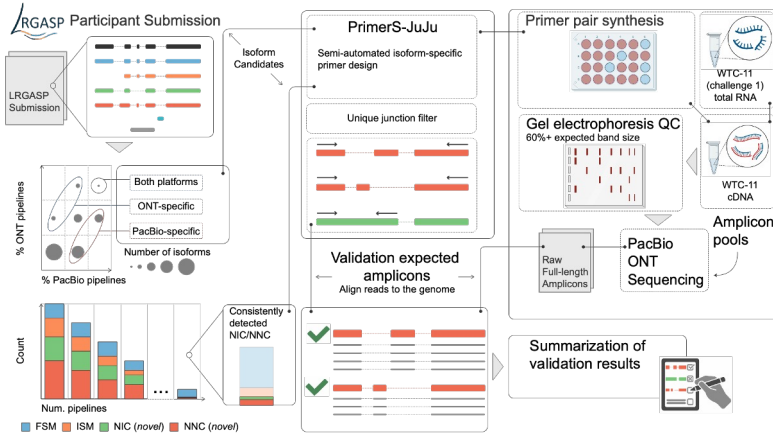


a

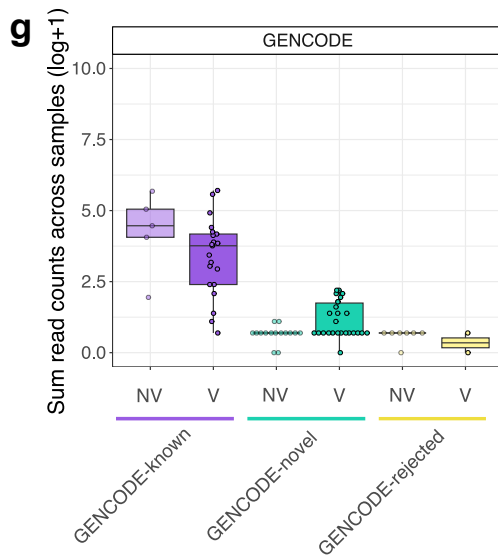
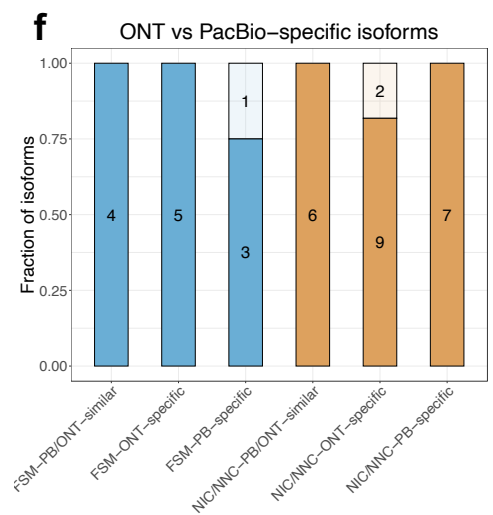
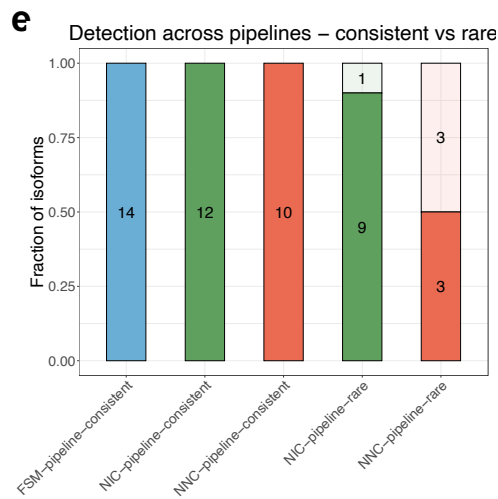
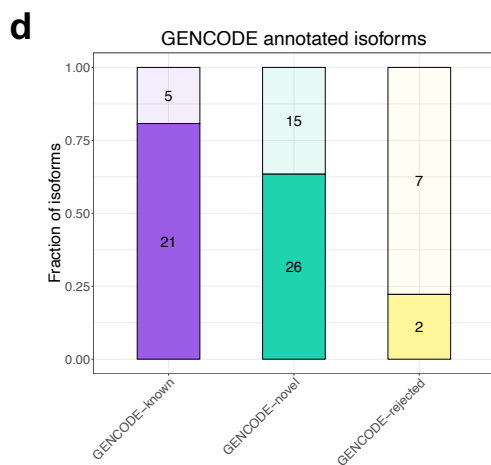
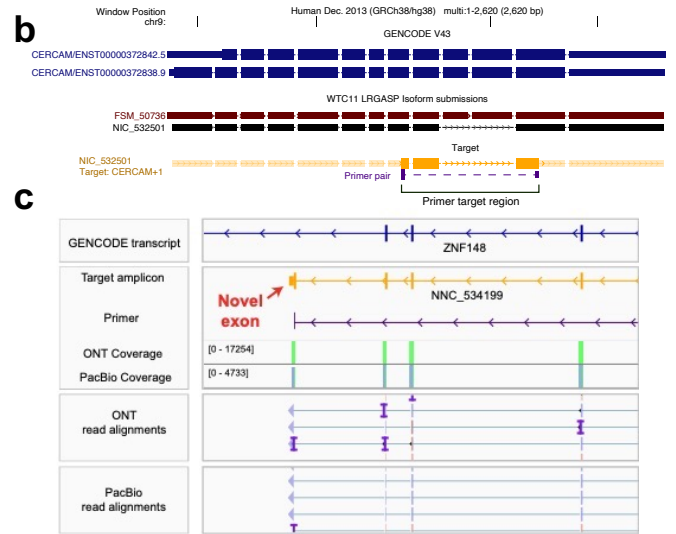


