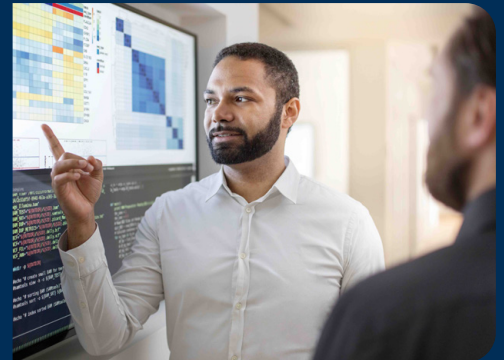


APPLICATION NOTE

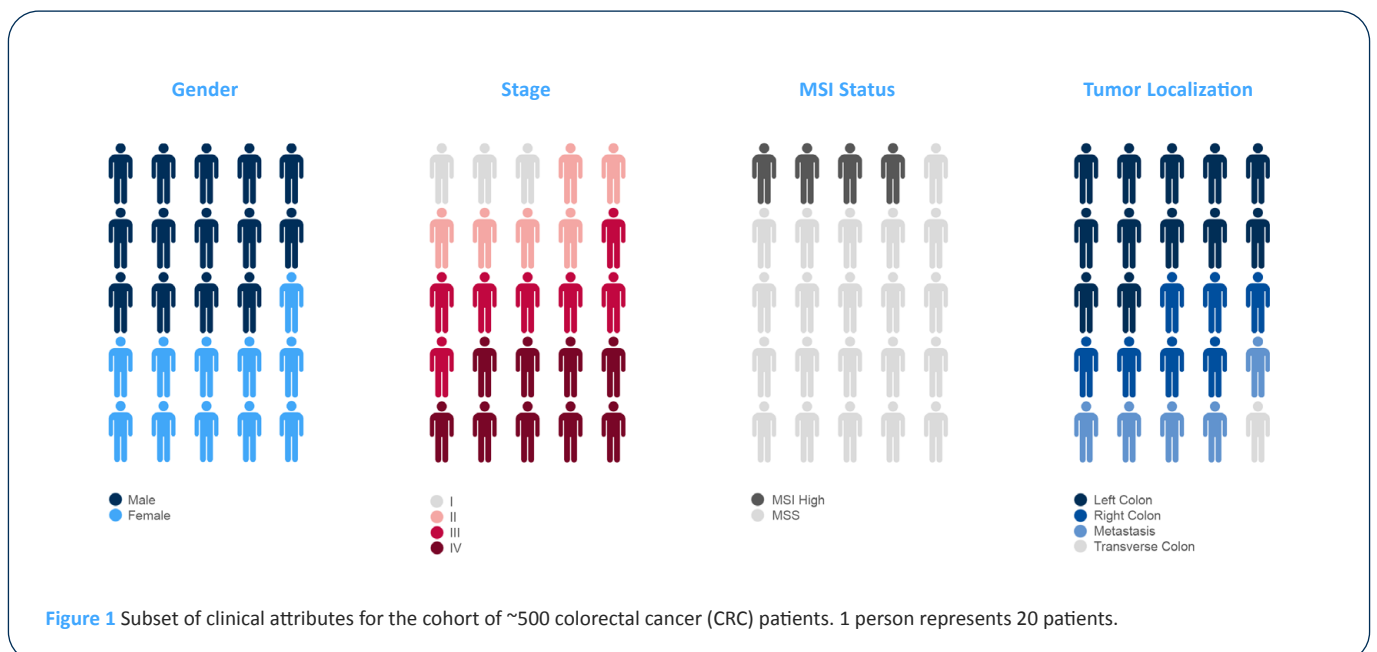
Proteogenomics Analysis of Colorectal Cancer Using the Indivumed Multi-omics Database



