Establishing baseline transcriptome profiling of ATCC's human and mouse cell lines



Ajeet P. Singh, PhD; Rula Khairi, MS; Amy L. Reese, MS; Jade Kirkland, MS; Corina Tabron, BS; Noah Wax, MS; Steve King, MS; Ana Fernandes, BS; John Bagnoli, BS; Briana Benton, BS; Jonathan L. Jacobs PhD ATCC, Manassas, VA 20110

Key points

1. Importance of cell lines in biomedical research:

- Cell lines are crucial tools in biomedical research for various laboratory experiments.
- Contamination and inaccuracies in public datasets can negatively impact research.

2. ATCC® and QIAGEN® collaboration:

- Established the ATCC® Cell Line Land database.
- The database provides authenticated cell lines with comprehensive data provenance.
- Focus on endogenous baseline transcriptome data traceable to highly utilized human and mouse cell lines in ATCC®'s repository.

3. Shift in cell line selection:

- Refine cell line selection by moving from traditional methods to decisions informed by genotype-phenotype correlations.
- Provide data to foster reproducible research and development of impactful therapeutics.

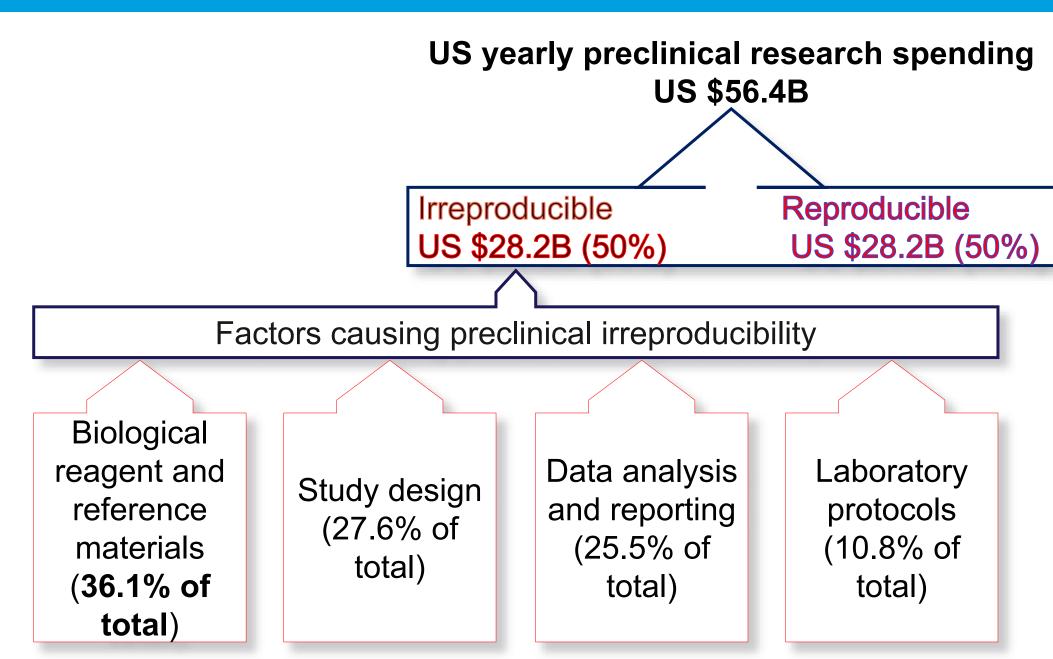
4. Leveraging RNAseq for unique insights:

- Over 300 cell lines from kidney, blood, and lung tissues sequenced.
- Encompasses primary, engineered, cancerous, and healthy cells.

5. Objective of establishing research standards:

- Offer traceable, standardized, and authenticated reference-quality transcriptomic data.
- Initiative to enhance the credibility and reproducibility of scientific endeavors.

Irreproducibility in research



Freedman LP, et al. PLoS Biol 13(6): e1002165, 2015. PubMed: 26057340.

Common challenges in R&D

- Finding the right cell line for the research is difficult
- Many cell types are not a good model for the disease of interest
- Pre-existing data are often not reproducible
- Search, analyze, and incorporate multi-omics data into cell line selection
- Cell lines associated OMICS data never produced/deposited from the repository

