

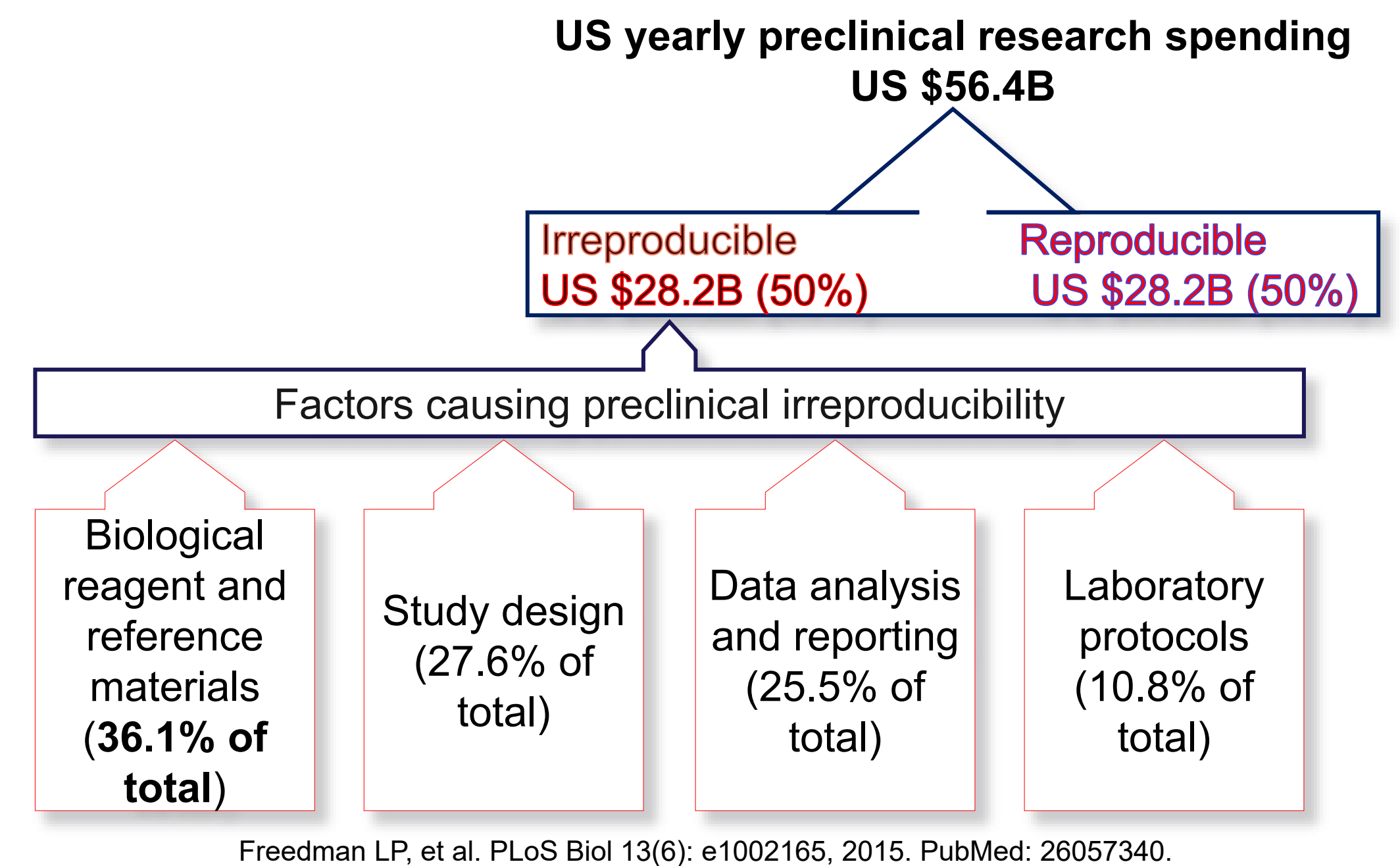
# Establishing baseline transcriptome profiling of ATCC's human and mouse cell lines

