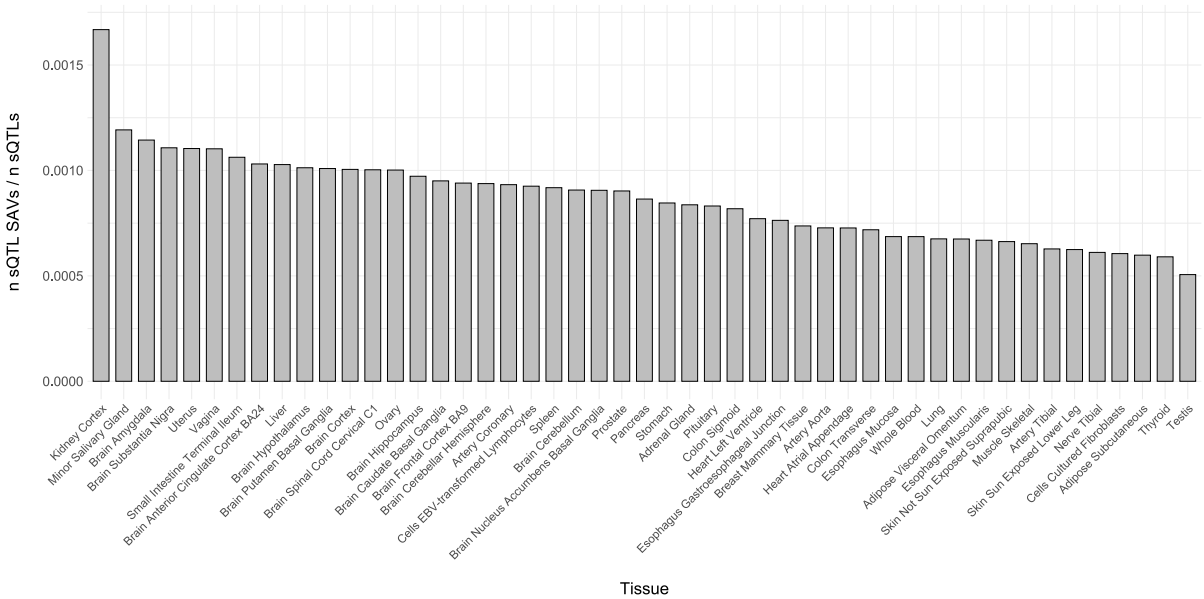


Supplementary Fig. 23 | The number of Browning et al. 2018 sQTL SAVs is associated with GTEx sample size.

