

# Creating High-quality RNA-sequencing Reference Data from ATCC's Verified Cell Lines

Ajeet P. Singh, PhD; Amy L. Reese, MS; Rula Khairi, MS; Jade Kirkland, MS; Corina Tabron, MS; 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

## Introduction

Access to reliable RNA-seq reference data from verified ATCC<sup>®</sup> cell lines is pivotal for progressing biomedical research. Nevertheless, there is a notable absence of datasets in the public domain that precisely mirrors the RNA characteristics linked with these cell lines. This kind of resource is fundamental for establishing a baseline control when utilizing these cell lines across diverse studies. Employing meticulous RNA sequencing techniques and rigorous quality checks, ATCC<sup>®</sup> has partnered with QIAGEN<sup>®</sup> Digital Insights and created the ATCC<sup>®</sup> Cell Line Land tool—an online platform featuring extensive transcriptomic and genomic data from the source material, serving as the benchmark reference standard.

Our ongoing sequencing endeavors to provide a thorough inventory of RNA transcripts shed light on the fundamental gene expression profiles within ATCC<sup>®</sup> cell lines. Through comparative analyses, distinct RNA expression patterns have emerged, delineating molecular discrepancies among various cell lines. Notably, our examination has delineated relative basal gene expression levels between genetically modified cells versus native ones and between healthy primary and cancerous cells. This information holds significance in the selection of cell lines for study and subsequent analysis. Integrative analyses of ATCC<sup>®</sup> Cell Line Land RNA-seq data with existing genomic and proteomic datasets in the public domain facilitates the validation of results across different data types, enabling comprehensive multi-omics analyses and mitigating ambiguity while enhancing clarity in data interpretation. Furthermore, comparative analysis of ATCC derived cell line RNA-seq data could aid in predicting cellular responses to treatment, facilitating drug screening, and anticipating potential outcomes. We advocate for collaborative endeavors to continually update and refine these high-quality reference databases, ensuring their pertinence and utility in the dynamic landscape of research.

## Materials & Methods

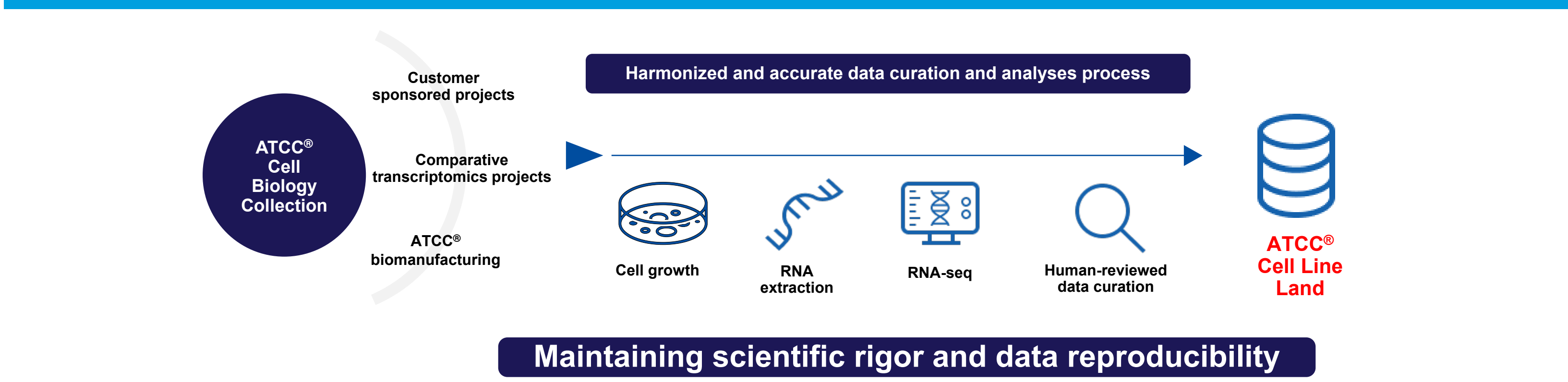


Figure 1: Schematic illustrating ATCC's workflow processes from cell culture to RNA extraction, sequencing, and data quality control, in accordance with ISO 9001 standards.