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

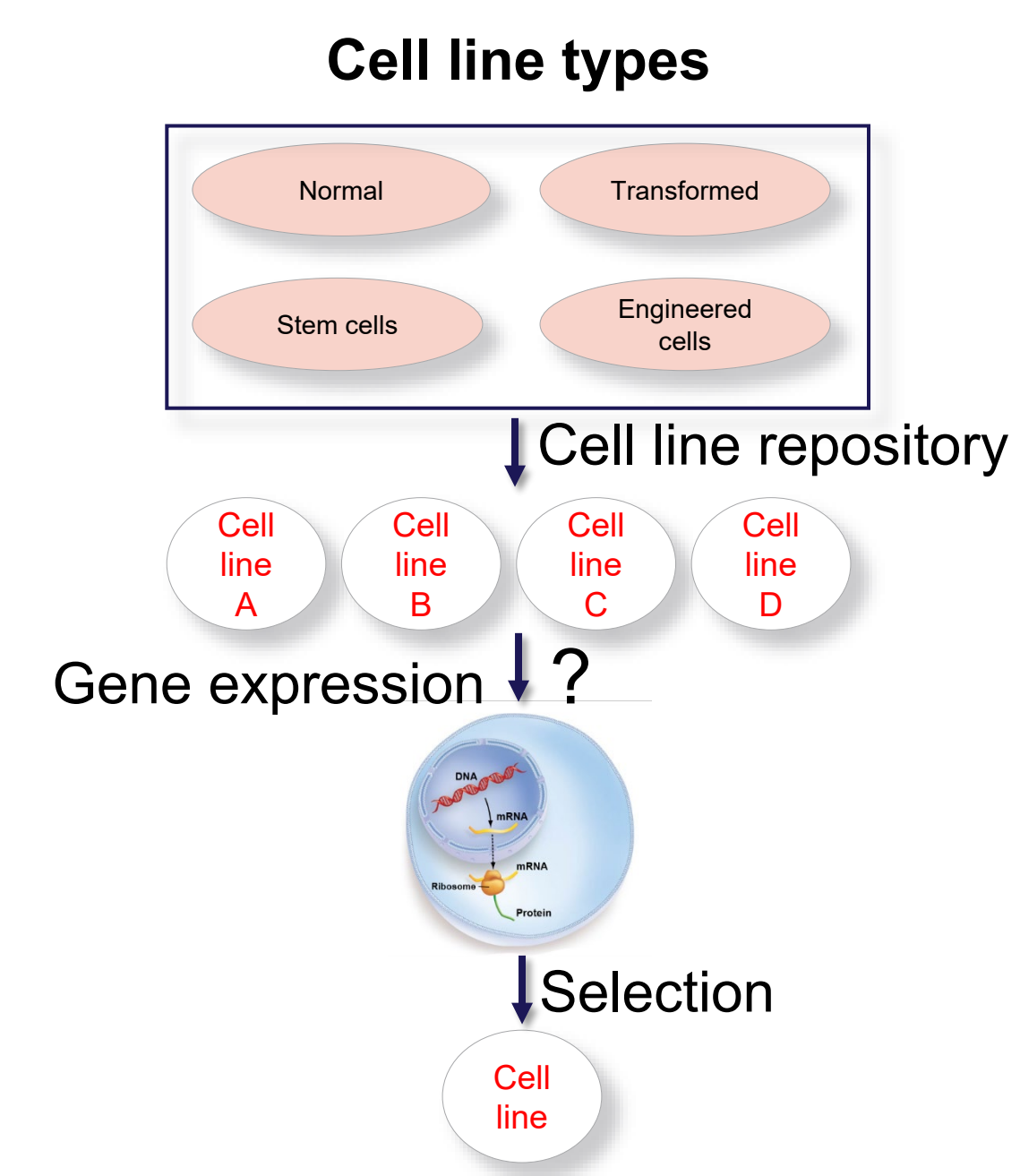
## Key points

- 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.
- ATCC<sup>®</sup> and QIAGEN<sup>®</sup> collaboration:**
  - Established the ATCC<sup>®</sup> 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.
- 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.
- Leveraging RNAseq for unique insights:**
  - Over 300 cell lines from kidney, blood, and lung tissues sequenced.
  - Encompasses primary, engineered, cancerous, and healthy cells.
- 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