Irreproducibility solutions



Physical Repository

3,000+ authenticated

mammalian cell lines

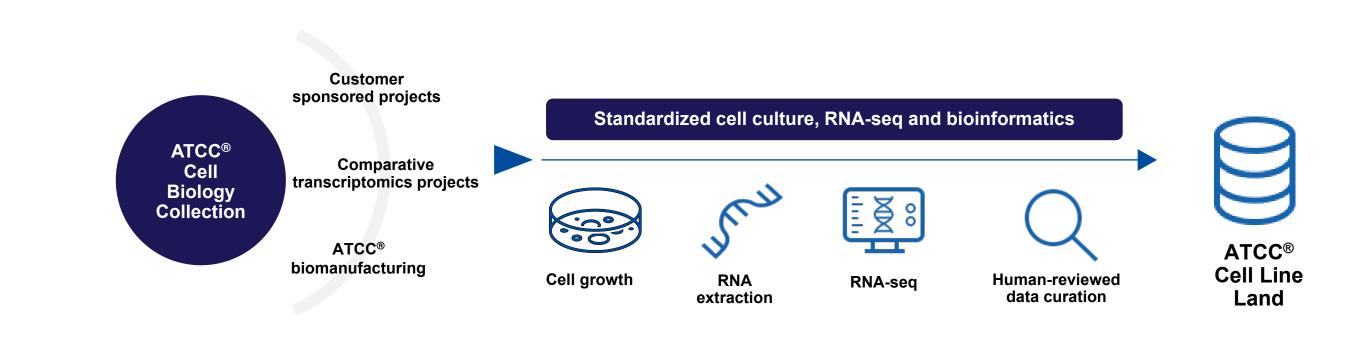
Cell culture

RNAseq data

ATCC® Cell Line land

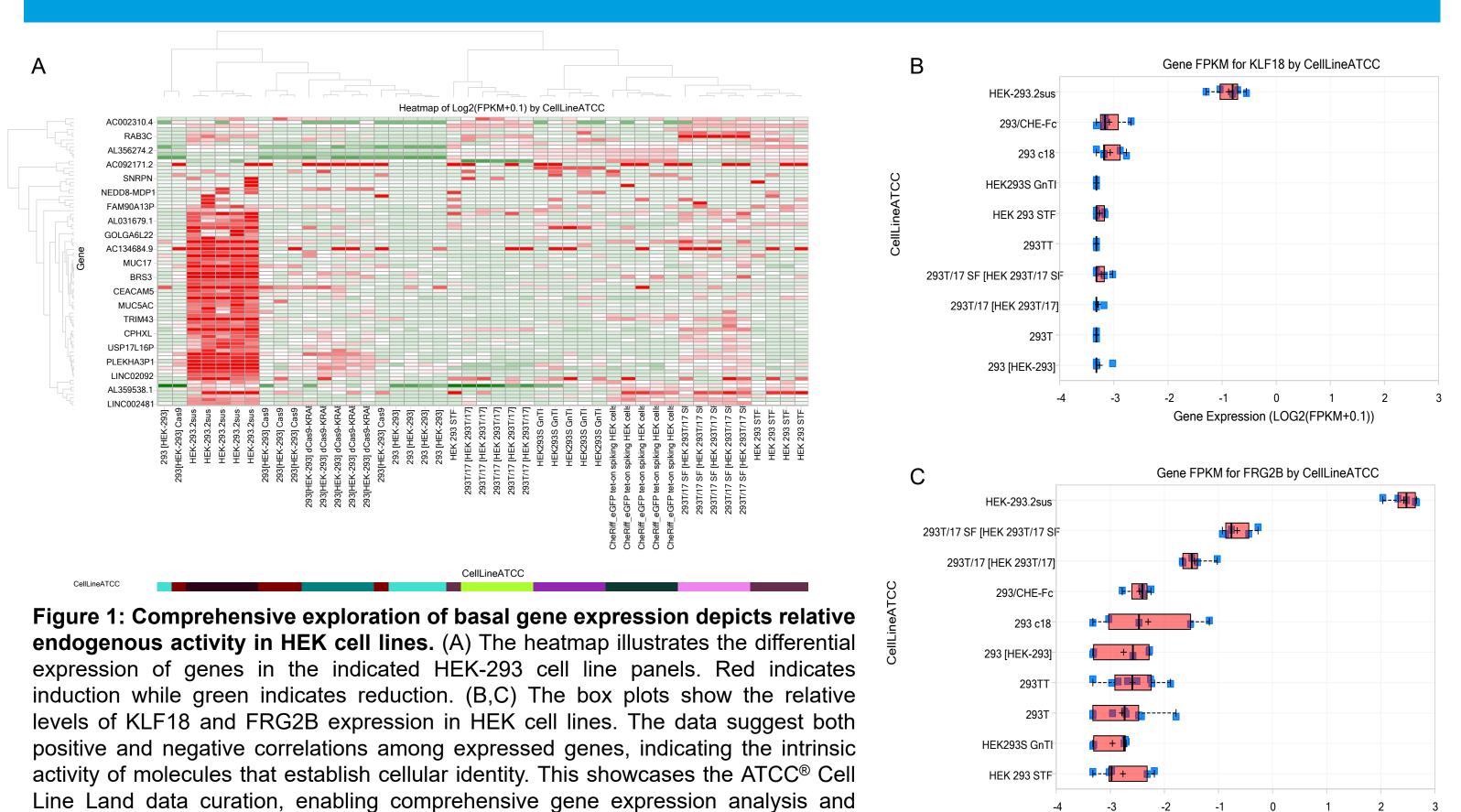
- High-quality reference database authenticated, traceable, reproducible, standardized
- Complete data provenance
- Bioinformatics solution
- Expanding content

