CERTIS™ ONCOLOGY INTELLIGENCE

Next-Generation Characterization of Patient-Derived Xenografts

Jonathan Nakashima, PhD[1]; 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]; Noah Federman, MD[1,2]; Arun Singh, MD[1,2]; Fritz C. Eilber, MD[1,2]; Brian Datnow, MD[1]

ABSTRACT

Patient-derived xenograft (PDX) models have provided the research community with dynamic preclinical model systems in which to study human cancers and accelerate drug development. To accelerate translational oncology, we have developed a library of PDX models and characterized them using next-generation sequencing. Here, we provide mutation and transcriptomic profiles of our PDX collection along with comparisons against matched normal and cancer patient profiles. Lastly, we provide a web portal providing comprehensive genomic profiles for every model to identify mutation, copy number, fusions, Microsatellite Instable (MSI) status, gene expression and enriched pathways to facilitate identification of clinically relevant models.

METHODS

Mutation Analysis: Whole exome sequencing (WES) paired-end reads were pre-processed to remove mouse contamination (Xenome v1.0) and poor-quality reads (fastp). STAR was used to map reads to the Human GRCh38 (hg38) genome followed by variant calling using GATK Mutect2 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 effects of genetic variants. The GATK somatic copy number variant (CNVs) tool was used to produce segmented copy number variant (CNV) data from WES, which was then subsequently mapped to genes to generate gene-level estimates.

Gene Expression Analysis: RNA Sequencing (RNA-Seq): 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. STAR was used to map stranded paired-end reads to the Human GRCh38 (hg38) genome, and gene expression was quantified using RSEM. Differential gene expression analysis was performed using edgeR between each model and matched, healthy tissue from the Genotype-Tissue Expression (GTEx) project. Differential gene expression from tissue-specific cancers was collected from GEPIA[3], an online resource comparing TGGA to matched normal tissue expression from GTEx, and used to compare against Certis PDX differential expression against a similarly matched normal from GTEx. The Gene Set Enrichment Analysis (GSEA) algorithm was used to find enriched pathways (KEGG).

CONCLUSIONS

Characterization of Certis' PDX models using NGS reveals high concordance with datasets from TCGA, which includes thousands of *omics* datasets from primary cancer samples. When compared to matched normal controls from GTEx, Certis' PDX models exhibit dysregulation of key pathways, indicative of cancer pathology. The quality and concordance of our models with known patient datasets ensure consistency and certainty, providing the research community and our customers with clinically relevant models for cancer research. To see detailed NGS data on all models in the Certis Tumor Bank, register for access to our searchable database.

RESULTS

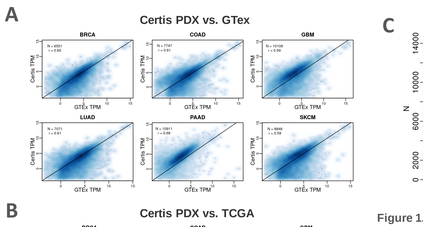


Figure 1. Certis PDX models show high concordance against matched normal and cancer gene expression datasets. **A**. Gene expression of normal tissue from the Genotype-Tissue Expression (GTEx) database shows positive correlation with matched tissue from Certis PDX models (Pearson's correlation average, r=0.62). **B**. Comparison of Certis PDX gene expression data against the Cancer Genome Atlas (TCGA) patient gene expression datasets shows even higher positive correlation (mean, r=0.82). **C.** Significant (FDR < 0.05) differential expression was calculated by comparing GTEx normal tissue to matched Certis PDX or GTEx vs. TCGA (from GEPIA).

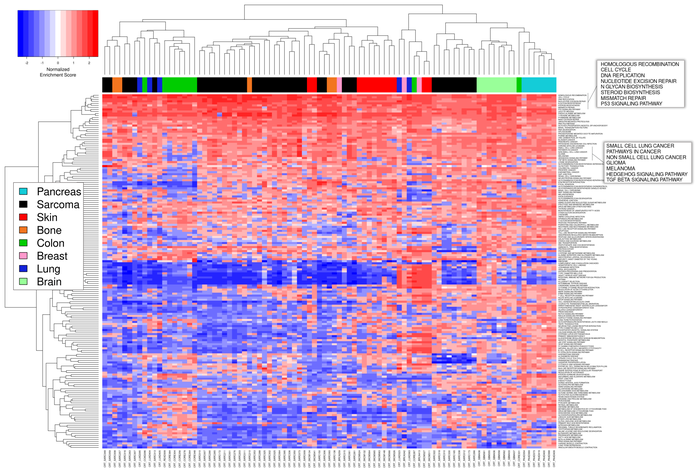


Figure 2. Gene Set Enrichment Analysis (GSEA) reviews dysregulated cancer pathways and concordance between similar disease models. GSEA analysis using matched normal controls from GTEx against Certis PDX models uncover pathways indicative of cancer pathology. Additionally, unsupervised hierarchical clustering review high concordance and clustering between similar models at the pathway level.

