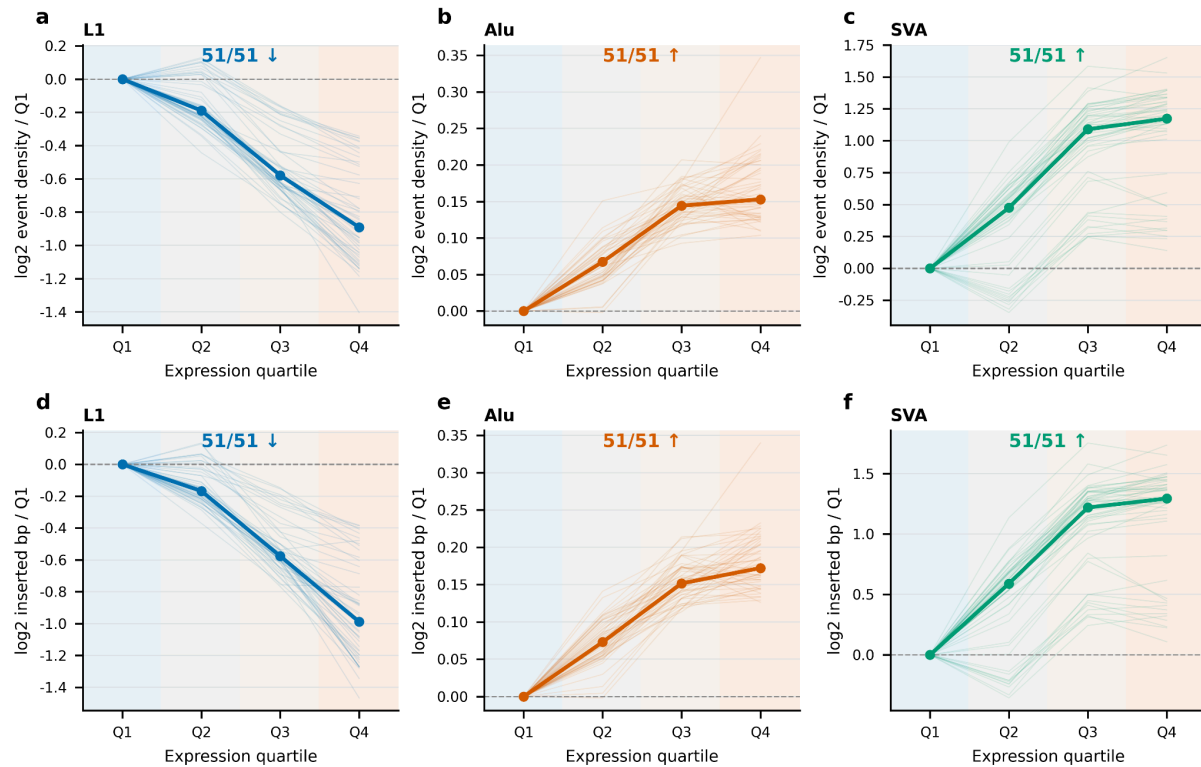
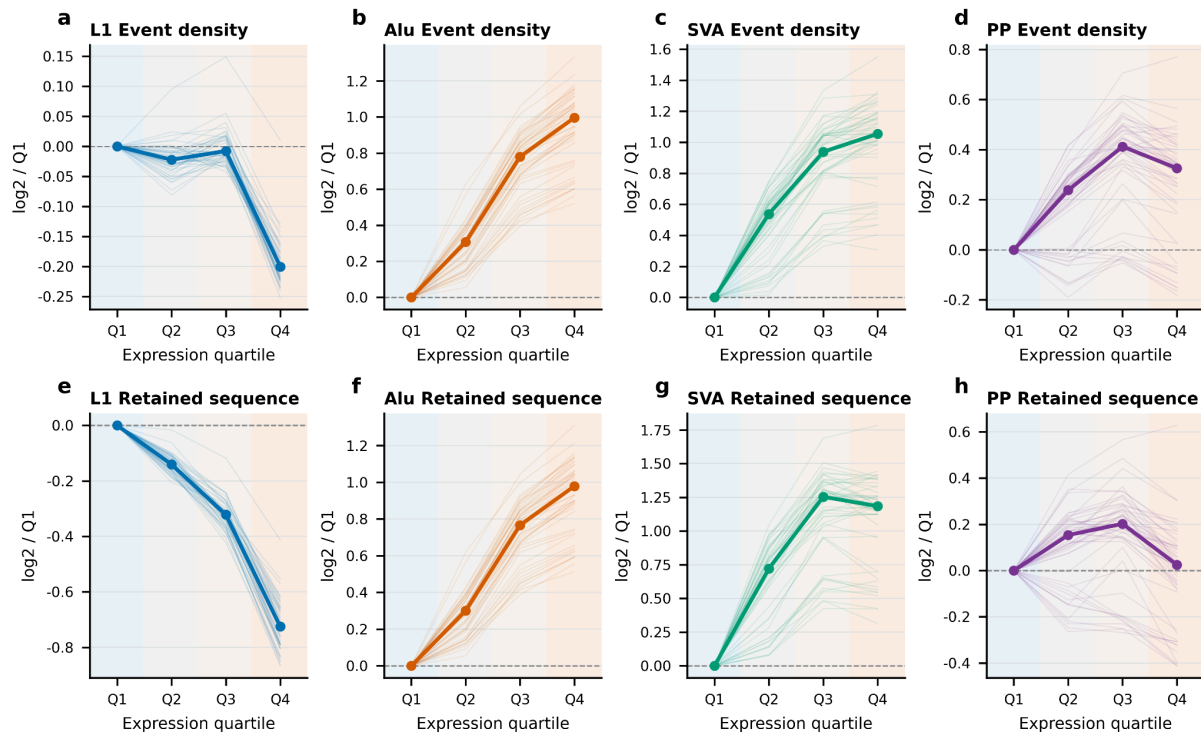


