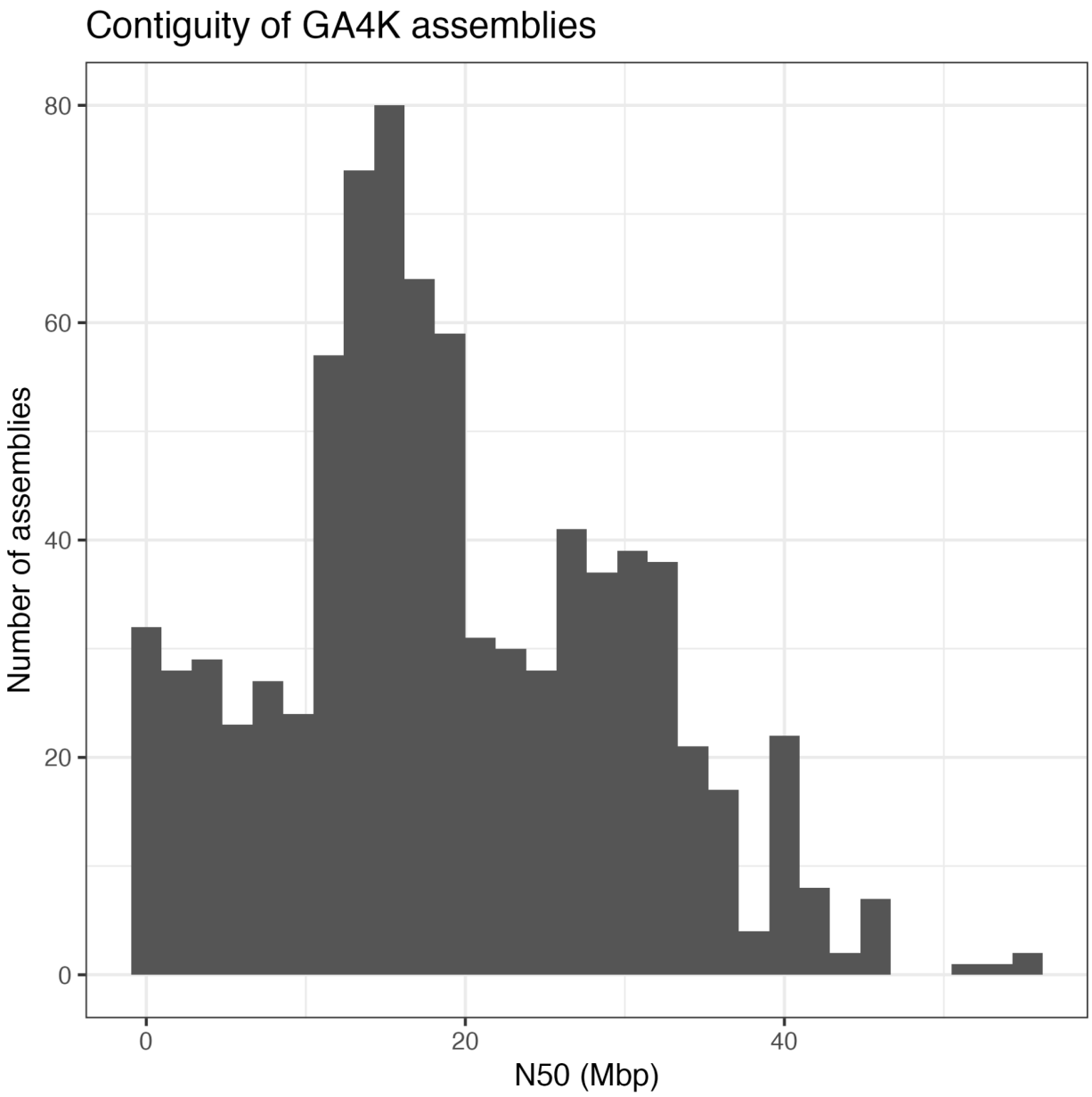

Fig S1: N50 distribution of the 782 GA4K assemblies that were used to construct the pangenome graph with minigraph.

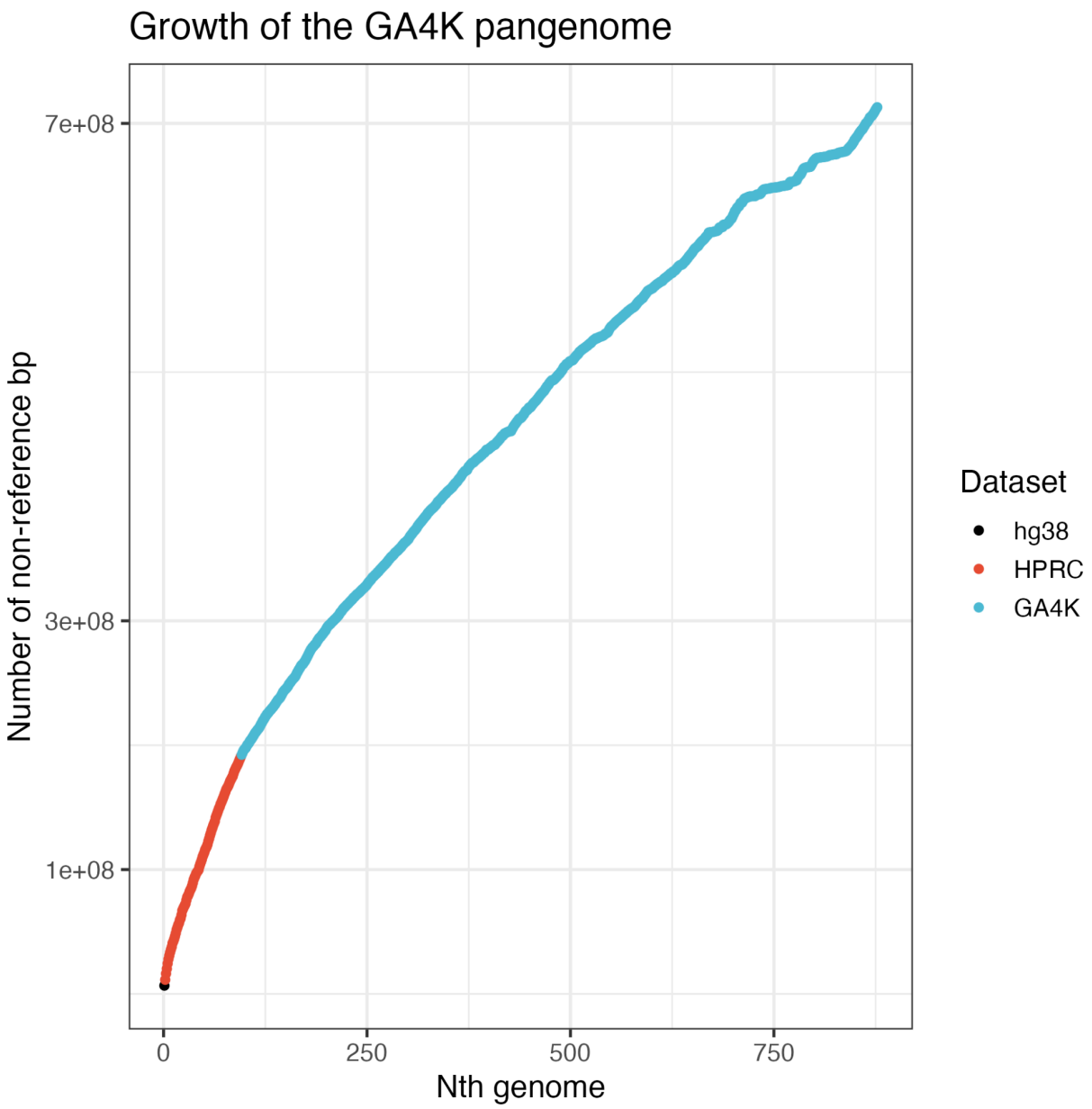

Fig S2: Growth of the pangenome during progressive construction with minigraph. The backbone reference is CHM13v2. Here, hg38 is the first genome added to the pangenome on top of the CHM13v2 backbone reference.

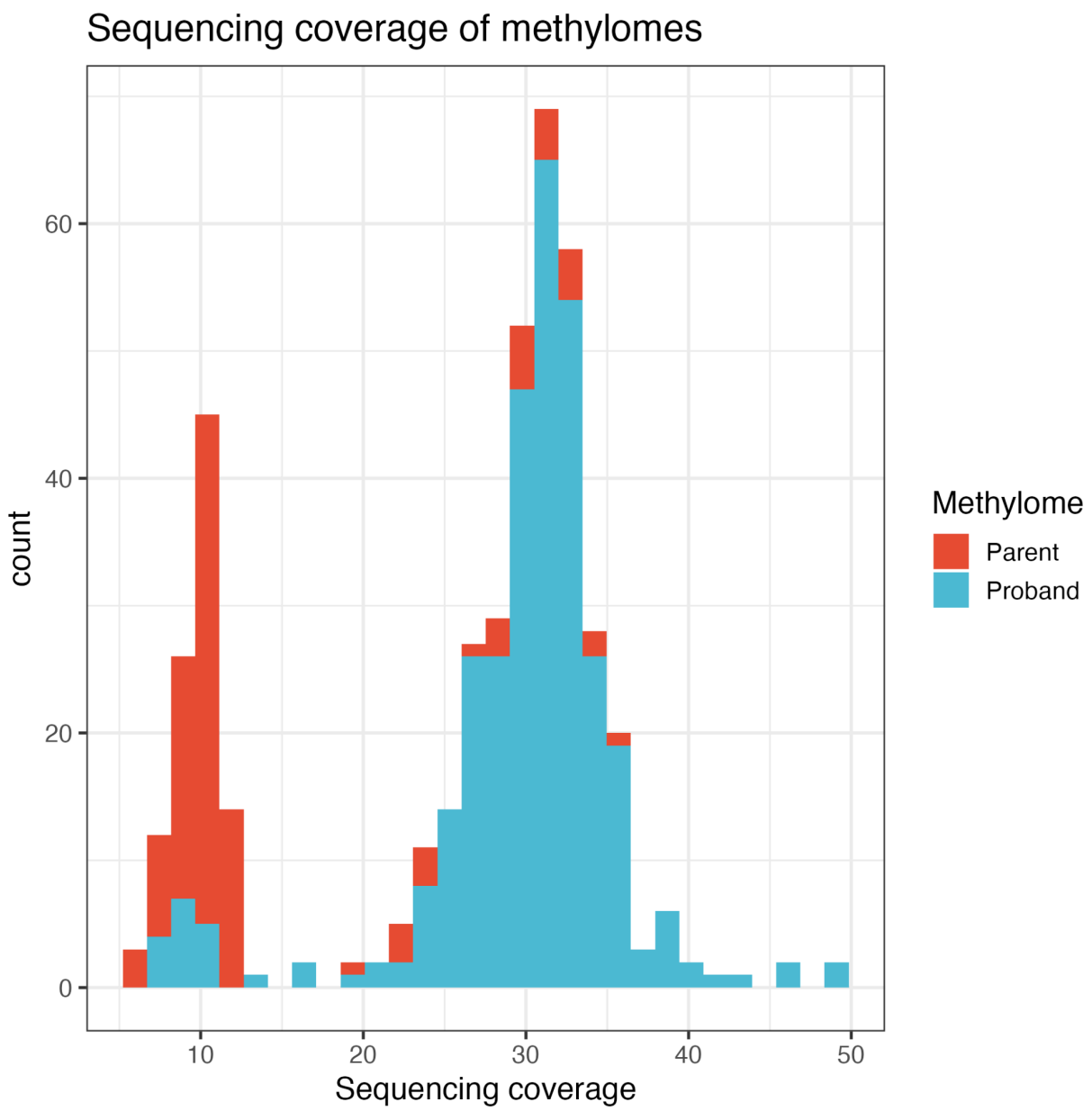

Fig S3: Sequencing depth of the 435 GA4K methylomes that were mapped to the GA4K pangenome with minigraph.