## Molecular characterization of ATCC's cell lines

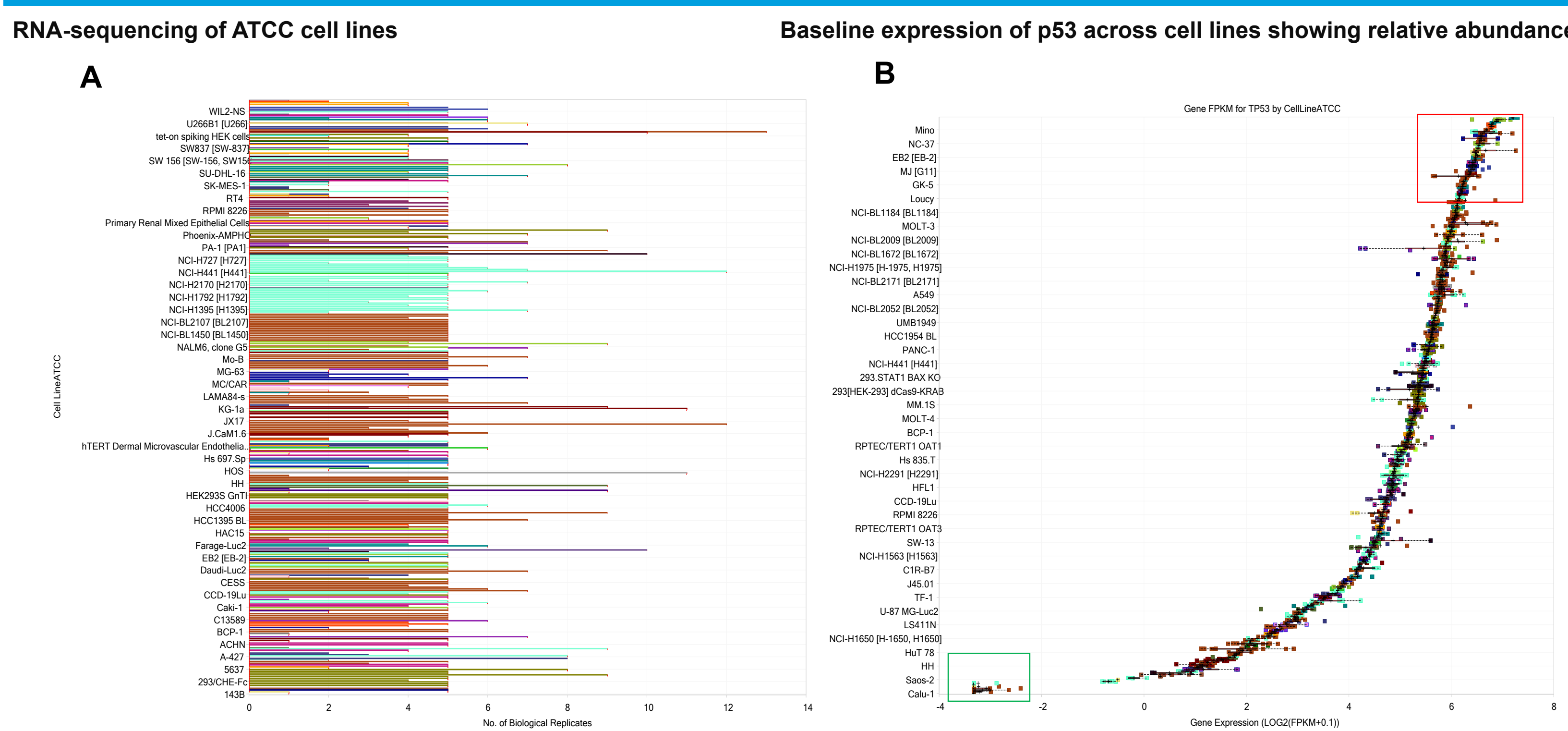


Figure 2: The ATCC<sup>®</sup> Cell Line Land (ACLL) project intends to map all human and mouse cell lines to promote reproducibility and scientific rigor in biomedical research. This perspective highlights the collaboration between QIAGEN<sup>®</sup> DIGITAL INSIGHTS and ATCC<sup>®</sup>. (A) ATCC cell lines characterized for molecular qualities derived from various tissue types and disease states. (B) Relative basal abundance of the tumor suppressor gene p53 in several cell lines. Green boxes indicate cell lines with negative expression, whereas red boxes indicate cell lines with high p53 mRNA levels. The X-axis in panel A indicates the number of RNA sequencing data sets produced from biological replicates of the respective cell lines.

