Unfolding the whole transcriptome of the ATCC Cell Line Land for its application as a molecular reference standard in the next-generation biomedicine research

Ajeet P. Singh, Amy L. Reese, Rula Khairi, Corina Tabron, James Duncan, Robert Marlow, Jade Kirkland, Steve King, Ana Fernandes, John Bagnoli, Briana Benton, Jonathan L. Jacobs

Sequencing & Bioinformatics Center, ATCC, Manassas, VA 20110

ATCC®

Credible leads to Incredible™

Abstract

In the cell biology research space, the use of unauthenticated biological reference materials, irregularities in crucial related metadata, and frequent gaps in data provenance often raise questions as to the credibility and reproducibility of experimental data. To address this problem, we have begun efforts to characterize the whole transcriptomes of all cell lines and primary cells held within ATCC's (American Type Culture Collection) cell biology collection to develop an authenticated application data resource to be used as a molecular reference standard. Here, we present our initial analysis of authenticated whole transcriptome (RNA-seq) datasets produced from 62 human kidney cell lines, including primary cells, found in ATCC's biorepository. These diverse datasets, represented by ATCC Cell Line Land, include the first comprehensive collection of reference transcriptomes for use by kidney biology researchers and are available through QIAGEN's OmicSoft platform. Multiple biological replicates for each cell line are included to help establish a baseline for a wide range of cell lines under typical cell culture conditions. ATCC Cell Line Land includes transcript and gene expression count data, RNA-seq variant calls, raw sequencing data, and detailed metadata on laboratory methods employed such as the total cell counts, RNA extraction methods, RIN scores, number of replicates, tissue/cell-type information, disease association, passage number, catalog number, lot number, growth media, culture conditions, and cryopreservation conditions. Furthermore, all biological replicate data included in this resource are produced using a common, standardized RNA extraction, library prep, sequencing, and bioinformatics workflow, thereby enabling comparative transcriptomics of these data at scale. ATCC Cell Line Land is a first-of-its-kind joint venture between ATCC and QIAGEN Digital Insights aimed at providing reference-grade whole transcriptome data that is authenticated, standardized, and traceable to physical source materials available in ATCC's biorepository.

Introduction

- An estimated 15-20% of all experiments found in the literature are using misidentified cell lines.[1]
- Short tandem repeat (STR) profiling represents the gold standard for cell line authentication; however, cell lines (eg, HEK293) with acquired mutations in mismatch repair genes can alter their characteristic repeats upon prolonged culture, leading to negative authentication results via STR.
- In order to combat negative authentication results, ATCC is completing RNA-seq analysis on their broad selection of human kidney cells using well-established protocols and stringent quality metrics (described below).
- All fully authenticated human kidney data will be available through ATCC Cell Line Land. For more information: https://digitalinsights.qiagen.com/atcc-cell-line-land.

Materials & Methods

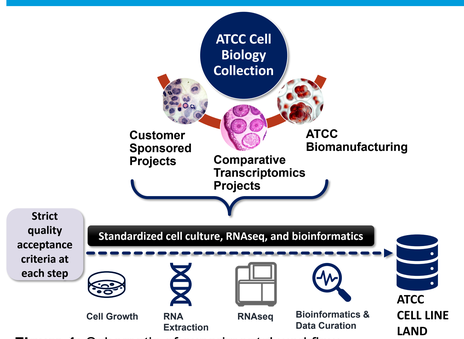

Figure 1: Schematic of experimental workflow.

Table 1: Sequencing and Bioinformatics QC Metrics

QC Metrics	Average Values	Metadata Available
RNA Integrity Number (RIN)	>6.5	• Passage # • Sex • Race • Age • Disease • Tissue/cell type • Growth media • Culture condition • Cryopreservation
260/280 Value	>1.8	
Input Read Number	$18x10^6$	
% Uniquely Mapped Reads	>80%	