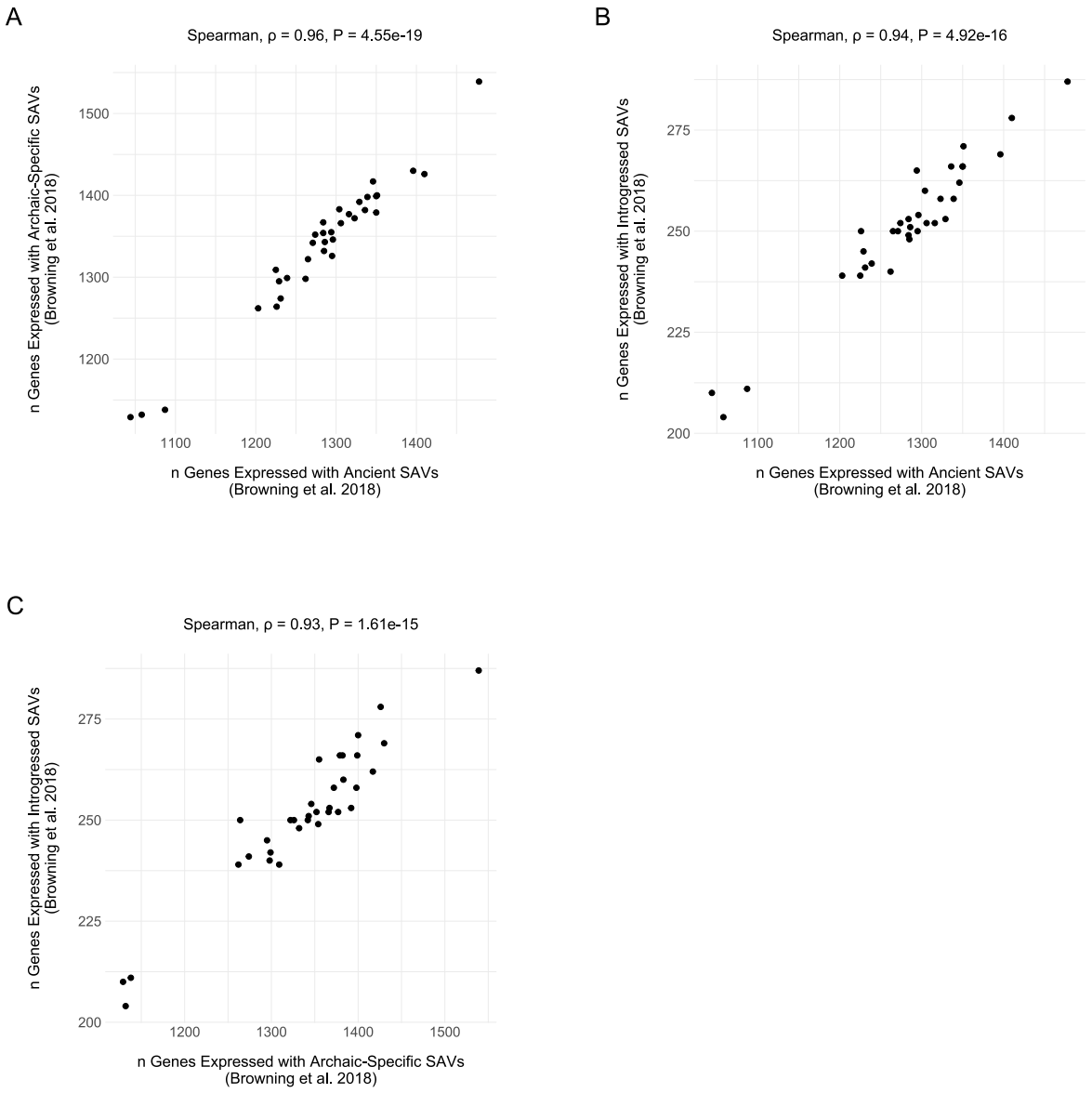
(A) The number of sQTL SAVs per tissue among [39] ancient SAVs ordered by decreasing GTEx sample size. **(B)** The number of sQTL SAVs per tissue among [39] introgressed SAVs ordered by decreasing GTEx sample size.



Supplementary Fig. 24 | The number of Vernot et al. 2016 sQTL SAVs is associated with GTEx sample size.
(A) The number of sQTL SAVs per tissue among [38] ancient SAVs ordered by decreasing GTEx sample size. **(B)** The number of sQTL SAVs per tissue among [38] introgressed SAVs ordered by decreasing GTEx sample size.