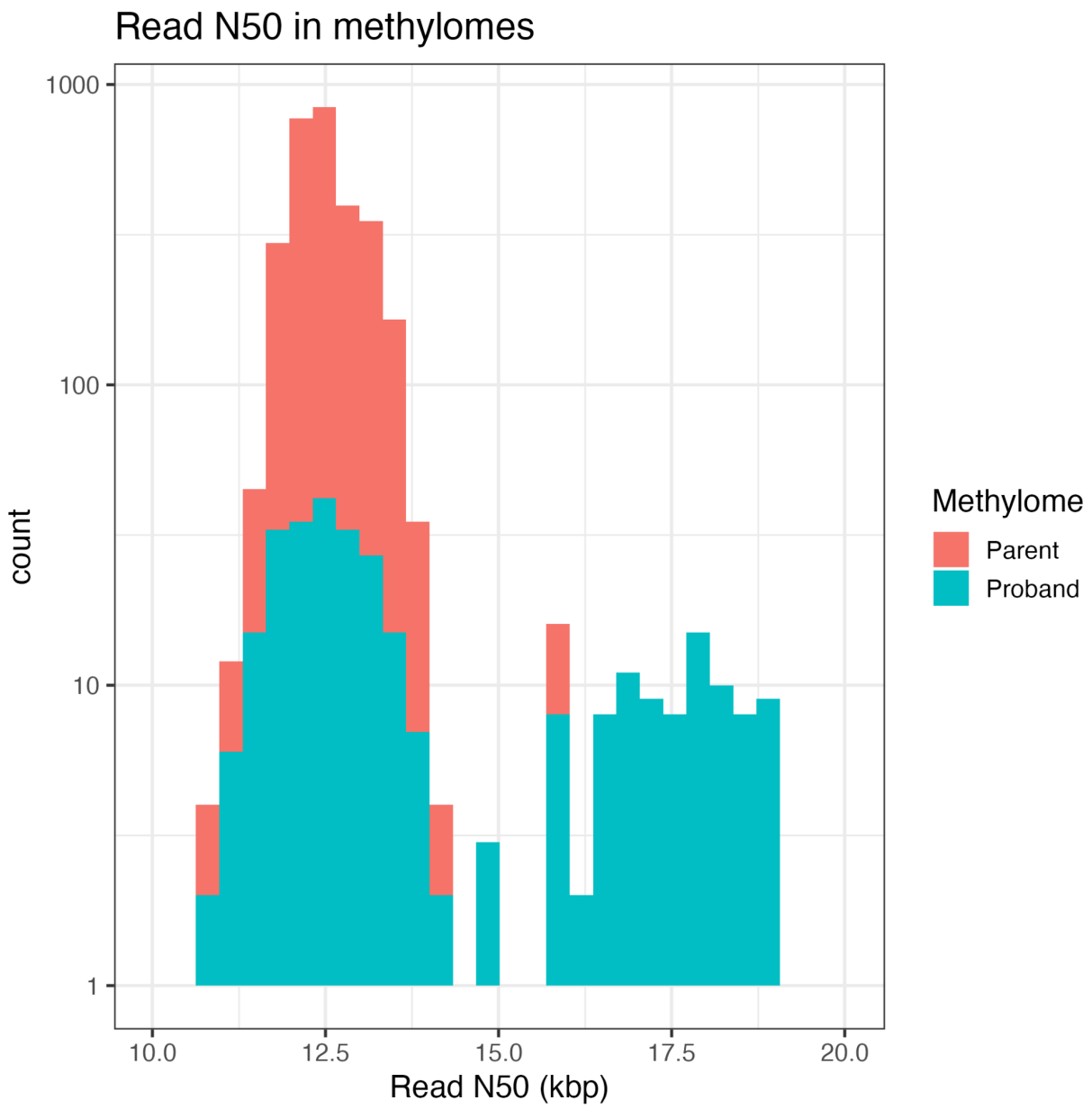

Fig S4: Read N50 distribution of the 435 GA4K methylomes that were mapped to the GA4K pangenome with minigraph.

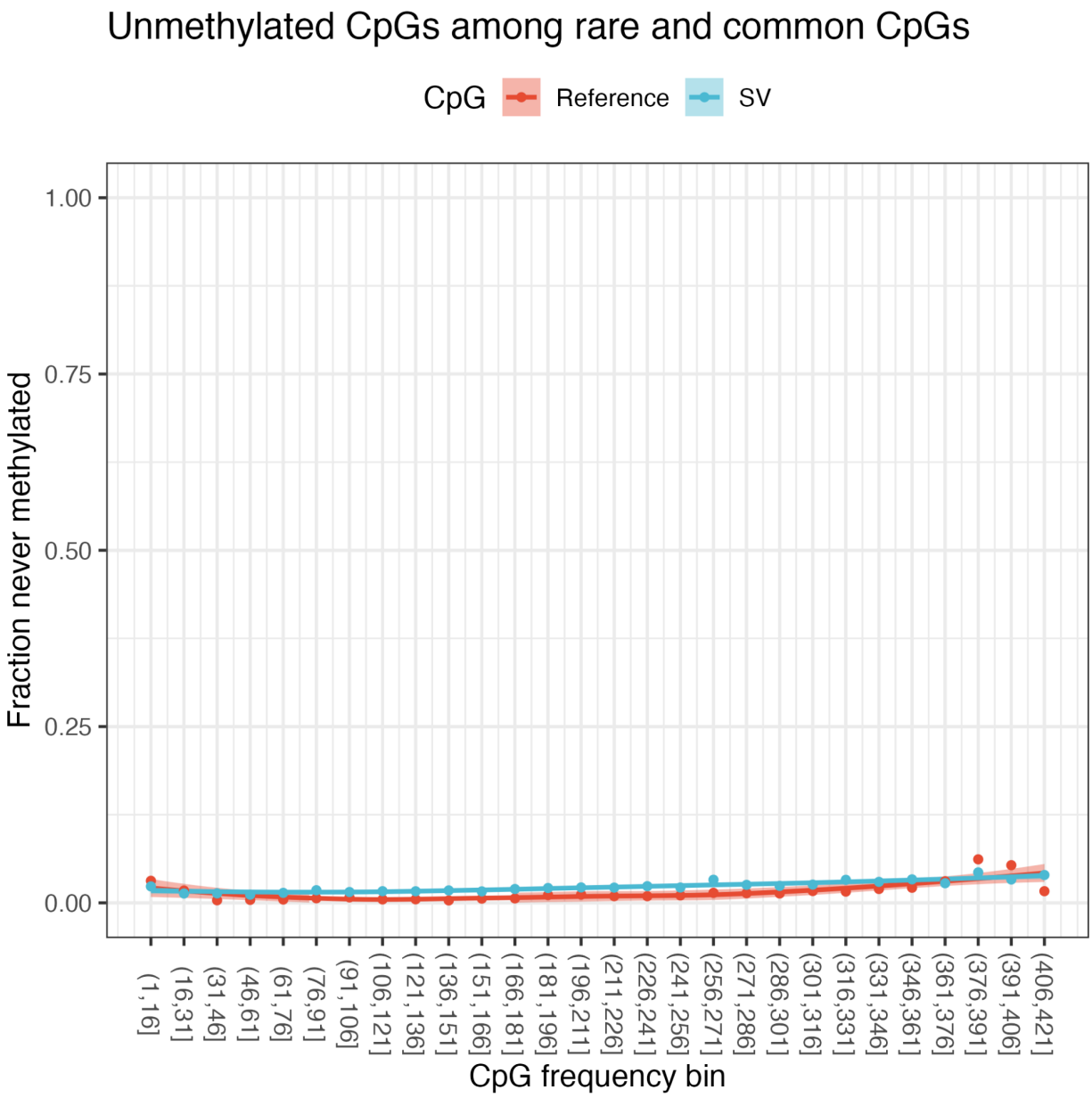

Fig S5: Proportion of CpGs that are never observed in a methylated state, stratified by the CpG frequency, and by SV-CpGs and reference CpGs. Error bands represent the 95% confidence interval as estimated by the loess regression fitted to the points.

