# TUSCO: Transcriptome Universal Single-isoform COntrol

Tianyuan Liu[1], Alejandro Paniagua[1], Fabian Jetzinger[2], Luis Ferrández-Peral[1,3], Adam Frankish[4], Ana Conesa[1]

[1]Institute for Integrative Systems Biology (I2SysBio), Spanish National Research Council (CSIC), Paterna 46980, Spain
[2]BioBam Bioinformatics S.L., Valencia, Spain | [3]Biozentrum, University of Basel, Switzerland | [4]EMBL-EBI, UK

Genomics of Gene Expression Lab

## Introduction & Background

Long-read sequencing enables comprehensive transcriptome profiling but faces accuracy challenges.

### Current Limitations:

- **BUSCO:** Misinterprets alternative splicing as duplications
- **Spike-ins:** Oversimplify complexity, neglect RNA degradation
- **Simulations:** Cannot recapitulate biological complexity

### TUSCO Solution:

**Single-isoform endogenous genes** as **internal ground truth** without external controls.

## TUSCO Gene Selection & Framework

**Four-step pipeline:** (1) Cross-annotation agreement (GENCODE, RefSeq, MANE), (2) No alternative junctions (recount3/IntroVerse), (3) Universal expression (≥95% human/≥90% mouse tissues), (4) AlphaGenome filtering (high expression, low novel junctions)

**Result:** Human **n=46**, Mouse **n=32** universal | Tissue: 36–461 genes

**SQANTI3 Integration:** TP (correct), PTP (boundary shifts), FP (artifacts), FN (undetected)

**Metrics:** Sensitivity, Precision, PDR, FDR, Redundancy

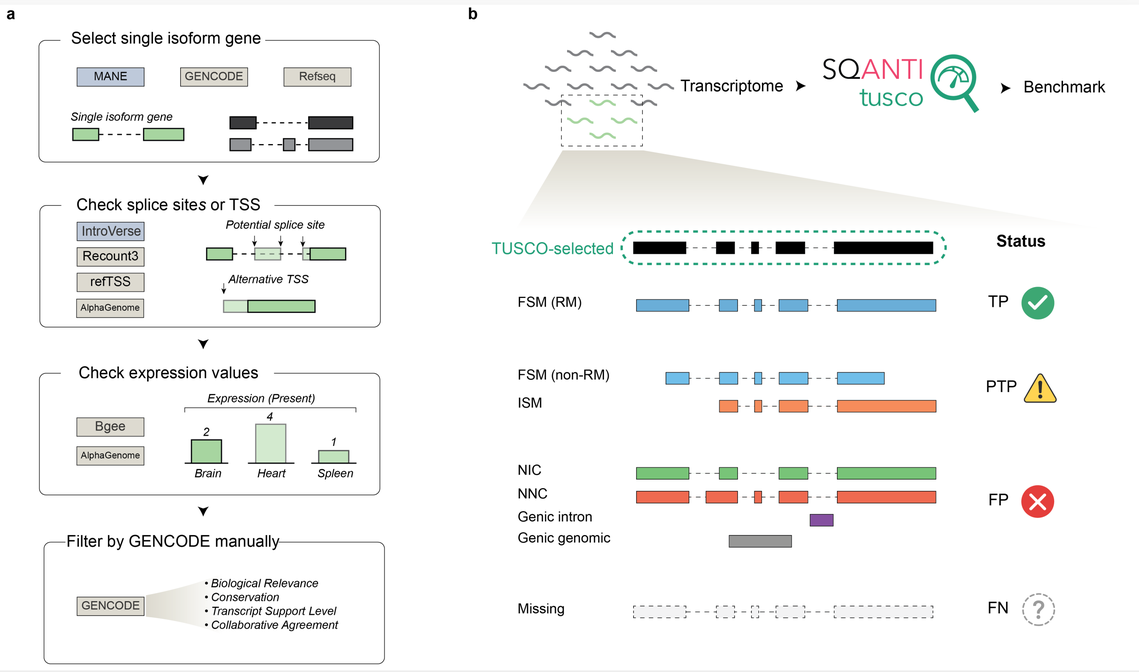


**Figure 1.** (a) Four-step pipeline. (b) SQANTI3 classification. (c) Comprehensive workflow.

## Results: TUSCO vs. SIRVs

High concordance (cosim 0.95–1.00), but TUSCO reveals **real biases:**

- **RIN correlation:** *R*=0.81 vs. Sequins *R*=0.075
- **Higher PTP/FN:** Sample quality issues
- **All genes detected** ≥1 dataset

