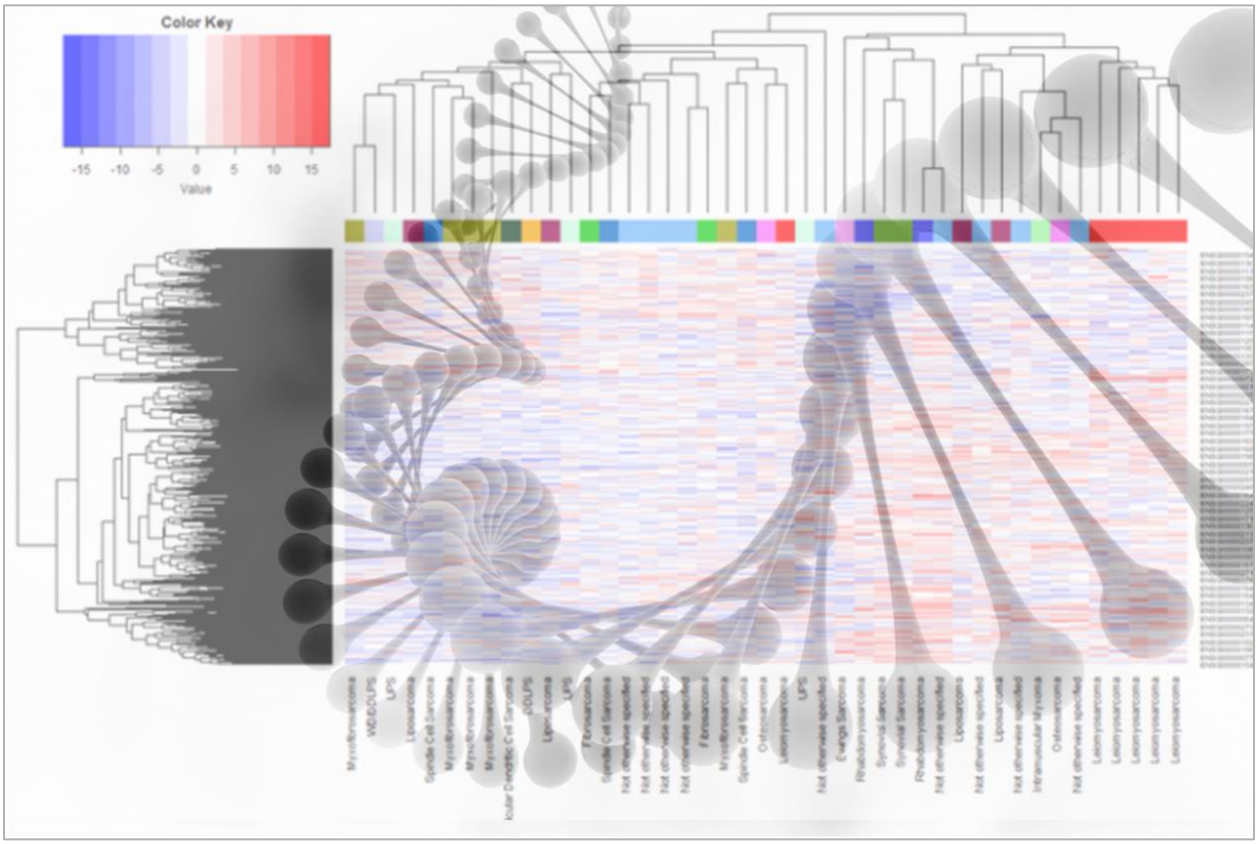
CERTIS™ | ONCOLOGY INTELLIGENCE

WHITE PAPER

BRINGING CERTAINTY TO PATIENT-DERIVED XENOGRAFT MODEL CHARACTERIZATION



AUTHORS AND CONTRIBUTORS

Long Do, PhD[1]; Warren Andrews, PhD[1]; Yuan-Hung Chien, PhD[1]; Luke Jervis[1]; Jantzen Sperry, PhD[1]; Bianca Carapia[1]; Deborah Yan[1]; Giovanni Rivera[1]; Elie Diner, PhD[1]; Noah Federman, MD[1,2]; Arun Singh, MD[1,2]; Fritz C. Eilber, MD[1,2]; Brian Datnow, MD[1]; Jonathan Nakashima, PhD[1]

[1]Certis Oncology Solutions, San Diego, CA; [2]UCLA, Los Angeles, CA

EXECUTIVE SUMMARY

Patient-derived xenograft (PDX) models are indispensable for translational cancer research. However, public and private PDX tumor banks may lack standard quality controls, which may lead to genomic and transcriptomic characterization inaccuracies, and the potential to reduce efficiencies with which oncology researchers and clinicians can mimic tumor pharmacology, growth, metastasis, resistance and relapse. To demonstrate the quality of the Certis tumor bank and to ensure the data in the BarneyOI Cancer Model Database™ accurately represents the growing collection of its PDX models, Certis analyzed mutation and transcriptomic profiles and compared gene expression results to both disease-free tissue and cancer patient profiles.

KEY TAKEAWAYS

- Gene expression from the Certis PDX tumor models is highly concordant with primary cancer samples in the Cancer Genome Atlas (TCGA), validating them as high-quality, valuable tools for preclinical research.
- Certis' PDX models exhibit dysregulation in typical cancer pathways compared to disease-free tissue (i.e., normal controls).
- Common cancer driver mutations in TP53, EGFR, PTEN, and KRAS are found at a high frequency across tumor types in Certis' PDX models.
- Mutational data for Certis' PDX models are highly concordant across whole-exome sequencing (WES) and RNA-seq data.
- The quality and concordance of Certis' PDX models with TCGA ensure consistency and certainty, providing drug developers with clinically relevant models for cancer research.
- The BarneyOI Cancer Model Database enables registered users to filter by mutation, copy number, fusions, microsatellite instability (MSI) and gene expression, to facilitate the identification of clinically relevant cancer models.