Supplementary Figure



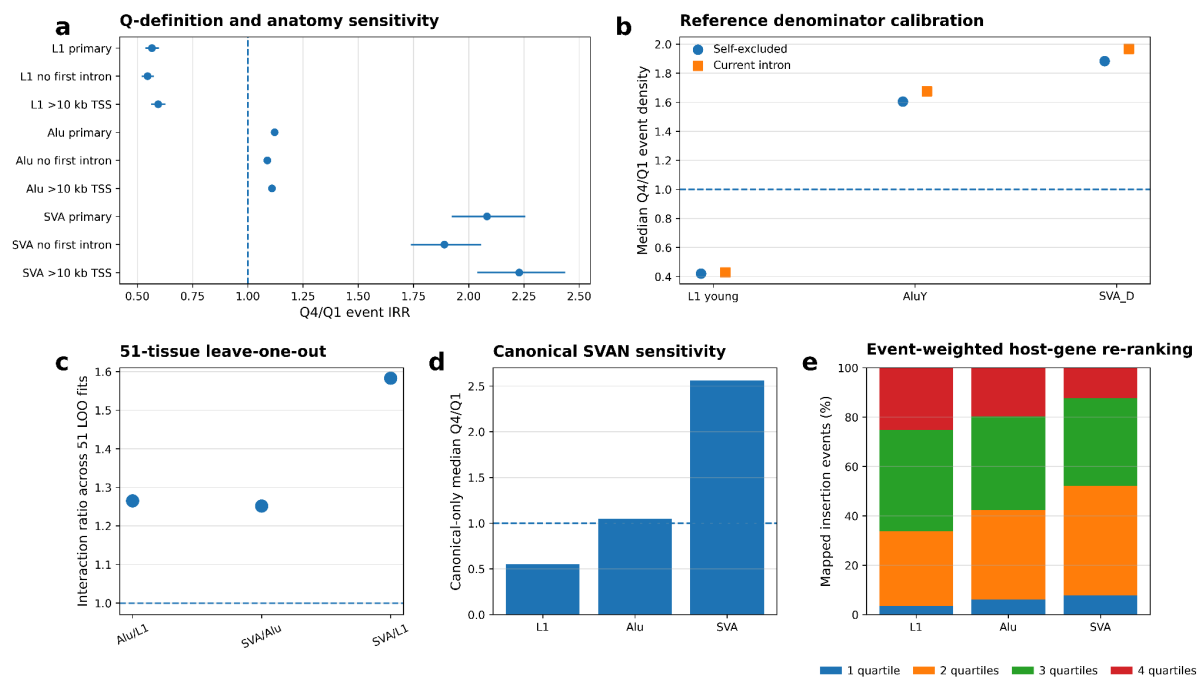
Supplementary Figure 1 | 1KGP polymorphic L1, Alu and SVA trajectories across 51 GTEx tissues

Long-read 1000 Genomes Project (1KGP) polymorphic L1, Alu and SVA insertions annotated by SVAN v1.3 in the native T2T-CHM13v2.0 (hs1) assembly. Individual Q1-Q4 event-density and inserted-sequence-burden trajectories are shown after host genes were ranked independently within each of the 51 retained GTEx anatomical tissues while the 1KGP insertion catalogue was held fixed. Thin lines show tissue-specific trajectories and overlaid lines show tissue medians.