INTRODUCTION

Here we present the power of our multi-omics discovery platform, nRavel®, to drive cutting edge analytics within oncology informatics. We combine genomics, transcriptomics, and proteomics datasets generated from high-quality patient tissue to accelerate precision medicine approaches. We used the nRavel® platform to drive an integrated proteogenomics approach to aid the identification of putative therapeutic targets in a colorectal cancer (CRC) cohort of 500 patients, consisting of 388 patients with primary tumor and matching

adjacent normal samples, as well as 112 patients with metastatic CRC (Figure 1). We identified hundreds of proteins dysregulated in CRC that are dependent on the genetic background of the patient. Further, we characterized these proteins using clinical parameters from the nRavel® database to prioritize targets for precision medicine approaches. This project highlights the utility of using multiple omics levels to better describe the molecular mechanisms underpinning cancer.



MATERIALS AND METHODS

Collection of Patient Tissue Samples

- Patients with colorectal cancer gave informed consent and institutional review board approval was obtained from partner clinical sites. The use of human tissue was approved, and anonymization of patient data and samples was in compliance with data protection regulations.
- Tumor tissue of CRC and matched normal adjacent tissue were collected according to Indivumed's standard operating procedures within 10 minutes of surgical resection and tissue samples were snap frozen in liquid nitrogen.

Genomics

- For each sample, ≤ 20 mg tissue was homogenized, and DNA was extracted using Qiagen's QIAamp DNA Mini Kit.
- The library was prepared using the Zymo clean up kit and Nextera DNA Flex kit.
- Whole genomes were sequenced to a depth of 35x (normal) and 70x (tumor/metastasis) using a NovaSeq 6000 sequencer (Illumina).
- Short somatic variation dataset was generated using a consensus calling approach from four variant callers.

Proteomics

- For each sample, ≤ 20 mg tissue was lysed and homogenized prior to trypsin and LysC protein digestion.
- Electrospray ionization was followed by MS/MS quadropole / using CID on FUSION LUMOS (Thermo Scientific™).
- Data dependent acquisition of ions was performed prior to peptide identification using MaxQuant software.

Transcriptomics

- For each sample, ≤ 20 mg tissue was homogenized, and RNA was extracted using Qiagen's miRNeasy Mini Kit.
- Library was prepared using TruSeq Stranded Total mRNA kit (Illumina).
- Sequencing was performed using Illumina's NovaSeq 6000 sequencer with 100bp paired end reads, finally obtaining ~ 70 million reads per sample.

qPCR Validation

- qPCR validation analysis was done with the same RNA samples used for RNA-Seq.
- cDNA synthesis was performed with oligo-dT and random primers.
- DPEP1 expression levels were measured in triplicates using a SYBR® Green-based assay on a CFX96 Real-Time Cycler (BioRad). EEF1A1, GAPDH and RPLP0 were used as reference genes.

Bioinformatics analysis

- All analysis was undertaken using R studio, a software environment for statistical computing and graphics, using custom built scripts.