INTRODUCTION

Patient-derived xenografts (PDX) have become an essential model for preclinical and clinical cancer research.[1] While there are several different PDX methodologies, generally, PDX models use tumor tissue from cancer patients that are subsequently implanted into an immunodeficient mouse and used for testing. Across various cancer types, they offer a more accurate clinical representative than other common models, such as cancer cell lines and genetically engineered mouse models (GEMMs). PDX models can mimic pharmacology, growth, metastasis, resistance, and relapse, characteristics typically seen in the clinic.[2-4] As a result, they are used widely in preclinical research to measure tumor response to drug candidates or combinations.[5-7] In clinical trials, they can facilitate co-clinical trials, where PDX models are generated from participants and used to predict clinical outcomes through treatment with the same regimens as patients. Clinicians can also use them to guide real-world patient treatments, allowing clinicians to test multiple therapeutic agents and select the one most likely to be efficacious for patients.[8]

Given their utility across the cancer research continuum, there has been an explosion in the availability of PDX models from public and private biorepositories.[9,10] Yet, depending on the PDX model selected and the tumor bank of origin, there can be gaps in metadata and inconsistent degrees of molecular characterization. Even with detailed genomic or transcriptomic data, proper quality control pipelines are required to ensure models are appropriately matched with initial annotations and free from sequencing

errors and artifacts. Standardized quality control practices should be firmly applied to ensure that model characteristics and performance are validated and maintained.[9] In addition, many tumor banks benchmark their expression data from each model against other tumor models, making it difficult to get an accurate measure of expression compared to normal tissue.[9,10] Together, this creates a significant gap in the precise characterization of PDX models and makes validation of mutation or expression data in the pathology of cancer challenging.

To more fully characterize the PDX models within the Certis tumor bank, we performed WES and transcriptomic profiling of each model. We used publicly available databases, including Genotype-Tissue Expression (GTEx)[11] and TCGA[12,13], to properly benchmark our tumor expression data against normal controls and compare tumors against other similar, cancer patients, respectively. Furthermore, we compare mutations identified by WES to those in TCGA and find high concordance between these models. Lastly, we demonstrate a high concordance between our WES and RNA-seq data. Taken together, we present a rigorous molecular characterization of our models by using publicly available datasets. This information will enable researchers to access trusted PDX models that deliver consistent and reproducible results that drive preclinical drug development.

MATERIALS AND METHODS

MUTATION ANALYSIS

WES paired-end reads were pre-processed to remove mouse contamination (Xenome v1.0) and poor-quality reads (fastp).[14] Spliced Transcripts Alignment to a Reference (STAR)[15] was used to map reads to the Human GRCh38 (hg38) genome followed by variant calling using GATK Mutect2[16] and filtered against Panel of Norm (PON) from the 1000 Genomes project. Additionally, a list of over 1.2 million human genome-aligned mouse alleles (HAMA) was used to filter any variants from mouse contamination. Lastly, SnpEff was used to annotate and predict the effects of genetic variants.[17] The GATK somatic copy number variant (CNVs) tool was used to produce segmented CNV data from WES, which was then subsequently mapped to genes to generate gene-level estimates.[16]

GENE EXPRESSION ANALYSIS

Gene expression was measured experimentally by bulk poly(A)-selected RNA-seq. Mouse contamination was removed (Xenome v1.0) and further processed for quality using fastp and FastQC.[18,19] STAR was used to map stranded paired-end reads to the Human GRCh38 (hg38) genome, and gene expression was quantified using RSEM.[20] Differential gene expression analysis was performed using edgeR between each model and healthy tissue from the GTEx project.[21] Differential gene expression from tissue-specific cancers was collected from GEPIA, an online resource comparing TCGA to normal controls from GTEx, and used to compare against Certis PDX differential expression against similarly normal controls from GTEx.[22] The Gene Set Enrichment Analysis (GSEA)[23] algorithm was used to find enriched pathways (KEGG).[24]

RESULTS AND DISCUSSION

TRANSCRIPTOMIC ANALYSIS REVEALS HIGH CONCORDANCE BETWEEN CERTIS PDX MODELS AND TCGA

We used NGS to characterize the transcriptome of over 150 PDX models in the BarneyOI Cancer Model database, including clinically meaningful cancer types such as sarcoma (SARC), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), glioblastoma multiforme (GBM), lung adenocarcinoma (LUAD), pancreatic adenocarcinoma (PAAD), and skin cutaneous melanoma (SKCM). To better understand how gene expression in Certis' PDX tumor models compares to normal controls and analogous tumor types from patients in TCGA, Certis performed pairwise correlation analysis for differentially expressed genes in each cancer type.

Compared to normal controls from GTEx, the Certis PDX models showed an intermediate positive correlation, with Person's correlation average of r=0.62 across all cancer types and genes of interest (Figure 1A). Certis found an even higher correlation between gene expression patterns in Certis' PDX models and those found in primary samples from the TCGA (Figure 1B; Pearson's correlation average r=0.82). To further validate concordance between the PDX models and those in TCGA, Certis used GEPIA to identify the total number of significant differentially expressed genes (false discovery rate (FDR) < 0.05) for specific cancer types in TCGA vs. GTEx (Figure 1C).[22] In a similar analysis with the PDX models vs. GTEx, Certis identified a similar number of differentially expressed genes across all cancer types, demonstrating agreement between transcriptional profiles in the PDX models and primary tumor data in TCGA (Figure 1C).

