Poster Number: 124
February 25, 2026

From Variability to Verifiability: Benchmarking Gene Expression in ATCC's Human and Mouse Cell Lines

ATCC®
Credible leads to Incredible®

Ajeet P Singh, PhD; Rula R. Khairi, MS; Amy L. Reese, MS; Jade L Kirkland, MS; Noah Wax, MS; Steve King, MS; James Duncan, BS; Robert Marlow, BS; Ana Fernandez, Jonathan L Jacobs PhD
ATCC, Manassas, VA 20110

Abstract

Reproducibility remains a significant challenge in biomedical research, especially when working with in vitro models such as immortalized cell lines. Variations in laboratory protocols, culture conditions, passage number, and experimental workflows can lead to substantial changes in gene expression, often resulting in inconsistent or contradictory findings. This variability undermines the reliability of data and impedes progress in both fundamental research and translational efforts.

To help address this issue, we established a comprehensive reference transcriptome for human and mouse cell lines originally sourced and maintained by ATCC. Using high-quality RNA sequencing and standardized bioinformatics workflows, we produced reproducible, well-annotated gene expression profiles across a wide array of cell lines representing diverse tissues, disease contexts, and research applications.

This dataset, available through QIAGEN OmicSoft's ATCC Cell Line Land and the ATCC Genome Portal, provides a robust benchmark for evaluating gene expressions commonly used in vitro models. It supports validation of experimental results, verification of cell line identity, detection of anomalies due to contamination or misidentification, and identification of cell line–specific biomarkers to inform model selection and experimental design.

By offering a consistent reference framework, the transcriptome enhances cross-study comparability, enables integration of datasets from multiple sources, and reinforces data quality. This resource promotes reproducibility and accelerates the translation of cell-based research into meaningful biomedical advances.

Our goal in delivering this reliable and accessible transcriptomic resource is to support the scientific community in moving from variability toward verifiability, advancing reproducible research, strengthening collaboration across sectors, and ultimately contributing to improved patient outcomes.

Why authenticated materials are essential

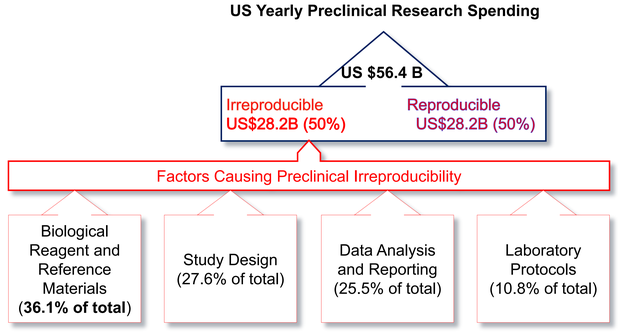


Figure 1: Financial impact and key drivers of irreproducibility. This figure summarizes the estimated annual cost of irreproducible preclinical research in the United States. Of the $56.4 billion invested each year, approximately 50% (~$28.2 billion) is attributed to studies that cannot be reliably reproduced. Major contributors to preclinical irreproducibility include issues related to biological reagents and reference materials (36.1%), study design (27.6%), data analysis and reporting (25.5%), and laboratory protocols (10.8%). These categories highlight critical areas where improved rigor, standardization, and quality control can substantially reduce scientific and economic losses. *Adapted from Freedman et al., PLOS Biology, 2015.*

- Contaminated cell lines create massive waste; for example, HEp-2 and Intestine 407 alone led to an estimated $990M loss across ~9,900 publications.
- The ICLAC registry lists 531 misidentified cell lines, resulting in billions in wasted research and follow-up studies.
- As research increasingly combines wet-lab experiments with AI and computational modeling, high-quality, authenticated datasets and standardized practices are critical.
- A major barrier to reproducibility is the lack of traceability linking genome assemblies to source materials, lab protocols, and bioinformatics workflows.
- These gaps slow scientific progress and weaken translational outcomes, making authentication essential.
- Reproducibility challenges grow as variables increase and machine learning-based analyses expand, emphasizing the need for verified datasets and reliable reference controls.

ATCC's standardized workflow for NGS

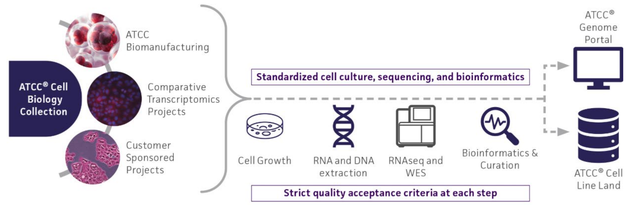
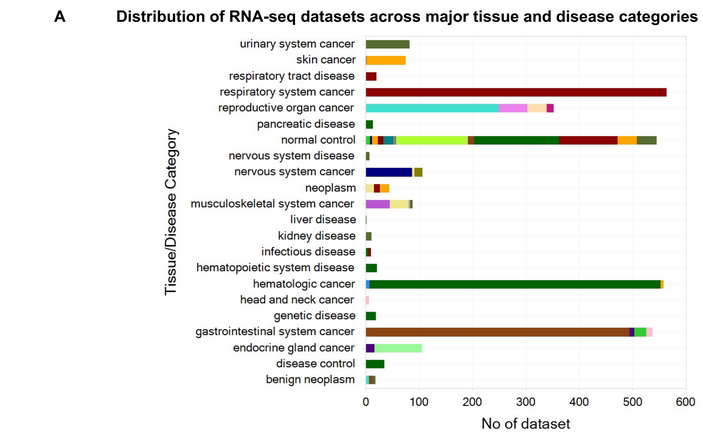