RESULTS

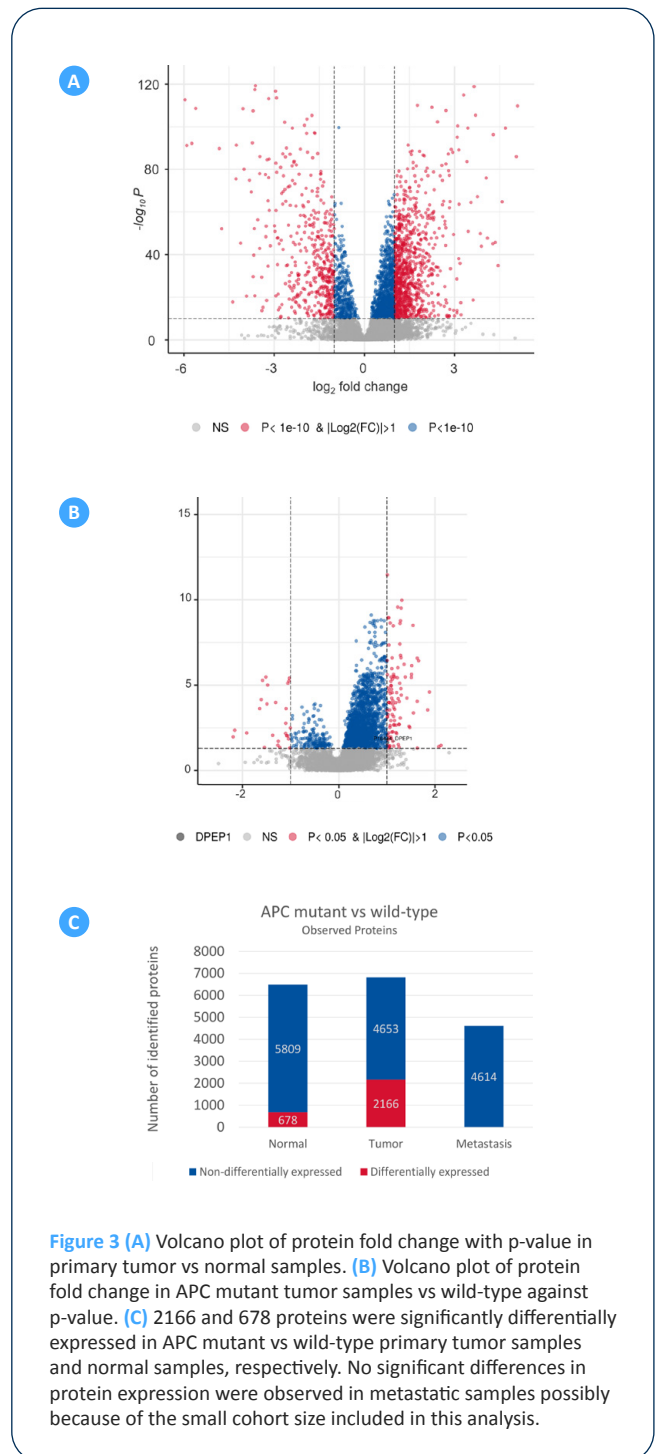
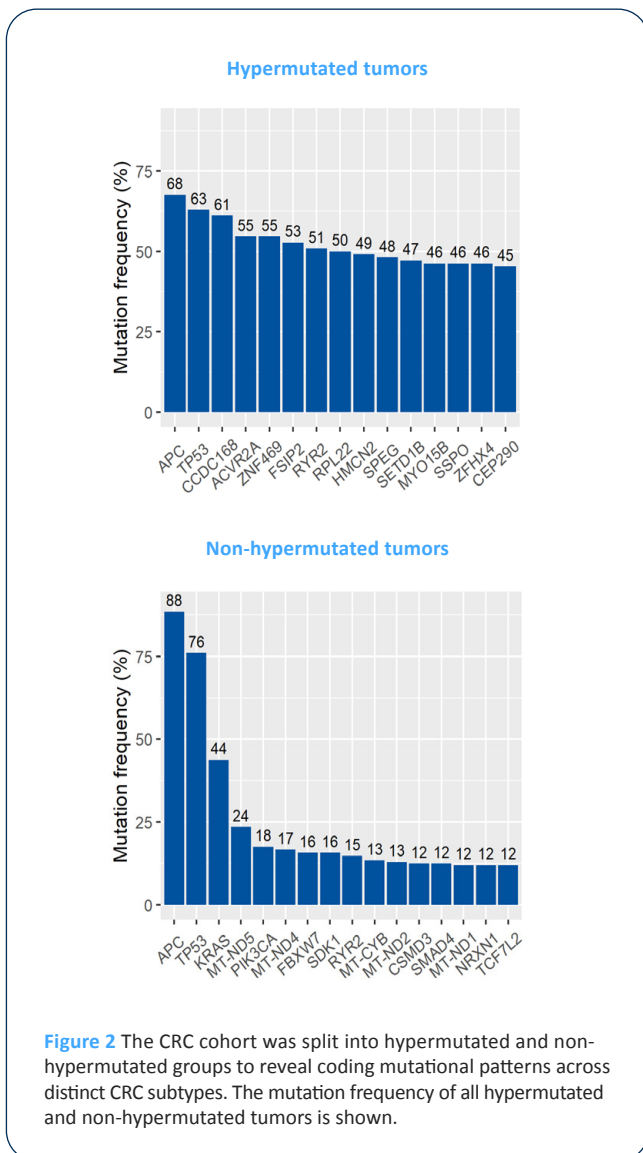
Genomics