FIGURE 1.A. **CERTIS PDX VS. GTEX:** GENE EXPRESSION OF NORMAL TISSUE FROM THE GENOTYPE-TISSUE EXPRESSION (GTEX) DATABASE SHOWS POSITIVE CORRELATION WITH SIMILAR TISSUE FROM CERTIS PDX MODELS (PEARSON'S CORRELATION AVERAGE, R=0.62).

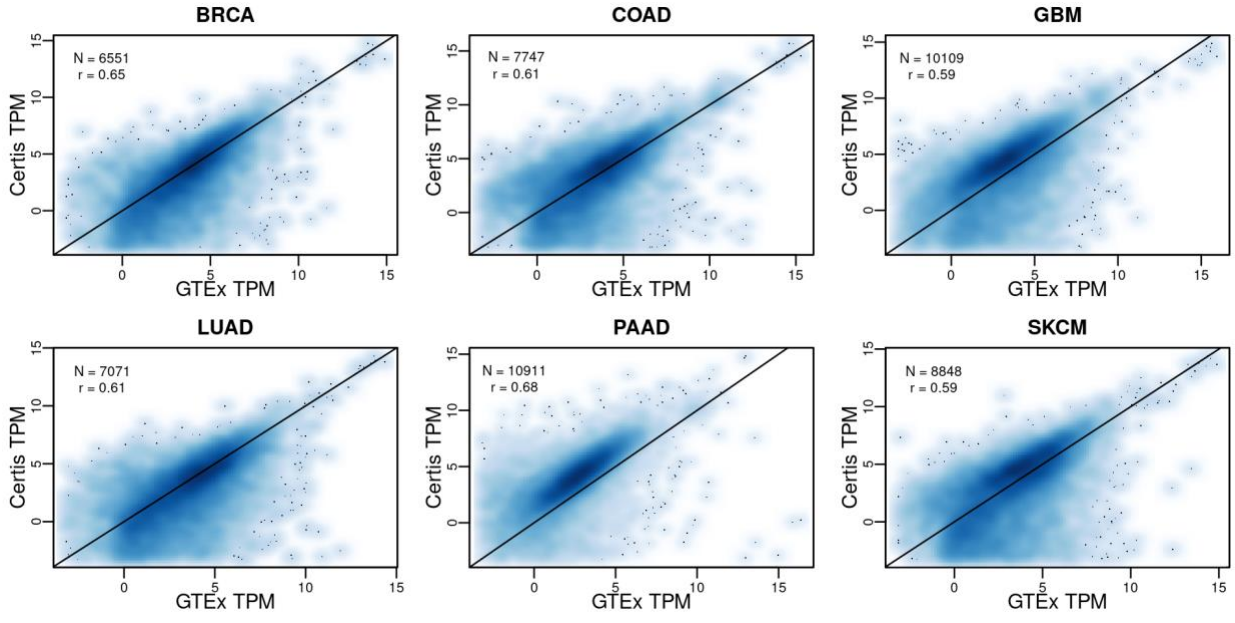


FIGURE 1.B. **CERTIS PDX VS. TCGA**: COMPARISON OF CERTIS PDX GENE EXPRESSION DATA AGAINST THE CANCER GENOME ATLAS (TCGA). CERTIS PDX GENE EXPRESSION SHOWS HIGH POSITIVE CORRELATION TO SIMILAR CANCER TYPES FROM TCGA (PEARSON'S CORRELATION AVERAGE, R=0.82).

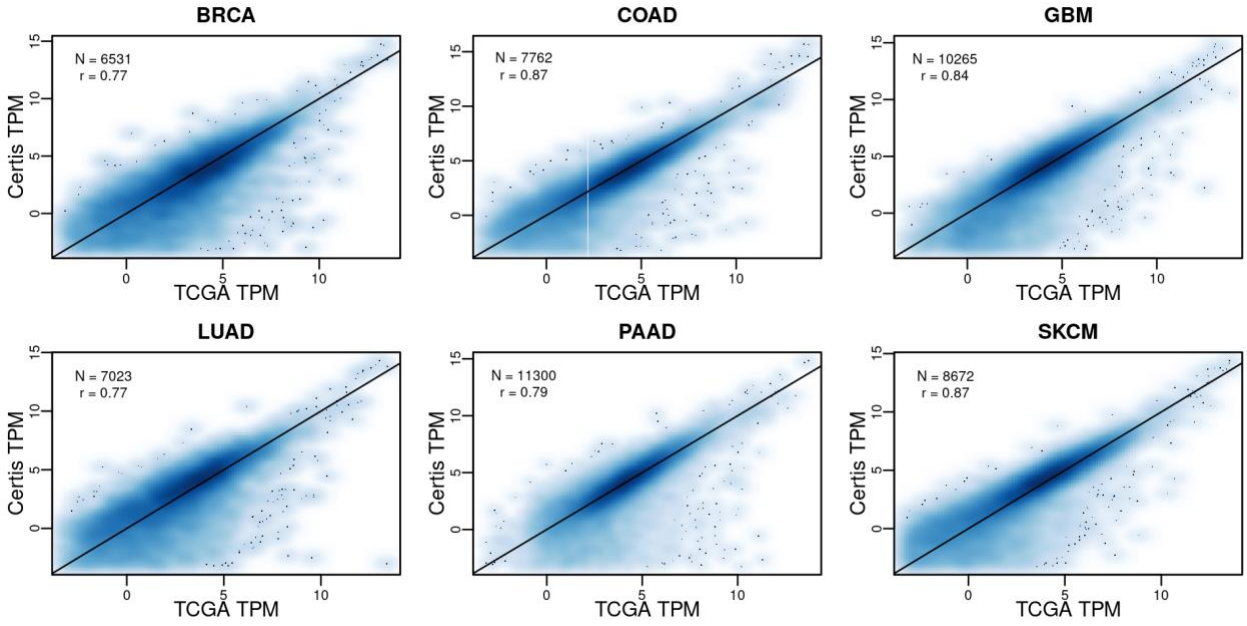
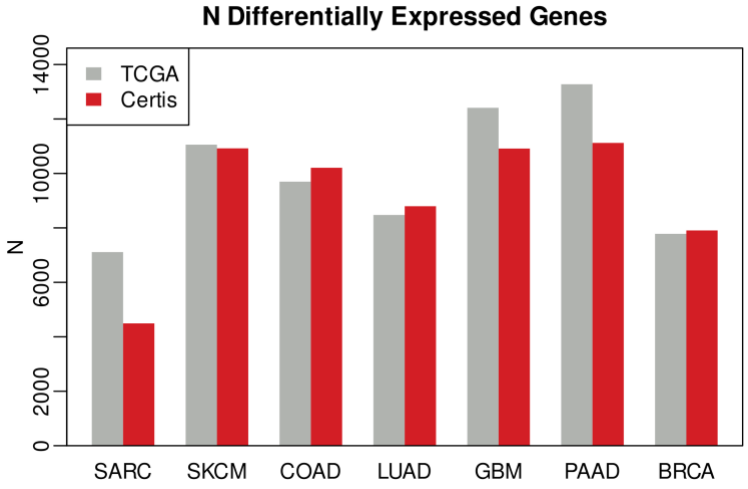