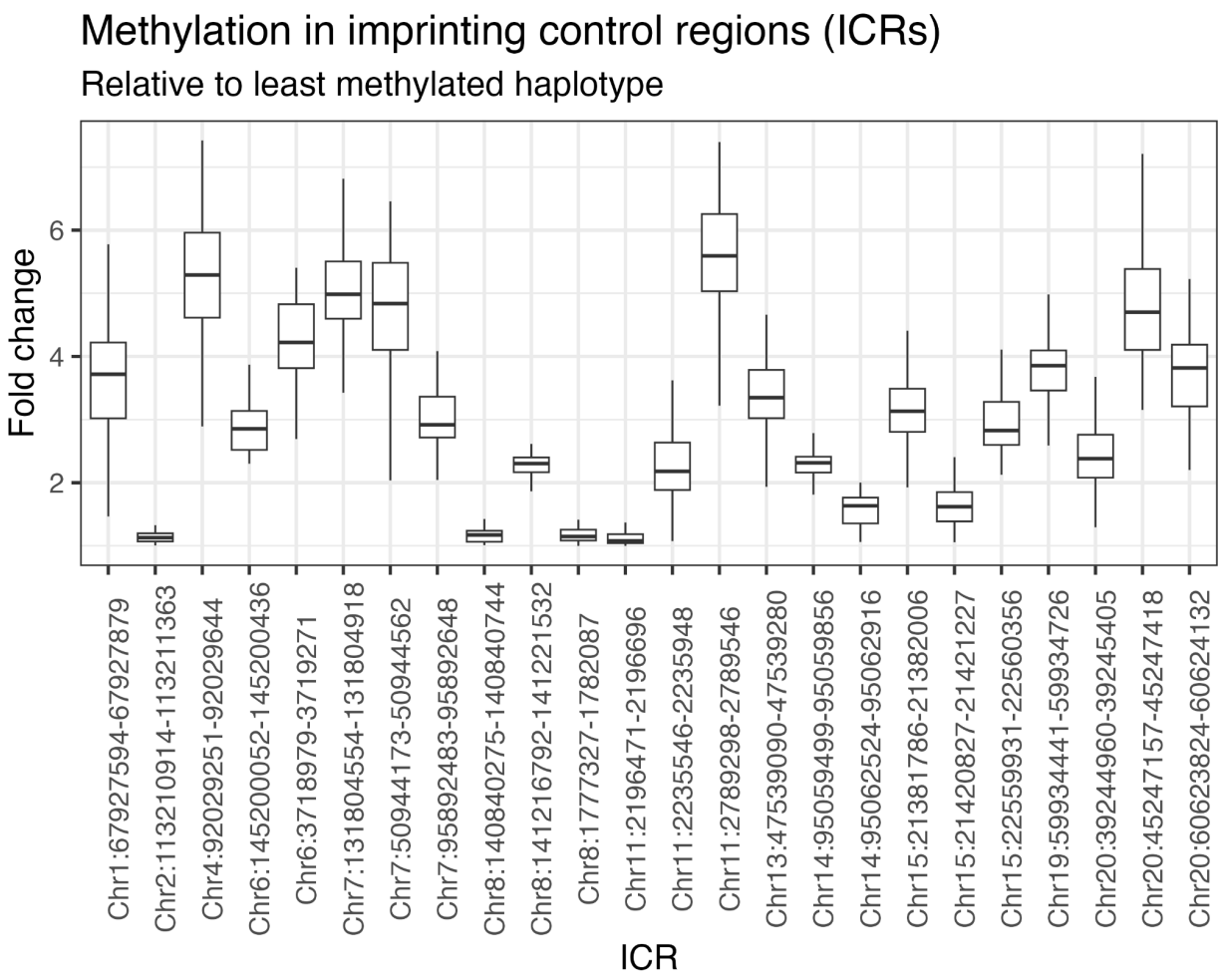

Fig S6: Methylation differences in imprinting control regions observed between the two haplotypes in methylomes as measured by panmethyl in the GA4K pangenome.

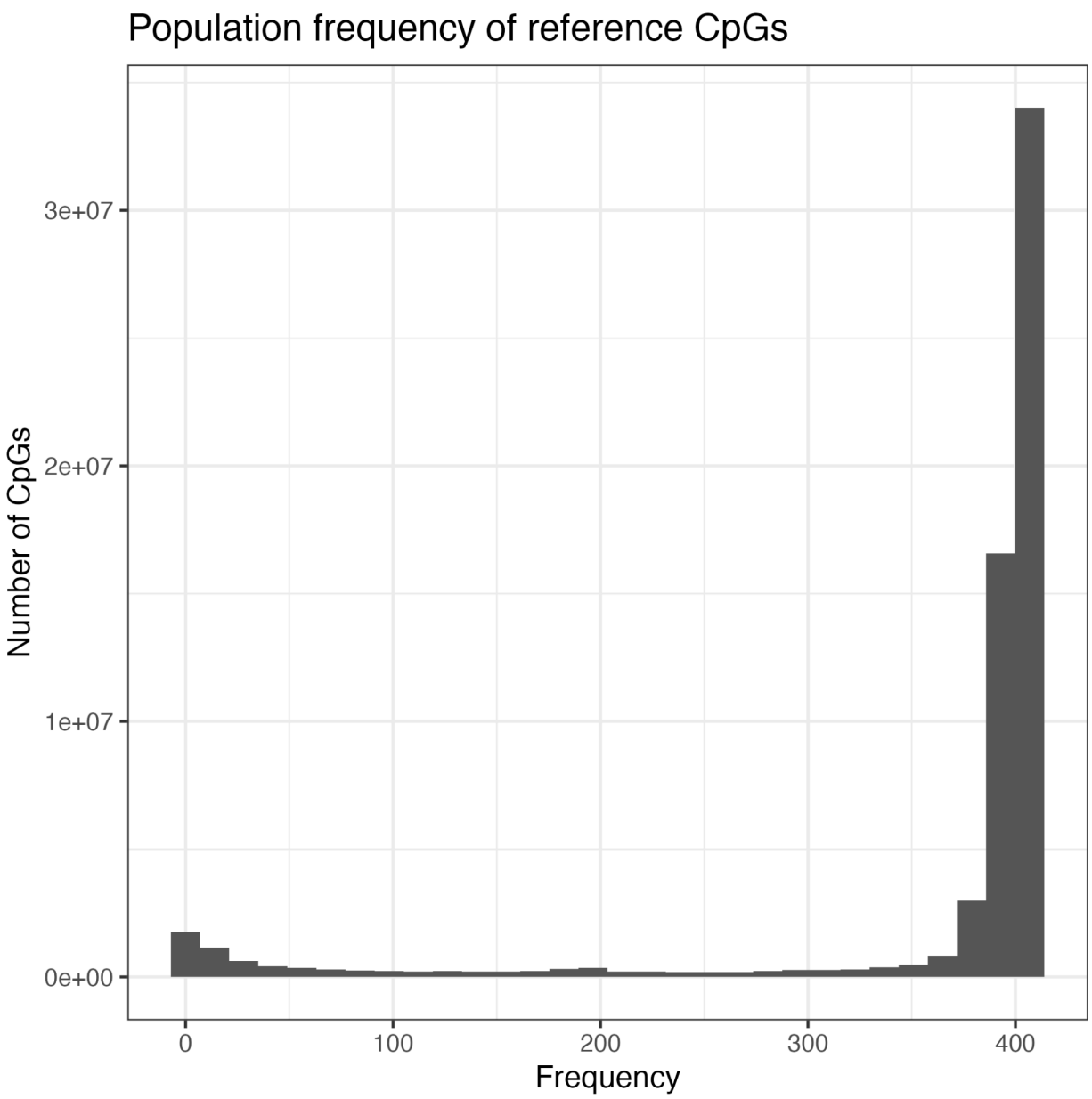

Fig S7: Population frequency of reference CpGs in the 435 GA4K methylomes.

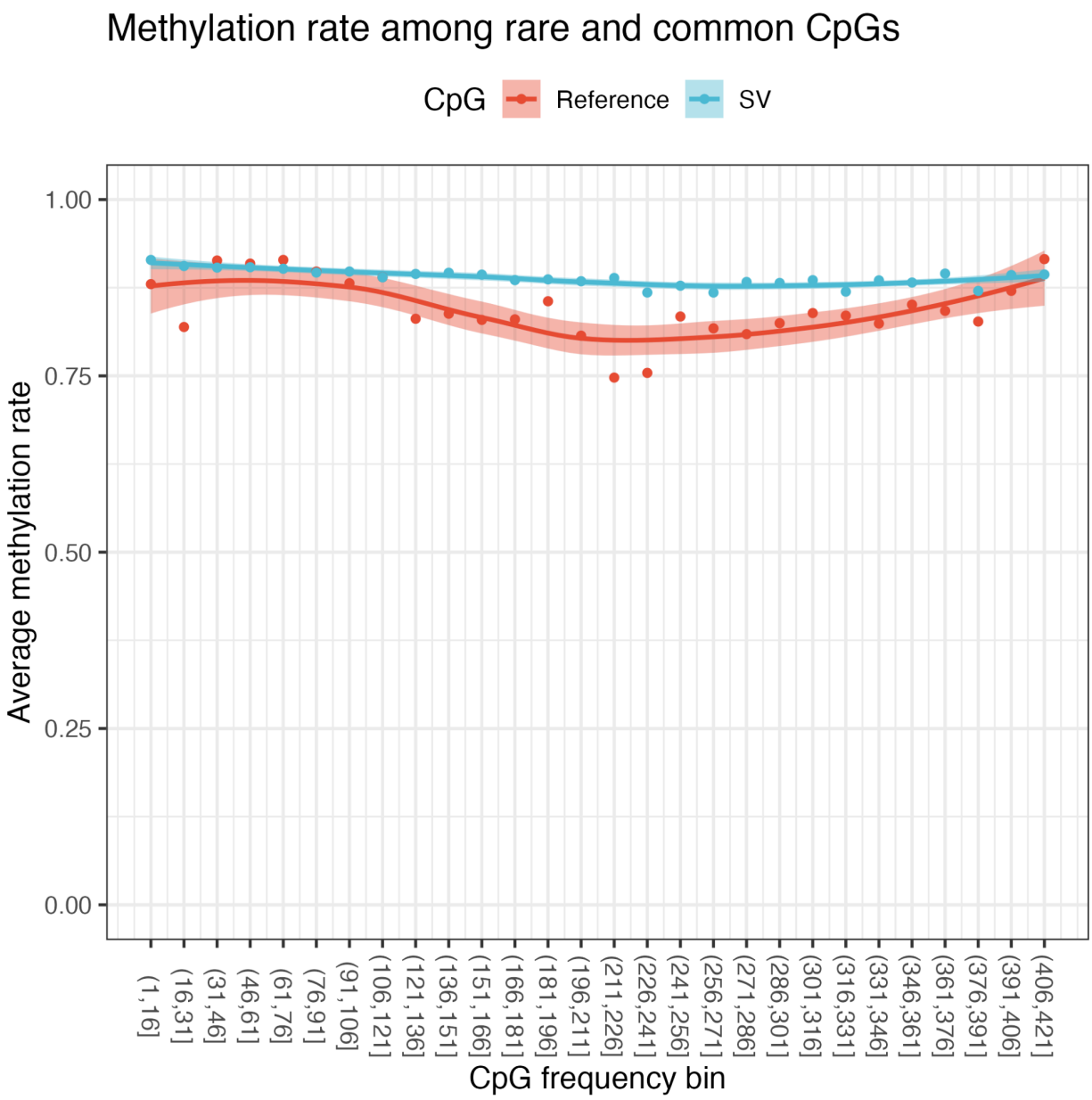

Fig S8: Mean methylation rate of SV and reference CpGs across frequency bins. Error bands represent the 95% confidence interval as estimated by the loess regression fitted to the points.