## Baseline expression in models of kidney cancer

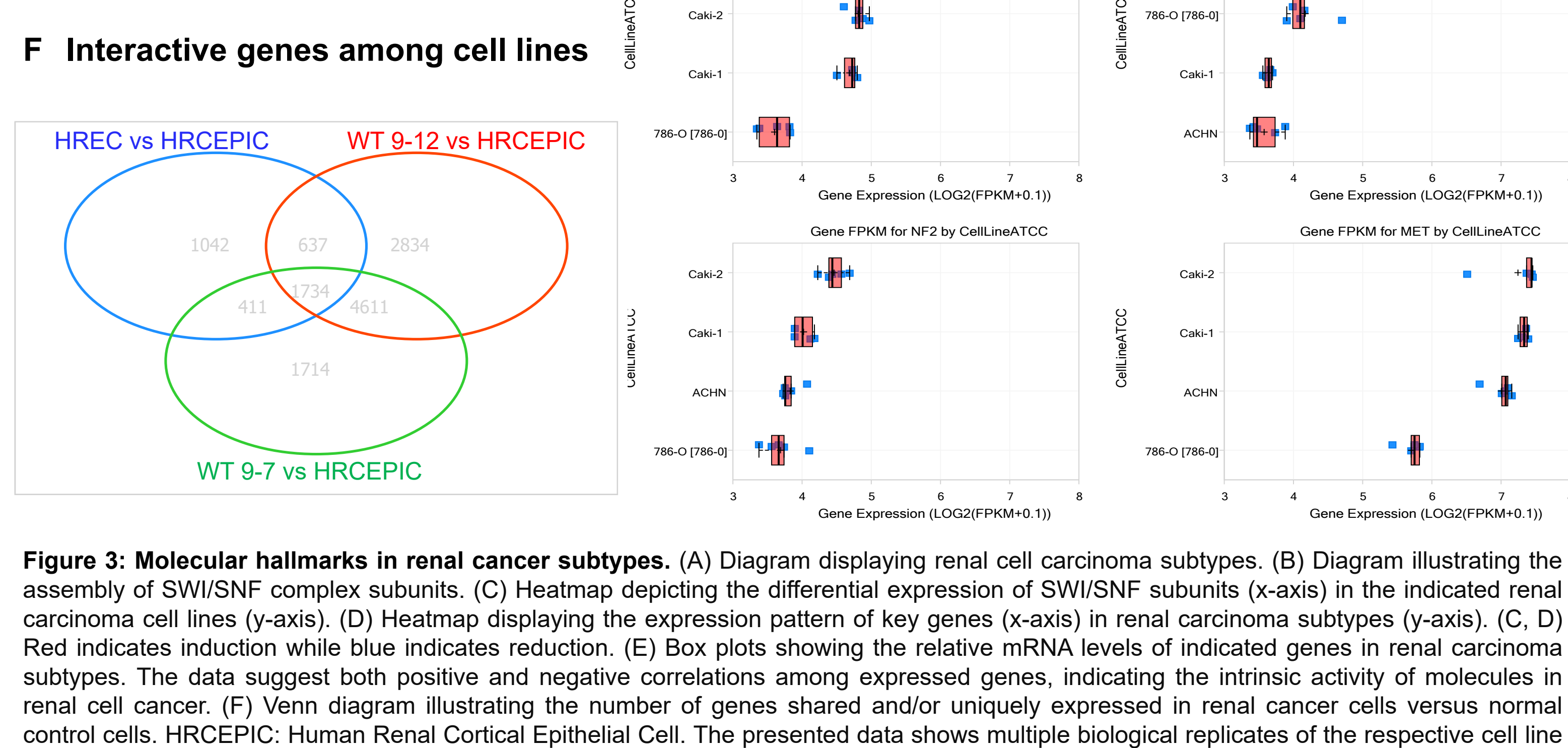
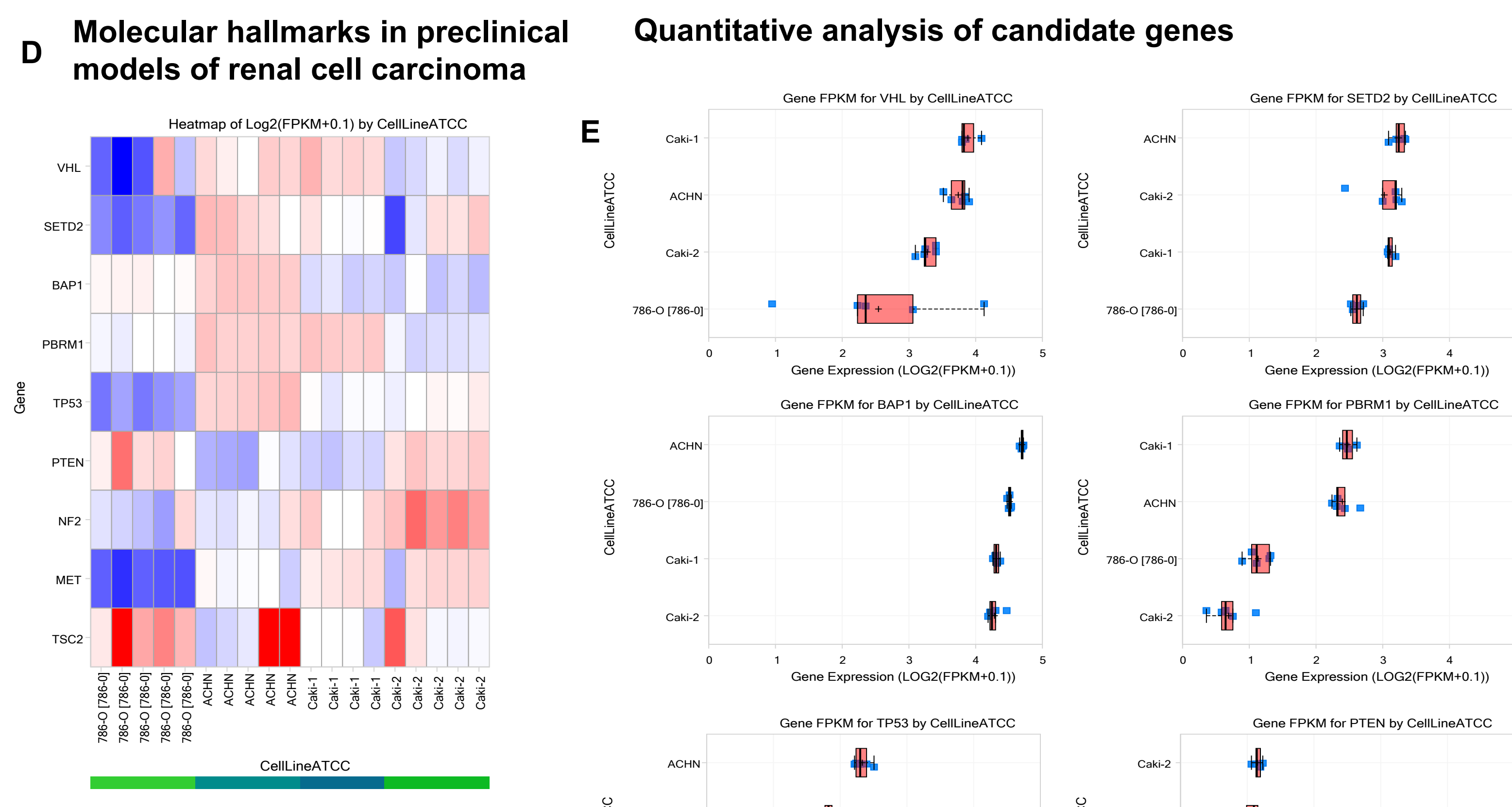