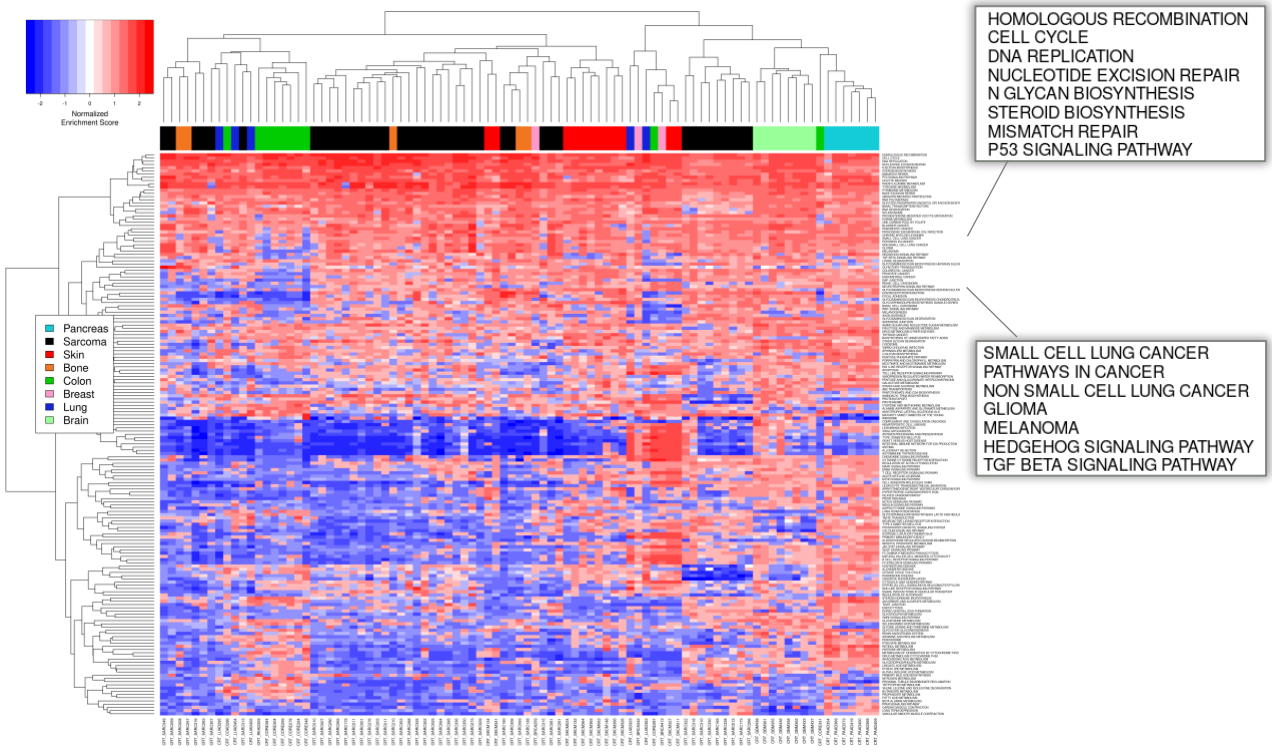
FIGURE 1.C. **SIGNIFICANT DIFFERENTIALLY EXPRESSED GENES**: SIGNIFICANT (FDR < 0.05) DIFFERENTIAL EXPRESSION WAS CALCULATED BY COMPARING GTEX NORMAL TISSUE TO CERTIS PDX OR GTEX VS. TCGA (FROM GEPIA).