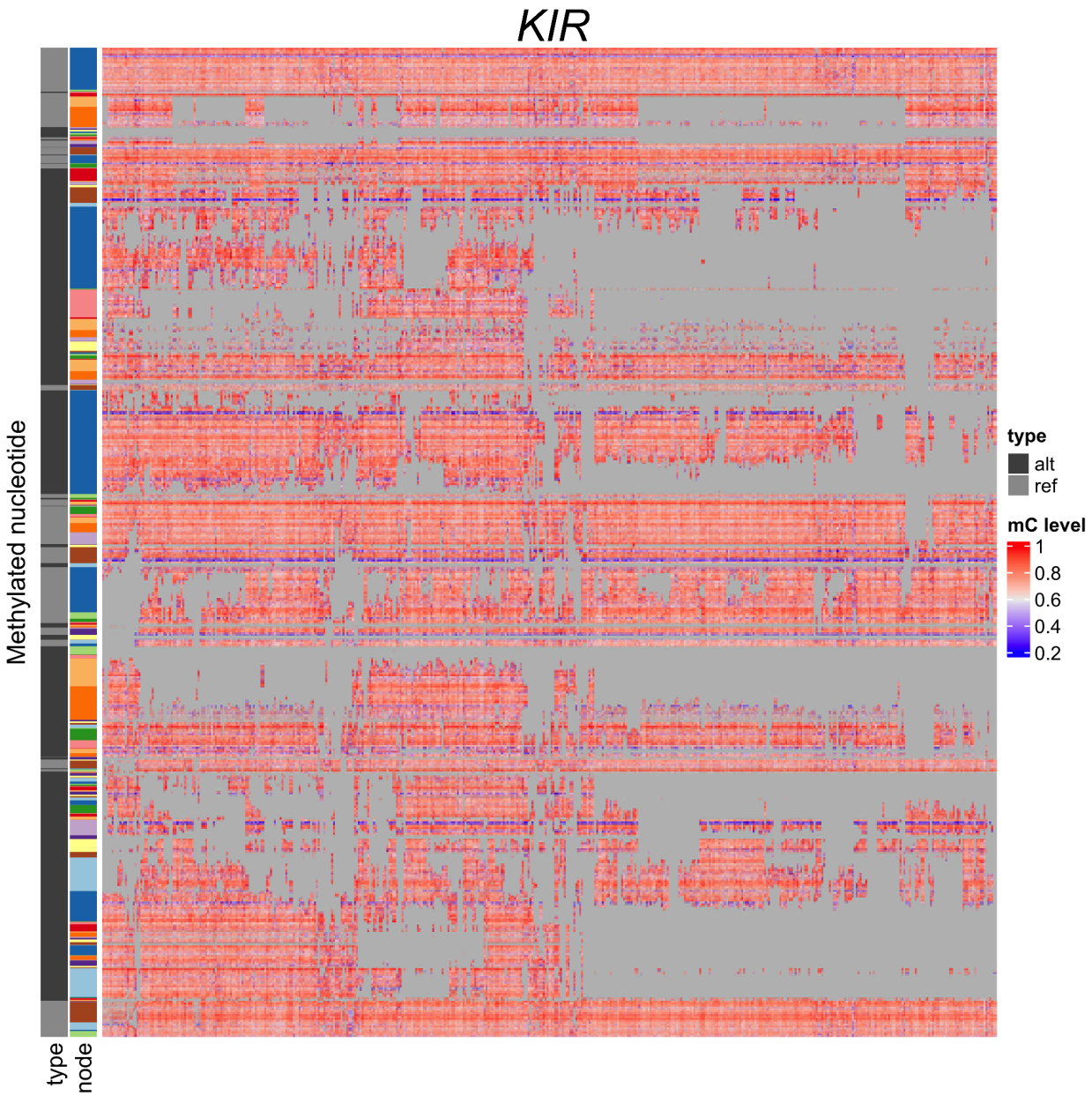

Fig S9: Heatmap visualization of methylation patterns in the *KIR* locus across 435 methylomes in the GA4K pangenome. Rows are CpGs and columns are methylomes. CpGs are ordered top to bottom, in the 5' to 3' direction as they appear in a haplotype, and are annotated by the node in the graph (the node row annotation) and whether it's a reference or non-reference CpG (the type row annotation, alt or ref).

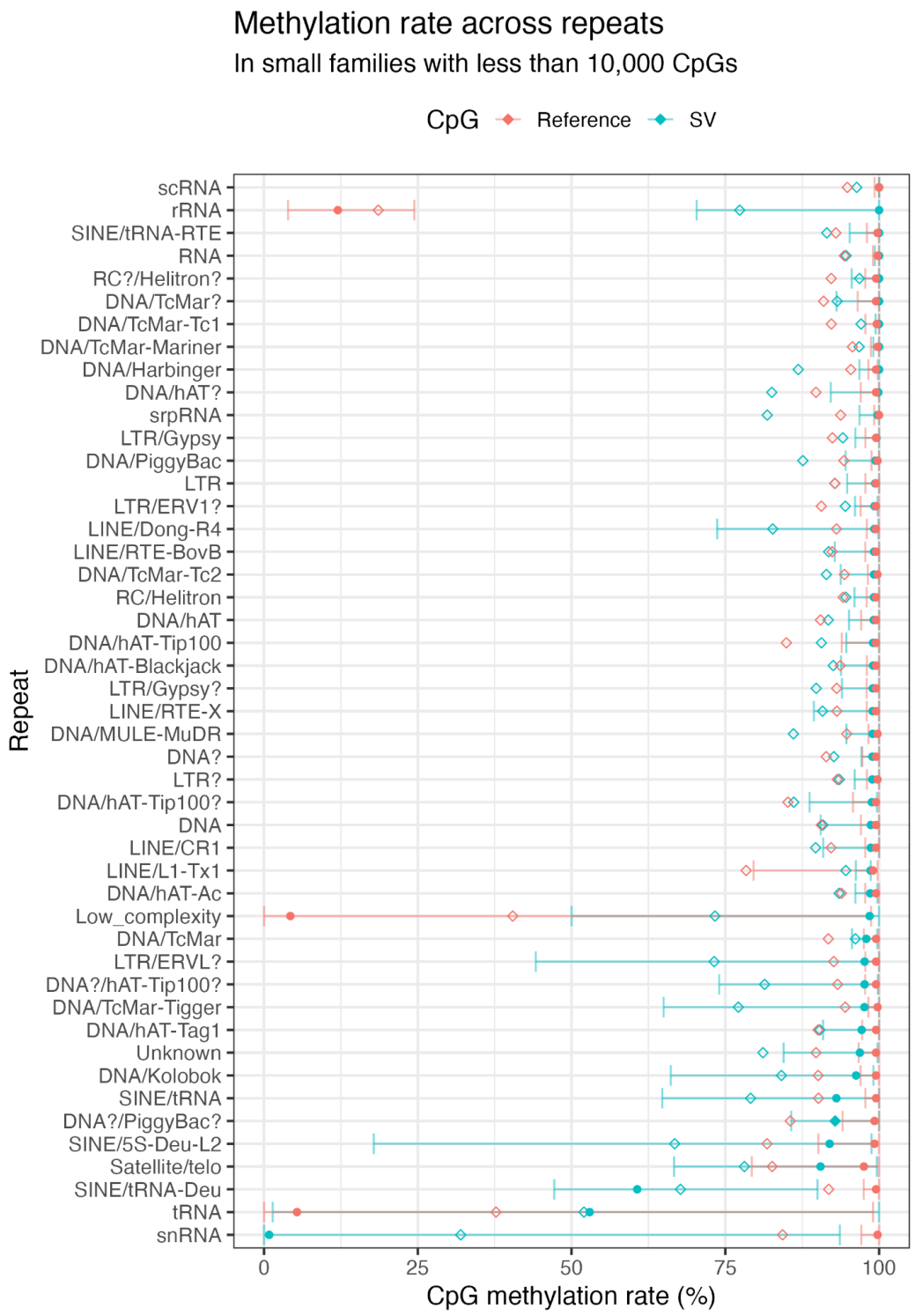

Fig S10: The observed methylation rates of CpGs in repeat families not shown in Fig 3D. Intervals denote the 25%, 50% (median, dot), and the 75% quantiles. The rhombi denote the means of the distributions.

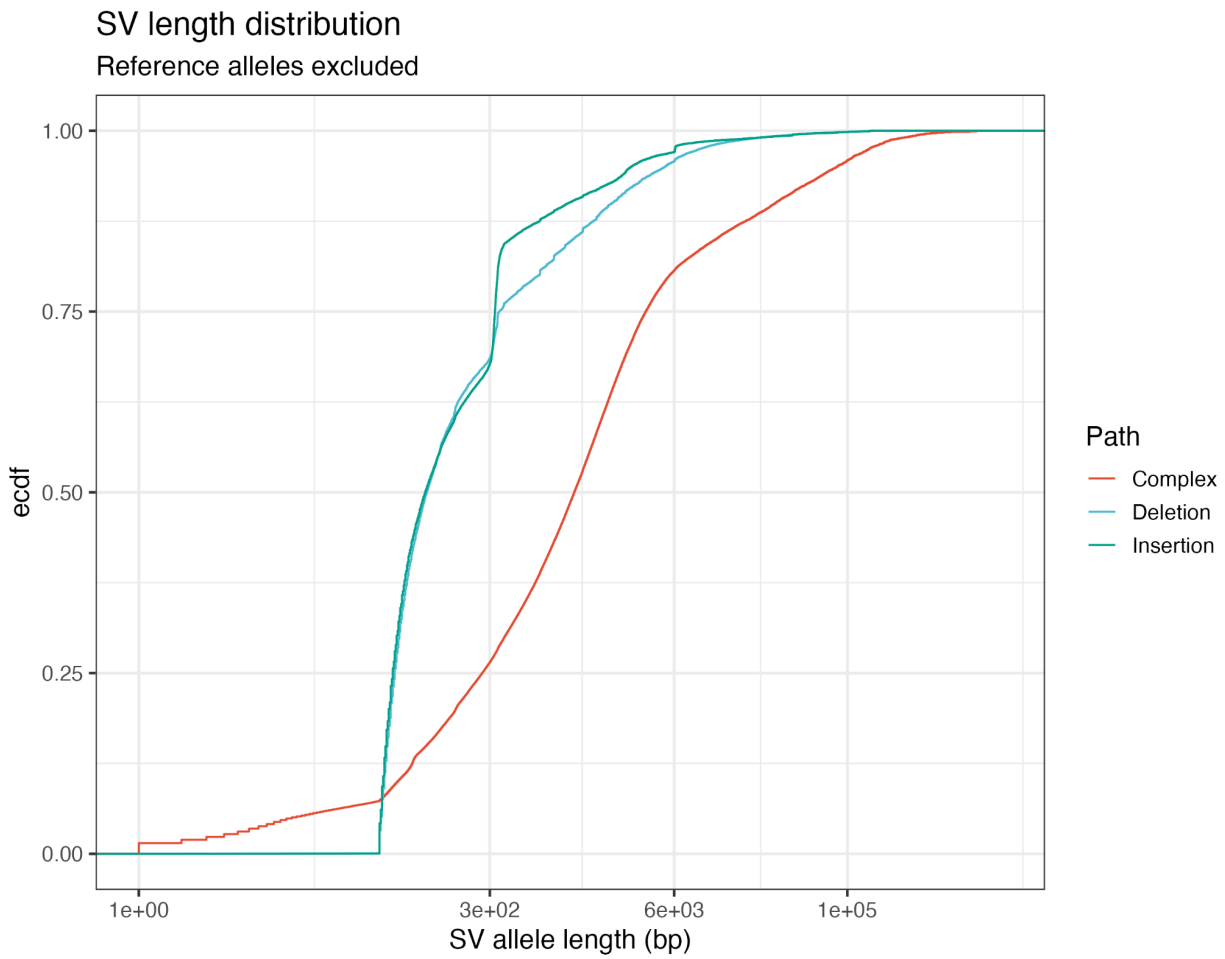

Fig S11: SV allele length distribution in the GA4K pangenome, stratified by insertions, deletions and complex SVs. Complex SVs are sequences that cannot be described by a single insertion, deletion or substitution event. These usually happen in loci where SVs from multiple genomes merge into bubbles, such as STRs or VNTRs.