Table 2: Commonly used HEK293 derivatives

HEK Cell Lines	Applications
HEK-293 (ATCC® CRL-1573™)	Efficacy testing, transfection host, virucide testing
Phoenix-AMPHO (ATCC® CRL-3213™)	Second-generation retrovirus producer cell line
Phoenix-GP (ATCC® CRL-3215™)	Second-generation retrovirus producer cell line
293[HEK-293] Cas9 (ATCC® CRL-1573Cas9™)	Carry out CRISPR genome editing applications with high efficiency
293T (ATCC® CRL-3216™)	Highly transfectable derivative of HEK293, used to produce retroviruses
293T/17 (ATCC® CRL-11268™)	293T/17; Embryonic Kidney; Human (*Homo sapiens*)
Phoenix-ECO (ATCC® CRL-3214™)	Second-generation retrovirus producer cell line
Tau RD P301S FRET Biosensor (ATCC® CRL-3275™)	Reporter-labeled cell
293.STAT1 BAX KO (ATCC® CRL-1573-VHG™)	Adeno-associated virus (AAV) production for gene therapy applications, viral vaccine production, virus propagation, production of high-titer viral stocks, viral vector transfection, and viral particle production.
GFPu-1 (ATCC® CRL-2794™)	Used as a reporter of proteasome activity
HEK-293.2sus (ATCC® CRL-1573.3™)	Grow in suspension; grow in serum-free media
HEK 293 STF (ATCC® CRL-3249™)	Luciferase reporter cell line, Assaying canonical Wnt signaling pathway
293 c18 (ATCC® CRL-10852™)	Very high frequencies of transformation are obtained with vectors containing the Epstein-Barr virus (EBV) oriP
293TT (ATCC® CRL-3467™)	Production of papillomavirus and polyomavirus-based reporter vectors.

Results

QC Analysis of RNA Sequencing Experiments

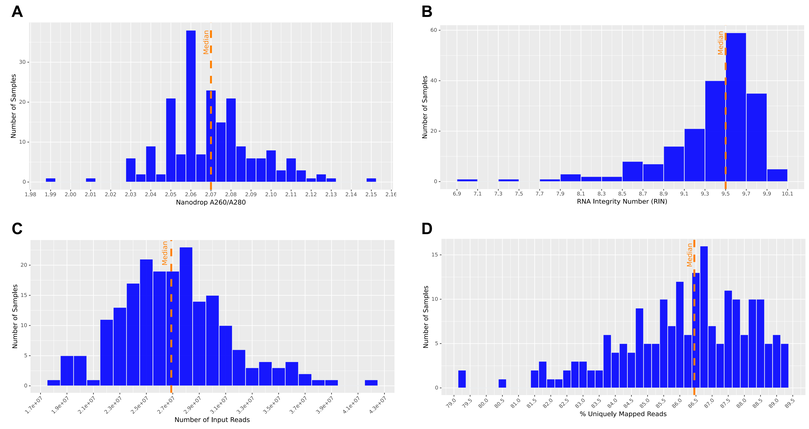

Figure 2: RNA-seq quality control metrics (199 libraries). (A) Range of RNA quality values as assessed by NanoDrop. (B) Range of RNA integrity values as assessed by TapeStation. (C) Distribution of the total number of input reads determined by STAR. (D) Percentage of uniquely mapped reads to human genome (GCh38). Median values are indicated in orange.

Transcriptome Profiling of 40 Human Kidney Cell Lines

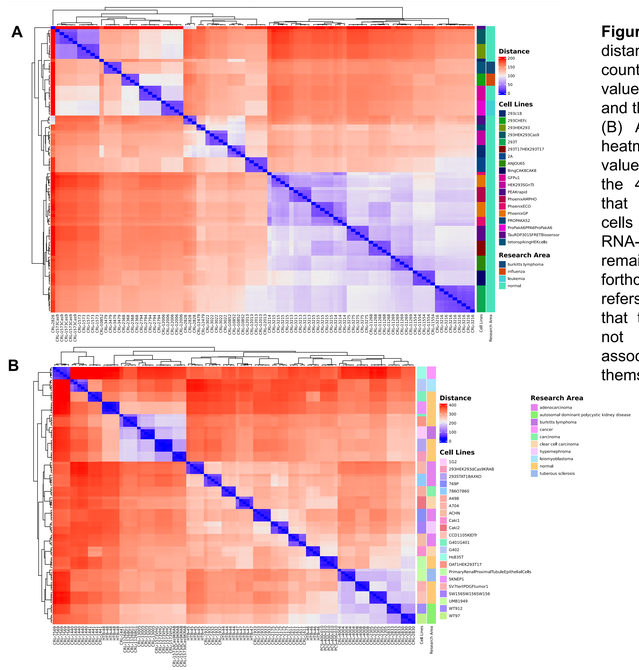

Figure 3: (A) A correlation distance heatmap of log2 counts per million (CPM) values for HEK-293 cells and their derivatives.
(B) A correlation distance heatmap of the log2 CPM values for cell lines, out of the 40 sequenced so far, that are neither HEK-293 cells nor their derivatives. RNA-seq data for the remaining 22 cell lines are forthcoming. Research area refers to the research areas that the cells are used in, not necessarily diseases associated with the cell lines themselves.