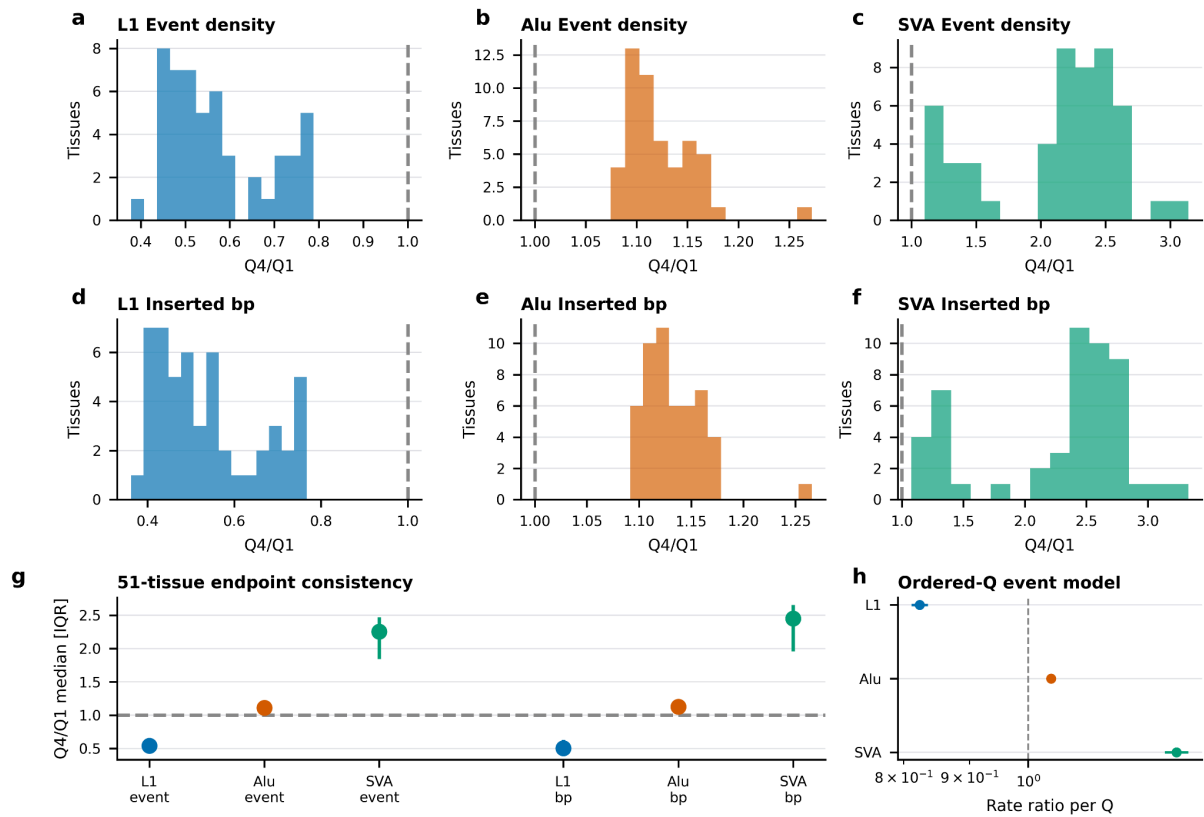
Supplementary Figure 2 | GRCh38 reference-retained cargo trajectories across 51 GTEx tissue rankings

Reference-retained L1, Alu and SVA events reconstructed from GRCh38 RepeatMasker annotations and processed pseudogenes (PPs) defined from GENCODE v48. Events were assigned to protein-coding introns and evaluated across the same 51 GTEx tissue-specific host-expression rankings. Tissue-resolved Q1-Q4 event-density and retained-sequence trajectories are shown for L1, Alu, SVA and PP. PP is treated as a cellular-mRNA-derived retrocopy rather than a transposable element.



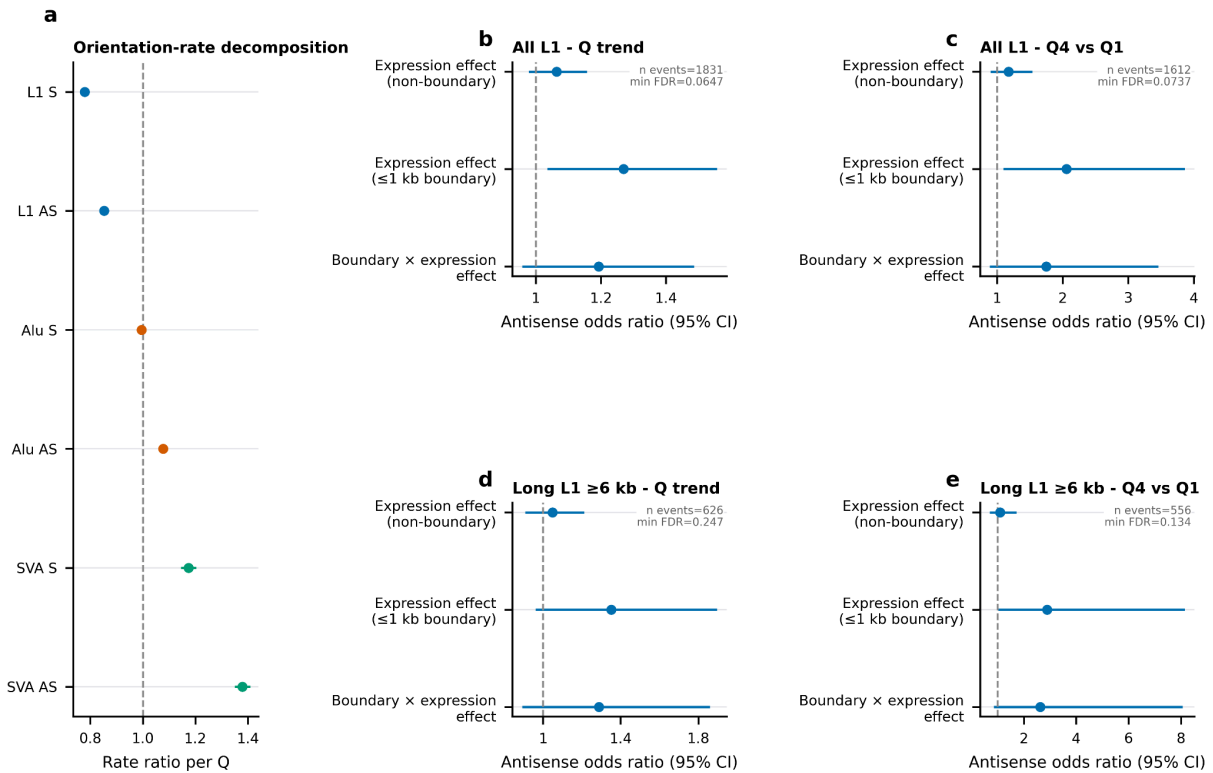
Supplementary Figure 3 | 1KGP robustness and host re-ranking with GRCh38 reference-denominator calibration