- 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

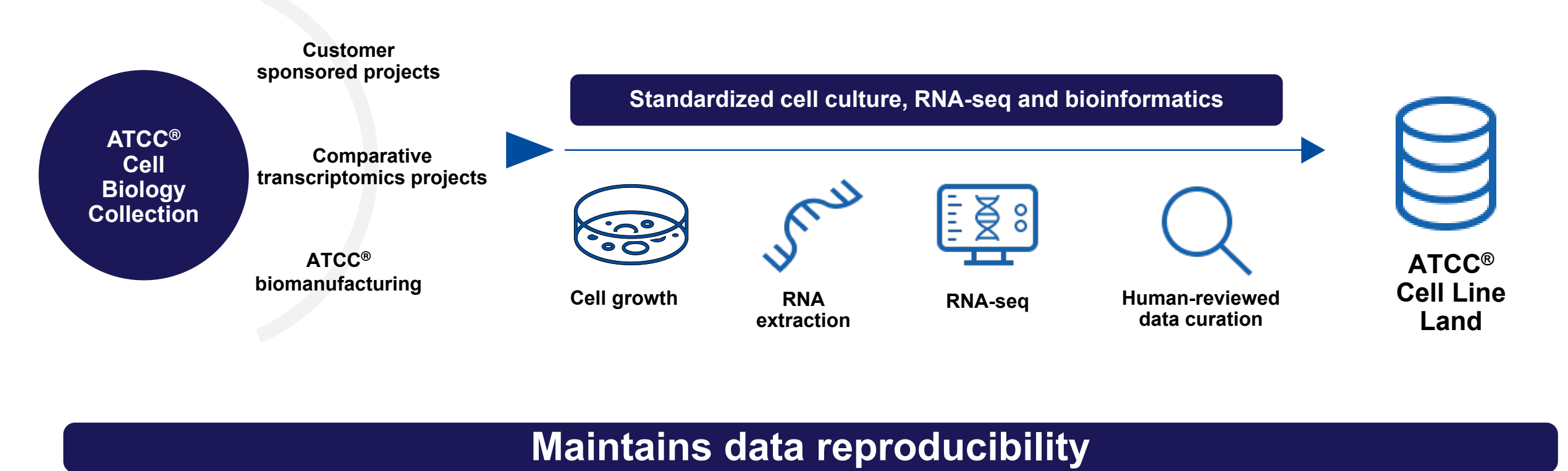


## Irreproducibility solutions

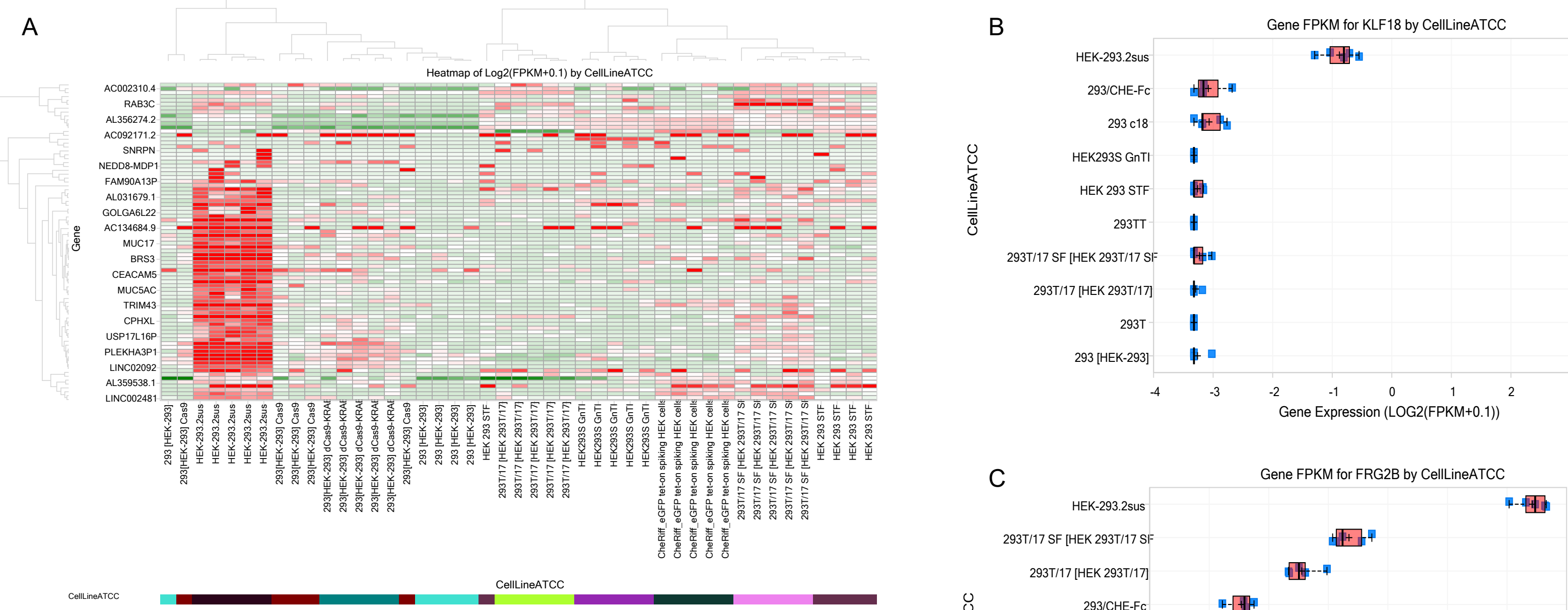


- Physical Repository**
  - 3,000+ authenticated mammalian cell lines
  - Cell culture
  - RNAseq data
- ATCC<sup>®</sup> Cell Line Land**
  - High-quality reference database - authenticated, traceable, reproducible, standardized
  - Complete data provenance
  - Bioinformatics solution
  - Expanding content

