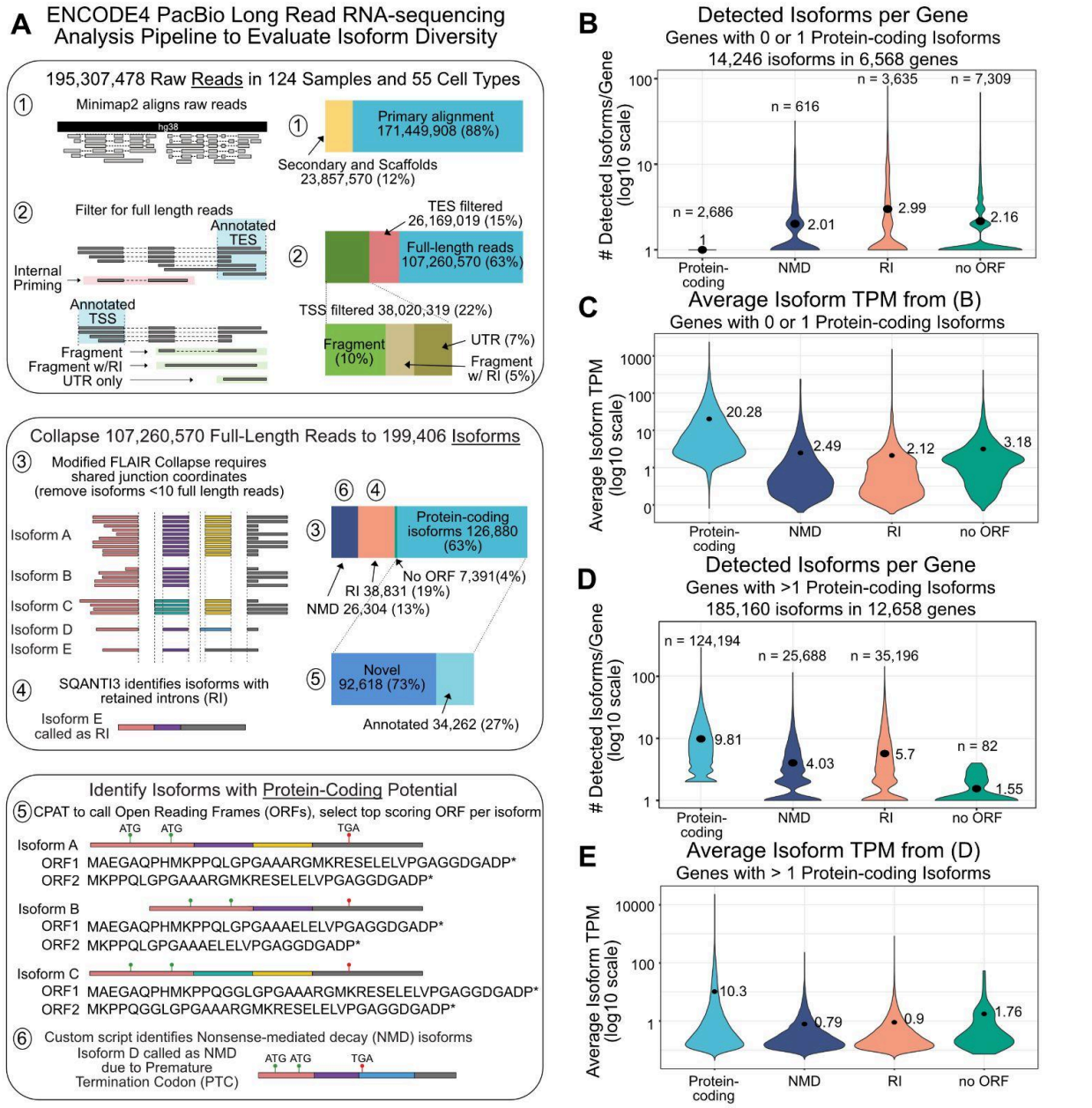

Fig S1. (A) Schematic of our custom analysis pipeline to process ENCODE4 PacBio LRS data. Raw reads were aligned to the reference genome (hg38) using Minimap2 (Step 1), followed by filtering for full-length reads matching annotated transcription start and end sites (TSS and TES; Step 2). Isoforms were collapsed using a modification of the *FLAIR Collapse* script requiring shared splice junctions and a minimum of 10 full-length supporting reads across samples (Step 3). SQANTI3 was used to classify retained introns (RI), which were then removed (Step 4). CPAT was applied to identify open reading frames (ORFs; Step 5), and a custom script removed isoforms predicted to undergo nonsense-mediated decay (NMD; Step 6). Barplots to the right of each schematic show the number of reads or isoforms at each step of processing. The circled numbers 1-6 show step(s) to which each barplot corresponds. Percentages and counts are indicated for raw reads, full-length reads, and isoforms. (B) The number of detected isoforms