a, Sensitivity of the long-read 1KGP polymorphic L1, Alu and SVA Q4/Q1 event effects to the expression-group definition and anatomical-tissue policy in hs1. b, GRCh38 reference-retained TE denominator calibration comparing self-excluded and conventional current-intron denominators. c, Direct cargo-interaction estimates for the 1KGP polymorphic insertion analysis from 51 leave-one-tissue-out refits. d, Canonical-SVAN-only sensitivity of median Q4/Q1 event-density effects in the 1KGP polymorphic catalogue. e, Event-weighted number of distinct GTEx expression quartiles occupied across 51 tissues by host genes linked to mapped 1KGP L1, Alu and SVA insertion events.



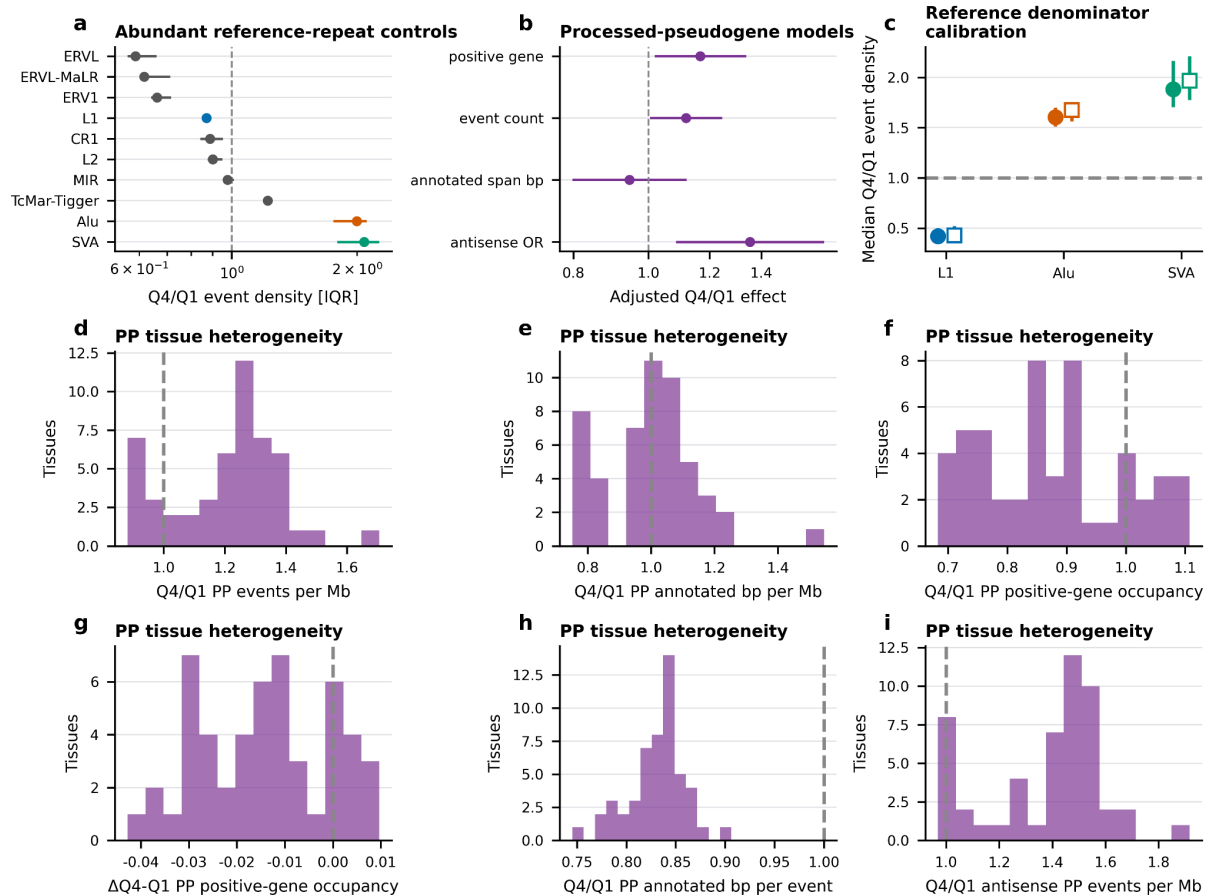
Supplementary Figure 4 | 1KGP tissue-wise endpoints and ordered-Q models

All panels use the long-read 1KGP polymorphic L1, Alu and SVA insertion catalogue in *hs1* with host genes independently ranked across 51 GTEx anatomical tissues. a-c, Tissue-wise Q4/Q1 event-density distributions for L1, Alu and SVA. d-f, Corresponding Q4/Q1 inserted-sequence-burden distributions. g, Tissue-wise endpoint consistency for event count and inserted sequence. h, Ordered-Q within-cargo rate ratios for event count and inserted sequence.