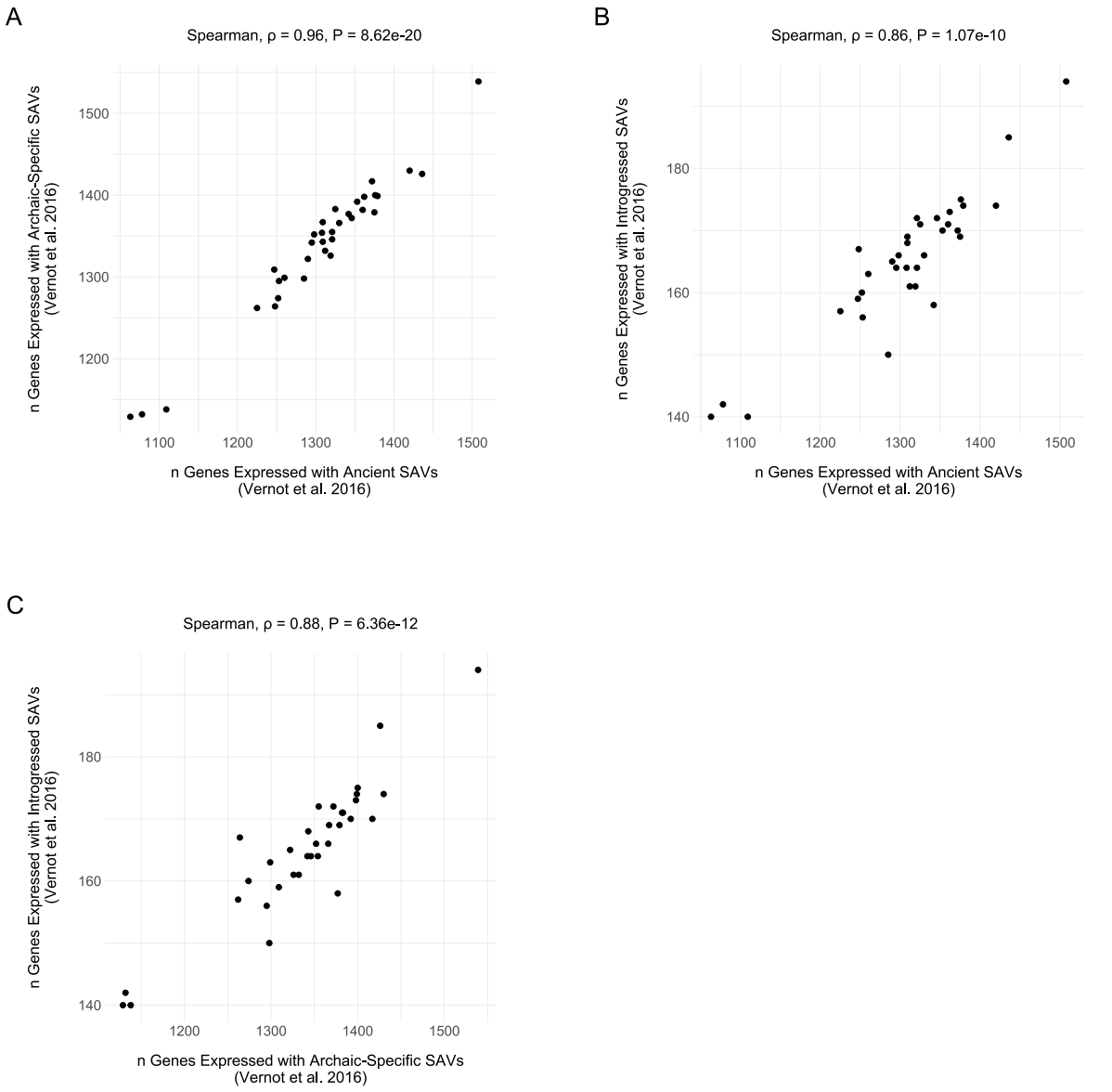


Supplementary Fig. 25 | The proportion of sQTL SAVs is similar across tissues. The proportion of sQTLs SAVs from all sQTLs per GTEx tissue, ordered by decreasing proportion.



Supplementary Fig. 26 | The number of genes expressed per tissue are similar across different [39] allele origins.

(A) The number of genes with TPM > 1 containing at least one SAV for [39] ancient and archaic-specific SAVs. Each dot represents a GTEx tissue. **(B)** The number of genes with TPM > 1 containing at least one SAV for [39] ancient and introgressed SAVs. **(C)** The number of genes with TPM > 1 containing at least one SAV for [39] archaic-specific and introgressed SAVs.



Supplementary Fig. 27 | The number of genes expressed per tissue are similar across different [38] allele origins.

(A) The number of genes with TPM > 1 containing at least one SAV for [38] ancient and archaic-specific SAVs. Each dot represents a GTEx tissue. **(B)** The number of genes with TPM > 1 containing at least one SAV for [38] ancient and introgressed SAVs. **(C)** The number of genes with TPM > 1 containing at least one SAV for [38] archaic-specific and introgressed SAVs.

5.3 Supplementary Tables

n Non-Reference Alleles	n
1	1,566,233
2	1,660
3	1

Supplementary Table 1 | Distribution of non-reference alleles. The number of variant positions with at least one non-reference allele among the archaics per number of non-reference alleles.

n Annotations	n
1	1,537,451
2	30,448
3	1,154
4	75
5	36
6	15
7	25
8	22
9	36
10	20
11	20
12	15
13	58
14	70
15	34
16	6
17	6
18	32
19	6
20	3
21	24

Supplementary Table 2 | Distribution of multiple annotations. The number of variants with the given number of annotations from GENCODE, Human Release 24, for hg19/GRCH37.