DIFFERENTIALLY EXPRESSED GENE EXPRESSION ANALYSIS REVEALS DYSREGULATION OF COMMON CANCER PATHWAYS

To further characterize gene expression in our tumor models, we performed GSEA using transcriptomics data from the Certis PDX models compared to normal control data from GTEx. GSEA analysis demonstrates strong dysregulation of cell cycle pathways and pathways for specific tumor types, a hallmark of cancer pathology (Figure 2). In addition, unsupervised hierarchical clustering based on GSEA resulted in models from the same tumor type clustering together, demonstrating the high quality of our models and the absence of batch effects in our RNA-seq data (Figure 2). We did not observe robust clustering for many of our sarcoma models, likely due to the molecular heterogeneity observed in this tumor type.[25]

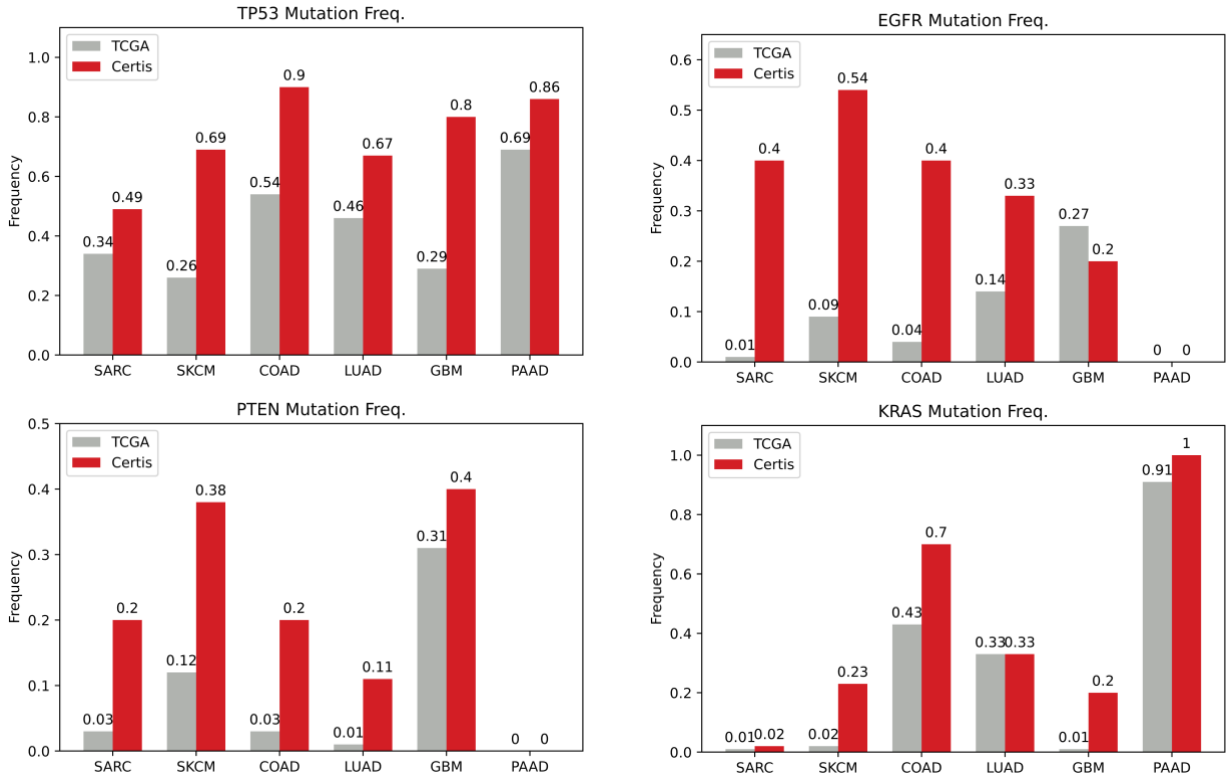
FIGURE 2. GENE SET ENRICHMENT ANALYSIS (GSEA) REVEALS DYSREGULATED CANCER PATHWAYS AND CONCORDANCE BETWEEN SIMILAR DISEASE MODELS. GSEA ANALYSIS USING NORMAL CONTROLS FROM GTEX AGAINST CERTIS PDX MODELS UNCOVER PATHWAYS INDICATIVE OF CANCER PATHOLOGY. ADDITIONALLY, UNSUPERVISED HIERARCHICAL CLUSTERING SHOWS HIGH CONCORDANCE AND CLUSTERING BETWEEN SIMILAR MODELS AT THE PATHWAY LEVEL.



MUTATIONAL PROFILING REVEALS COMMON CANCER DRIVER MUTATIONS

The TCGA has performed molecular characterization of over 20,000 primary cancer and normal control samples, across 33 cancer types. Given this substantial dataset, Certis characterized the mutational profile of the tumors and compared them to those found in the TCGA. Certis performed WES on the PDX models and analyzed the frequency of common driver mutations in *TP53*, *EGFR*, *PTEN*, and *KRAS*, commonly mutated in various types of cancer.[26] The most frequently mutated gene in the PDX models was *TP53,* and for genes like *KRAS* and *TP53*, the PDX models shared higher frequencies than TCGA (Figure 3). Furthermore, there are high frequencies of driver mutations in the Certis PDX tumor bank when high frequencies are seen in TCGA data, as would be expected in a comprehensive dataset such as the TCGA.

FIGURE 3. COMMON CANCER DRIVER MUTATIONS. MUTATION FREQUENCY OF COMMON CANCER GENES BETWEEN TCGA AND CERTIS PDX MODELS FROM WHOLE EXOME SEQUENCING (WES) REVEAL P53 AS THE MOST COMMON MUTATED GENE PAN-CANCER AND KRAS AS THE MOST COMMON MUTATED GENE IN PANCREATIC CANCER (PAAD).



In addition, Certis characterized the copy number variations (CNVs) in different tumor types, including amplifications and deletions (Table 1). Strong associations between CNVs and a wide range of human cancers have been extensively described. Certis found significant evidence of gene amplifications and deletions in the Certis PDX tumor bank.[27] Across all cancer types studied, gene amplification was more common than gene deletion, and SARC and LUAD showed the highest amplification (> 1000 genes).