per gene, for genes with 0-1 protein-coding isoforms, grouped by transcript biotype. NMD = nonsense-mediated decay; RI = retained intron. Numbers next to each violin are the average, and ‘n’ values above indicate the number of isoforms per biotype. (C) Averaged isoform TPMs across samples for genes in (B). Numbers next to the violins are the average TPM per transcript biotype. (D-E) Analogous plots to B-C but for genes with multiple (>1) protein-coding isoforms.

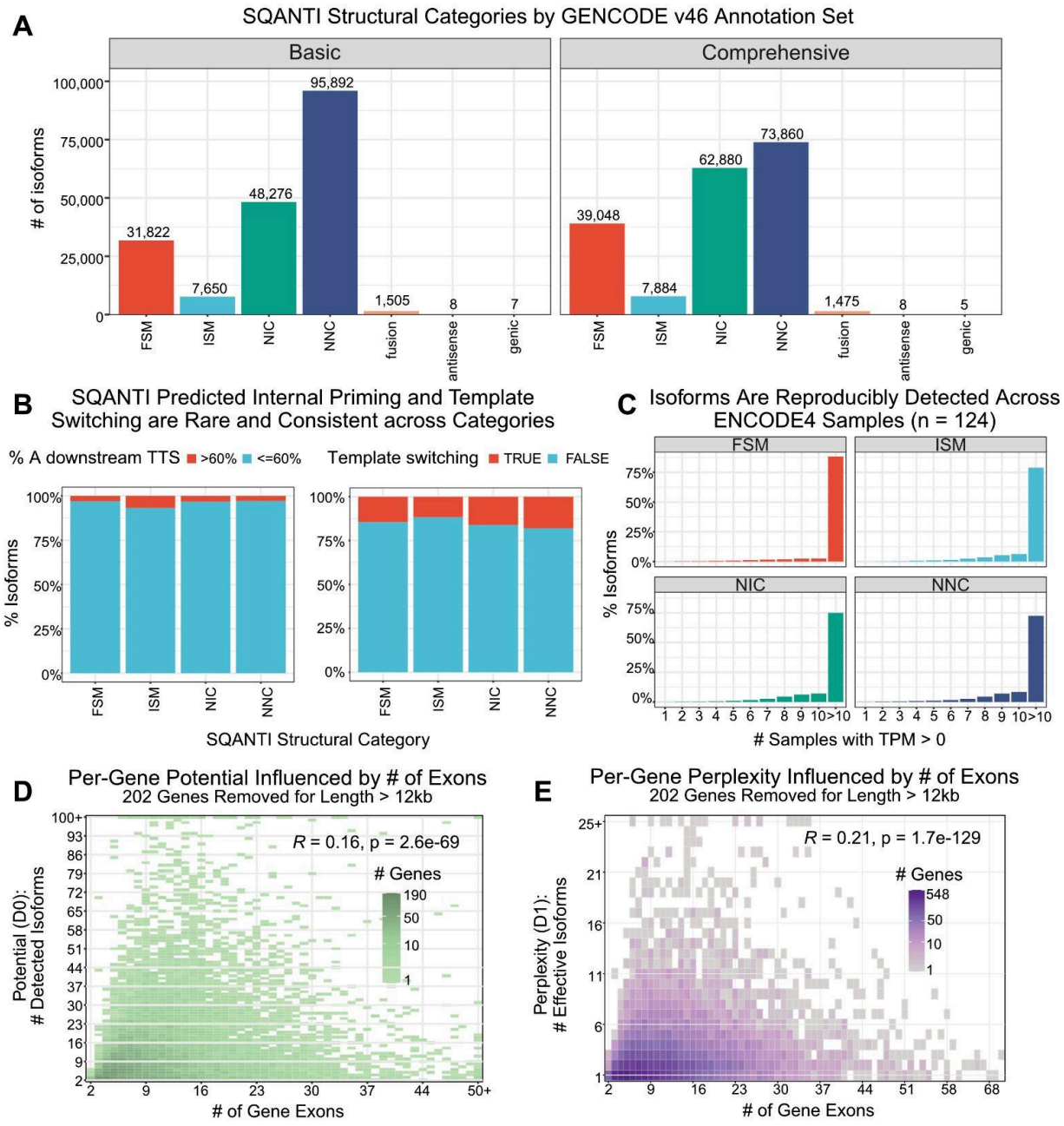

Fig S2. (A) SQANTI3 structural categories for all 185,160 isoforms using GENCODE v46 basic (left) and comprehensive (right) annotations. FSM = full splice match, ISM = incomplete splice match, NIC = novel in catalog, NNC = novel not in catalog. The number of isoforms is shown above each bar. (B) SQANTI3 quality metrics across structural categories for isoforms passing the technical artifact filtering described in Methods. As expected, predicted internal priming, defined as >60% A content in the genomic window downstream of the transcription termination site (TTS; left) and predicted RT template switching (right) are rare and consistent across categories. (C) Reproducibility of isoform detection across the 124 ENCODE4 samples, binned by the number of samples in which each isoform is detected (TPM > 0), faceted by structural category. 75% of NIC and 72.5% of NNC isoforms are detected in more than 10 samples, and

only 0.1% are detected in a single sample. (D-E) Relationship between number of gene exons and per-gene potential (D) and per-gene perplexity (E). Genes with an annotated isoform longer than 12kb were excluded (n = 202) due to the known length bias in long-read sequencing. Color intensity corresponds to the number of genes at each point. Pearson's *R* and p-values are shown.

