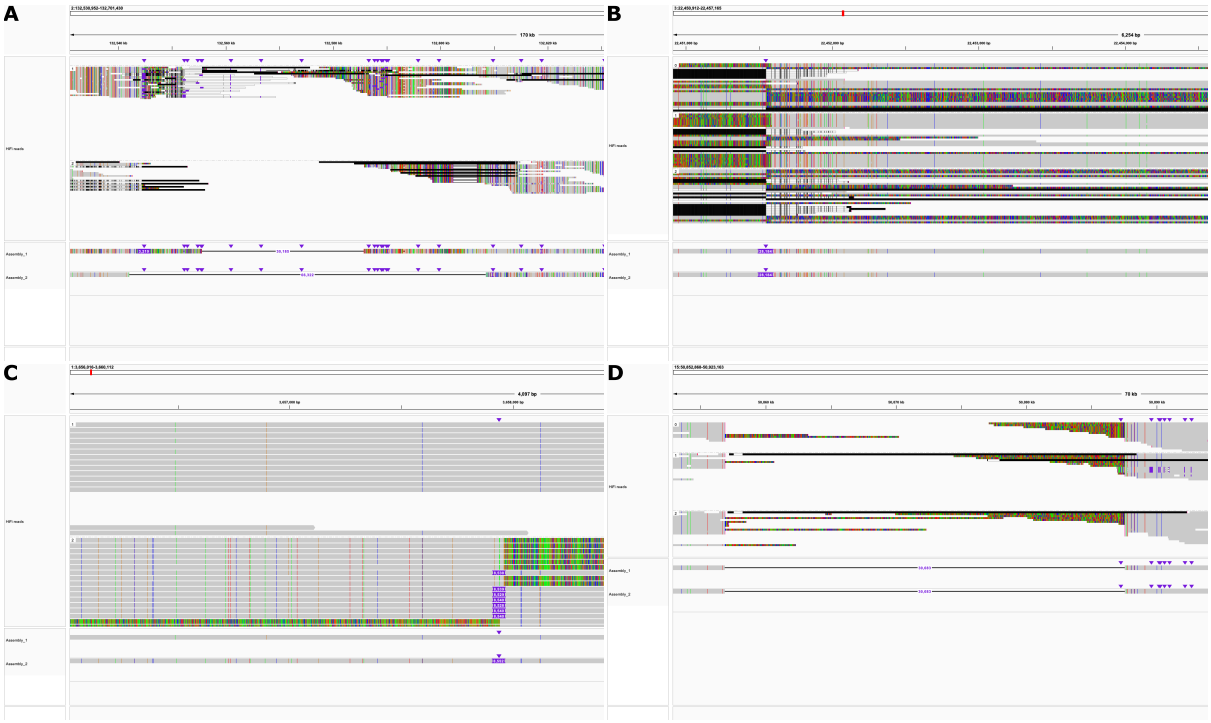
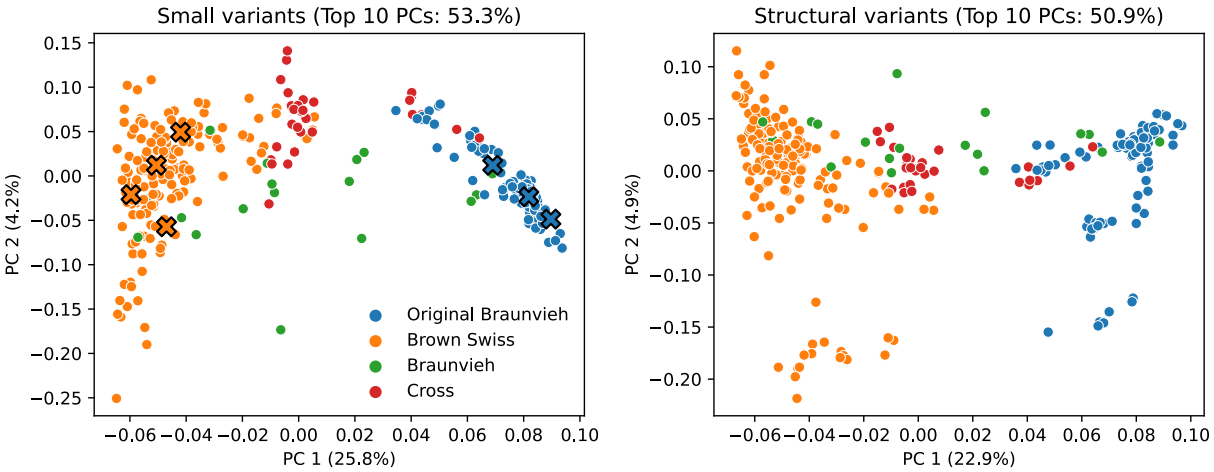
Supplementary material

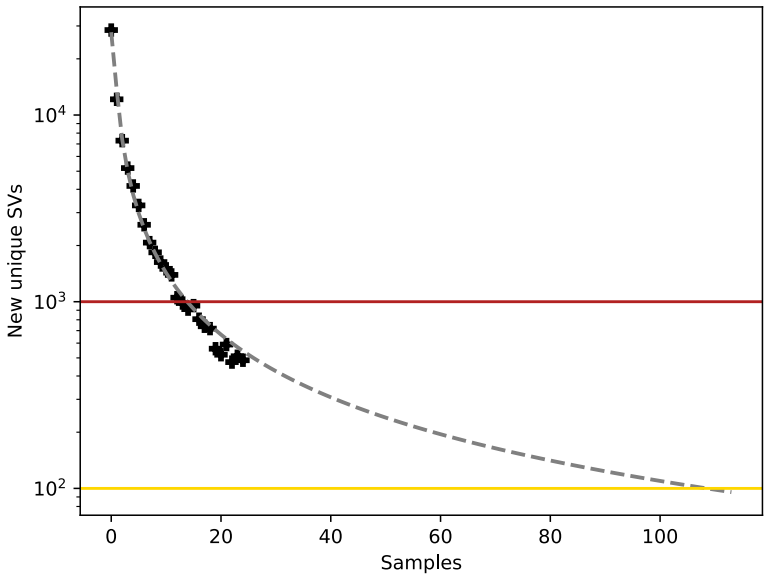


Supplementary Figure 1. Series of 4 IGV screenshots for one sample with haplotype-binned HiFi reads (upper panel) and haplotype-resolved assemblies (lower panels). (a-b) Two examples where haplotype-resolved assemblies can confidently identify an SV but HiFi read-based SV calling fails, and (c-d) two examples where read-based calling matches the SVs called by assemblies.