GENCODE status

- GENCODE-known
- GENCODE-novel
- GENCODE-rejected

SQANTI category FSM NNC
 NIC  NIC/NNC

validation  not_validated
 validated



NV: not validated
V: validated

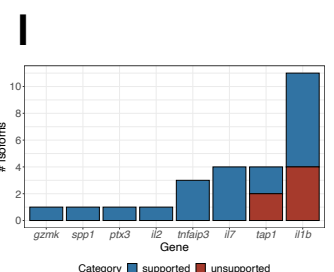
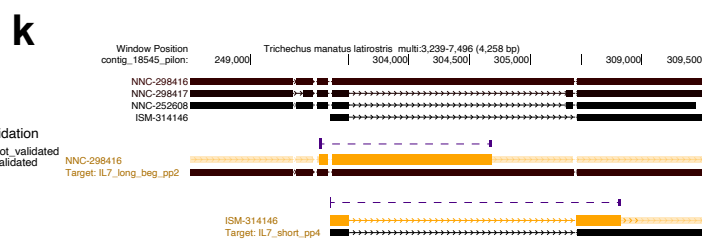
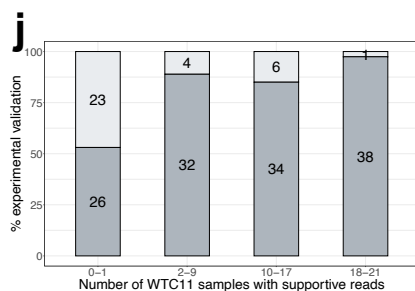
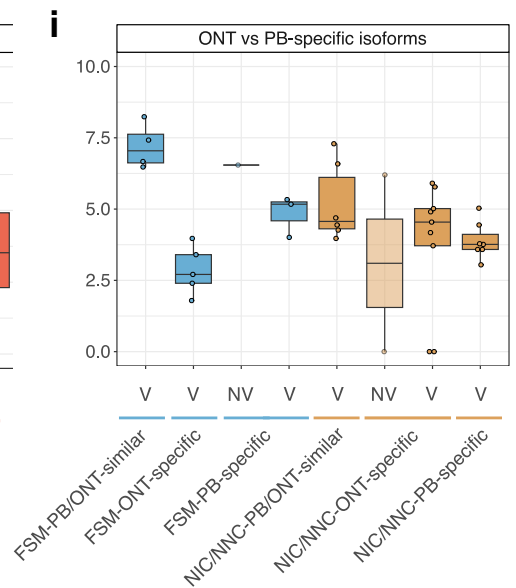
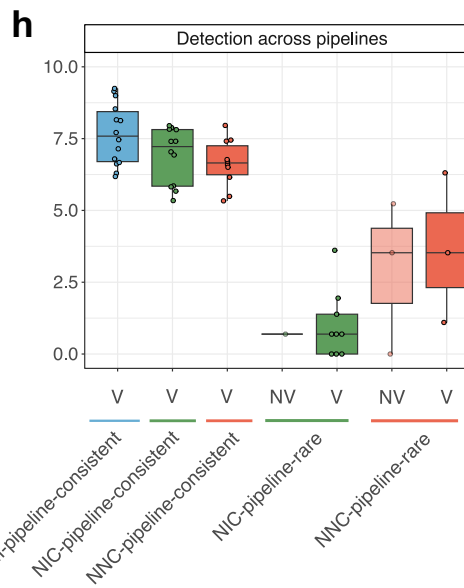


Figure 5. Experimental validation of known and novel isoforms. **a)** Schematic for the experimental validation pipeline. **b)** Example of a consistently detected NIC isoform (detected in over half of all LRGASP pipeline submissions) which was successfully validated by targeted PCR. The primer set amplifies a novel event of exon skipping (NIC). Only transcripts above ~5 CPM and part of the GENCODE Basic annotation are shown. **c)** Example of a successfully validated novel terminal exon, with ONT amplicon reads shown in the IGV track (PacBio produce similar results). **d)** Recovery rates for GENCODE annotated isoforms that are reference-matched (known), novel, and rejected. **e)** Recovery rates for consistently versus rarely detected isoforms, for known and novel isoforms. **f)** Recovery rates between isoforms that are more frequently identified in ONT versus PacBio pipelines. **g-i)** Relationship between estimated transcript abundances (calculated as the sum of reads across all WTC11 sequencing samples) and validation success for GENCODE (**g**), consistent versus rare (**h**), and platform-preferential (**i**) isoforms. **j)** Fraction of validated transcripts as a function of the number of WTC11 samples in which supportive reads were observed. **k)** Example of two *de novo* isoforms in Manatee validated through isoform-specific PCR amplification, blue corresponds to supported transcripts and red to unsupported transcripts. **l)** PCR validation results for manatee isoforms for seven target genes.