All Variants				
	./.	0/0	1/0	1/1
Altai	6,226	591,371	126,440	883,313
Chagyrskaya	37,156	583,787	98,076	888,331
Denisovan	21,574	546,792	139,774	899,210
Vindija	29,672	570,175	119,918	887,585
Variants with Δ Max $\geq$ 0.2				
	./.	0/0	1/0	1/1
Altai	18	2,227	538	3,167
Chagyrskaya	266	2,202	397	3,085
Denisovan	157	2,122	587	3,084
Vindija	163	2,160	496	3,131
Variath with Δ Max $\geq$ 0.5				
	./.	0/0	1/0	1/1
Altai	4	386	101	558
Chagyrskaya	69	394	61	525
Denisovan	45	376	107	521
Vindija	35	361	108	545

Supplementary Table 3 | Archaic genotype distribution. The number of genotypes per individual for all variants and both delta thresholds. ./. = missing genotypes, 0/0 = homozygous for reference allele, 1/0 = heterozygote, 1/1 = homozygous for alternate allele.

1KG Superpopulation	1KG Population	Sample Name	n SAVs
AFR	ESN	HG00137	3,266
	GWD	HG00338	3,344
	LWK	HG00619	3,320
	MSL	HG01198	3,514
	YRI	HG01281	3,493
AMR	CLM	HG01524	3,363
	MXL	HG01802	3,363
	PEL	HG02142	3,349
	PUR	HG02345	3,385
EAS	CDX	HG02629	4,120
	CHB	HG03060	4,160
	CHS	HG03190	4,142
	JPT	HG03708	3,368
	KHV	HG03711	3,379
EUR	CEU	HG03800	3,435
	FIN	HG04014	3,421
	GBR	NA11830	3,200
	GIH	NA18552	3,382
	IBS	NA18868	4,102
	TSI	NA19011	3,321
SAS	BEB	NA19452	4,078
	ITU	NA19741	3,284
	PJL	NA20537	3,168
	STU	NA21141	3,487

Supplementary Table 4 | The average number of SAVs per randomly sampled individual in 1KG. We randomly sampled one individual per 1KG population and ran SpliceAI on their combined variants. There were 4,582,422 total SpliceAI annotated variants among autosomal SNVs and 14,006 variants with a Δ max $\geq$ 0.2. These N SAVs reported here include any site where an individual had at least one alternate allele present (i.e., heterozygotes and homozygotes for the alternate allele). Superpopulation and population labels follow 1KG convention.

n SAVs	n Genes	Genes
1	3,111	
2	769	
3	246	
4	71	
5	19	
6	14	
7	6	*CDH13, CDH23, CSMD1, DNAH17, HLA-DPA1, LAMA5*
8	1	*ADARB2*
9	2	*SDK1, WWOX*
10	1	*HLA-DPB1*
11	2	*CNTNAP2, SSPO*

Supplementary Table 5 | Distribution of SAVs per gene. The distribution of SAVs per gene with the specific genes listed for those with $\geq$ 7 SAVs.

Δ	Gene Characteristic	ρ	P Value
0.2	N Exons	0.3161	$< 2.23 \times 10^{-308}$
	CDS/Exon Length	0.1851	1.07×10^{-135}
	Gene Length	0.2920	$< 2.23 \times 10^{-308}$
	N Isoforms	0.1767	1.87×10^{-113}
0.5	N Exons	0.1330	2.20×10^{-70}
	CDS/Exon Length	0.0625	1.01×10^{-16}
	Gene Length	0.1017	8.99×10^{-42}
	N Isoforms	0.0769	1.24×10^{-22}

Supplementary Table 6 | Physical gene characteristic associations with the number of SAVs per gene. Spearman correlation between four variables: 1) the number of exons, 2) length of coding sequence in bp, 3) gene body length in bp, 4) the number of isoforms and the N SAVs per gene for both Δ thresholds.

Protein Effect	All	Vernot et al. 2016				Browning et al. 2018			
		Ancient	Archaic-Specific	Introgressed	Low-Confidence Ancient	Ancient	Archaic-Specific	Introgressed	Low-Confidence Ancient
No effect	2,391	1,092	729	69	501	1,061	729	148	453
Longer protein w/ PTCs	1,897	634	753	91	419	620	753	133	391
Longer protein w/o PTCs	333	95	153	14	71	90	153	25	65
Longer UTR	479	151	206	26	96	149	206	25	99
Single missense	396	142	137	21	96	139	137	28	92
Truncated protein	332	93	155	18	66	93	155	18	66
Truncated UTR	116	44	48	5	19	42	48	9	17
Unknown	6	1	5	0	0	1	5	0	0

Supplementary Table 7 | SAV effects on resulting protein. The number of SAVs per effect on the resulting transcript and protein. PTC = premature termination codon, UTR = 5' or 3' untranslated region.