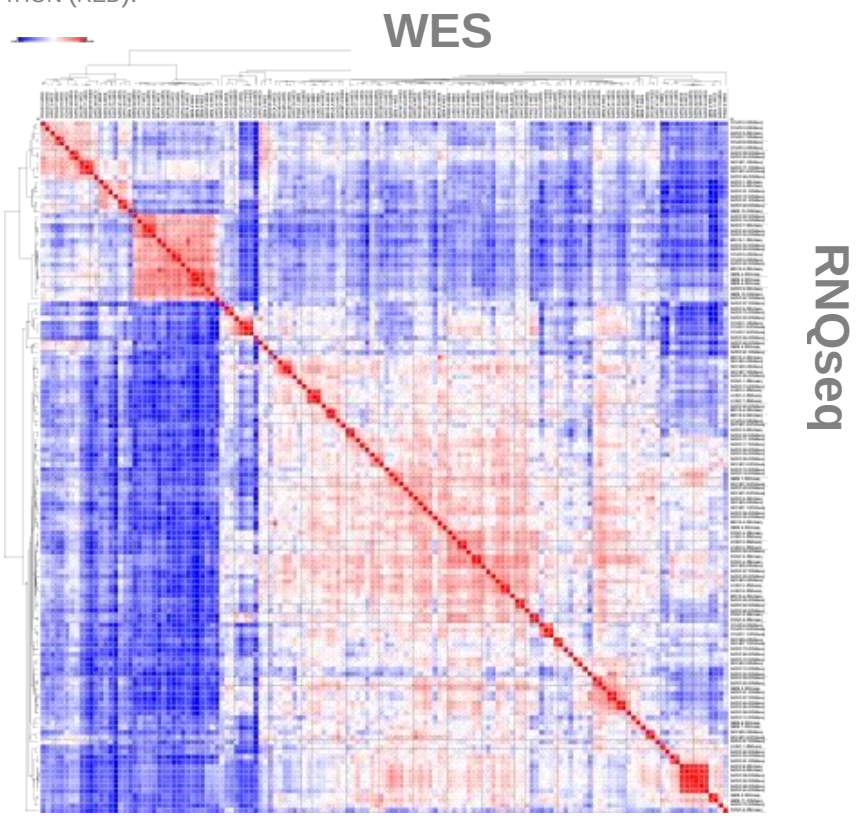
TABLE 1. CERTIS PDX COPY NUMBER VARIATION (CNV) ANALYSIS. CNV ANALYSIS ON CERTIS PDX MODELS FROM WES REVEALS A GREATER NUMBER OF GENE AMPLIFICATIONS ACROSS MULTIPLE CANCER MODELS COMPARED TO GENE DELETIONS.

Cancer	Category	Average # genes	Top 3 genes
SARC	Amplification	1243	GRRIN2, CDKN2A, CDKN2B
	Deletion	457	CDKN2A, RB1, DDR1
SKCM	Amplification	693	B2M, GRRIN2, FOXL2
	Deletion	161	CDKN2A, LENG8, CDKN2B
COAD	Amplification	978	GRRIN2, BMI1, CDK8
	Deletion	577	DDR1, DAXX, ATP6V1B2
LUAD	Amplification	1590	KLF5, MAFA, DDR1
	Deletion	504	EPHA3, MSH3, PIK3R1
GBM	Amplification	966	TERT, LTK, FGFR3
	Deletion	220	CDKN2B, CDKN2A, MTAP
PAAD	Amplification	1030	LENG8, FOXL2, FGF4
	Deletion	453	MXRA5, CDKN2A, CDKN2B

STRONG CONCORDANCE BETWEEN WES AND RNA-SEQ TUMOR MUTATION DATA

To supplement our mutational analysis, Certis performed a correlation analysis between WES and RNA-seq mutation data for each Certis PDX model. Certis found high concordance between datasets for tumor models, adding certainty to the molecular characterization and further validating the RNA-seq, WES, and bioinformatics methodology (Figure 4).

FIGURE 4. MUTATION CORRELATION ANALYSIS BETWEEN WES AND RNA-SEQ IN CERTIS PDX MODELS. OVERLAP ANALYSIS BETWEEN OUR RNA-SEQ AND WES REVEAL HIGH CONCORDANCE OF MUTATIONS BETWEEN THE NGS DATASETS. THE SAME PDX MODEL CAN BE SEEN ON THE DIAGONAL WITH HIGH CORRELATION (RED).