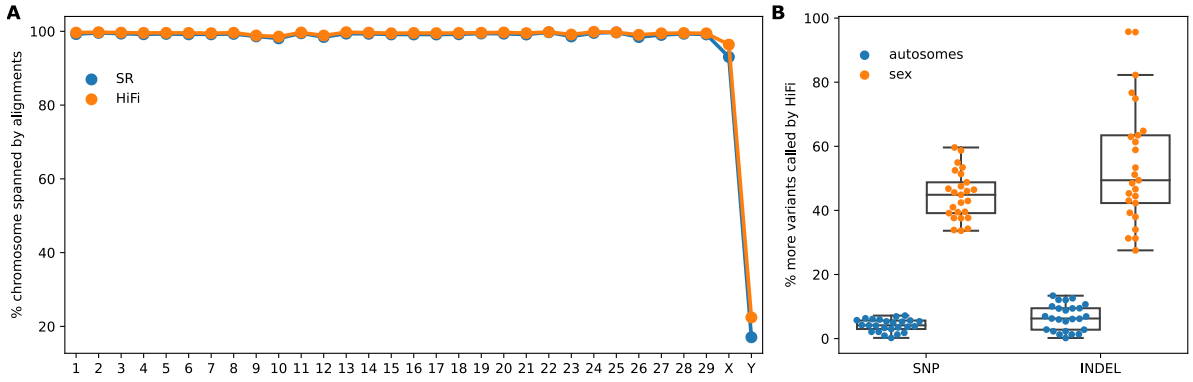


Supplementary Figure 2. PCA for small (A) and structural (B) variants genotyped in the 307 samples (circles). Slightly more variance was explained by the top 10 PCs for small variants compared to structural variants. Breed structure is revealed for Brown Swiss and Original Braunvieh, with crosses between these breeds (Braunvieh) or crosses with a different breed (Cross) distributed between the two primary clusters. (A) The assembly samples (crosses) in used in creating the pangenome

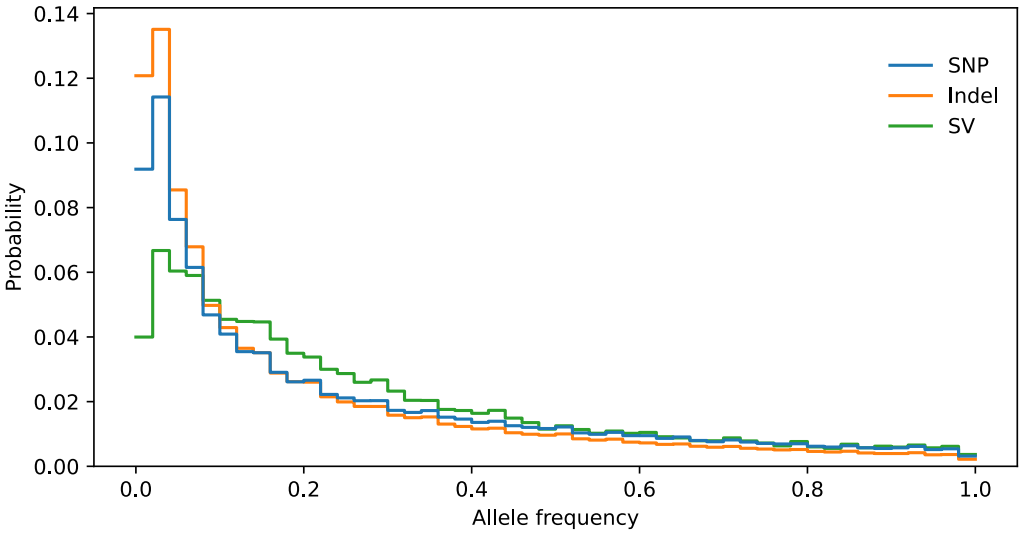
panel are broadly representative for their respective Original Braunvieh or Brown Swiss population. Short read sequencing was not available for three Brown Swiss haplotypes and the Piedmontese haplotype.