Strict QA criteria of data curation



Maintains data reproducibility

Comparative gene signature in HEK cell lines



Cell line selection

Identify cell lines based on gene expression level

The need: Cell lines with high or low expression of the gene of interest for use as controls with primary cell lines

Steps:

- 1. ATCC® Cell Line Land
- 2. Search for KRAS
- 3. Highlight samples with high or low expression from the interactive plot
- 4. Export summary table with expression values and links to product pages

KRAS expression Gene FPKM for KRAS by Cellune High KRAS expressing cell COCO260M REH SNR622 CORFCEM PANCT PANCT

Figure 2: Basal levels of KRAS mRNA in ATCC® cell lines. The box plot illustrates the relative abundance and endogenous activity of the KRAS gene among ATCC® cell lines, representing various tissue, cell, and disease types.

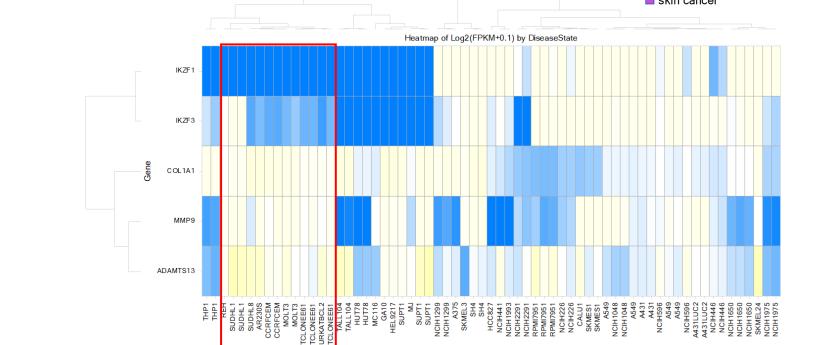


Figure 3: Basal expression depicts the endogenous activity of a gene set across various cell lines. The heatmap illustrates the expression correlation of multiple genes among cell lines representing various tissue and disease types. Dark colors indicate high expression while light colors indicate low expression.

Find new cell lines based on gene expression correlation

The need: Cell lines that have high expression of two genes but not three other genes

Steps:

- 1. ATCC® Cell Line Land
- 2. Type multiple genes and search
- 3. Highlight samples with desired expression profile
- 4. Export summary table with expression values and links to product pages

Identify cell lines with a specific coding mutation

The need: Cell lines that express a specific substitution gene mutation

Steps:

- 1. ATCC® Cell Line Land
- 2. Search for AREG
- 3. Switch view to RNA-Seq Mutation
- 4. Filter to desired mutation
- 5. Export summary table with expression values and links to product pages

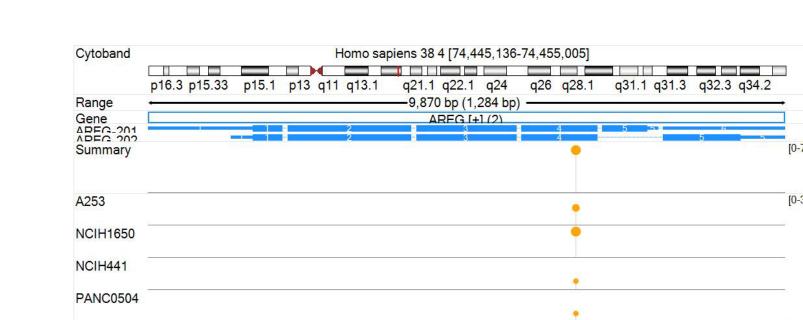


Figure 4: Genome browser view depicting specific coding mutations in selected genes across various cell lines. Alternative allele frequency in the general population of 0.00519

Summary

- There are numerous challenges that impact research quality, including inaccuracies in public datasets and sourcing relevant cell models.
- ATCC® Cell Line Land provides a resource of OMICS data that are traceable to ATCC® 's verified and authenticated cell lines.
- This database aims to enhance reproducibility and therapeutic development through a shift in cell line selection that is based on genotype-phenotype correlations.
- This study provides valuable data for optimizing cell line use, improving therapeutic outcomes and promoting credibility and reproducibility in scientific research.

ATCC 10801 University Boulevard, Manassas, Virginia 20110-2209

Cell line types

Gene expression

Transformed

Selection

line

Cell line repository

Phone: 800.638.6597

Email: sales@atcc.org

Web: www.atcc.org

facilitating the selection of the right cell line.