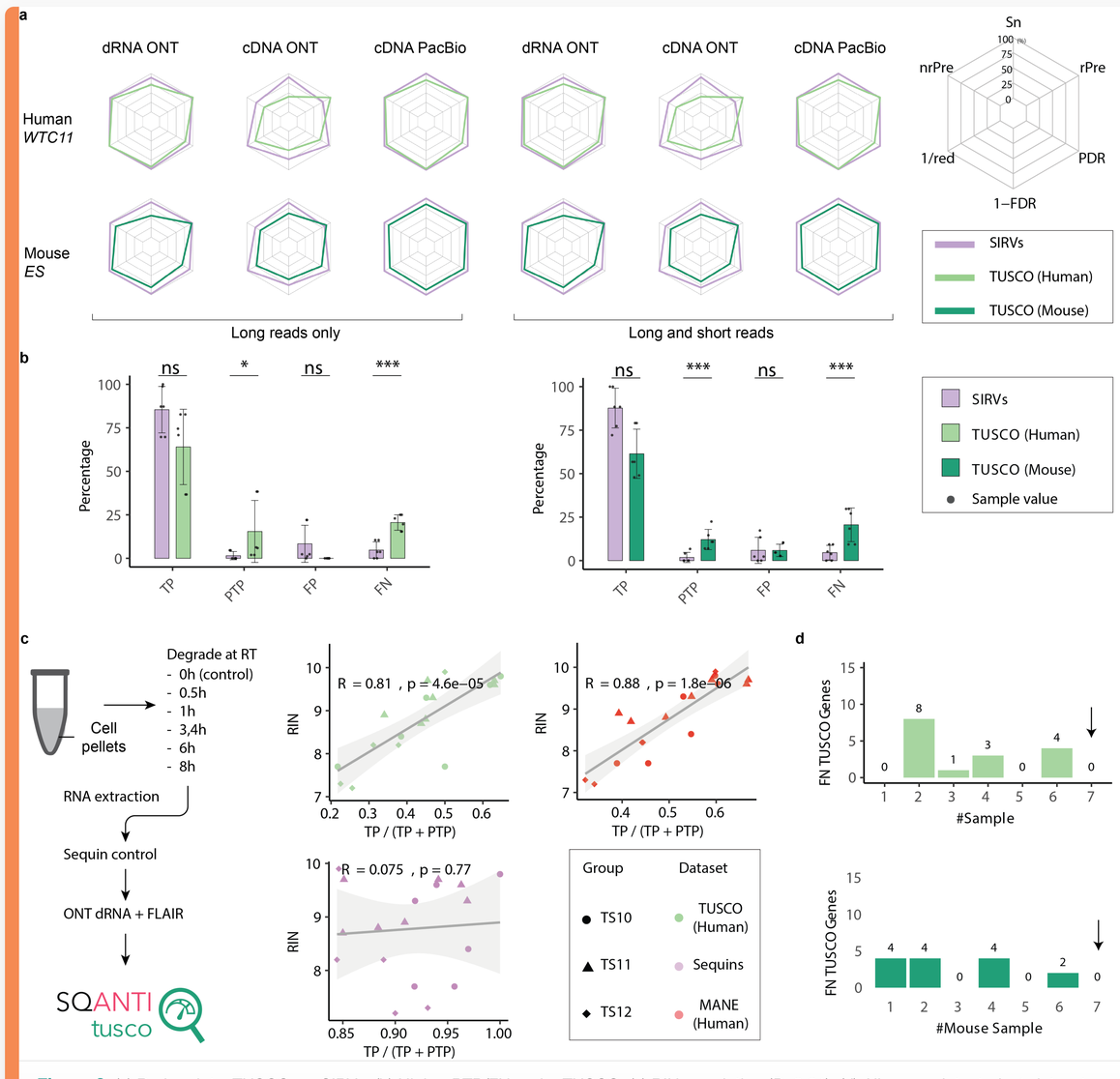


**Figure 2.** (a) Radar plots: TUSCO vs. SIRVs. (b) Higher PTP/FN under TUSCO. (c) RIN correlation (*R*=0.81). (d) All genes detected ≥1 dataset.