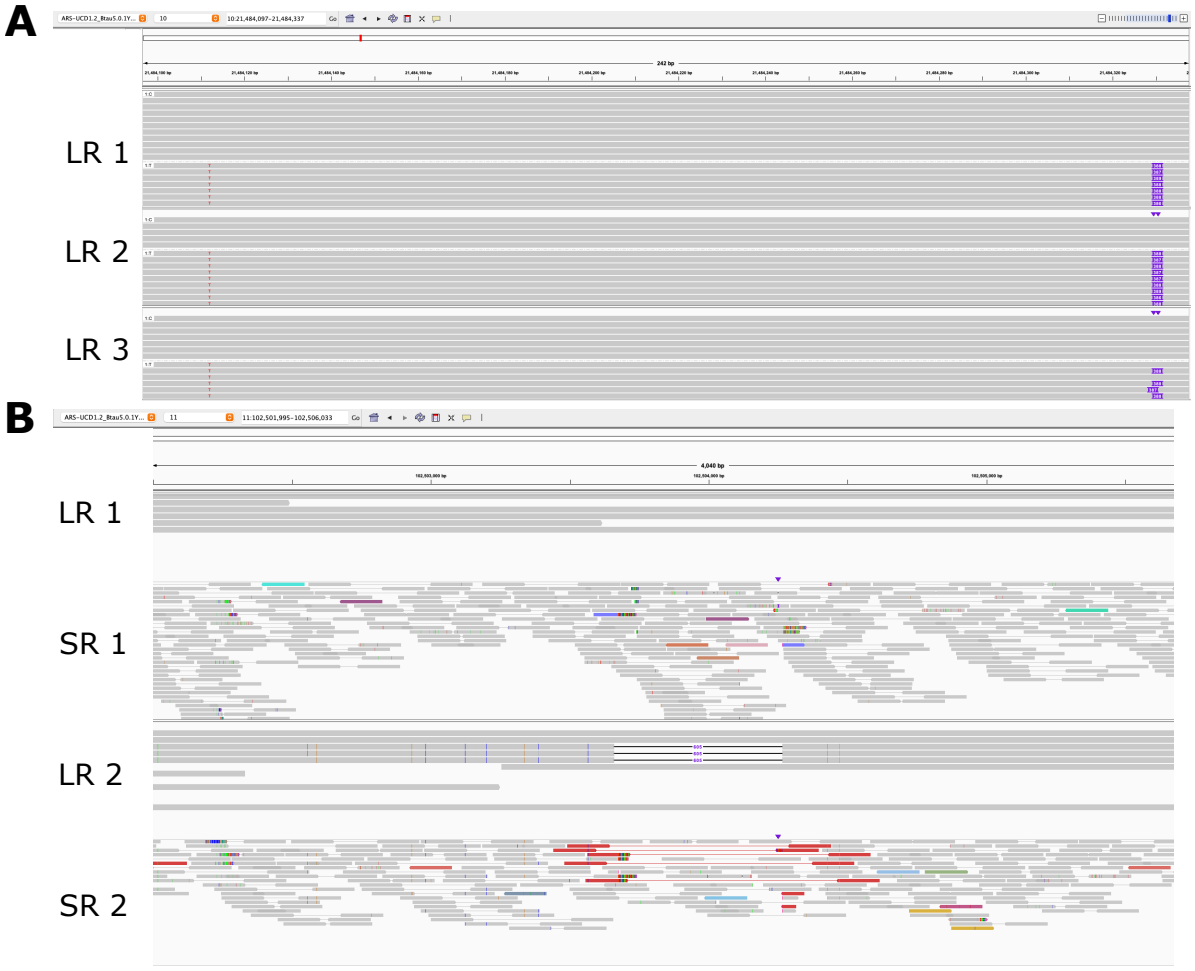
Supplementary Figure 3. New unique SVs per additional sample added to the cohort (black markers). The grey fitted line is taken from Figure 2b. The red and orange horizontal lines respectively indicate expected discovery of 1000 and 100 new unique SVs per sample, intersecting the fitted line at approximately 14 and 108 samples.



Supplementary Figure 4. (A) Both HiFi and short reads (SR) alignments confidently span almost the entirety of the autosomes, while HiFi performs better for the typically more complex and repetitive sex chromosomes (X and Y). (B) HiFi-based alignments called more variants than short read-based alignments at approximately 10-fold coverage, particularly in the sex chromosomes.

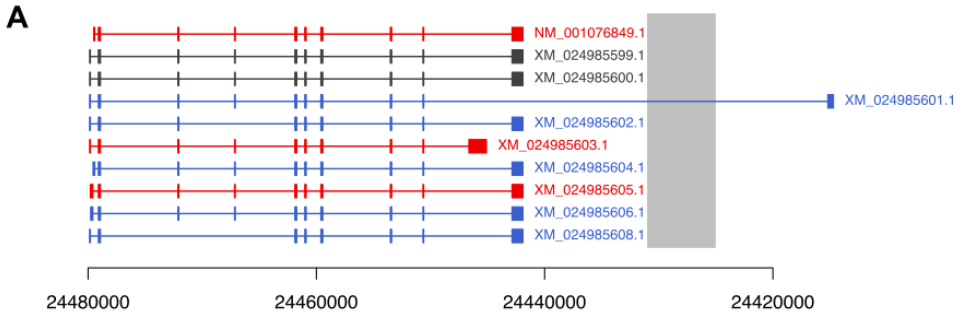


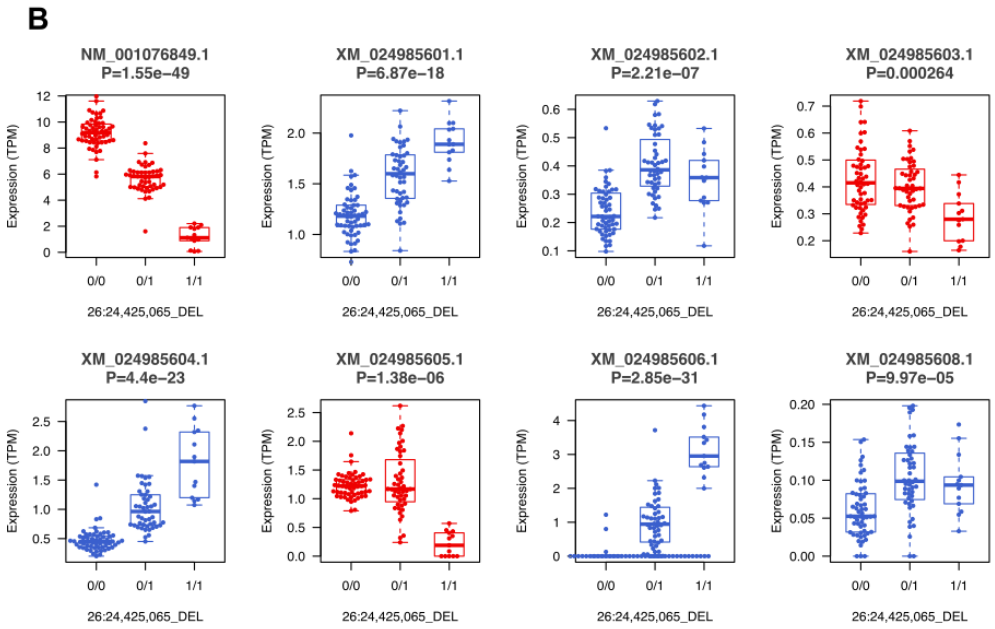
Supplementary Figure 5. Allele frequency distribution for SNPs, Indels, and SVs in the PanGenie+ variant set, across the 307 samples.