Delta	n Unique SAVs	n Unique Non SAVs	n Shared SAVs	n Shared Non SAVs	OR	95% CI Lower Bound	95% CI Upper Bound	P Value	Bonferroni
0.2	2,416	615,666	1,933	571,264	1.160	1.092	1.231	0.000	Y
0.3	1,175	616,907	945	572,252	1.153	1.059	1.257	0.001	Y
0.4	685	617,397	533	572,664	1.192	1.064	1.335	0.002	Y
0.5	437	617,645	328	572,869	1.236	1.071	1.426	0.004	Y

Supplementary Table 8 | Enrichment tests for the sum of lineage-specific SAVs at different Δs. Input data for Fisher's exact tests. OR = odds ratio; 95% CI Lower Bound and 95% CI Upper Bound = the lower and upper bounds of the 95% confidence interval, respectively; P Value = unadjusted p-value.

Δ	Distribution	n Unique SAVs	n Unique Non SAVs	n Shared SAVs	n Shared Non SAVs	OR	95% CI Lower Bound	95% CI Upper Bound	P Value	Bonferroni
0.2	Altai	399	81,517	1,933	571,264	1.447	1.298	1.612	0.000	Y
	Chagyrskaya	218	53,457	1,933	571,264	1.205	1.047	1.386	0.011	Y
	Denisovan	1,492	410,000	1,933	571,264	1.075	1.005	1.151	0.036	N
	Vindija	307	70,692	1,933	571,264	1.283	1.138	1.448	0.000	Y
	Total	2,416	615,666	1,933	571,264	1.160	1.092	1.231	0.000	NA
0.5	Altai	75	81,841	328	572,869	1.601	1.245	2.057	0.000	Y
	Chagyrskaya	35	53,640	328	572,869	1.140	0.804	1.615	0.452	N
	Denisovan	254	411,238	328	572,869	1.079	0.916	1.271	0.377	N
	Vindija	73	70,926	328	572,869	1.798	1.395	2.317	0.000	Y
	Total	437	617,645	328	572,869	1.236	1.071	1.426	0.004	NA

Supplementary Table 9 | Enrichment tests for lineage-specific SAVs. Input data for Fisher's exact tests. OR = odds ratio; 95% CI Lower Bound and 95% CI Upper Bound = the lower and upper bounds of the 95% confidence interval, respectively; P Value = unadjusted p-value.

Δ	≥ 0.2			≥ 0.5		
Distribution	Archaic-Specific n SAVs	Browning Introgressed n SAVs	Vernot Introgressed n SAVs	Archaic-Specific n SAVs	Browning Introgressed n SAVs	Vernot Introgressed n SAVs
Altai	300	17	7	57	4	1
Chagyrskaya	169	0	0	27	0	0
Denisovan	916	2	5	165	2	1
Late Neanderthal	94	0	0	10	0	0
Neanderthal	236	207	143	36	38	38
Other	128	38	21	27	4	5
Shared	112	122	68	27	23	14
Vindija	231	0	0	55	0	0

Supplementary Table 10 | Distribution of archaic-specific and introgressed SAVs. The number of SAVs for both Δ thresholds among archaic-specific and introgressed SAVs per [39] and [38]. Other denotes any combination of archaics not already listed.