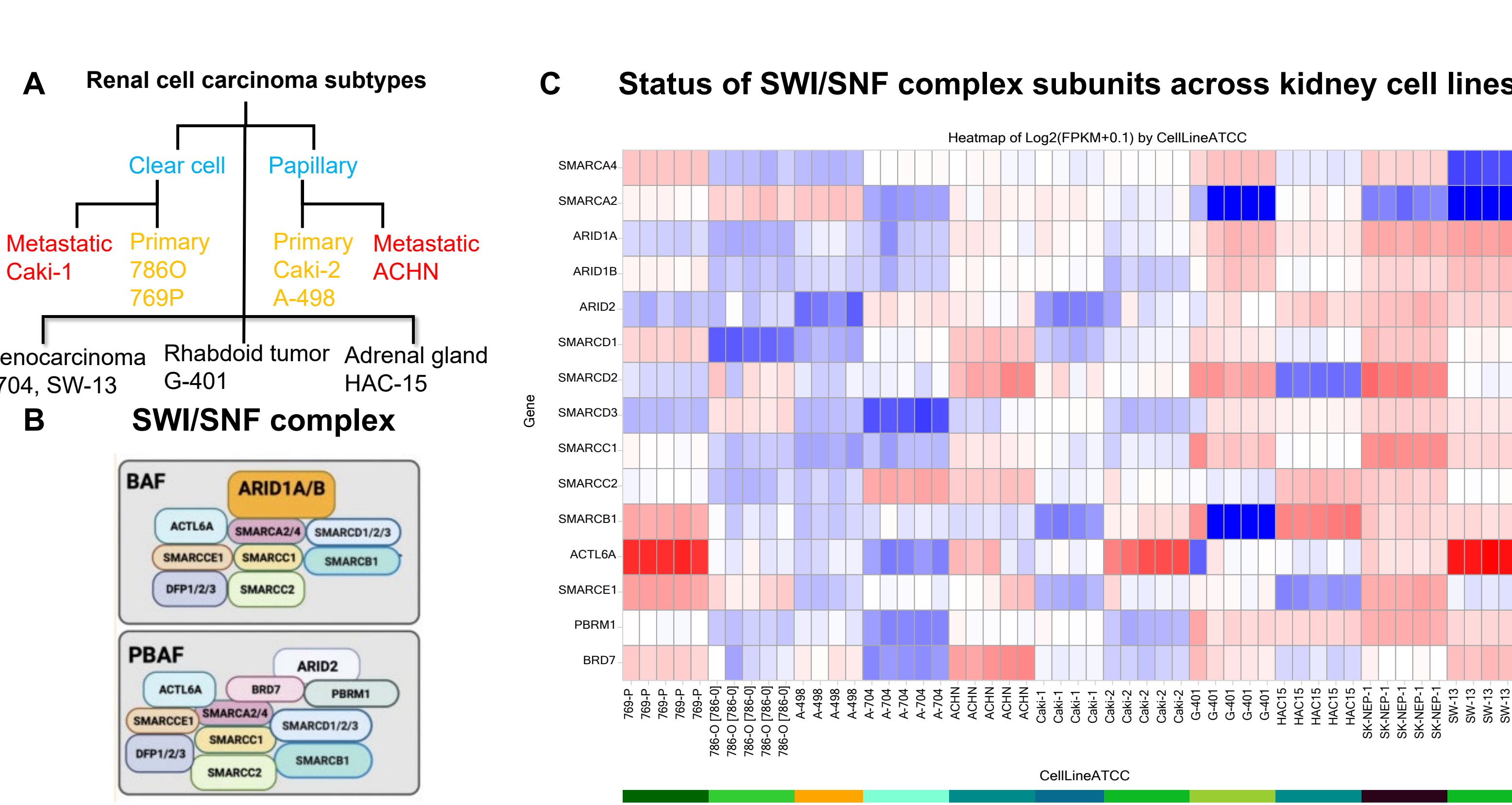