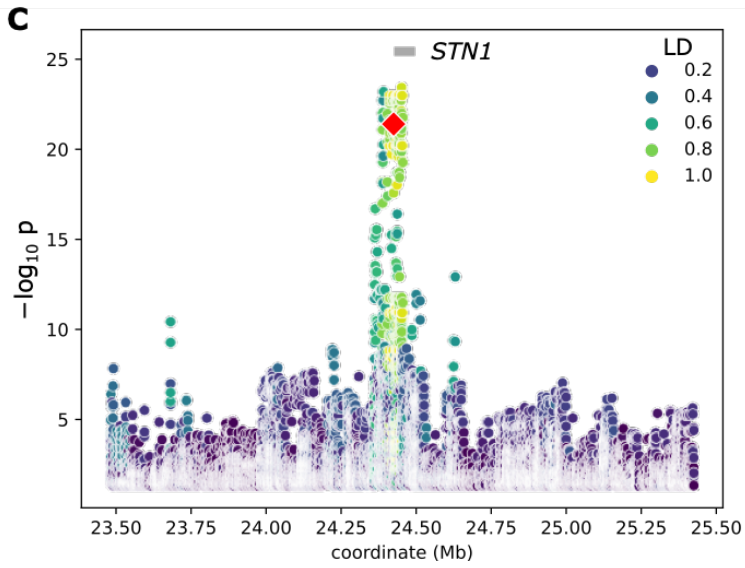


Supplementary Figure 6. (A) HiFi alignments for three heterozygous carriers of the 388 bp structural variant (right) that was a top associated hit for MYH7. *InSurVeyor called these samples as homozygous alternate and mis-identified LR2 and LR3 as 458 and 130 bp insertions respectively. The tagging SNP (left) supports the heterozygous call of the long read truth set and*

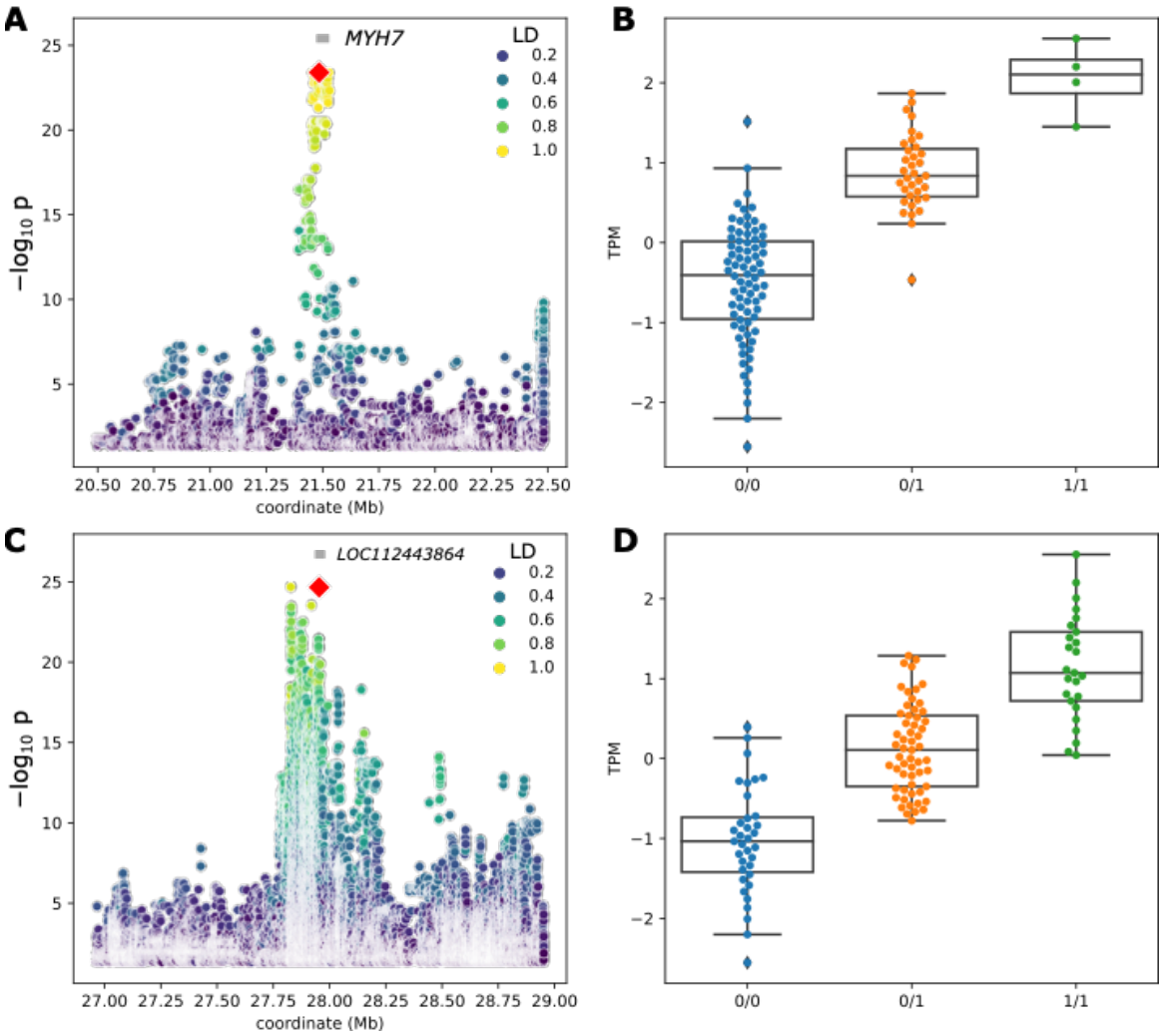
PanGenie. (B) HiFi alignments clearly reveal a heterozygous 605 bp deletion for LR 2 that is not found in LR 1, while short read SV calling results in a 605 bp deletion heterozygous in SR 2 as well as an unsupported heterozygous call in SR 2.