The selected CRC cohort was characterized at the genomic level, identifying many known and novel protein coding somatic variants. We found 307 of 388 primary tumor cases and 96 of 112 metastatic tumor cases to be APC mutant and so decided to use APC as an example to demonstrate a proteogenomics workflow for target discovery (Figure 2).

Proteomics

The CRC cohort was characterized by proteomics analysis and hundreds of dysregulated proteins were identified in the tumor samples as compared to the matching normal samples (Figure 3A), encompassing many known and novel proteins underlying CRC biology. When the sub-cohorts of

patients defined by the mutational status of APC were compared, we observed the impact of the genetic background on protein expression (Figure 3B). We found over three-fold more differentially expressed proteins in tumor samples that were dependent on genetic background, in comparison to the corresponding normal samples (Figure 3C).



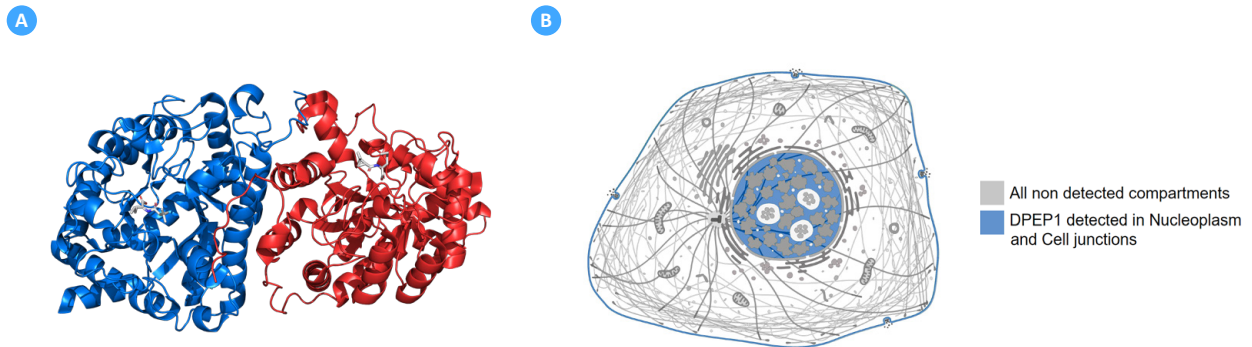


Figure 4 (A) DPEP1 dimer structure (red, blue; PDBID:1ITU) with bound cilastatin molecule (grey) [1]. (B) DPEP1 localizes in nucleoplasm and cell junctions (blue; modified from Human Genome Atlas).

Target Discovery and Therapeutic Potential

Next, we used multiple open source resources (e.g. CanSar, OpenTarget) to systematically prioritize APC dependent differentially expressed proteins for their actionability using either antibody or small molecule-based approaches.

Example of APC Mutant-Dependent Expression, DPEP1

One of the top targets identified was Dipeptidase 1 (DPEP1). DPEP1 is involved in metabolism of glutathione by dipeptide hydrolysis. There is an approved drug for DPEP1, cilastatin, which is used to treat bacterial infections (Figure 4A). Cilastatin

is being independently tested in CRC [2]. DPEP1 over-expression is reported to cause significant increase in colon cancer cell adhesion, invasion and metastasis (Figure 4B) [2].