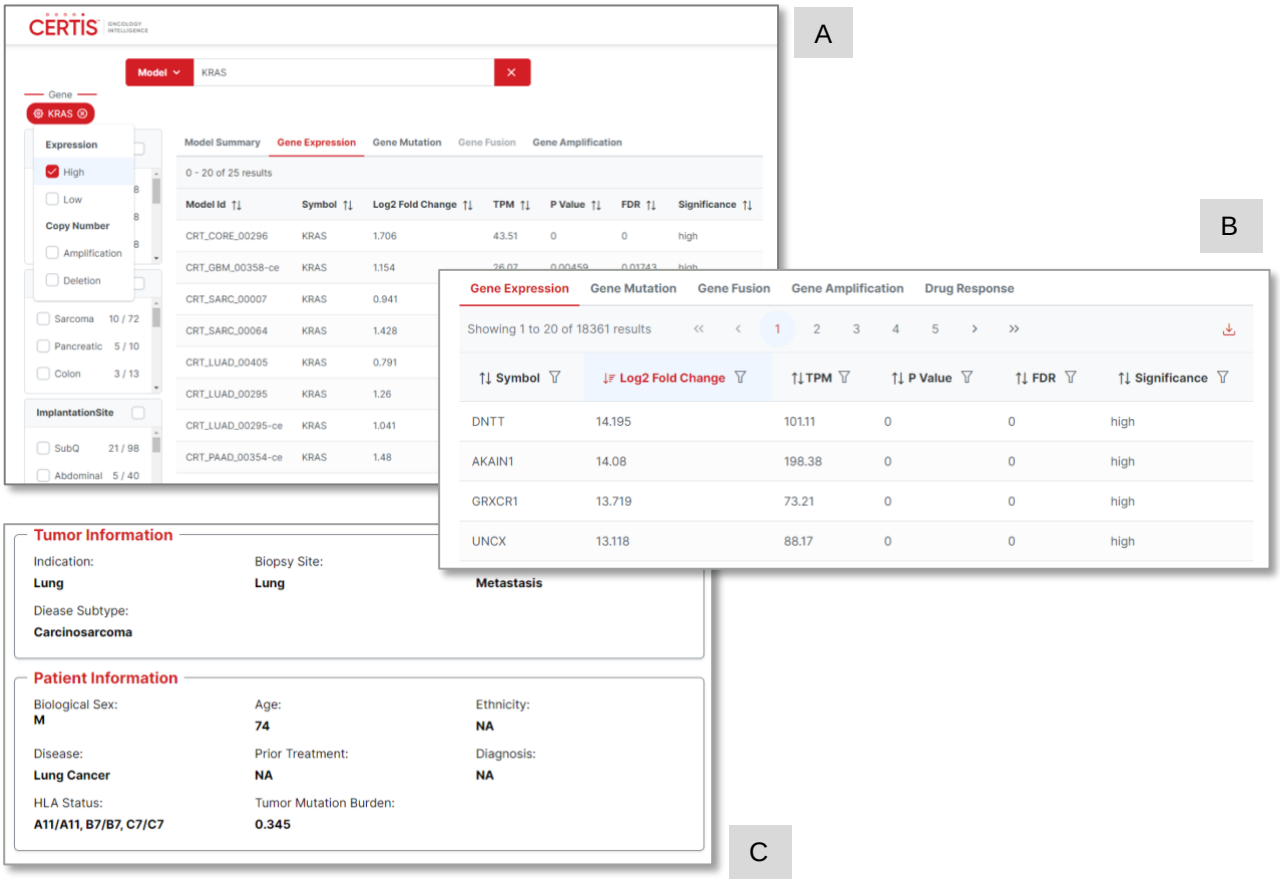
SUMMARY

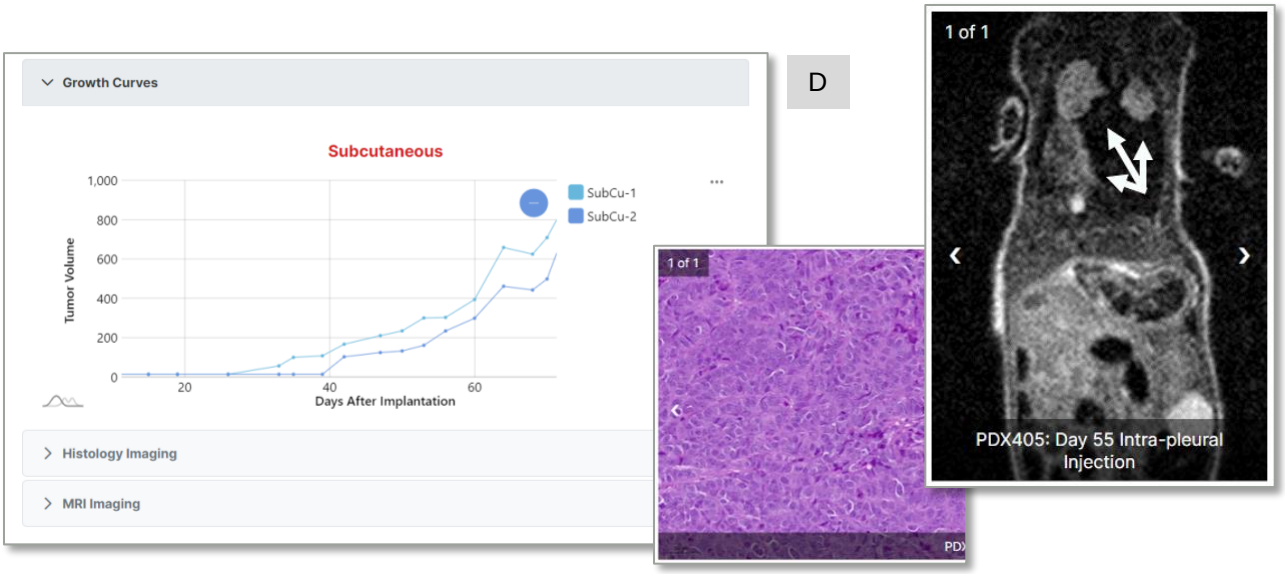
ACCESS HIGHLY CHARACTERIZED, RICHLY ANNOTATED PDX TUMOR MODELS

Characterization of Certis PDX models using NGS reveals high concordance with datasets from TCGA, including thousands of genomics and transcriptomics datasets from primary cancer samples. Compared to normal controls from GTEx, Certis PDX models exhibit dysregulation of crucial pathways indicative of cancer pathology. The quality and concordance of the PDX models with known patient datasets ensure consistency and certainty, providing the research community and pharmaceutical scientists with clinically relevant models for cancer research.

FIGURE 5. THE BARNEYOI CANCER MODEL DATABASE IS A WEB-BASED APPLICATION THAT ALLOWS USERS TO SEARCH A GROWING COLLECTION OF HIGHLY CHARACTERIZED, RICHLY ANNOTATED, AND LOW-PASSAGE PATIENT-CONSENTED XENOGRAFT MODELS AND MATCHED CANCER CELL LINES.