A
NM_001076849.1
XM_024985599.1
XM_024985600.1
XM_024985601.1
XM_024985602.1
XM_024985603.1
XM_024985604.1
XM_024985605.1
XM_024985606.1
XM_024985608.1
24480000
24460000
24440000
24420000

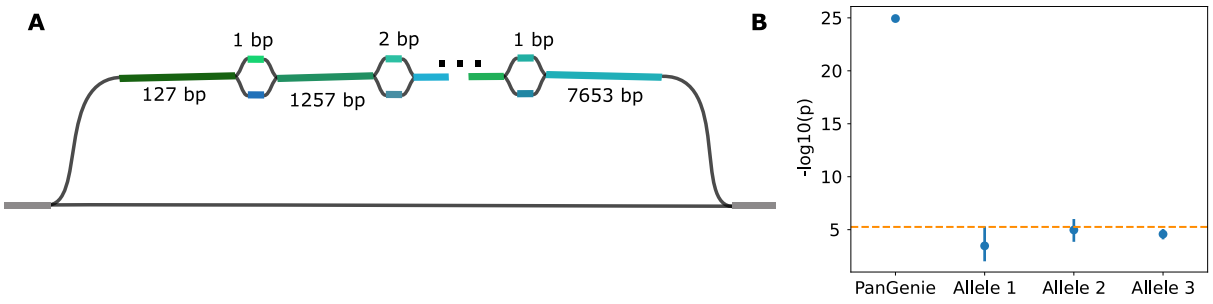

B
NM_001076849.1
P=1.55e−49
XM_024985601.1
P=6.87e−18
XM_024985602.1
P=2.21e−07
XM_024985603.1
P=0.000264
XM_024985604.1
P=4.4e−23
XM_024985605.1
P=1.38e−06
XM_024985606.1
P=2.85e−31
XM_024985608.1
P=9.97e−05
Expression (TPM)
0/0
0/1
1/1
26:24,425,065_DEL


C
STN1
LD
0.2
0.4
0.6
0.8
1.0
−log10 p
coordinate (Mb)
23.50
23.75
24.00
24.25
24.50
24.75
25.00
25.25
25.50

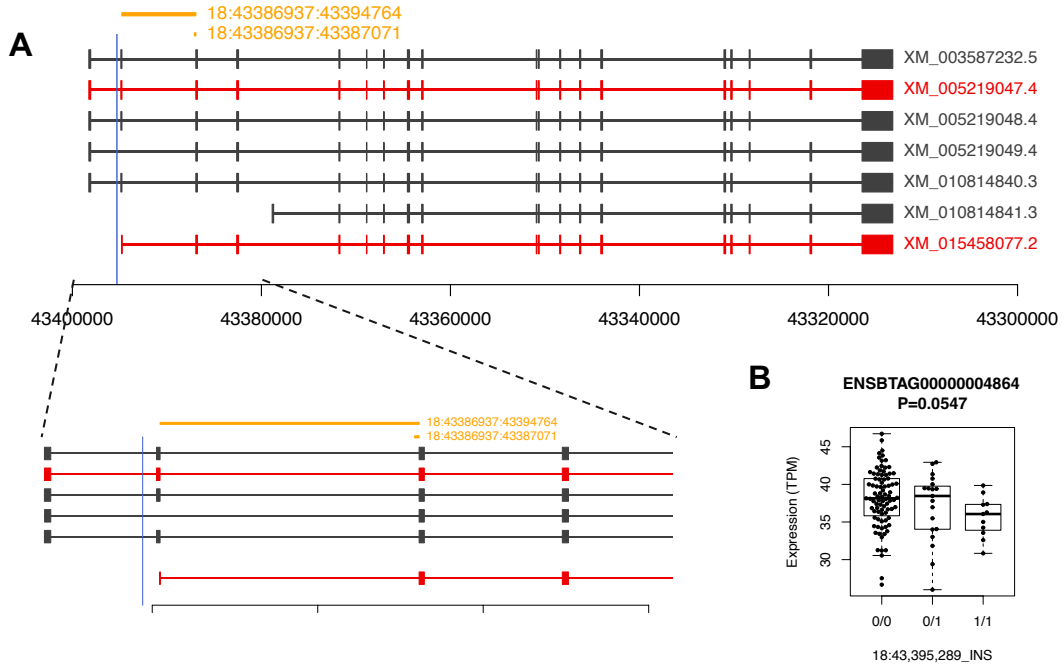
Supplementary Figure 7. A 5.9 kb deletion is associated with STN1 *expression. (A) Structure of ten* STN1 *isoforms that are annotated in Refseq (version 106). The grey box indicates the position of the deletion; it is downstream of nine isoforms, and within the last intron of XM_024958601.1. Boxes represent exons. Red and blue color indicates transcripts whose expression is respectively reduced and increased by the deletion. Transcripts that are not expressed are grey-colored. (B) Impact of the deletion on the expression of eight* STN1 *isoforms. Red and blue color indicates transcripts whose expression is respectively reduced and increased by the deletion. The P values are from a linear regression of TPM values on the genotype (coded as 0, 1, 2). (C) The 5.9 kb deletion (red diamond) is also strongly associated with a splicing QTL for* STN1, *although it is not the top associated variant.*