## Replication Optimization

- **1→2 reps:** FDR drops 13–20%
- **3 reps:** Optimal (plateau)
- **Universal vs. tissue:** High agreement (cosim >0.999)

**3 biological replicates optimal for most pipelines.**

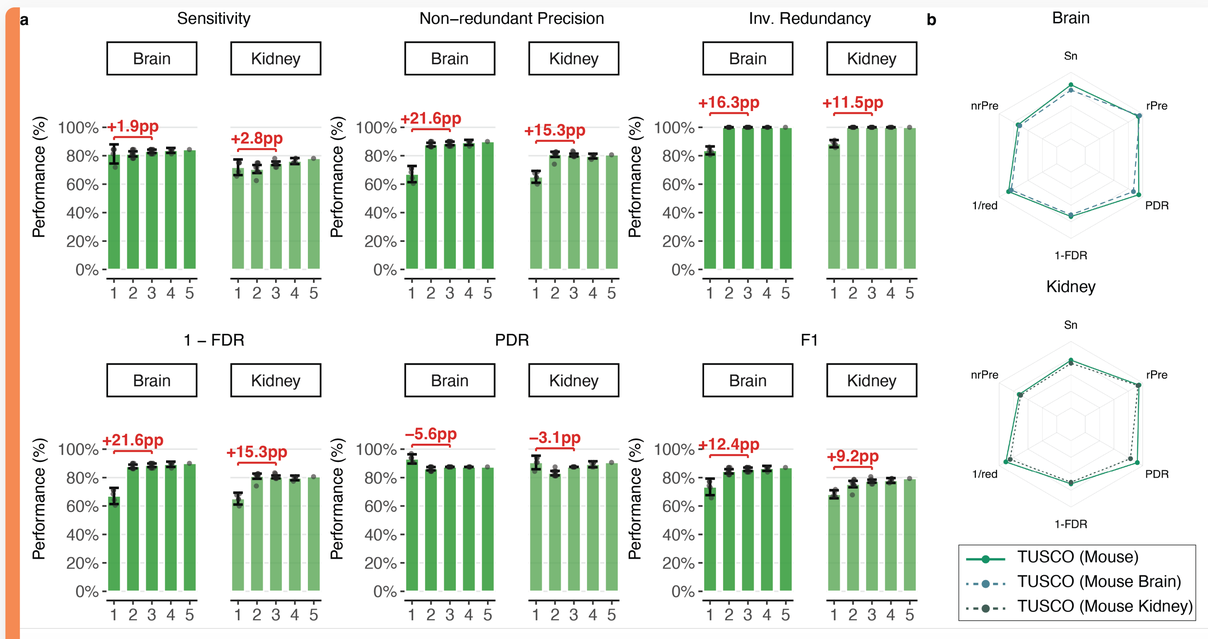


**Figure 3.** (a) Consensus evaluation. (b) FDR drops 1→2 reps, plateaus at 3. (c) Universal vs. tissue-specific: high agreement.

## TUSCO-novel: Novel Isoforms

Masks true transcripts, inserts artificial splice variants.

- **Bambu/StringTie2:** 55–81% performance drop (annotation-dependent)
- **FLAIR:** 39–44% drop (partial recovery)
- **Iso-Seq+ML:** 0 FP, minimal change (data-driven)