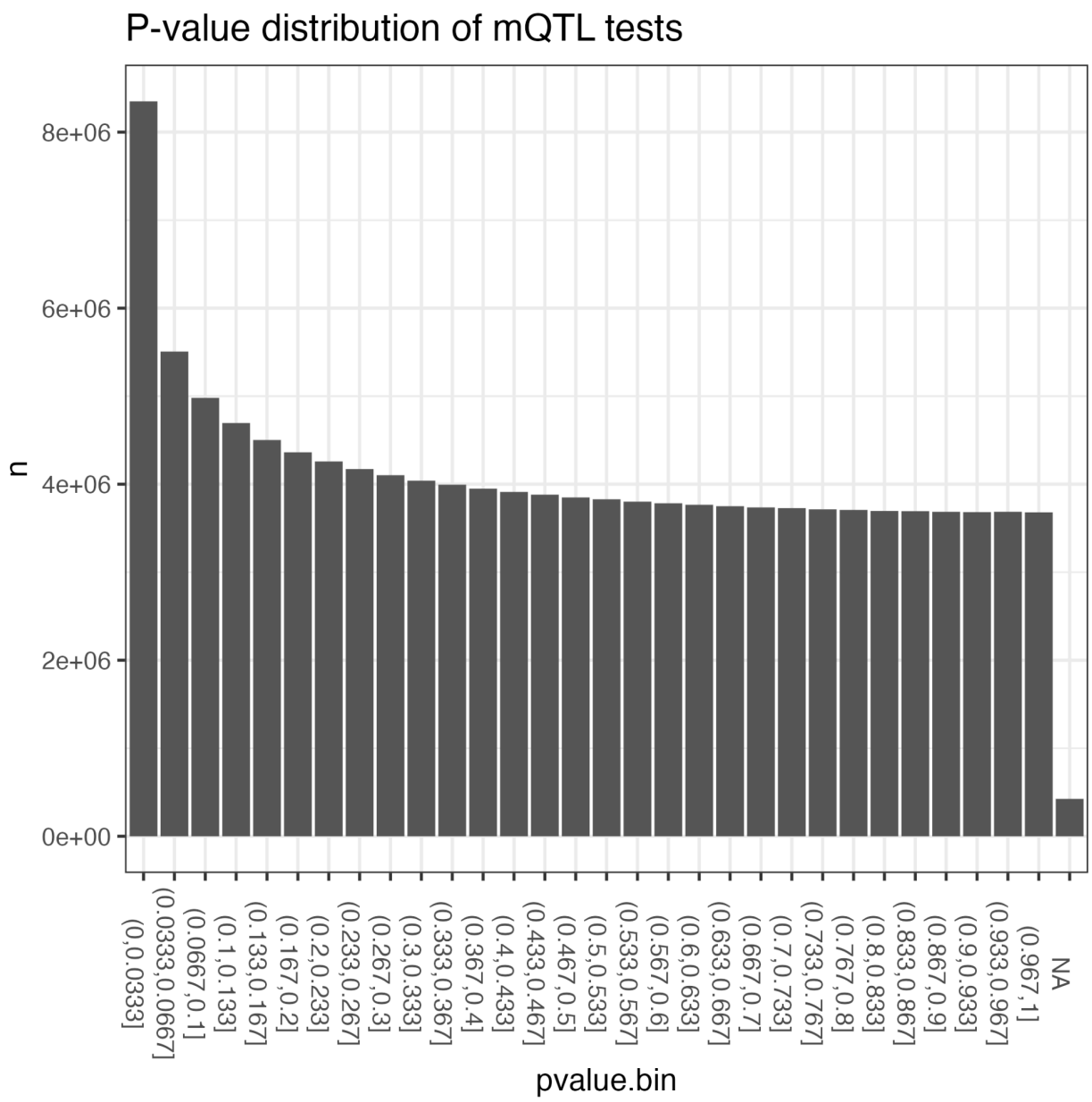

Fig S12: P-value distribution of SV-mQTL tests. NA p-values are tests where almost all samples had missing data.

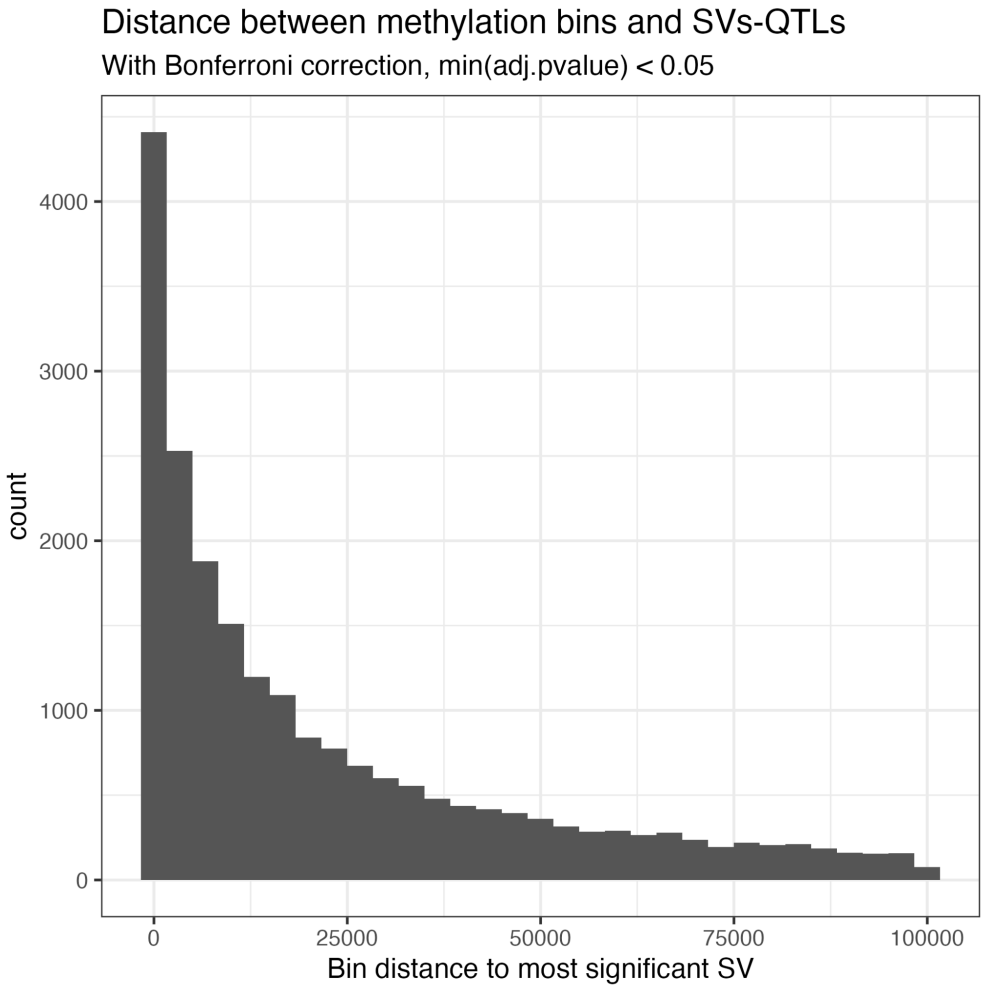

Figure S13: Range of statistical significant interactions between methylation bins and SV-mQTLs, corrected for multiple testing with the Bonferroni method.

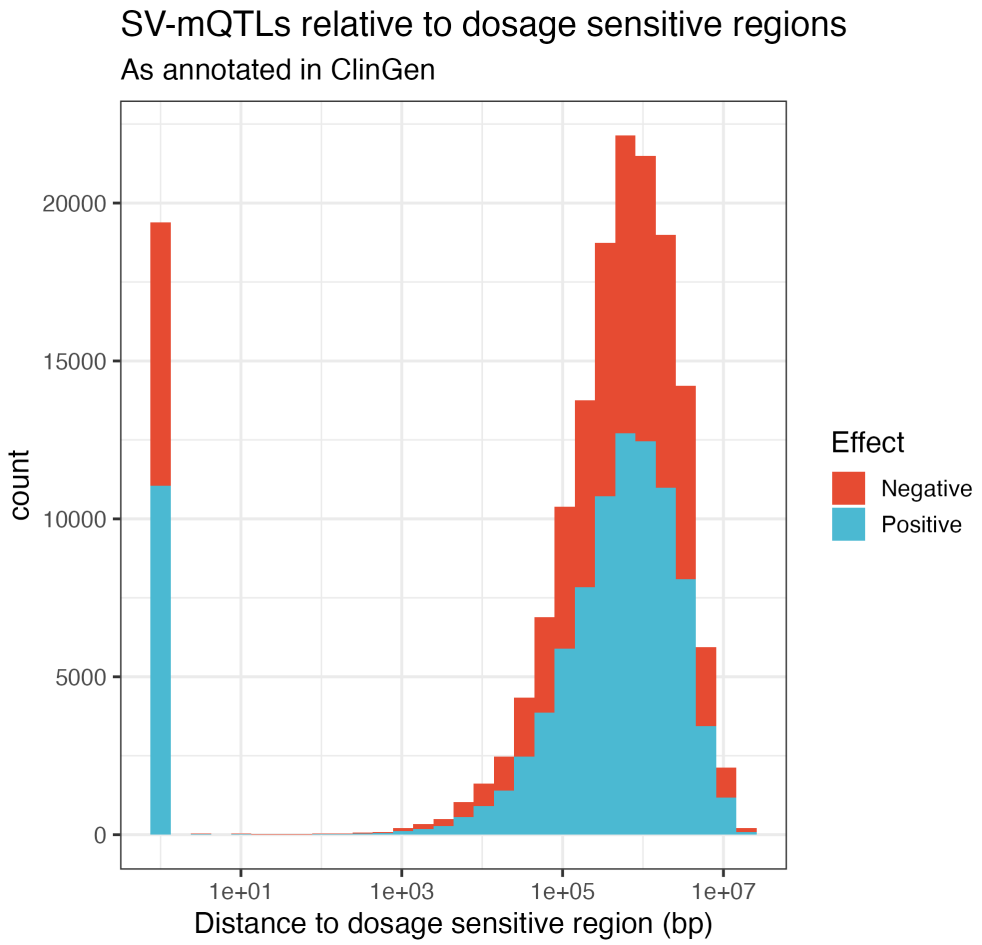

Fig S14: Distribution of SV-mQTLs relative to dosage sensitive regions in ClinGen. Alleles with both positive and negative effects are counted twice.