## Strict QA criteria of data curation



## Comparative gene signature in HEK cell lines

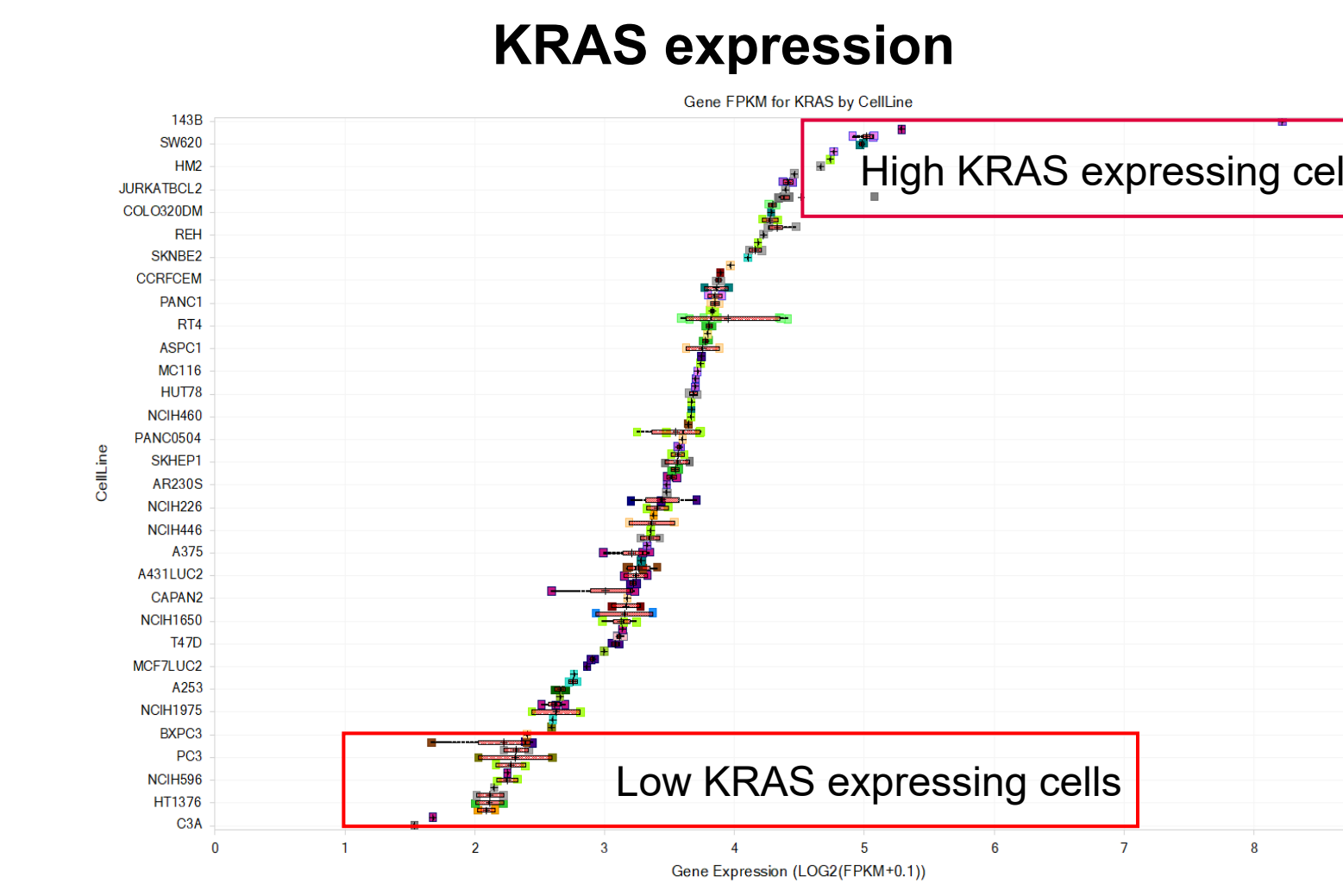


## Cell line selection

### 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

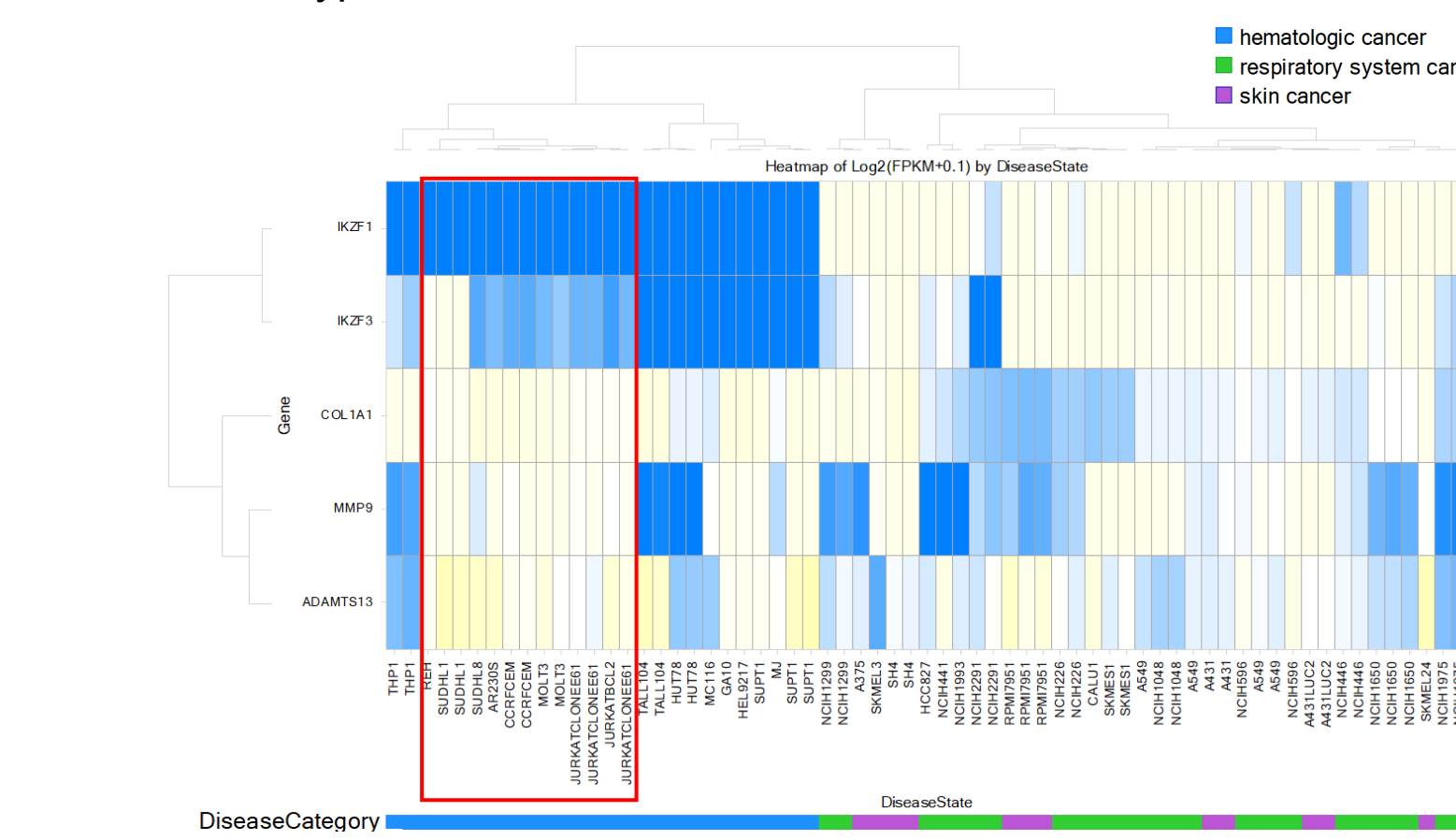
- Steps:**
- ATCC<sup>®</sup> Cell Line Land
  - Search for KRAS
  - Highlight samples with high or low expression from the interactive plot
  - Export summary table with expression values and links to product pages