**Distinguishes annotation-dependent vs. *de novo* discovery tools.**

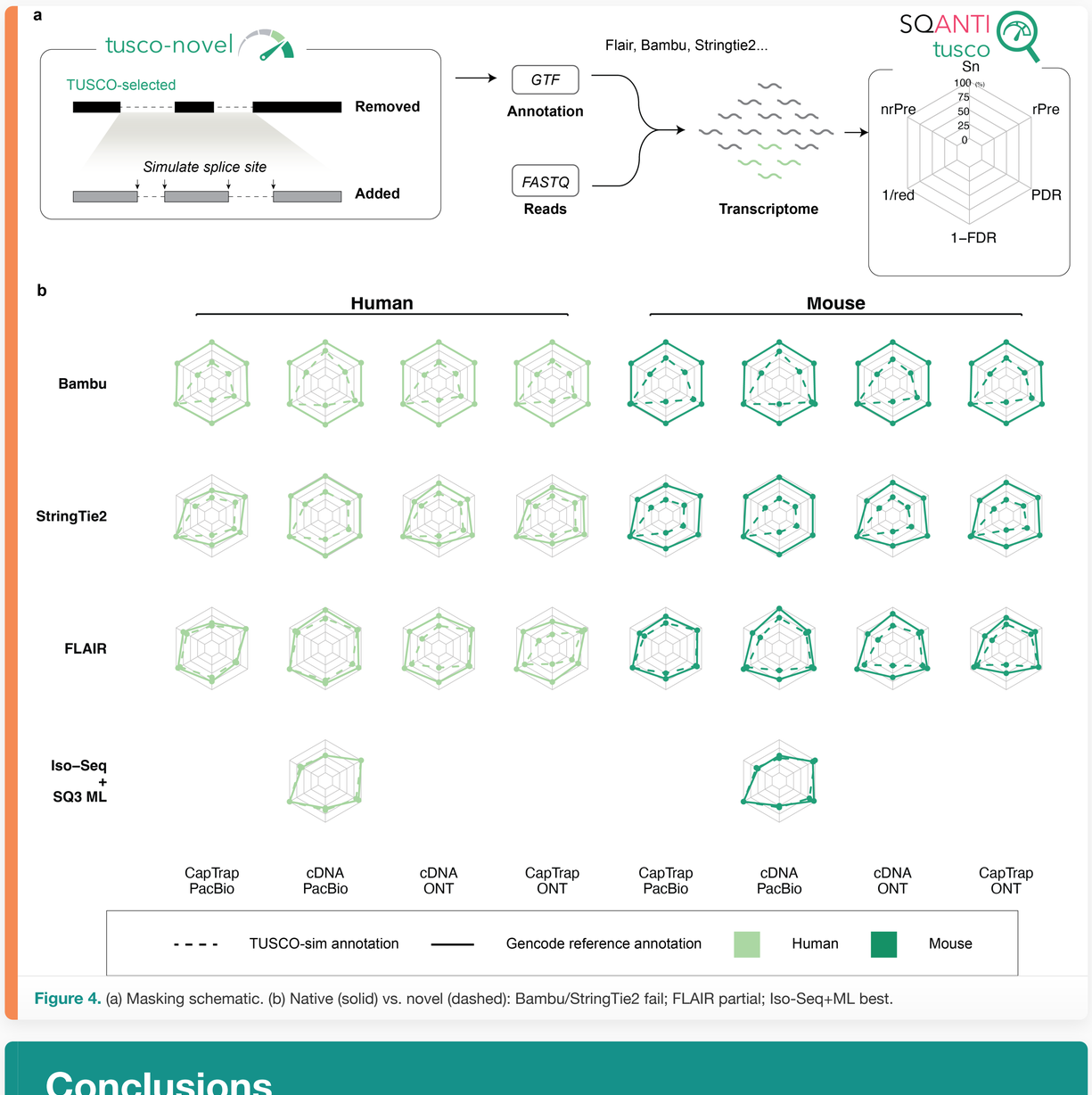


**Figure 4.** (a) Masking schematic. (b) Native (solid) vs. novel (dashed): Bambu/StringTie2 fail; FLAIR partial; Iso-Seq+ML best.

## Conclusions

- **Endogenous ground truth** (no external controls)
- **Realistic:** Captures degradation, bias, sample quality
- **Dual modes:** Known + novel isoform benchmarking
- **Practical:** Informs experimental design
- **Integrated:** Built into SQANTI3

**Availability:** github.com/ConesaLab/SQANTI3 | tusco.uv.es

**TUSCO bridges synthetic controls and biological complexity for realistic LRS benchmarking.**

SUPPORTED BY:

I2SysBio Institute for Integrative Systems Biology
CSIC Consejo Superior de Investigaciones Científicas
Funded by the European Union
Marie Curie Actions
LongTREC