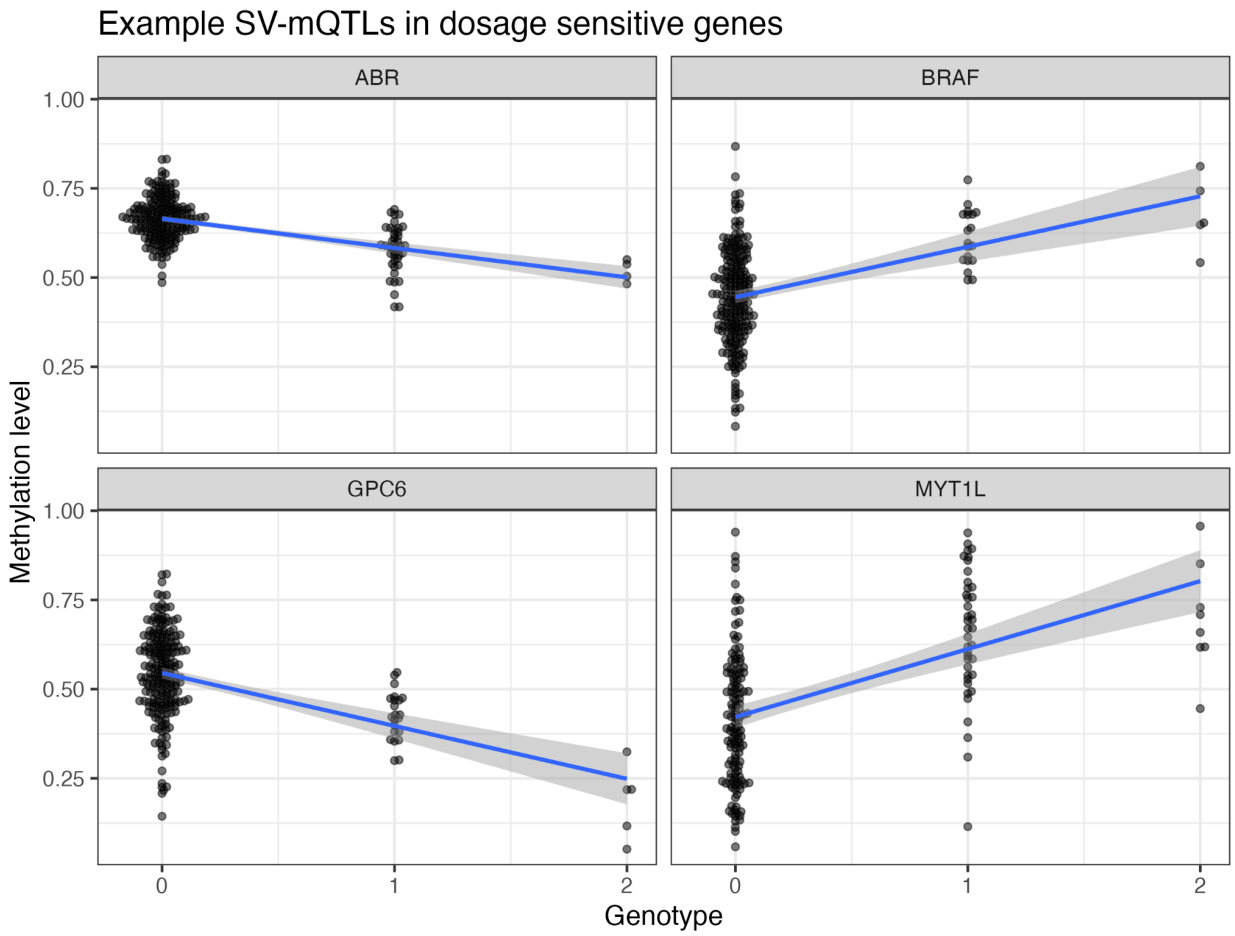

Fig S15: Example methylation bins within four dosage sensitive genes. The involved bins are Chr17:1047600-1047799 (*ABR*), Chr7:142218400-142218599 (*BRAF*), Chr13:92714600-92714799 (*GPC6*), Chr2:1805800-1805999 (*MYT1L*). The involved SVs are Chr17:1045448-1047678 (*ABR*), Chr7:142222985-142223202 (*BRAF*), Chr13:92714752-92714907 (*GPC6*), Chr2:1805709-1806032 (*MYT1L*).

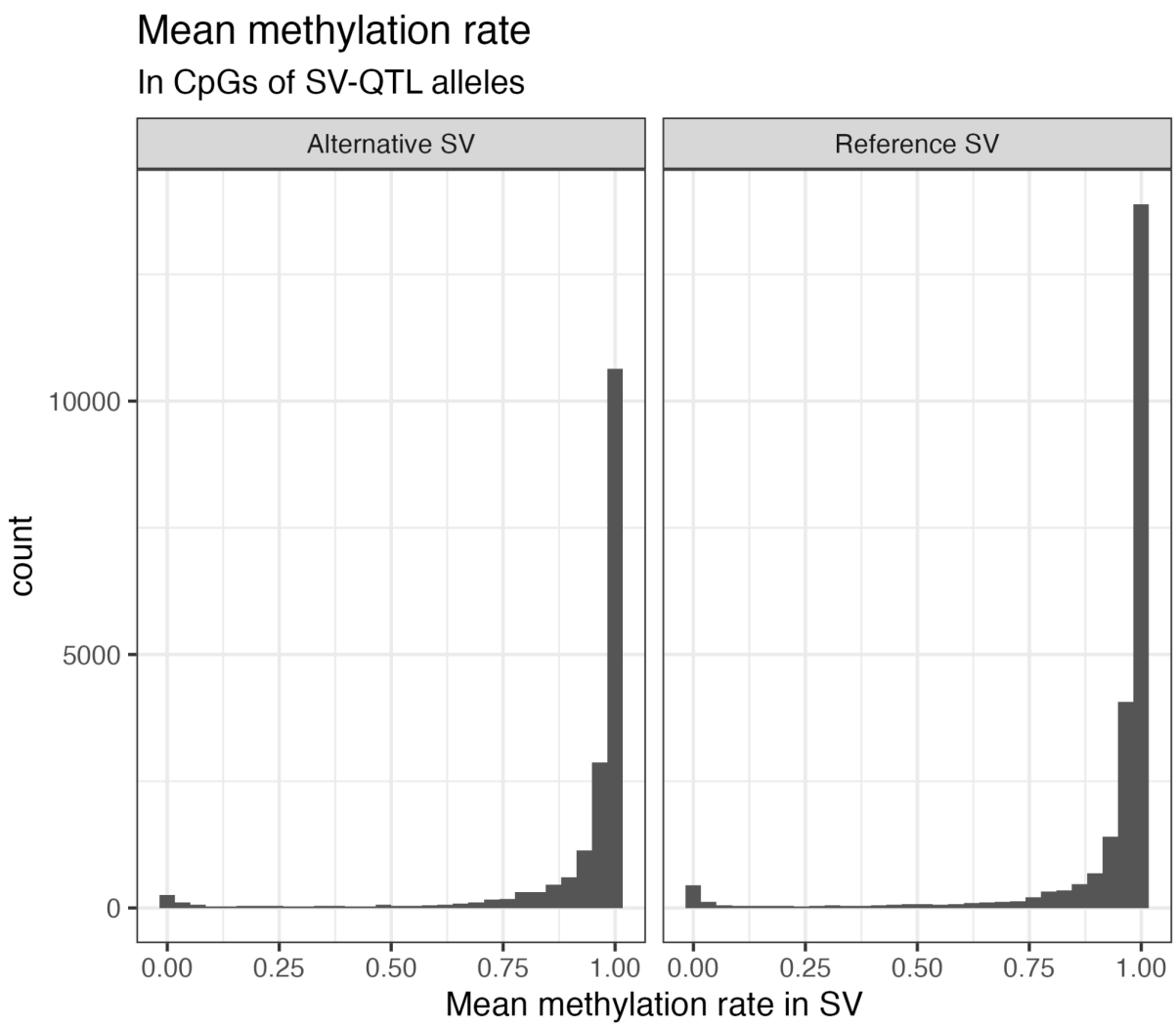

Fig S16: Average methylation rate of CpGs in SVs that are QTL for methylation, stratified by reference and non-reference SV alleles.

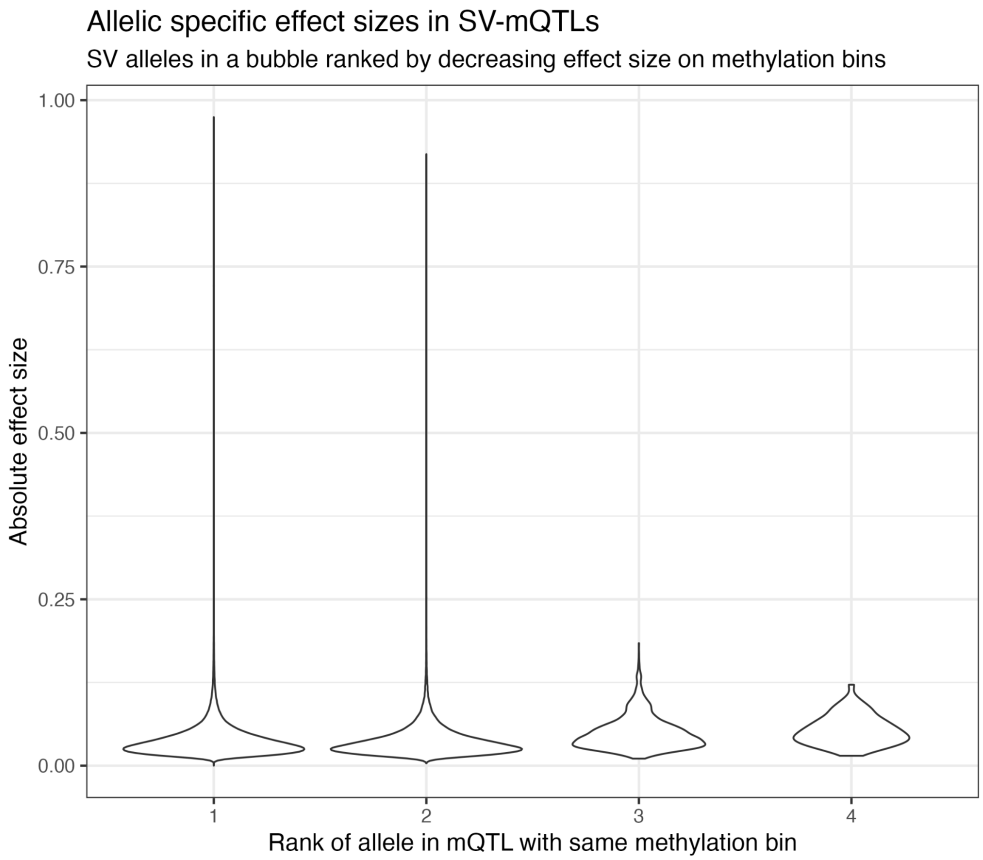

Fig S17: Distribution of allelic specific effect sizes in mQTLs. For each methylation bin and multi-allelic bubble, we rank the SV alleles in the bubble in decreasing order of effect size on the methylation bin. Then, we plot the distribution of effect sizes at each rank.