### 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

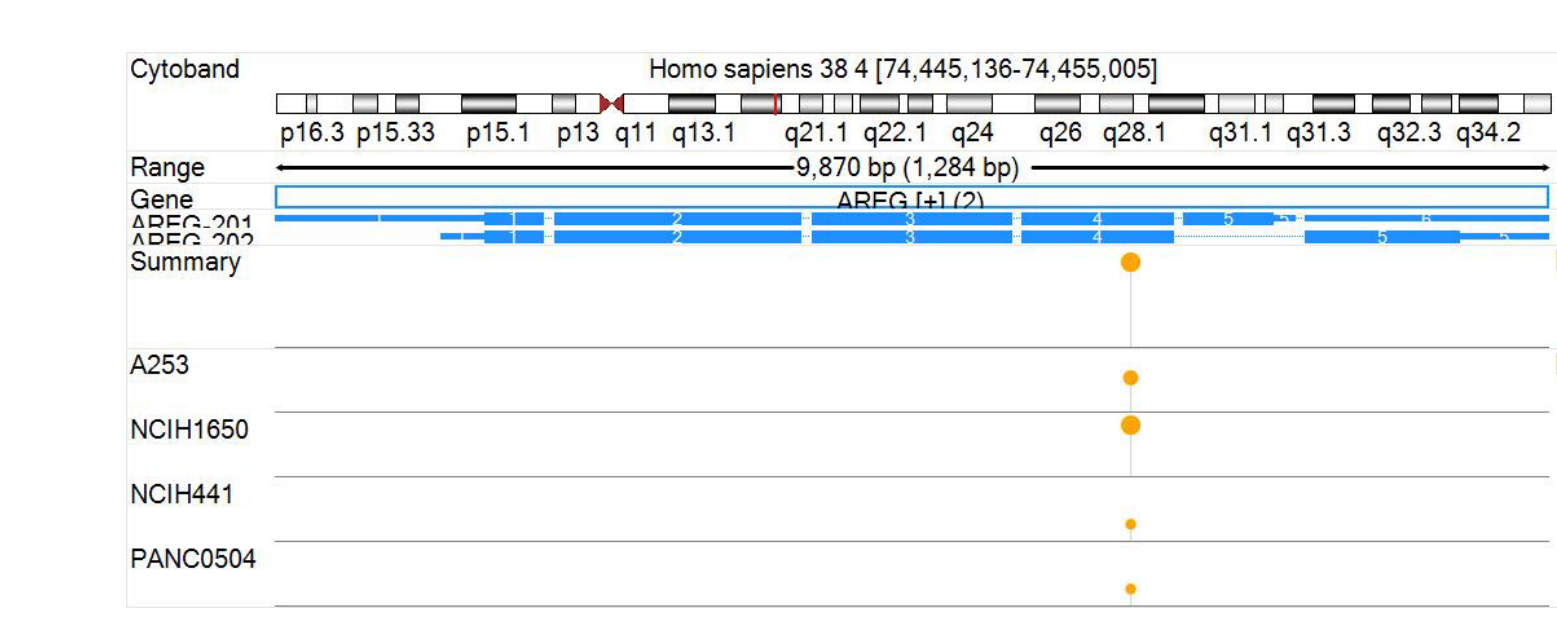
- Steps:**
- ATCC<sup>®</sup> Cell Line Land
  - Type multiple genes and search
  - Highlight samples with desired expression profile
  - Export summary table with expression values and links to product pages



### Identify cell lines with a specific coding mutation

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

- Steps:**
- ATCC<sup>®</sup> Cell Line Land
  - Search for AREG
  - Switch view to RNA-Seq Mutation
  - Filter to desired mutation
  - Export summary table with expression values and links to product pages



## Summary

- There are numerous challenges that impact research quality, including inaccuracies in public datasets and sourcing relevant cell models.
- ATCC<sup>®</sup> 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.