Figure 3: Molecular hallmarks in renal cancer subtypes. (A) Diagram displaying renal cell carcinoma subtypes. (B) Diagram illustrating the assembly of SWI/SNF complex subunits. (C) Heatmap depicting the differential expression of SWI/SNF subunits (x-axis) in the indicated renal carcinoma cell lines (y-axis). (D) Heatmap displaying the expression pattern of key genes (x-axis) in renal carcinoma subtypes (y-axis). (E, F) Box plots showing the relative mRNA levels of indicated genes in renal carcinoma subtypes. The data suggest both positive and negative correlations among expressed genes, indicating the intrinsic activity of molecules in renal cell cancer. (F) Venn diagram illustrating the number of genes shared and/or uniquely expressed in renal cancer cells versus normal control cells. HRCEPIC: Human Renal Cortical Epithelial Cell. The presented data shows multiple biological replicates of the respective cell line models.

## Tumor paired normal cell line comparison

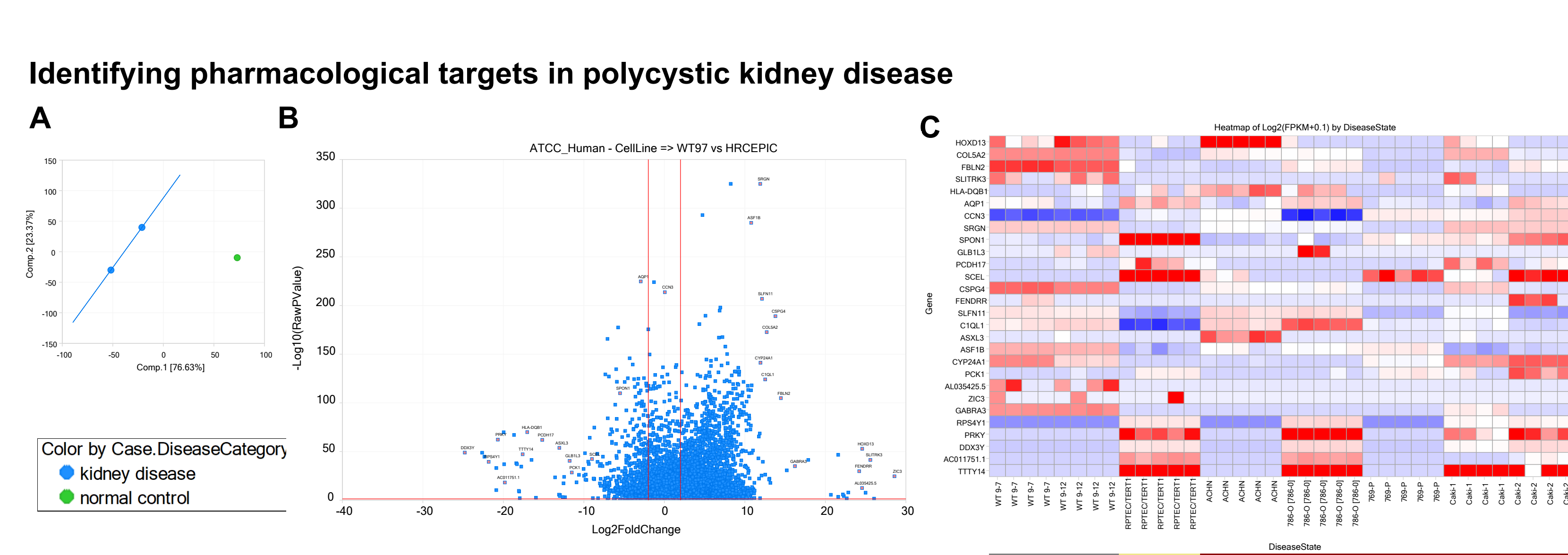


Figure 4: Comparative expression analysis reveals pathways and regulators in polycystic kidney disease. (A) PCA plot displaying the segregation of normal and kidney cancer samples. (B) Volcano plot showing genes differentially expressed in WT 9-7 cells compared to normal controls. (C) Expression pattern of genes differentially expressed in the WT 9-7 cell line across various renal carcinoma cell lines. (D) Heatmap depicting the activity of indicated pathways in polycystic kidney and normal cells. (E, F) IPA analysis revealing molecular mechanisms in polycystic kidney disease. (G) Differentially expressed genes in the WT 9-7 cell line compared to normal control cell lines.