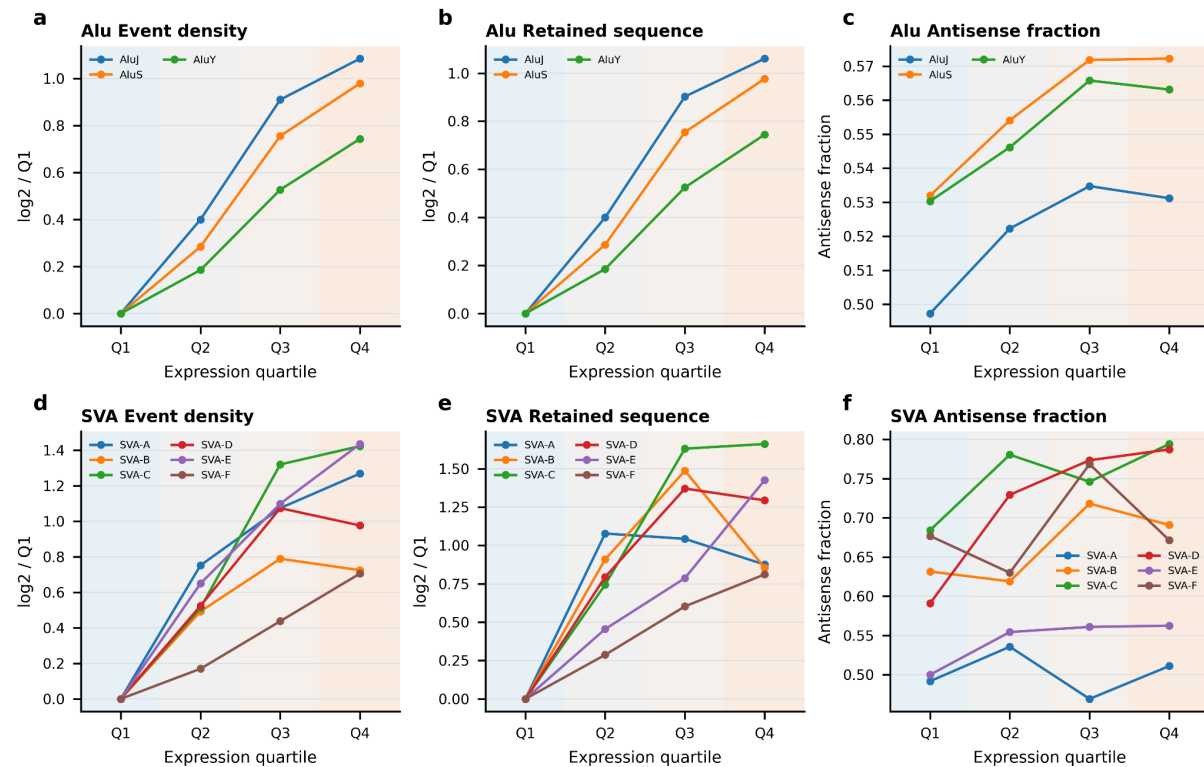
Supplementary Figure 5 | 1KGP orientation-rate decomposition and L1 exon-boundary tests

All panels use orientation-resolved long-read 1KGP polymorphic insertions in hs1 and the 51 GTEx tissue-specific host-expression rankings. a, Ordered-Q decomposition of sense- and antisense-specific event-density rate ratios for L1, Alu and SVA. b,c, Adjusted antisense-odds models for all 1KGP L1 insertions testing host-gene expression, proximity within 1 kb of the nearest exon boundary and the boundary-by-expression interaction under ordered-Q and Q4-versus-Q1 contrasts. d,e, Corresponding models restricted to long 1KGP L1 insertions of at least 6 kb.



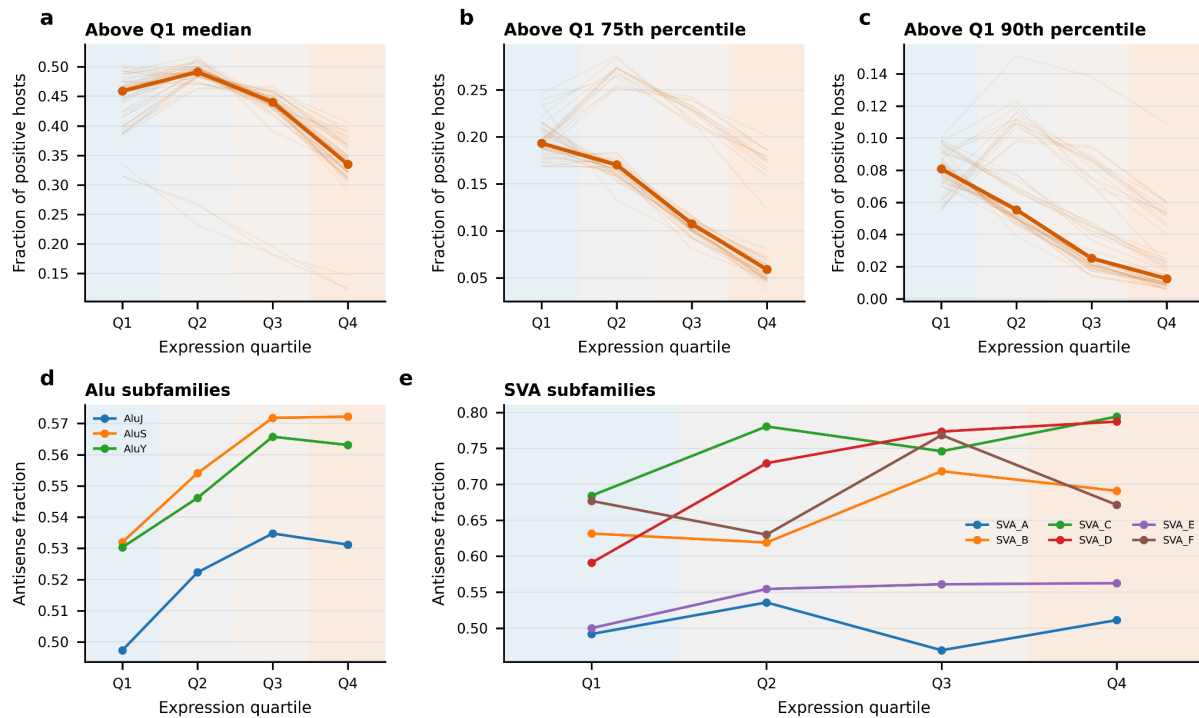
Supplementary Figure 6 | GRCh38 reference-retained controls and processed-pseudogene analyses