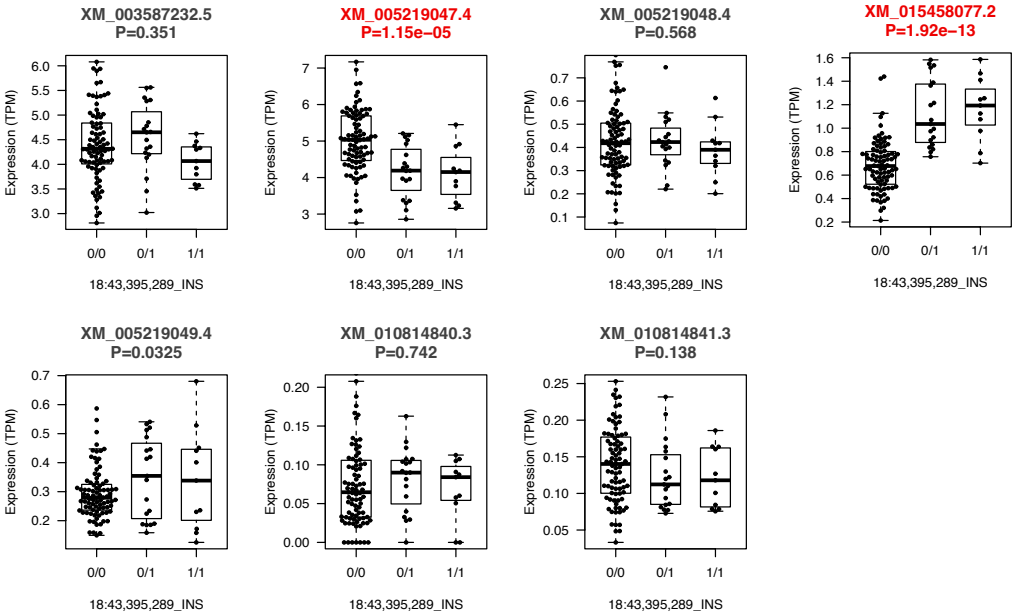
Supplementary Figure 8. Nominal eQTL association significance (left) and normalized TPM values for the expressed gene (right) for (A,B) MYH7 *and (C,D)* LOC112443864.

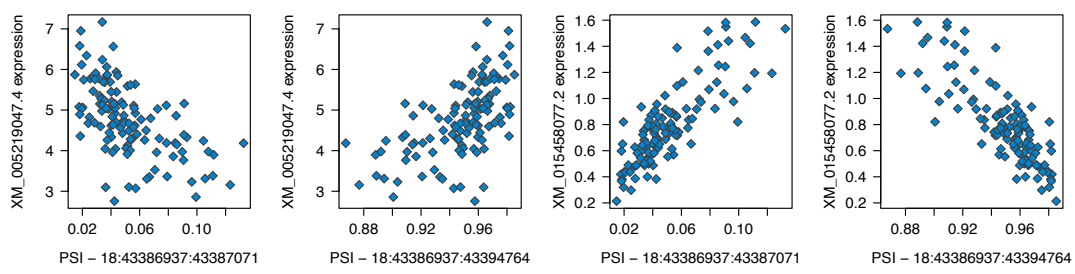


Supplementary Figure 9. (A) Bandage plot for an 11.6 Kb insertion present in seven out of eight samples. There were three distinct alleles, but with >99.9% sequence identify, differing only by six SNPs and one 2 bp indel across the entire SV. (B) Nominal significance of association for the SV and ENSBTAG00000053433. *Given the high sequence identity and limited number of unique* k-*mers, PanGenie could not consistently genotype the 3 different alleles, and so mean and standard deviation are calculated for 5 replicates. The dashed line indicates the conditional significance threshold calculated from the PanGenie+ dataset. Only the merged PanGenie variant was significant, while the three near-identical alleles were all insignificant due to association dilution.*

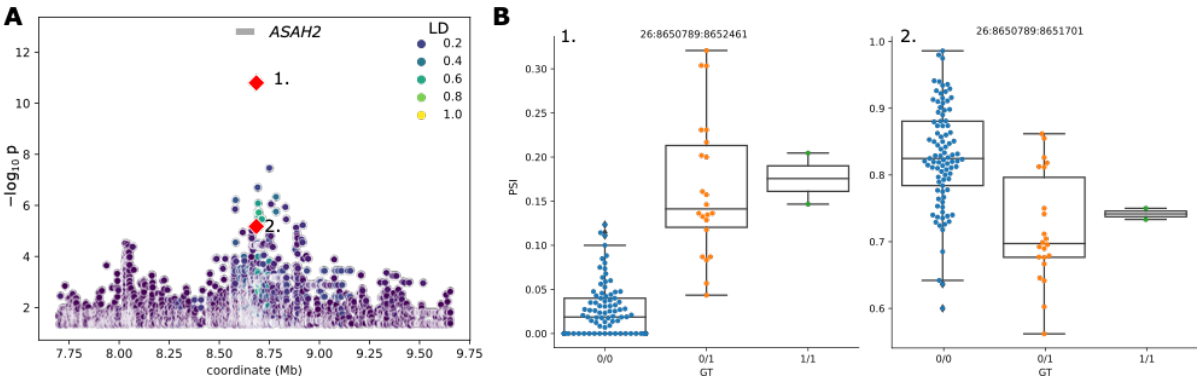

A
18:43386937:43394764
18:43386937:43387071
XM_003587232.5
XM_005219047.4
XM_005219048.4
XM_005219049.4
XM_010814840.3
XM_010814841.3
XM_015458077.2
43400000
43380000
43360000
43340000
43320000
43300000
18:43386937:43394764
18:43386937:43387071