## Canonical pathways in kidney cell line

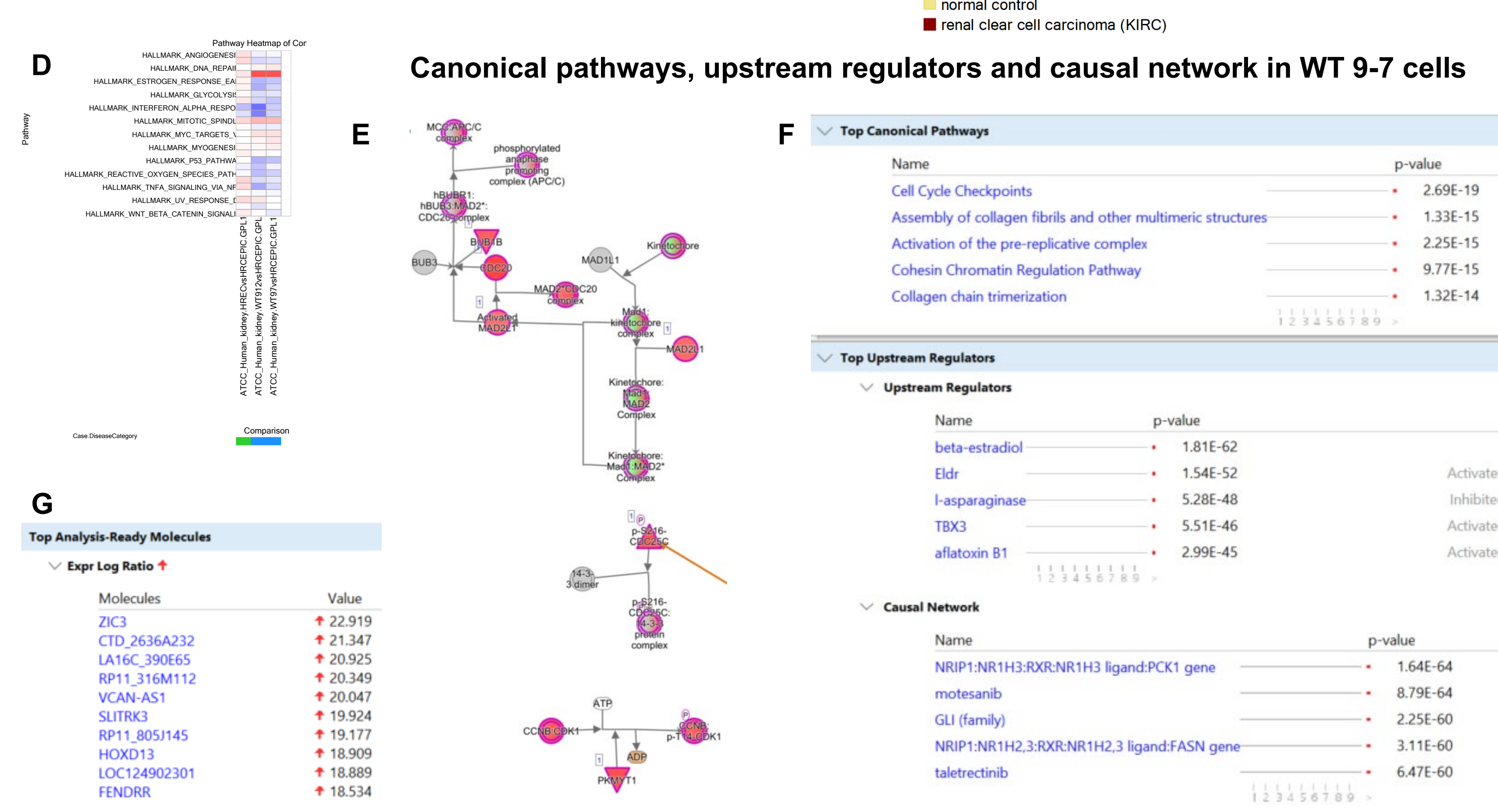


Figure 4: Comparative expression analysis reveals pathways and regulators in polycystic kidney disease. (A) PCA plot displaying the segregation of normal and kidney cancer samples. (B) Volcano plot showing genes differentially expressed in WT 9-7 cells compared to normal controls. (C) Expression pattern of genes differentially expressed in the WT 9-7 cell line across various renal carcinoma cell lines. (D) Heatmap depicting the activity of indicated pathways in polycystic kidney and normal cells. (E, F) IPA analysis revealing molecular mechanisms in polycystic kidney disease. (G) Differentially expressed genes in the WT 9-7 cell line compared to normal control cell lines.

## Summary

- ATCC<sup>®</sup> Cell Line Land provides a resource of OMICS data for ATCC's verified and authenticated cell lines that are physically traceable to the repository for cross-validation and independent verification.
- Utilizing ACLL data enhances reliability, accuracy, and reproducibility in pre-clinical research.
- ACLL data facilitates suitable model cell line selection for research purposes.
- Additionally, ACLL data expedite exploratory research and target identification.
- These data can be combined with publicly available data for sanity checks and to ensure scientific rigor and accuracy. ACLL is accessible through QIAGEN<sup>®</sup> Digital Insights upon subscription. <https://digitalinsights.qiagen.com/atcc-cell-line-land/>