All panels belong to the GRCh38 reference layer evaluated across the 51 GTEx tissue-specific host-expression rankings. Reference TE panels use reconstructed RepeatMasker events, whereas processed-pseudogene panels use GENCODE v48 PP annotations. a, Broad reference-repeat controls for Q4/Q1 event-density effects. b, Adjusted PP associations for positive-host occupancy, event count, annotated-span burden and antisense orientation. c, Reference-TE denominator calibration comparing self-excluded and current-intron definitions. d-i, Tissue-level distributions of PP event density, annotated-span burden, positive-host occupancy, occupancy change, annotated span per event and antisense-event metrics. PP is a cellular-mRNA-derived retrocopy and is not classified as a transposable element.



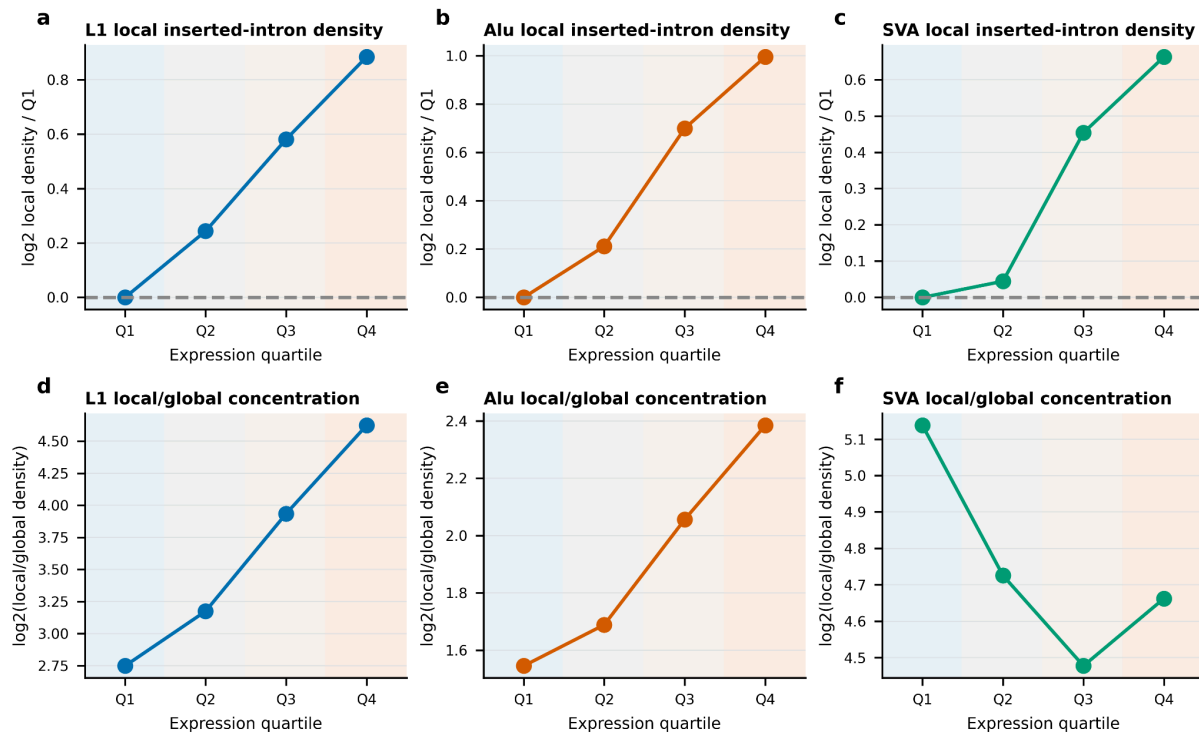
Supplementary Figure 7 | GRCh38 reference-retained Alu and SVA subfamily trajectories

GRCh38 reference-retained Alu and SVA events were reconstructed from RepeatMasker annotations, assigned to GENCODE v48 protein-coding introns and evaluated across 51 GTEx tissue-specific host-expression rankings. a-c, Q1-Q4 event-density, retained-sequence-burden and antisense-fraction trajectories for AluJ, AluS and AluY. d-f, Corresponding trajectories for SVA_A-SVA_F.