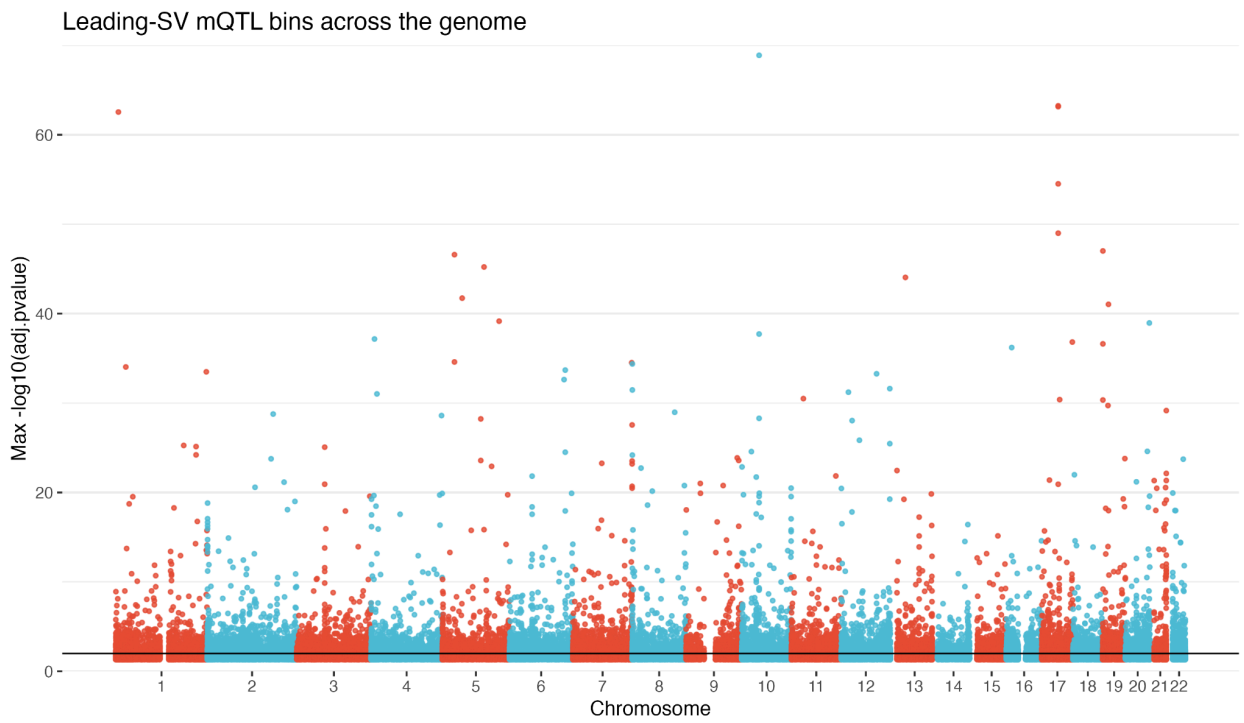

Fig S18: Manhattan plot of QTL methylation bins associated with a leading SV that surpasses nearby SNPs at FDR < 0.05 over the entire reference genome backbone.

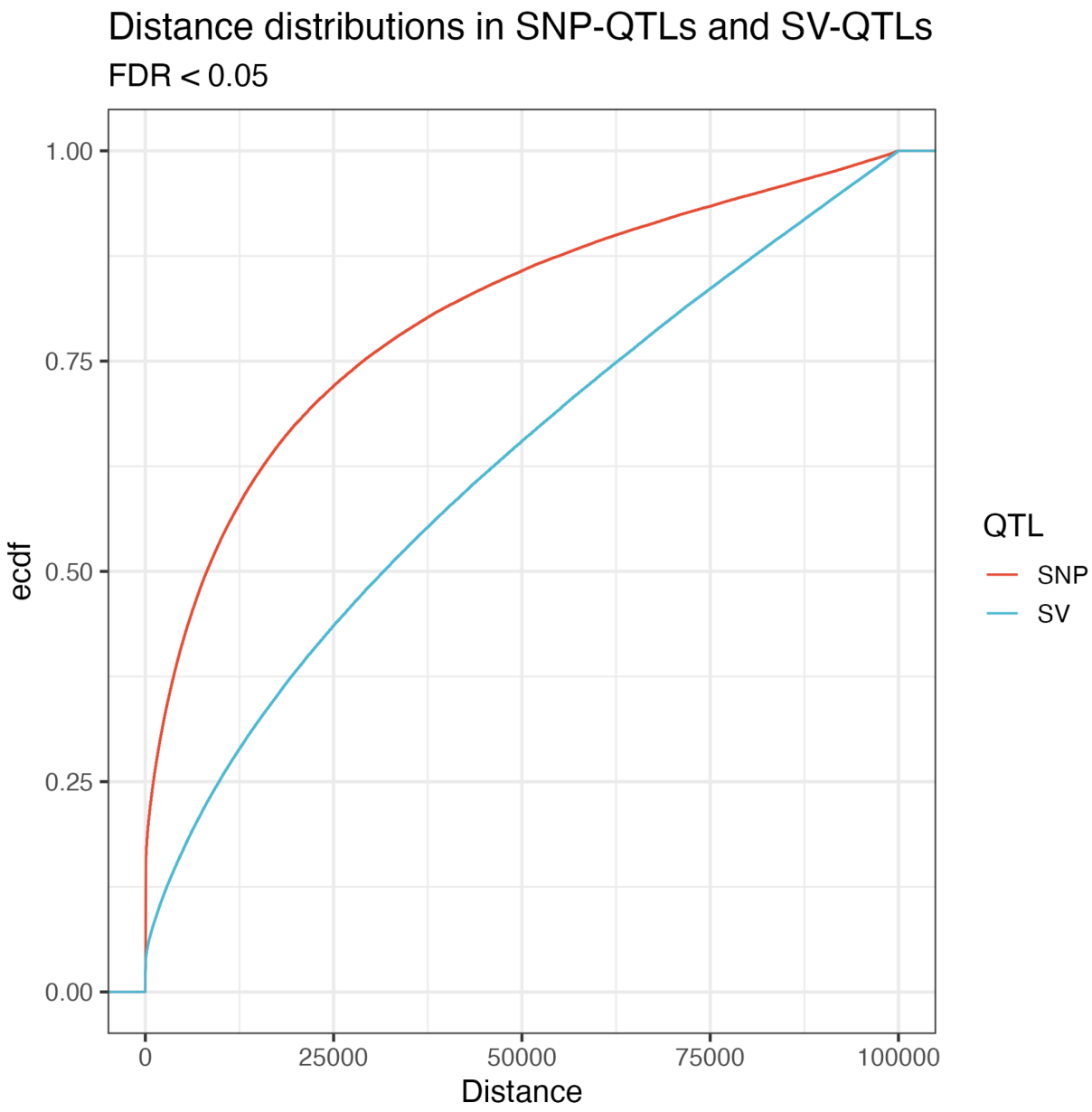

Fig S19: The distribution of distances between SV-mQTLs, SNP-QTLs and their methylation bins.

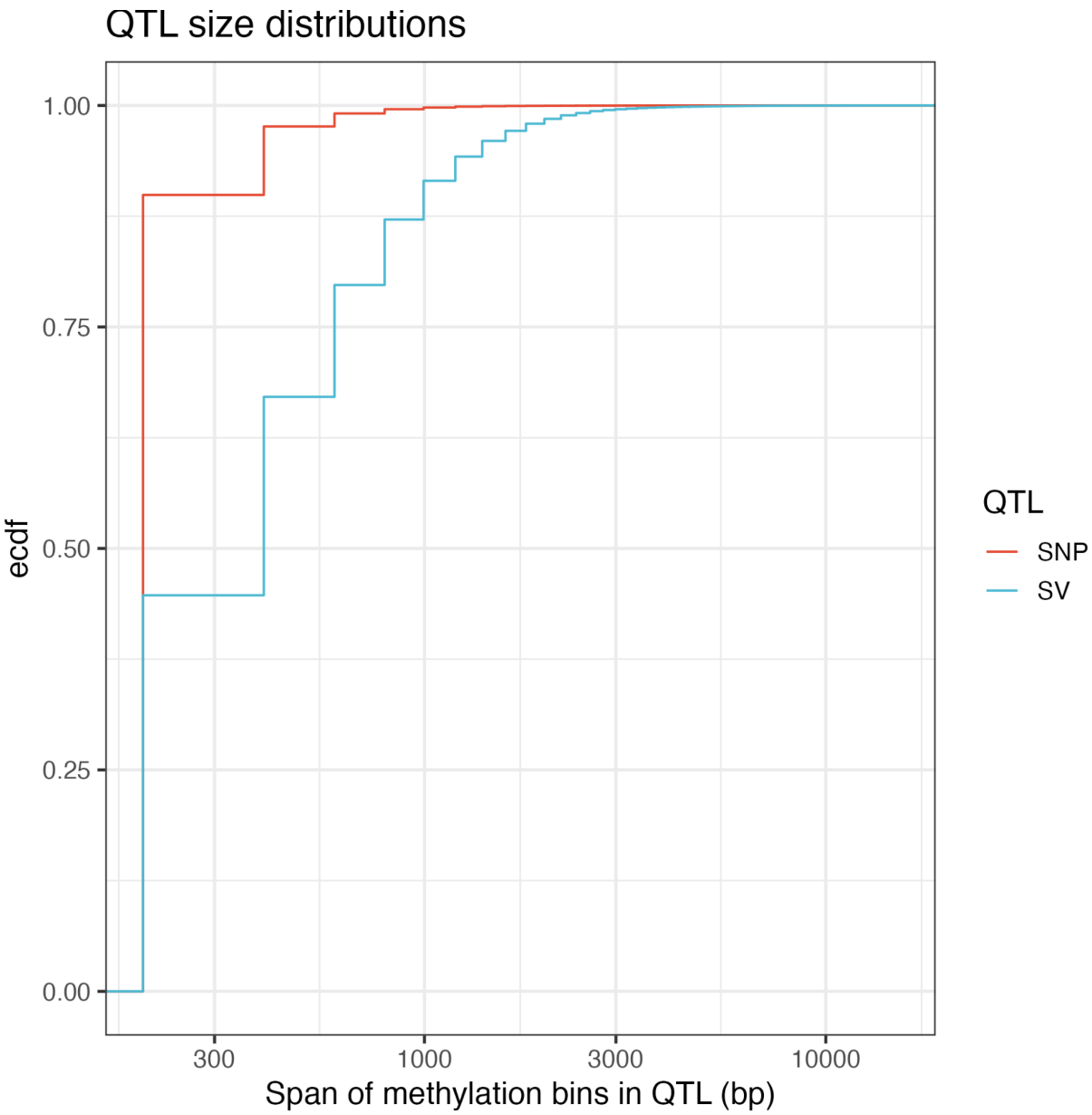

Fig S20: Distribution of the number of methylation bins (in bp) that are in QTL (FDR < 0.05) with each SV and each SNP.