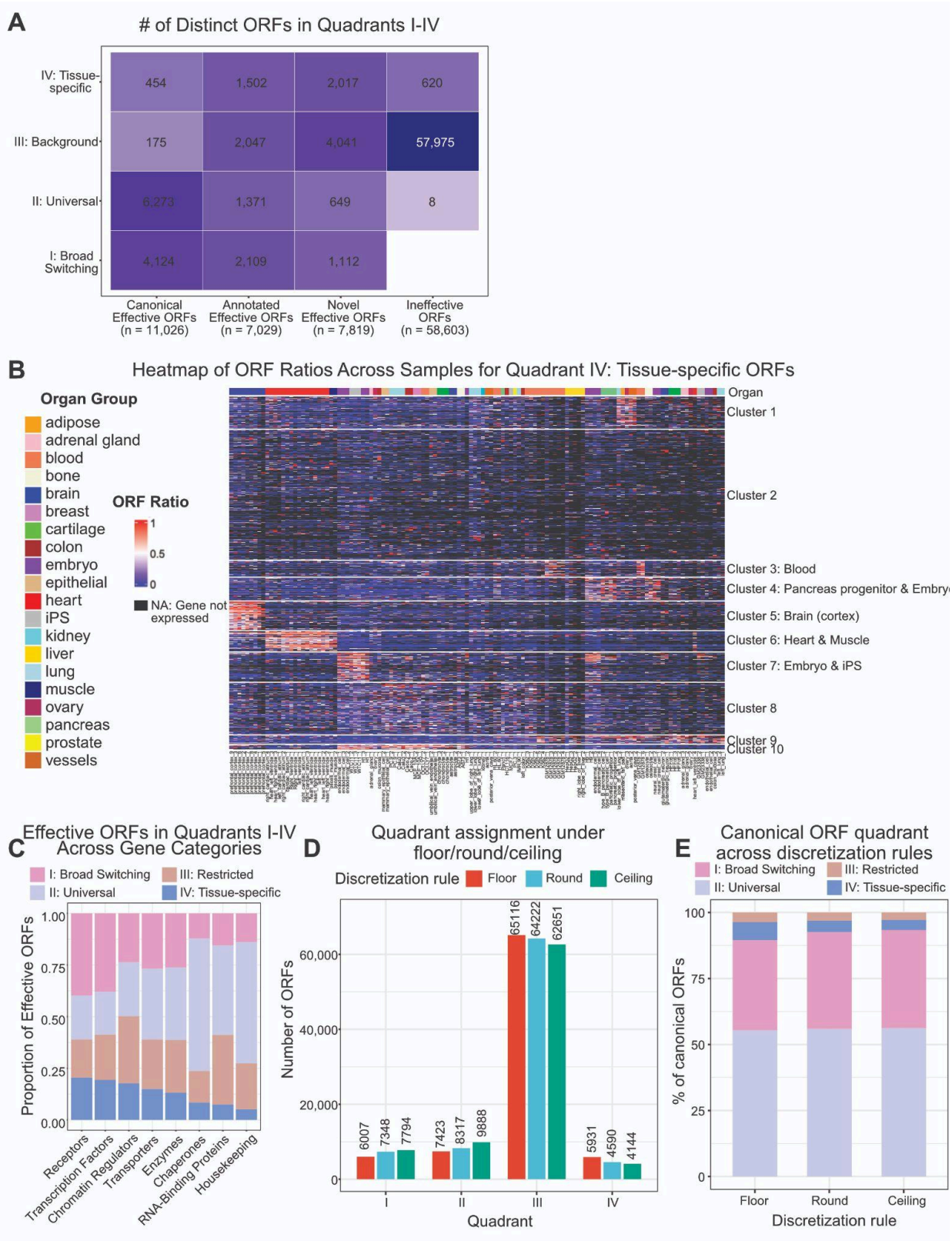
A
# of Distinct ORFs in Quadrants I-IV
IV: Tissue-specific
III: Background
II: Universal
I: Broad Switching
454
1,502
2,017
620
175
2,047
4,041
57,975
6,273
1,371
649
8
4,124
2,109
1,112
Canonical Effective ORFs (n = 11,026)
Annotated Effective ORFs (n = 7,029)
Novel Effective ORFs (n = 7,819)
Ineffective ORFs (n = 58,603)
B
Heatmap of ORF Ratios Across Samples for Quadrant IV: Tissue-specific ORFs
Organ Group
adipose
adrenal gland
blood
bone
brain
breast
cartilage
colon
embryo
epithelial
heart
iPS
kidney
liver
lung
muscle
ovary
pancreas
prostate
vessels
ORF Ratio
1
0.5
0
NA: Gene not expressed
Organ
Cluster 1
Cluster 2
Cluster 3: Blood
Cluster 4: Pancreas progenitor & Embry
Cluster 5: Brain (cortex)
Cluster 6: Heart & Muscle
Cluster 7: Embryo & iPS
Cluster 8
Cluster 9
Cluster 10
C
Effective ORFs in Quadrants I-IV Across Gene Categories
I: Broad Switching
III: Restricted
II: Universal
IV: Tissue-specific
Proportion of Effective ORFs
1.00
0.75
0.50
0.25
0.00
Receptors
Transcription Factors
Chromatin Regulators
Transporters
Enzymes
Chaperones
RNA-Binding Proteins
Housekeeping
D
Quadrant assignment under floor/round/ceiling
Discretization rule
Floor
Round
Ceiling
Number of ORFs
60,000
40,000
20,000
0
6007
7348
7794
7423
8317
9888
65116
64222
62651
5931
4590
4144
I
II
III
IV
Quadrant
E
Canonical ORF quadrant across discretization rules
I: Broad Switching
III: Restricted
II: Universal
IV: Tissue-specific
% of canonical ORFs
100
75
50
25
0
Floor
Round
Ceiling
Discretization rule

Fig S3. (A) The number of distinct ORFs across ORF types and quadrants I-IV. X-axis numbers are sums of columns. There are no ORFs that are both Broad Switching and Ineffective. (B) Tissue-specific ORFs are hierarchically clustered by Euclidean distance of their ORF usage ratios, resulting in ten groups. Organ types are marked at the top of the heatmap. For samples where the gene is not expressed, the ORF ratio is NA and colored black on the heatmap whereas a value of 0 indicates that the gene is expressed but the ORF is not. (C) The proportion of effective ORFs from each quadrant across the different gene categories–ordered on the x-axis by the proportion in type IV: Tissue-specific. For example, Receptors have the highest proportion of tissue-specific ORFs, while housekeeping has the lowest. (D) Number of ORFs assigned to each quadrant when per-sample ORF perplexity is discretized by floor, round, or ceiling. Counts are shown above each bar. (E) Quadrant distribution of canonical ORFs under each discretization rule. Canonical ORFs remain predominantly in the broadly-expressed quadrants (I and II) under all three rules.