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).

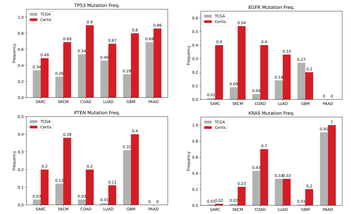


Table 1. Certis PDX copy number variation (CNV) analysis. CNV analysis on Certis PDX models from WES reveals a greater number of gene amplifications across multiple cancer models compared to gene deletions.

Cancer	Category	Average # genes	Top 3 genes
SARC	Amplification	1243	GRB1N2, CDKN2A, CDKN2B
	Deletion	457	CDKN2A, RB1, DDR1
SKCM	Amplification	693	B2M, GRB1N2, FOXL2
	Deletion	161	CDKN2A, LENG8, CDKN2B
COAD	Amplification	978	GRB1N2, 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

Figure 4. Mutation correlation analysis between WES and RNA-Seq in Certis PDX models. Overlap analysis between our RNA-Seq and WES review high concordance of mutations between the NGS datasets. The same PDX model can be seen on the diagonal with high correlation (red).

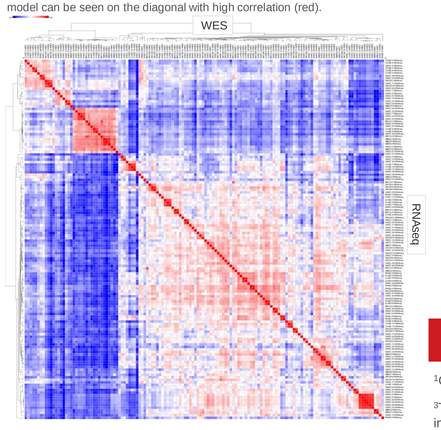
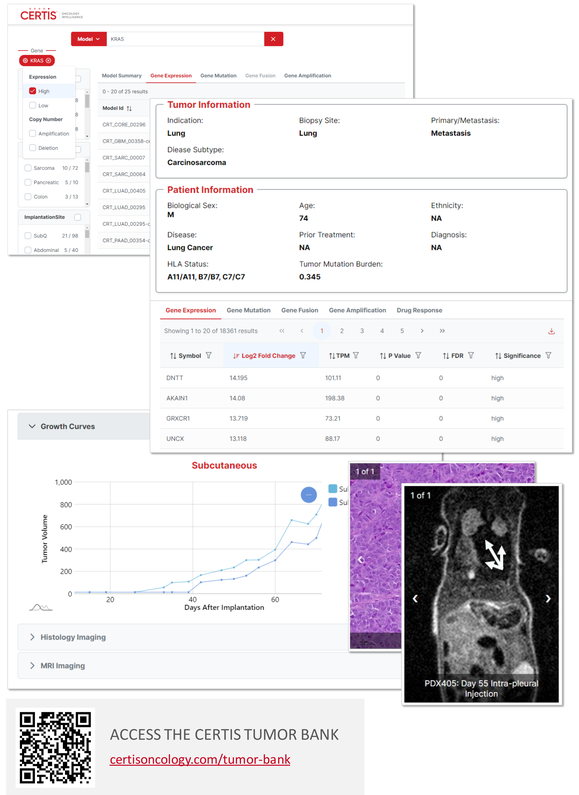


Figure 5. Certis Oncology Tumor Bank. The Certis Tumor Bank database is a web-based application with granular search and filtering features, allowing for robust and easy access to NGS characterization and model information from our commercially available collection of tumor models.



ACCESS THE CERTIS TUMOR BANK
certisoncology.com/tumor-bank

CITATIONS & ACKNOWLEDGEMENTS

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

[3]Tang, Zefang et al. "GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses." Nucleic acids research vol. 45,W1 (2017)