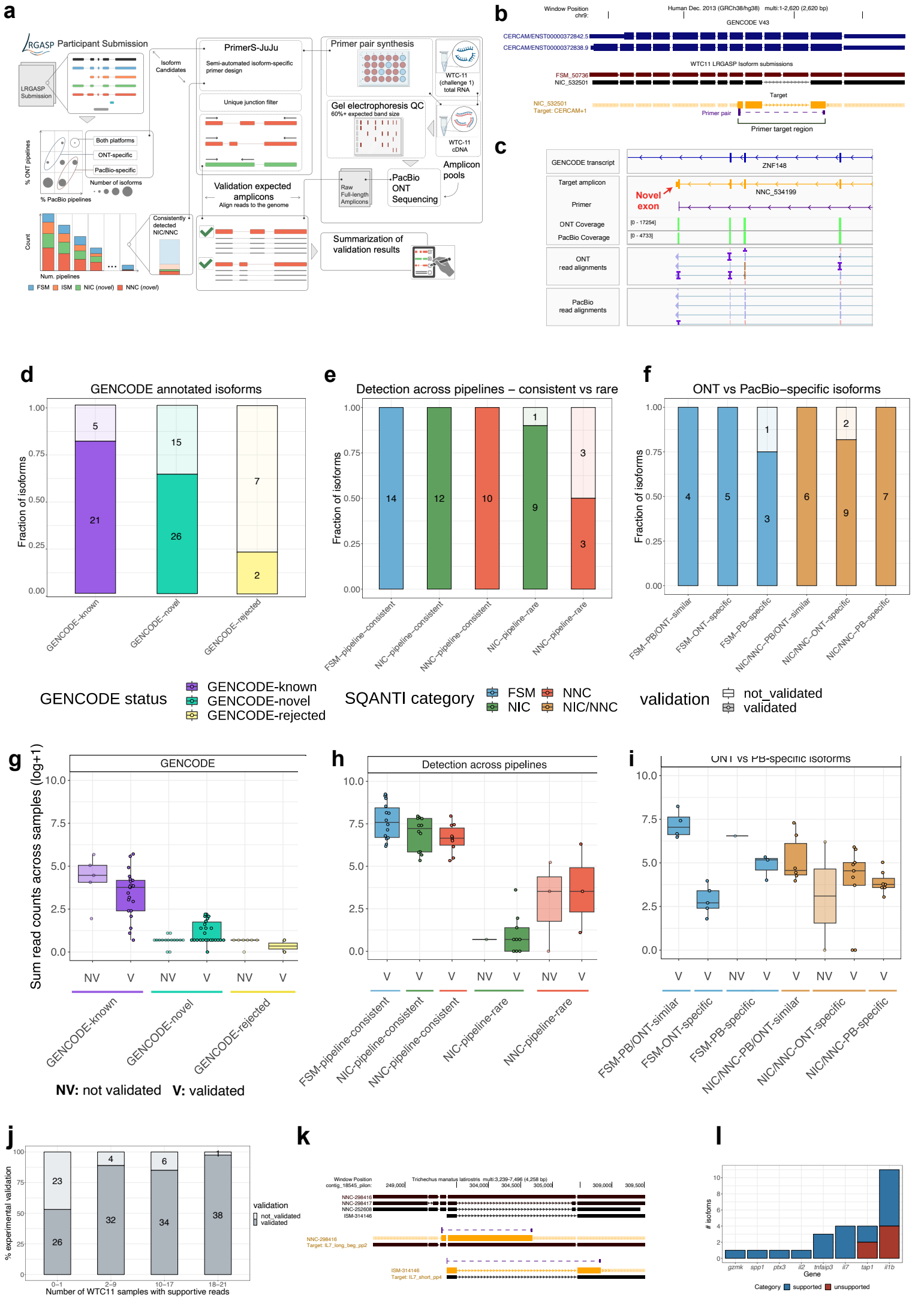
a
LRGASP Participant Submission
LRGASP Submission
Isoform Candidates
PrimerS-JuJu
Semi-automated isoform-specific primer design
Unique junction filter
Validation expected amplicons
Align reads to the genome
Primer pair synthesis
WTC-11 (challenge 1) total RNA
Gel electrophoresis QC
60%+ expected band size
WTC-11 cDNA
Amplicon pools
PacBio ONT Sequencing
Raw Full-length Amplicons
Summarization of validation results
Both platforms
ONT-specific
PacBio-specific
% ONT pipelines
% PacBio pipelines
Number of isoforms
Consistently detected NIC/NNC
Count
Num. pipelines
FSM ISM NIC (novel) NNC (novel)
b
Window Position chr9:
Human Dec. 2013 (GRCh38/hg38) multi:1-2,620 (2,620 bp)
GENCODE V43
CERCAM/ENST00000372842.5
CERCAM/ENST00000372838.9
WTC11 LRGASP Isoform submissions
FSM_50736
NIC_532501
Target
NIC_532501
Target: CERCAM+1
Primer pair
Primer target region
c
GENCODE transcript
ZNF148
Target amplicon
NNC_534199
Primer
Novel exon
ONT Coverage
[0 - 17254]
PacBio Coverage
[0 - 4733]
ONT read alignments
PacBio read alignments
d
GENCODE annotated isoforms
Fraction of isoforms
GENCODE-known
GENCODE-novel
GENCODE-rejected
e
Detection across pipelines – consistent vs rare
FSM–pipeline–consistent
NIC–pipeline–consistent
NNC–pipeline–consistent
NIC–pipeline–rare
NNC–pipeline–rare
f
ONT vs PacBio–specific isoforms
FSM–PB/ONT–similar
FSM–ONT–specific
FSM–PB–specific
NIC/NNC–PB/ONT–similar
NIC/NNC–ONT–specific
NIC/NNC–PB–specific
GENCODE status
GENCODE-known
GENCODE-novel
GENCODE-rejected
SQANTI category
FSM
NNC
NIC
NIC/NNC
validation
not_validated
validated
g
GENCODE
Sum read counts across samples (log+1)
NV: not validated V: validated
h
Detection across pipelines
i
ONT vs PB-specific isoforms
j
% experimental validation
Number of WTC11 samples with supportive reads
validation
not_validated
validated
k
Trichechus manatus latirostris multi:3,239-7,496 (4,258 bp)
contig_18545_pilon:
NNC-298416
NNC-298417
NNC-252606
ISM-314146
NNC-298416
Target: IL7_long_beg_pp2
ISM-314146
Target: IL7_short_pp4
l
isoforms
Gene
gzmk spp1 ptx3 il2 tnfaip3 il7 tap1 il1b
Category supported unsupported

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 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. Purple corresponds to the designed primers, orange to the possible amplification product associated to one isoform, and black to the predicted isoforms. **l)** PCR validation results for manatee isoforms for seven target genes. Blue corresponds to supported transcripts and red to unsupported transcripts.