B
ENSBTAG00000004864
P=0.0547
Expression (TPM)
0/0
0/1
1/1
18:43,395,289_INS

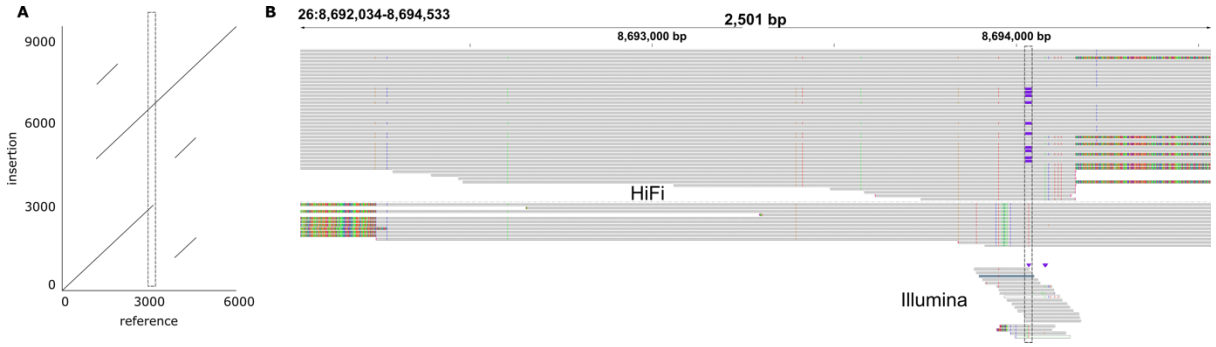

C
XM_003587232.5
P=0.351
XM_005219047.4
P=1.15e−05
XM_005219048.4
P=0.568
XM_015458077.2
P=1.92e−13
XM_005219049.4
P=0.0325
XM_010814840.3
P=0.742
XM_010814841.3
P=0.138
Expression (TPM)
0/0
0/1
1/1
18:43,395,289_INS


D
XM_005219047.4 expression
PSI – 18:43386937:43387071
XM_005219047.4 expression
PSI – 18:43386937:43394764
XM_015458077.2 expression
PSI – 18:43386937:43387071
XM_015458077.2 expression
PSI – 18:43386937:43394764

Supplementary Figure 10. An intronic insertion is associated with CEP89 *splicing and transcript-level expression. A) Structure of seven* CEP89 *isoforms that are annotated in Refseq (version 106). The blue vertical line between the first two exons indicates the position of the insertion. Boxes represent exons. Red color indicates two transcripts whose expression is impacted by the insertion. The orange lines represent the two splice junctions that are associated with the insertion. B) Gene-level expression (quantified in TPM) of* CEP89 *(*ENSBTAG00000004864*) is not affected by the insertion. The P value is from a linear regression of TPM values on the genotype (coded as 0, 1, 2). C) Impact of the insertion on the expression of seven* CEP89 *isoforms. Red color indicates two isoforms (XM_005219047.4, XM_015458077.2) whose expression is significantly (P<0.05/7) associated with the insertion. The P values are from a linear regression of TPM values on the genotype (coded as 0, 1, 2). D) Correlation between the expression of the two significant transcripts and the percent-spliced-in (PSI)-values of the two splice junctions.*



Supplementary Figure 11. (A) Nominal association significance for ASAH2*, where the two red diamonds indicate the same variant affecting two separate junction splicings within the sQTL cluster. (B) PSI (percent spliced in) across the two significantly associated junctions (indicated by number from (A)) within the sQTL cluster.*



Supplementary Figure 12. a) Dotplot of 6 Kb of the reference sequence (26:8691035-8697035) against the syntenic region containing the 3.6 Kb insertion, where the dotted box indicates the start of the SV. The additional off-diagonal elements indicate duplications. b) IGV screenshot for HiFi (top) and Illumina (bottom) sequencing for the same individual who was heterozygous for the SV, where the dotted box again indicates the start of the SV. Within the dotted box, purple rectangles indicate the insertion SV allele, and red lines indicate the erroneous C-to-T SNP. All HiFi reads not containing the insertion have clear rainbow-banding patterns at their ends, indicating soft-clipping and lower quality alignment compared to reads clearly showing the insertion. Illumina reads were generally too short to confidently state the SNP-reads were misaligned, and so the SNP looks valid.

*Supplementary Table 1. Short read-based structural variation detection and genotyping compared to long read detection using the same 25 samples for 81 cis-SV-QTL that were not detected using short reads. For genotype accuracy, * indicates there were multiple distinct alleles and the highest accuracy was taken, while a dash indicates the variant was not discovered by short reads.*

QTL type	Gene	Chromosome	Variant position	Size (bp)	Genotype accuracy
Expression	*ENSBTAG00000014441*	1	56280167	53	92%*
Expression	*ENSBTAG00000015198*	1	131145050	-154	100%
Expression	*ENSBTAG00000007039*	1	146551120	250	-
Expression	*ENSBTAG00000043956*	2	15839284	-142	100%
Expression	*ENSBTAG00000024288*	3	39781161	209	100%*
Expression	*ENSBTAG00000047605*	4	99065240	143	83%