Using our bioinformatics platform nRavel[®], we observed DPEP1 protein levels in APC mutant tumor samples were higher as compared to wild-type (Figure 5A), a trend replicated in RNA-Seq data (Figure 5B). In order to experimentally validate DPEP1 expression data, we performed qPCR on both tumor and normal samples. DPEP1 fold changes were highly correlated between the RNA-Seq and qPCR results (Figure 5C).

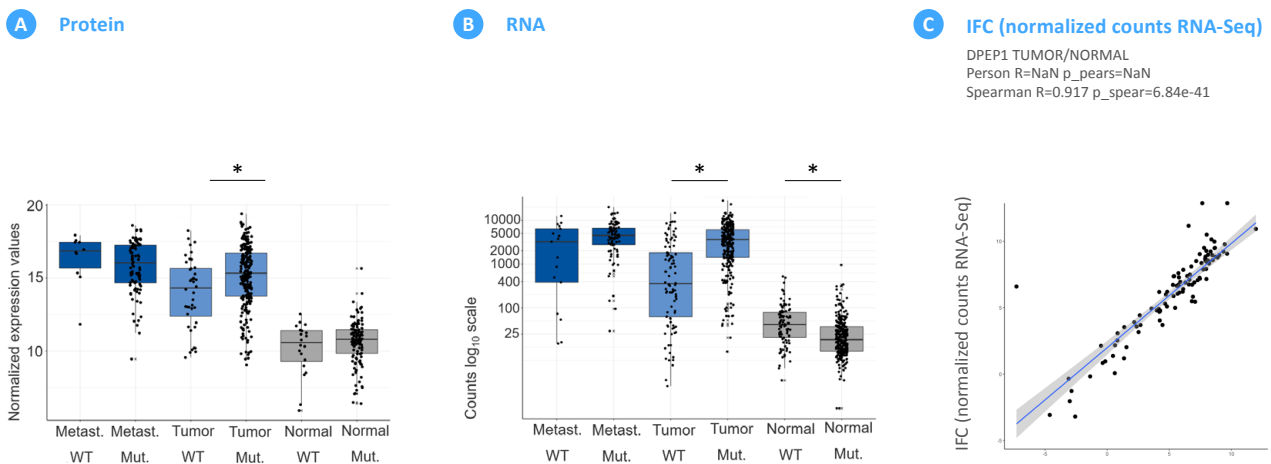
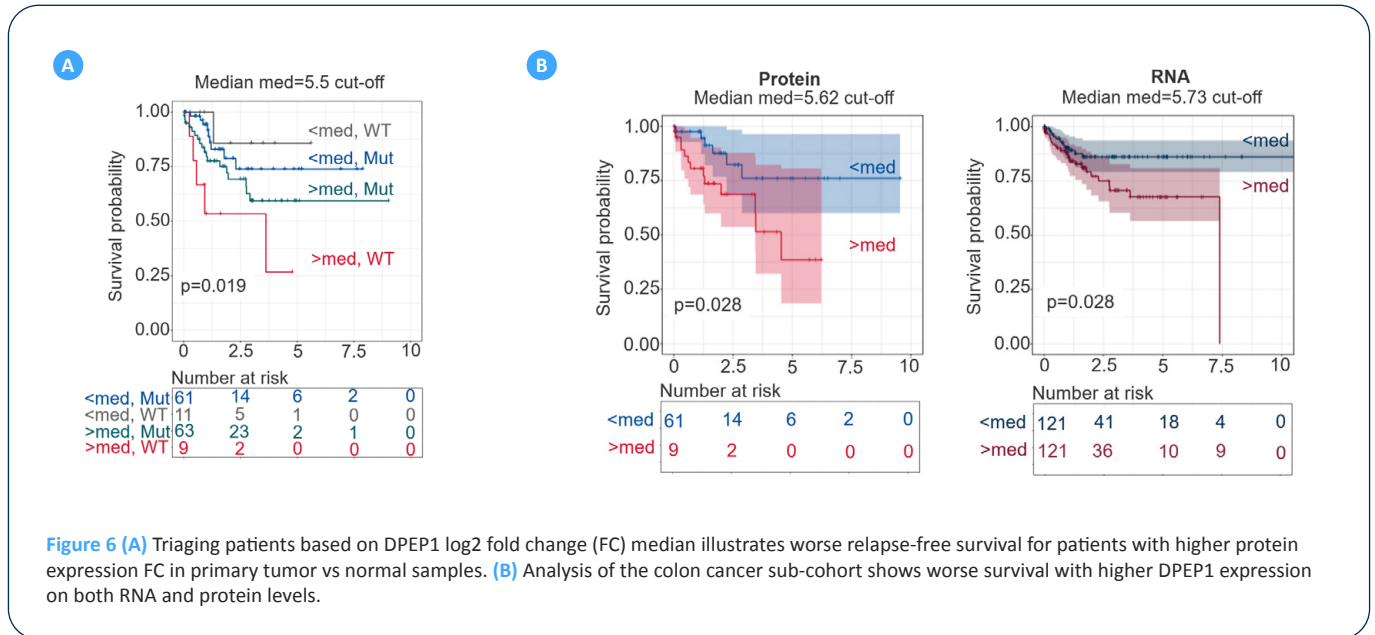