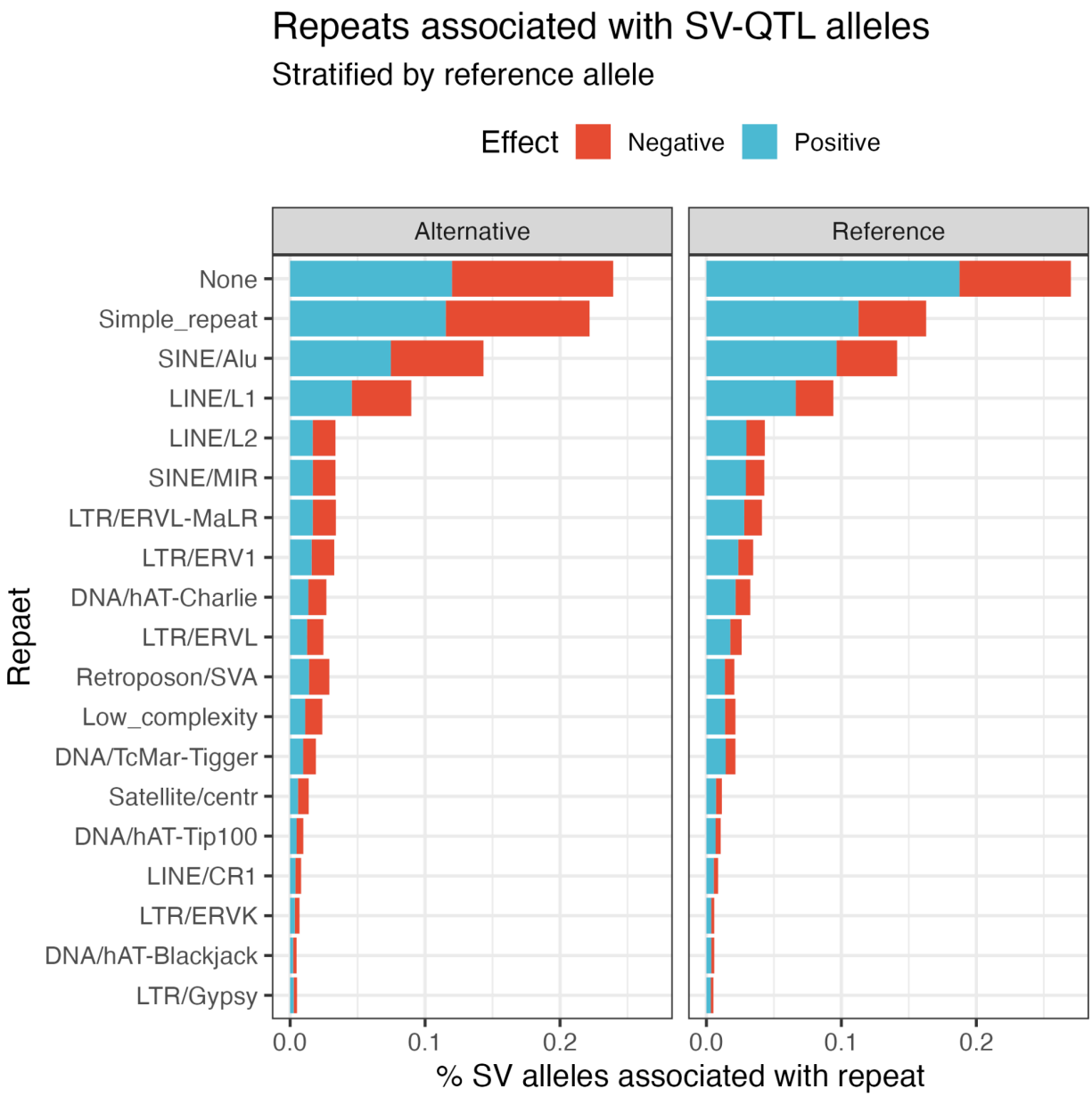

Fig S21: Repeat annotation of SV-QTL alleles, stratified by reference and non-reference SV alleles, and the direction of the effect of methylation.

Table S1: t-tests involving the rate of discovering new SV-CpGs, SV-5mCpGs, reference CpGs and reference 5mCpGs. The standard deviations are estimated from 10 random permutations of the order of genomes in Fig 1B.

Rate A	Rate B	Mean rate A	Mean rate B	Standard deviation A	Standard deviation B	p-value
SV-CpGs	Reference CpGs	2758	615	2680	594	0.0334
SV-CpGs	SV-5mCpG	2758	2053	2680	2397	0.5431
Reference CpGs	Reference 5mCpGs	615	1535	594	1208	0.0498

Table S2: Mann Whitney *U* testing of differences in methylation rates between SV-CpGs and reference CpGs in each repeat family.

statistic	p.value	repeat	adj.p.value	statistic	p.value	repeat	adj.p.value
57286567338873	<2.225074e-308	None	<2.225074e-308	56314	3.39E-04	DNA?/hAT-Tip100?	5.84E-04
9510935047	5.24E-18	DNA/hAT-Charlie	1.41E-17	12170354	1.20E-08	LTR/Gypsy?	2.57E-08
2962406132	<2.225074e-308	DNA/TcMar-Tigger	<2.225074e-308	2171472	4.48E-04	DNA	7.51E-04
1093369350894	<2.225074e-308	Simple_repeat	<2.225074e-308	9371612.5	0.1839913762	LTR	0.2236757907
1115255185920	<2.225074e-308	LINE/L1	<2.225074e-308	4560901	0.04970469719	DNA/hAT	0.06556789843
8212186720795	<2.225074e-308	SINE/*Alu*	<2.225074e-308	525552	0.05771781786	LTR/ERV1?	0.07303070831
68914462.5	0.3627159072	LTR/Gypsy	0.4015783259	1234778	0.3573041794	SINE/tRNA-RTE	0.4015783259
23573998295	<2.225074e-308	LINE/L2	<2.225074e-308	2290940	1.37E-221	snRNA	6.55E-221
12454902318	2.33E-122	SINE/MIR	9.03E-122	581212	9.18E-04	DNA/PiggyBac	0.001498329329
177325565	1.94E-47	LINE/CR1	6.34E-47	437591	0.006223109941	DNA/hAT-Tag1	0.009410556496
57481909943	<2.225074e-308	Retroposon/SVA	<2.225074e-308	245768	0.0559357693	RC/Helitron	0.07225036869

42488594204	0.2437992524	LTR/ERVL-MaLR	0.2851991255	1068633	0.02374337613	LINE/RTE-BovB	0.032713096
42776042.5	1.71E-16	LINE/RTE-X	4.42E-16	801997.5	7.59E-09	RNA	1.68E-08
1237425758927	<2.225074e-308	Satellite/centr	<2.225074e-308	1452569817	<2.225074e-308	rRNA	<2.225074e-308
1306603720	<2.225074e-308	Low_complexity	<2.225074e-308	256318	1.57E-06	DNA/hAT-Ac	3.03E-06
70276249881	<2.225074e-308	LTR/ERV1	<2.225074e-308	13846	0.1992345487	LINE/Dong-R4	0.2375488849
18586158681	2.03E-275	LTR/ERVL	1.05E-274	513178	0.001259019626	DNA/hAT-Tip100?	0.00200151838
102187748108	7.27E-27	Satellite	2.15E-26	891960	0.1029223885	DNA?	0.1276237617
13464712	5.70E-77	tRNA	1.96E-76	98678	0.5448178204	DNA/hAT?	0.5537492601
2413777282	1.02E-215	LTR/ERVK	4.52E-215	281462	9.03E-05	DNA/TcMar	1.60E-04
393870442.5	2.87E-16	Satellite/acro	7.13E-16	654515	2.65E-06	DNA/MULE-MuDR	4.98E-06
2950	0.3814289631	DNA/Kolobok	0.4148876441	30805	0.002657287909	SINE/tRNA	0.004118796259
145481947	4.94E-116	Satellite/telo	1.80E-115	920144	6.95E-08	srpRNA	1.39E-07
14301101	5.52E-08	DNA/hAT-Blackjack	1.14E-07	27050	0.01299233485	SINE/5S-Deu-L2	0.01917916096
3471823.5	2.58E-23	scRNA	7.26E-23	904	0.02266036194	SINE/tRNA-Deu	0.03193051001
10838476	3.85E-11	LTR?	9.17E-11	1402	0.8812534885	DNA?/PiggyBac?	0.8812534885
237286530	0.3987371374	DNA/hAT-Tip100	0.4262362504	6766	0.321598534	DNA/TcMar?	0.3692427612
3074393	4.29E-28	LTR/ERVL?	1.33E-27	19885	0.04186459267	DNA/TcMar-Tc1	0.05642619012
4717742	8.23E-06	DNA/TcMar-Tc2	1.50E-05	10696	0.40737546	RC?/Helitron?	0.4280894664
1799572	2.78E-10	Unknown	6.39E-10	4438	0.4627516554	LINE/L1-Tx1	0.4781767106
162903338	3.53E-167	DNA/TcMar-Mariner	1.46E-166	6138	0.01807889122	DNA/Harbinger	0.0260672385

Table S3: Number of tested SNPs and SVs, the number of SNPs and SVs that were a statistically significant QTL, and the number of SNPs and SVs that were the leading QTL variant. For SVs, we give the number of alleles and loci, since some SVs are multi-allelic.

Variants	Tested alleles	QTL (FDR <0.05)	Leading QTL
SVs	160,064 alleles 97,746 loci	76,677 alleles 59,872 loci	32,947 alleles 31,312 loci
SNPs	5,617,307	156,047	145,453