Chromosome	Position	Reference Allele	Alternate Allele	Delta Max	Browning Allele Origin	Vernot Allele Origin	Annotation
chr1	22,174,518	G	T	0.98	introgressed	introgressed	*HSPG2*
chr1	55,537,474	C	G	0.33	low-confidence ancient	introgressed	*USP24*
chr1	161,681,848	C	T	0.20	introgressed	introgressed	*FCRLA*
chr1	212,985,592	G	A	0.52	introgressed	introgressed	*TATDN3*
chr11	86,159,859	G	A	0.26	ancient	introgressed	*ME3*
chr12	133,272,470	G	T	0.26	introgressed	introgressed	*PXMP2*
chr12	133,272,470	G	T	0.26	introgressed	introgressed	*RP13-672B3.2*
chr15	85,403,496	G	A	0.33	introgressed	introgressed	*ALPK3*
chr16	88,924,425	C	G	0.41	introgressed	introgressed	*TRAPPC2L*
chr19	40,913,595	G	A	0.23	introgressed	low-confidence ancient	*PRX*
chr22	50,684,852	G	C	0.25	introgressed	introgressed	*HDAC10*
chr4	170,990,750	G	A	0.22	introgressed	introgressed	*AADAT*
chr5	6,753,013	C	T	0.24	introgressed	introgressed	*PAPD7*
chr5	167,919,825	G	A	0.38	introgressed	introgressed	*RARS*
chr6	33,264,115	A	C	0.34	low-confidence ancient	introgressed	*RGL2*
chr8	130,763,783	C	T	0.87	introgressed	introgressed	*GSDMC*

Supplementary Table 11 | SAVs that exhibit allele-specific expression in modern humans. 16 SAVs from our dataset that matched variants from [59].

Distribution	n sQTLs
Altai	2
Chagyrskaya	3
Denisovan	8
Late Neanderthal	2
Neanderthal	9
Other	14
Shared	74

Supplementary Table 12 | Core sQTL SAV distribution among the archaics. The distribution of core sQTL SAVs (n = 1,145) among the archaics. Core sQTLs were defined as those variants detected in $>$ 40 tissues. All the variants represented here are either low or high-confidence ancient.

Chromosome	Position	Reference Allele	Alternate Allele	Delta Max Alternate Allele	Delta Max Reference as Alternate
chr1	52,820,680	C	T	0.23	0.23
chr1	152,944,501	A	T	0.23	0.23
chr1	183,750,131	C	A	0.26	0.26
chr1	**216,945,756**	**C**	**A**	**0.20**	**0.19**
chr10	17,373,518	T	C	0.22	0.22
chr11	84,191,111	T	C	0.40	0.40
chr11	99,345,359	C	G	0.67	0.67
chr17	14,109,360	T	G	0.20	0.21
chr18	60,003,587	G	A	0.31	0.30
chr19	33,410,289	T	C	0.71	0.71
chr20	62,224,595	G	A	0.41	0.41
chr3	52,486,376	T	C	0.34	0.35
chr4	38,805,942	G	C	0.38	0.37
chr4	100,460,531	A	G	0.35	0.35
chr4	135,122,662	G	A	0.80	0.80
chr5	118,669,326	G	C	0.22	0.22
chr5	126,989,007	A	G	0.91	0.91
chr5	148,430,560	A	C	0.32	0.32
chr6	51,889,358	T	C	0.42	0.42
chr6	52,138,226	C	G	0.21	0.21
chr7	119,524,266	C	T	0.30	0.30
chr7	126,797,408	T	A	0.43	0.43
chr7	140,801,296	T	C	0.25	0.25
chr7	140,958,656	A	G	0.26	0.26
chr7	**157,177,273**	**C**	**T**	**0.31**	**0.16**
chr7	158,543,832	C	A	0.21	0.22

Supplementary Table 13 | Maximum Δ for SAVs with an introgressed reference allele. Δ maximum for 26 SAVs whose reference allele rather than the alternate allele matched an introgressed tag SNP from [38]. We switched the reference and alternate alleles for these (and non-SAV) loci and reran SpliceAI. Delta Max Alternate Allele = the originally predicted Δ where the reference allele is introgressed, Delta Max Reference as Alternate = the Δ when the reference and alternate alleles are switched. Bolded variants are those whose values did not pass the SAP threshold after switching alleles.