Figure 5 DPEP1 is differentially expressed in APC mutant vs wild-type in primary tumor samples, but not in normal samples, in (A) protein expression data, and (B) RNA expression data. (C) High correlation of DPEP1 expression levels in RNA-Seq compared to qPCR.

Relapse-Free Survival

Finally, we assessed if DPEP1 expression levels could be a prognostic factor for CRC relapse-free survival. We performed survival analysis on protein expression in both, the entire colon cohort and the sub-cohort dependent on APC genetic

background (Figure 6A). Patients with higher DPEP1 levels showed worse prognosis for both APC mutant and wild-type sub-cohorts. DPEP1 expression also had prognostic value at the RNA level when it was assessed for the colon cohort (Figure 6B).



CONCLUSIONS

We have demonstrated how nRavel®'s proteogenomics workflow successfully identified potential drug targets for colorectal cancer. Availability of high-quality multi-omics data enabled the in-depth characterization of the CRC cohort used in this study. The integration of genomics, transcriptomics, proteomics, and clinical data makes nRavel® a powerful tool to advance target and biomarker discovery. As DPEP1 has pre-existing therapeutics that are being evaluated for CRC, it was not taken forward into experimental functional validation. A prioritized shortlist of proprietary therapeutic targets based on this analysis was moved forward into experimental validation.

REFERENCES

- [1] Nitanai et al.; Crystal structure of human renal dipeptidase involved in beta-lactam hydrolysis., *Journal of Molecular Biology*, 2002, Aug 1; 321(2):177-184, [http://doi.org/10.1016/s0022-2836\(02\)00632-0](http://doi.org/10.1016/s0022-2836(02)00632-0).
- [2] Park et al.; Dehydropeptidase 1 promotes metastasis through regulation of E-cadherin expression in colon cancer., *Oncotarget*, 2016, Feb 23;7(8):9501-12, <https://doi.org/10.18632/oncotarget.7033>.

Contact Information

Indivumed GmbH · Falkenried 88, Bldg. D · D-20251 Hamburg · Germany

PHONE: +49 40 4133 83 0

EMAIL: application-notes@indivumed.com

www.indivumed-therapeutics.com

© Copyright 2019–2023 by Indivumed GmbH. All Rights Reserved. No copying and/or distribution of copies allowed without prior approval by Indivumed GmbH. This material contains Confidential Information.