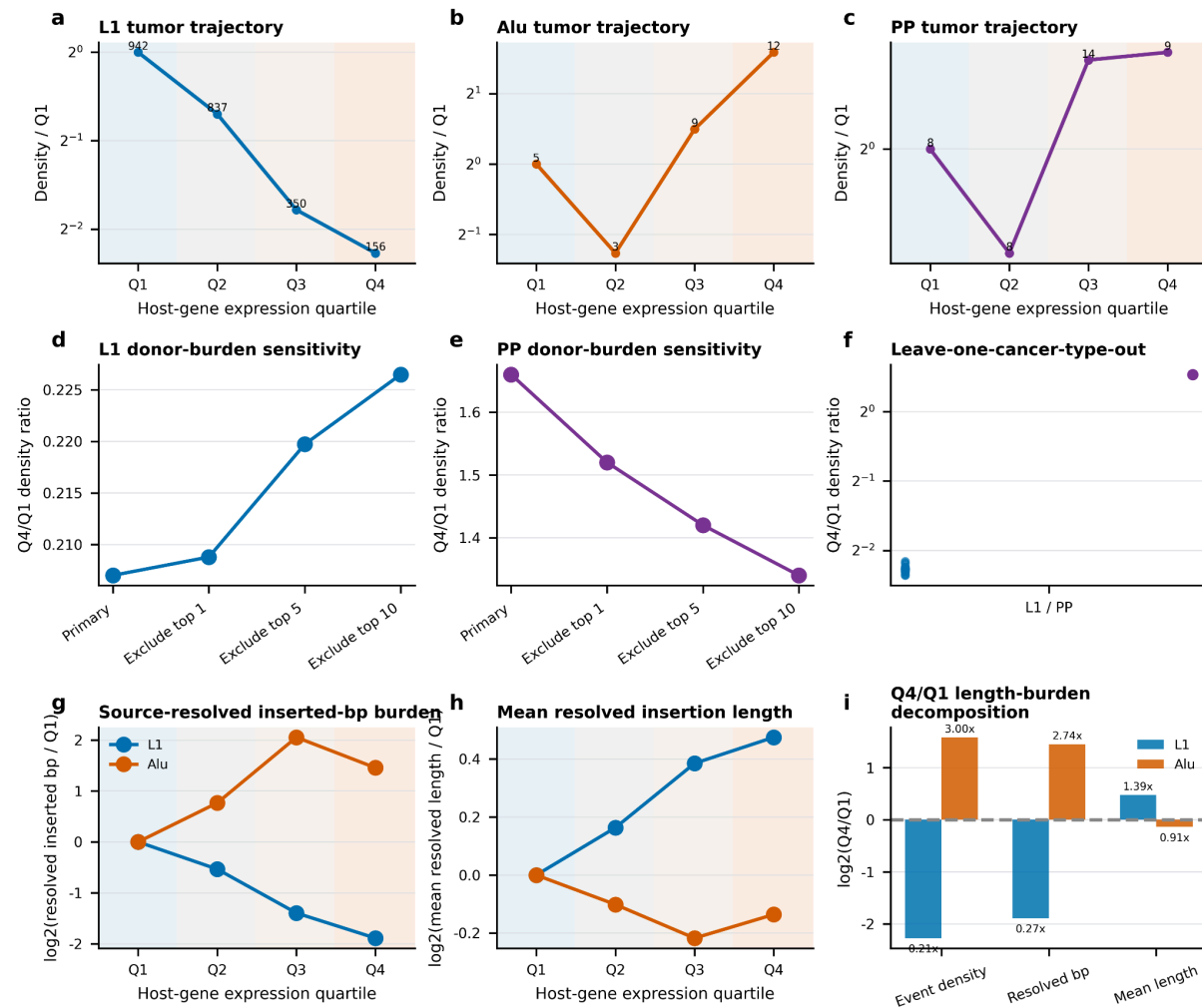
Supplementary Figure 8 | 1KGP Alu multiplicity sensitivity and GRCh38 Alu/SVA subfamily antisense trajectories

a-c, Long-read 1KGP polymorphic Alu insertions in hs1. The fraction of Alu-positive hosts exceeding the tissue-specific Q1 median, 75th-percentile and 90th-percentile event-count thresholds is shown across Q1-Q4 for the 51 GTEx anatomical tissues. d,e, GRCh38 reference-retained Alu and SVA subfamilies reconstructed from RepeatMasker. Antisense-fraction trajectories are shown across the same 51 GTEx tissue-specific expression rankings. Thus, panels a-c are the 1KGP polymorphic layer, whereas panels d,e are the GRCh38 reference-retained layer.