1KG Superpopulation	1KG Population	Sample	n Transcripts	n Variants	n Variants PolyPhen	n Variants SIFT	n Variants Both
		Altai	34	19	2	5	2
		Chagyrskaya	27	15	1	2	0
		Denisovan	43	25	2	5	0
		Vindija	24	14	1	1	0
AFR	ESN	HG00137	71	23	3	5	2
	GWD	HG00338	57	23	1	3	1
	LWK	HG00619	68	27	2	2	1
	MSL	HG01198	67	28	2	6	1
	YRI	HG01281	56	22	2	2	1
AMR	CLM	HG01524	69	31	3	7	3
	MXL	HG01802	42	21	0	3	0
	PEL	HG02142	49	19	0	1	0
	PUR	HG02345	43	22	0	3	0
EAS	CDX	HG02629	70	26	1	6	1
	CHB	HG03060	59	23	1	6	0
	CHS	HG03190	56	23	1	4	1
	JPT	HG03708	66	24	0	3	0
	KHV	HG03711	52	25	1	5	1
EUR	CEU	HG03800	55	17	1	3	1
	FIN	HG04014	59	24	1	2	1
	GBR	NA11830	83	26	1	5	1
	GIH	NA18552	56	24	1	5	1
	IBS	NA18868	53	21	0	3	0
	TSI	NA19011	59	25	0	4	0
SAS	BEB	NA19452	65	27	2	6	1
	ITU	NA19741	83	29	1	7	1
	PJL	NA20537	41	23	1	4	1
	STU	NA21141	56	23	1	5	1

Supplementary Table 14 | Missense variants among 147 spliceosome genes. The number of variant positions with at least one non-reference allele among all four archaics and 24 randomly sampled 1KG samples that was identified as a missense variant by Ensembl's Variant Effect Predictor. Superpopulation and population labels follow 1KG convention. n transcripts = number of transcripts with a missense variant, n Variants = number of missense variant, n Variants PolyPhen = number of damaging variants as measured by PolyPhen, n Variants SIFT = number of deleterious variants as measured by SIFT, n Variants Both = number of variants that were considered both damaging and deleterious by PolyPhen and SIFT, respectively.

Tissue	n Vernot Ancient	n Vernot Introgressed	n Browning Ancient	n Browning Introgressed	GTEx Sample Size
Adipose Subcutaneous	530	25	511	45	663
Adipose Visceral Omentum	451	12	436	27	541
Adrenal Gland	279	6	275	11	258
Artery Aorta	433	16	423	27	432
Artery Coronary	259	5	258	6	240
Artery Tibial	525	19	508	40	663
Brain Amygdala	120	2	119	3	152
Brain Anterior cingulate cortex BA24	158	3	158	3	176
Brain Caudate basal ganglia	226	7	225	8	246
Brain Cerebellar Hemisphere	283	10	275	19	215
Brain Cerebellum	319	11	317	13	241
Brain Cortex	254	11	252	12	255
Brain Frontal Cortex BA9	190	7	188	9	209
Brain Hippocampus	153	4	150	7	197
Brain Hypothalamus	186	5	184	7	202
Brain Nucleus accumbens basal ganglia	214	8	213	9	246
Brain Putamen basal ganglia	172	4	170	6	205
Brain Spinal cord cervical c-1	137	2	135	4	159
Brain Substantia nigra	109	0	108	1	139
Breast Mammary Tissue	467	15	450	34	459
Cells Cultured fibroblasts	477	16	465	29	504
Cells EBV-transformed lymphocytes	231	9	222	18	174
Colon Sigmoid	390	8	386	12	373
Colon Transverse	384	13	377	20	406
Esophagus Gastroesophageal Junction	368	15	362	22	375
Esophagus Mucosa	465	13	453	26	555
Esophagus Muscularis	463	19	453	31	515
Heart Atrial Appendage	348	7	336	19	429
Heart Left Ventricle	295	6	292	9	432
Kidney Cortex	70	0	70	0	85
Liver	208	1	206	3	226
Lung	526	16	509	38	578
Minor Salivary Gland	207	2	203	6	162
Muscle Skeletal	521	15	504	34	803
Nerve Tibial	557	25	538	45	619
Ovary	233	4	230	8	180
Pancreas	274	6	270	10	328
Pituitary	349	12	343	18	283
Prostate	307	5	303	9	245
Skin Not Sun Exposed Suprapubic	523	20	507	39	604
Skin Sun Exposed Lower leg	572	18	555	39	701
Small Intestine Terminal Ileum	254	5	252	7	187
Spleen	319	11	311	20	241
Stomach	321	7	319	9	359
Testis	681	20	660	44	361
Thyroid	581	21	566	38	653
Uterus	185	2	184	3	142
Vagina	188	1	188	1	156
Whole Blood	377	8	363	24	755

Supplementary Table 15 | sQTL SAVs by allele origin. The number of sQTL SAVs by allele origin for 49 tissues in GTEx, v8.