WITH ENHANCED FILTERING FEATURES, USERS CAN: A. FINE-TUNE CUSTOM KEYWORD SEARCHES AND DRILL DOWN DEEP INTO GENOMIC PROFILES USING ADDITIVE GENE FILTERS, B. COMPARE GENE EXPRESSION RESULTS TO DISEASE-FREE TISSUE, C. OBTAIN A DETAILED MODEL PROFILE OF PATIENT AND TUMOR CHARACTERISTICS, D. VIEW INTERACTIVE GRAPHICAL REPRESENTATIONS OF MODEL SETS AND EXPORT AND SHARE YOUR SELECTED ANALYSIS.





CLOSING THE TRANSLATION GAP STARTS WITH USING THE MOST CLINICALLY RELEVANT CANCER MODEL

With Certis tumor models having a strong average positive correlation (r=0.82) to the thousands of primary cancer samples in the Cancer Genome Atlas (TCGA), cancer researchers can be assured that each Certis model and its genomic characterization data is accurate, ultimately helping to save time, money, and lives.

Learn more about the searchable cancer model database and the detailed NGS data on all Certis models in the tumor bank.

Log in to the Barney OI Cancer Model Database™ now using your existing username and password or request access.

REFERENCES

1 Jung J, Seol HS, Chang S. The Generation and Application of Patient-Derived Xenograft Model for Cancer Research. *Cancer Res Treat.* 2018;50:1–10.

2 Julien S, Merino-Trigo A, Lacroix L *et al*. Characterization of a Large Panel of Patient-Derived Tumor Xenografts Representing the Clinical Heterogeneity of Human Colorectal Cancer. *Clin Cancer Res.* 2012;18(19):5314–5328.

3 DeRose YS, Wang G, Lin YC *et al.* Tumor Grafts Derived from Women with Breast Cancer Authentically Reflect Tumor Pathology, Growth, Metastasis and Disease Outcomes. *Nat Med.* 2011;17(11):1514–1520.

4 Patrizii M, Bartucci M, Pine SR, Sabaawy HE. Utility of Glioblastoma Patient-Derived Orthotopic Xenografts in Drug Discovery and Personalized Therapy. *Front Oncol.* 2018;8(23).

5 Hidalgo M, Amant F, Biankin AV *et al*. Patient-Derived Xenograft Models: An Emerging Platform for Translational Cancer Research. *Cancer Discovery*. 2014;4(9):998–1013.

6 Tentler JJ, Tan AC, Weekes CD *et al.* Patient-Derived Tumour Xenografts as Models for Oncology Drug Development. *Nat Rev Clin Oncol.* 2012;9(6):338–350.

7 Ice RJ, Chen M, Sidorov M *et al.* Drug Responses are Conserved Across Patient-Derived Xenograft Models of Melanoma Leading to Identification of Novel Drug Combination Therapies. *Br J Cancer.* 2020;122(5):648–657.

8 Vargas R, Gopal P, Kuzmishin GB *et al.* Case Study: Patient-Derived Clear Cell Adenocarcinoma Xenograft Model Longitudinally Predicts Treatment Response. *NPJ Precis Oncol.* 2018;2(14).

9 Craciun L, Spinette SA, Rassy M *et al*. Tumor Banks: A Quality Control Scheme Proposal. *Front Med* (Lausanne). 2019;6:225.

10 Sun H, Cao S, Mashl RJ *et al*. Comprehensive Characterization of 536 Patient-Derived Xenograft Models Prioritizes Candidates for Targeted Treatment. *Nat Commun*. 2021;12(1):5086.

11 GTEx Consortium. The Genotype-Tissue Expression (GTEx) Project. *Nat Genet*. June 2013;45(6):580-5.

12 Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): An Immeasurable Source of Knowledge. *Contemp Oncol* (Pozn). 2015;19(1A):A68-77.

13 National Cancer Institute. The Cancer Genome Atlas (TCGA) Program. NIH. Accessed on July 1, 2022.

14 Conway T, Wazny J, Bromage A *et al*. Xenome--A Tool for Classifying Reads from Xenograft Samples. *Bioinformatics*. June 15, 2012;28(12):i172-8.

15 Dobin A, Davis CA, Schlesinger F, Drenkow J *et al*. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics*. Jan. 1, 2013;29(1):15-21.

16 McKenna A, Hanna M, Banks E *et al*. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res*. 2010;20:1297-303.

17 Cingolani P, Platts A, Wang le L *et al*. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain w1118; iso-2; iso-3. *Fly* (Austin). Apr-Jun 2012;6(2):80-92.

18 Shifu Chen, Yanqing Zhou, Yaru Chen, Jia Gu. fastp: An Ultra-Fast-All-In-One FASTQ Preprocessor. *Bioinformatics.* Sep. 2018;34(17):i884-i890.

19 Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. *Babraham Bioinformatics*. 2010.

20 Li B, Dewey CN. RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinformatics*. 2011;12(323).

21 Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics*. 2010; 26(1):139-140.

22 Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A Web Server for Cancer and Normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res*. 2017;45(W1): W98–W102.

23 Subramanian A, Tamayo P, Mootha VK *et al.* Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc Natl Acad Sci USA*. Oct 25, 2005;102(43):15545-50.

24 Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. Jan 1, 2000;28(1):27-30.

25 Skubitz, KM, Pambuccian S, Manivel JC, Skubitz AP. Identification of Heterogeneity Among Soft Tissue Sarcomas by Gene Expression Profiles from Different Tumors. *J Transl Med.* 2008;6(23).

26 Bailey MH, Tokheim C, Porta-Pardo E, *et al*. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell.* 2018;173(2):371–385.e18.

27 Shlien A, Malkin D. Copy Number Variations and Cancer. *Genome Med. 2009;*1(6):62.