Increased Viral Production in 293.STAT1 BAX KO Cells

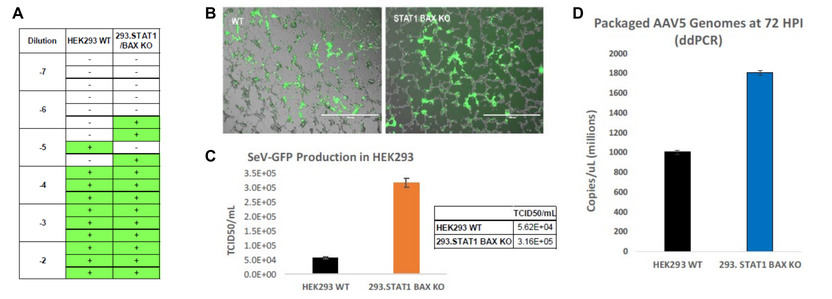

Figure 4: Increased viral production in 293.STAT1 BAX KO cells. (A) $TCID_{50}$ of GFP Sendai viral supernatants produced by WT parental and 293.STAT1 BAX KO cells. Cells infected with GFP Sendai virus at an MOI of 0.01. Supernatants collected 48 hours after infection and used to reinfect WT HEK293 cells at the indicated dilution. (B) Cells were imaged 48 hours post GFP Sendai viral infection. (C) $TCID_{50}$ of GFP Sendai viral supernatants produced at 48 hours post infection were calculated. (D) Cells were transfected with AAV5 viral vector. Supernatants were collected 48 hours after transfection and used to re-infect WT HEK293 cells. Droplet digital PCR quantification of AAV5 viral genomes produced in WT parental and 293.STAT1 BAX KO cells at 72 h.p.i.

Differential Gene Expression

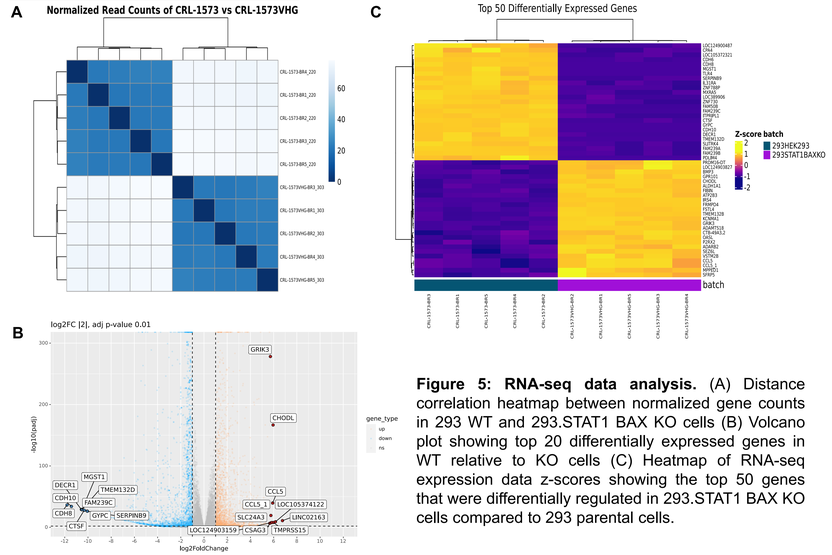

Figure 5: RNA-seq data analysis. (A) Distance correlation heatmap between normalized gene counts in 293 WT and 293.STAT1 BAX KO cells (B) Volcano plot showing top 20 differentially expressed genes in WT relative to KO cells (C) Heatmap of RNA-seq expression data z-scores showing the top 50 genes that were differentially regulated in 293.STAT1 BAX KO cells compared to 293 parental cells.

Summary

In this study, ATCC characterized the whole transcriptome of 62 authenticated human kidney cell lines from the ATCC biorepository. The data produced in this study are intended to be used as molecular reference standards and will be available to the scientific community within the ATCC Cell Line Land.

LEARN MORE

References

1. Dunham JH, Guthmiller P. Doing good science: Authenticating cell line identity. Promega Notes 101: 15-18, 2012.

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