Figure 2: Schematic illustrating ATCC's standardized workflow processes from cell culture to RNA extraction, sequencing, and data quality control in accordance with ISO 9001 standards.

Whole-exome and RNA sequencing



C
VHL gene expression across kidney cancer subtypes and controls

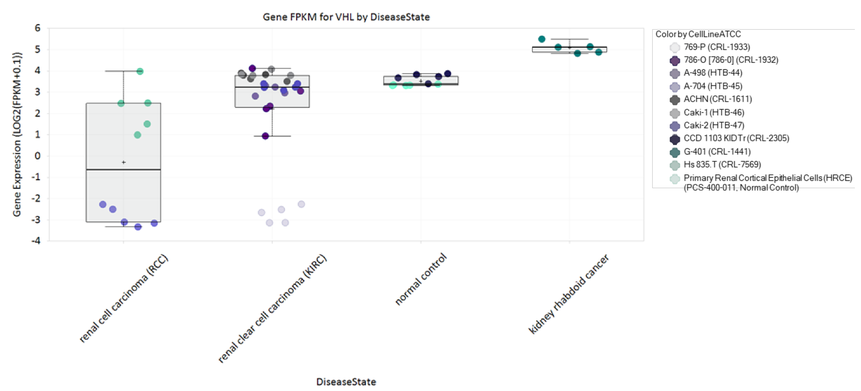
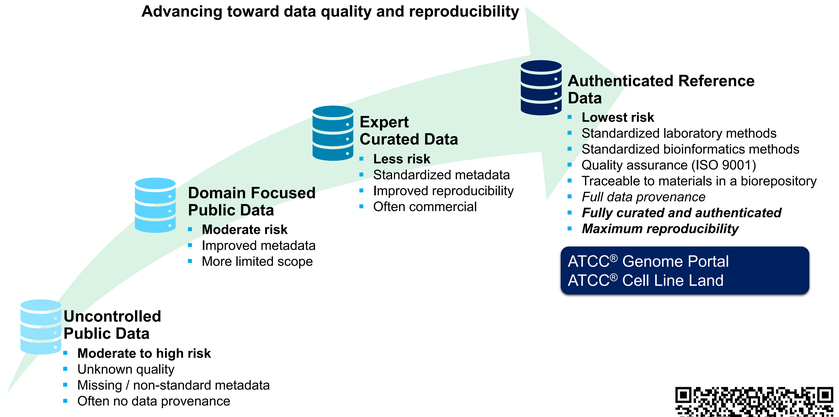


Figure 2: Landscape of cell line whole-transcriptome sequencing data derived representing various tissue/diseases. (A) Overview of whole-transcriptome RNA-seq datasets generated from cell lines across diverse tissue and disease categories. The Y-axis shows the tissue/disease classifications represented, and the X-axis shows the total number of datasets per category. Only cell lines with ≥3 RNA-seq datasets were included to ensure rigor and reproducibility. (B) Baseline expression of RAS family genes and CD274 across pancreatic cancer cell lines. The heatmap compares transcript levels across multiple models, supporting selection of appropriate in vitro systems for drug screening and target validation. The X-axis indicates the number of biological replicate RNA-seq datasets per cell line (≥3 per line). Color scale: purple = higher expression; dark blue = lower expression. (C) VHL tumor-suppressor gene expression across normal kidney tissue and kidney-derived cancer cell lines. Reduced VHL mRNA levels in clear cell renal cell carcinoma reflect frequent genomic inactivation. Expression values are log-transformed; each disease or cell-line category includes ≥3 datasets for robust comparison.

Conclusions



- Progression shows increasing data reliability from public datasets to authenticated references.
- Standardized methods and curated metadata improve consistency.
- Proven provenance enhances traceability and trust.
- ATCC authenticated datasets provide the most reproducible foundation for research.

ATCC 10801 University Boulevard, Manassas, Virginia 20110-2209 **Phone:** 800.638.6597 **Email:** sales@atcc.org **Web:** www.atcc.org

© 2026 American Type Culture Collection. ATCC product identifiers marked with the TM symbol are trademarks owned by the American Type Culture Collection. QIAGEN is a registered trademark of QIAGEN GmbH.