Expression	*ENSBTAG00000003864*	4	99481913	-877	100%
Expression	*ENSBTAG00000026977*	5	10071225	-698	-
Expression	*ENSBTAG00000002227*	5	54954146	-51	-
Expression	*ENSBTAG00000001057*	5	113638212	-63	-
Expression	*ENSBTAG00000006683*	6	116283443	670	-
Expression	*ENSBTAG00000003103*	7	37944012	-61	-
Expression	*ENSBTAG00000008962*	8	91917309	76	100%
Expression	*ENSBTAG00000009703*	10	21484330	388	-
Expression	*ENSBTAG00000032869*	10	26332517	-562	92%
Expression	*ENSBTAG00000006470*	10	37917838	57	86%
Expression	*ENSBTAG00000011494*	10	43976374	-2181	96%
Expression	*ENSBTAG00000020630*	11	50581272	-863	96%
Expression	*ENSBTAG00000048781*	11	104137996	-53	-
Expression	*ENSBTAG00000052185*	12	75339058	64	83%
Expression	*ENSBTAG00000000291*	13	12190635	198	83%
Expression	*ENSBTAG00000006278*	13	64812316	108	-
Expression	*ENSBTAG00000049782*	13	73587243	-59	100%
Expression	*ENSBTAG00000013955*	14	3342962	119	-
Expression	*ENSBTAG00000050111*	15	49001256	-80	79%
Expression	*ENSBTAG00000014847*	15	57829264	-87	20%*
Expression	*ENSBTAG00000005913*	15	82676875	177	75%
Expression	*ENSBTAG00000024539*	16	44118432	-95	-
Expression	*ENSBTAG00000027182*	17	9926879	545	67%
Expression	*ENSBTAG00000034522*	17	11258716	917	100%

Expression	*ENSBTAG00000015811*	17	18449267	-3105	100%
Expression	*ENSBTAG00000019891*	17	72370701	-53	71%
Expression	*ENSBTAG00000022783*	18	29948163	-409	92%
Expression	*ENSBTAG00000011616*	19	9427531	138	96%
Expression	*ENSBTAG00000054102*	19	62391084	251	-
Expression	*ENSBTAG00000001061*	19	62832442	1863	-
Expression	*ENSBTAG00000006775*	20	70350801	-844	-
Expression	*ENSBTAG00000003132*	21	31041012	135	63%*
Expression	*ENSBTAG00000007109*	21	58568918	-324	-
Expression	*ENSBTAG00000004117*	22	2807984	110	54%
Expression	*ENSBTAG00000001889*	22	39001060	-118	92%
Expression	*ENSBTAG00000053433*	23	27953867	11556	-
Expression	*ENSBTAG00000012870*	23	4565407	846	71%*
Expression	*ENSBTAG00000011622*	24	21134872	-90	-
Expression	*ENSBTAG00000002374*	25	27604763	112	62%*
Expression	*ENSBTAG00000015019*	26	24425065	-5902	-
Expression	*ENSBTAG00000019792*	28	16085648	37	-
Expression	*ENSBTAG00000002670*	28	44601251	-116	-
Expression	*ENSBTAG00000016470*	29	47880247	-74	72%
Expression	*ENSBTAG00000018274*	29	48053399	-60	-
Expression	*ENSBTAG00000022160*	29	48547803	273	-
Splicing	*ENSBTAG00000018502*	2	118926093	-884	-
Splicing	*ENSBTAG00000050039*	4	69754876	-2599	-
Splicing	*ENSBTAG00000037534*	5	100296038	1954	67%*

Splicing	*ENSBTAG00000052571*	7	17329610	-43206	-
Splicing	*ENSBTAG00000011402*	8	59609182	1647	-
Splicing	*ENSBTAG00000013991*	9	41961852	-780	100%
Splicing	*ENSBTAG00000013176*	9	102776565	-68	-
Splicing	*ENSBTAG00000025324*	10	90423226	69	-
Splicing	*ENSBTAG00000018596*	10	100171914	-50	-
Splicing	*ENSBTAG00000008448*	11	63907254	547	87%*
Splicing	*ENSBTAG00000020791*	11	101716574	53	87%
Splicing	*ENSBTAG00000009415*	12	15183906	67	79%*
Splicing	*ENSBTAG00000000939*	13	10221508	53	100%
Splicing	*ENSBTAG00000008812*	13	73810018	138	-
Splicing	*ENSBTAG00000008812*	13	81632248	-68	-
Splicing	*ENSBTAG00000012242*	14	133719	143	96%
Splicing	*ENSBTAG00000021472*	14	1210137	396	-
Splicing	*ENSBTAG00000051547*	14	19659293	221	-
Splicing	*ENSBTAG00000017694*	14	48339115	-70	88%
Splicing	*ENSBTAG00000012433*	16	76538065	146	-
Splicing	*ENSBTAG00000016813*	17	6760179	-100	-
Splicing	*ENSBTAG00000022022*	17	71490201	-1297	86%
Splicing	*ENSBTAG00000004864*	18	43395289	1293	96%
Splicing	*ENSBTAG00000002844*	18	55500165	1290	92%
Splicing	*ENSBTAG00000017275*	19	7835372	203	92%*
Splicing	*ENSBTAG00000035544*	21	64949347	-175	-
Splicing	*ENSBTAG00000012067*	24	45470681	76	50%*

Splicing	*ENSBTAG00000024884*	25	2011244	-77	-
Splicing	*ENSBTAG00000003529*	26	8694035	3629	-
Splicing	*ENSBTAG00000032730*	29	41435647	57	87%

Supplementary Table 2. SV-QTL candidates for expression and splicing phenotypes, with conditional p-values and effect sizes.

External file

*Supplementary Table 3. Percentages of repeat elements across all ARS-UCD1.2 autosomes (Genome-wide) or the top-associated SV-e/SQTL. Statistical significance was calculated using a one-sided Fisher's exact test on the total number of bases annotated as the respective elements between genome-wide and eQTL and genome-wide and sQTL. Significant p-values are indicated by *.*

	DNA transposons	SINE	LINE	LTR	nonLTR-RTE	Total
Genome-wide	1.23	8.11	15.08	3.05	4.94	40.40
eQTL	1.36*	6.47	34.44*	3.07	6.13*	59.40*
sQTL	0.55	5.83	39.57*	17.65*	5.83*	78.26*