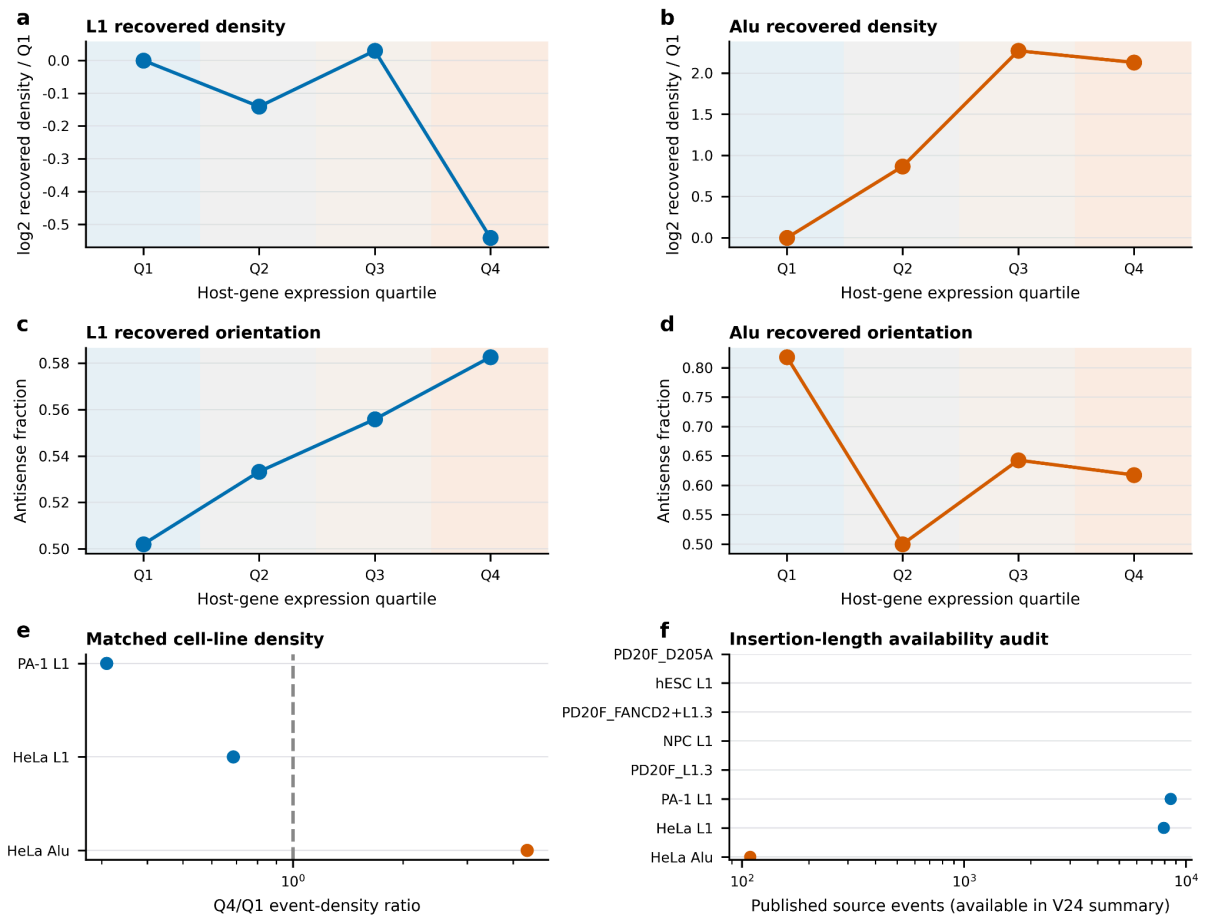
Supplementary Figure 9 | 1KGP local inserted-intron density and concentration

All panels use the long-read 1KGP polymorphic L1, Alu and SVA insertion catalogue in hs1 across the 51 GTEx tissue-specific host-expression rankings. a-c, Local event density using only the introns that actually contain the focal 1KGP cargo as the denominator. d-f, Corresponding local-to-global concentration metrics across Q1-Q4.



Supplementary Figure 10 | PCAWG somatic retrotransposition trajectories, sensitivity and source-length analyses

PCAWG somatic retrotransposition calls were analyzed in GRCh37 with GENCODE v19 protein-coding introns. Donor-matched tumor RNA from E-MTAB-5200 defined donor-specific Q1-Q4 host-gene expression strata. A-C, Somatic L1, Alu and source-designated PP event-density trajectories. D,E, Donor-burden trimming sensitivity for L1 and PP. F, Leave-one-cancer-type-out Q4/Q1 density ratios for L1 and PP. G,H, Source-resolved reported retrotransposed-base-pair burden and mean resolved insertion length for L1 and Alu. I, Q4/Q1 decomposition of event density, resolved-base-pair burden and mean resolved insertion length. Formal pan-cancer inference is restricted to L1. Sparse Alu and PP results are supporting or descriptive.



Supplementary Figure 11 | Experimentally recovered L1 and Alu insertion analyses

Published experimentally recovered L1 and Alu insertion maps were reannotated against a common GENCODE v19 protein-coding intron background. L1 maps derive from the recovered-insertion datasets of Flasch et al. (2019), and Alu maps derive from Wagstaff et al. (2012). Human Protein Atlas cell-line expression was used to define host-gene Q1-Q4 strata in the corresponding cell line. A,B, Recovered L1 and Alu event-density trajectories. C,D, Antisense-fraction trajectories for recovered L1 and Alu. E, Matched cell-line Q4/Q1 event-density ratios for HeLa Alu, HeLa L1 and PA-1 L1. F, Audit of insertion-length-field availability across the published recovered-insertion source sets. These are recovered insertion landscapes after cell growth, survival, selection and ascertainment and are not insertion-time snapshots.