# Proteomic study of 2,002 tumors identifies 11 pan-cancer molecular subtypes across 14 types of cancer
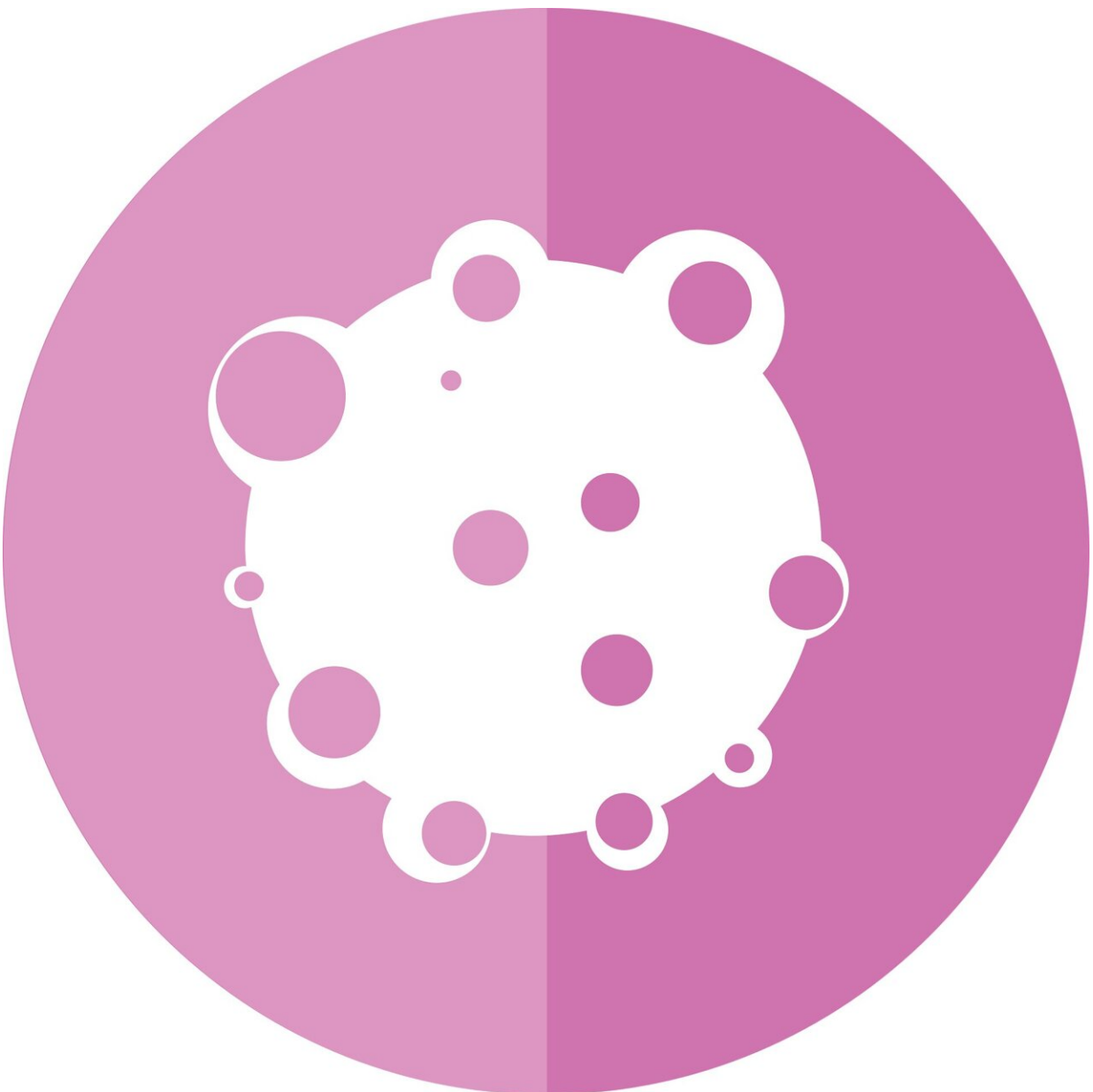
June 10 2022

A new study that analyzed protein levels in 2,002 primary tumors from 14 tissue-based cancer types identified 11 distinct molecular subtypes, providing systematic knowledge that greatly expands a searchable online database that has become a go-to platform for cancer data analysis by users worldwide.

The University of Alabama at Birmingham Cancer Data analysis portal, or UALCAN, was developed and released to public use in 2017 as a user-friendly portal for pan-cancer omics data analysis, including transcriptomics, epigenetics and proteomics. UALCAN has had nearly 920,000 site visits from researchers in more than 100 countries, and it has been cited more than 2,750 times.

"UALCAN is an effort to distribute comprehensive cancer data to researchers and clinicians in a user-friendly format to make discoveries and find needles in the haystack," said Sooryanarayana Varambally, Ph.D., professor in the UAB Department of Pathology Division of Molecular and Cellular Pathology and director of UAB's Translational Oncologic Pathology Research program. "Cancer detection, diagnosis, treatment, cure and research need a global team effort, and making sense of the huge amount of data involved needs a way to analyze and interpret these data."

Cancer is a complex disease, and its initiation, progression and metastasis, the spread to distant organs, involves dynamic molecular changes in each type of cancer. Individual cancer patients show variations apart from some of the common genomic events.

In the new study, Varambally worked with longtime collaborator Chad Creighton, Ph.D., Baylor College of Medicine, Houston, Texas. Creighton led the proteomic study, published in *Nature Communications,* "Proteogenomic characterization of 2002 [human cancers](#) reveals pan-cancer molecular subtypes and associated pathways." This extends two early proteomics studies published in 2019 and 2021.

Previously the team performed RNA transcripts analysis, providing the data to researchers through UALCAN, to determine which pathways the myriad forms of cancer use to aid growth, spread and aggressiveness. With this recent study, the team performed and incorporated large-scale proteomics analysis. The data and results provide new ideas for further research and possible therapeutic interventions.

A proteome is the complement of proteins expressed in a cell or tissue, and these can be measured quantitatively through recent technological advances in mass-spectrometry. In cells, DNA makes mRNA, and mRNA makes protein, processes known as the central dogma of molecular biology. Proteins are major functional moieties of cells, crucial in cell metabolism, structure, growth, signaling and movement.

The cancer types represented in the UALCAN proteomic dataset include breast, colorectal, gastric, glioblastoma, head and neck, liver, lung adenocarcinoma, lung squamous, ovarian, pancreatic, pediatric brain, prostate, renal, and uterine cancers. The number of tumors in each cancer type in the study ranged from 76 to 230, with an average of 143. Intriguingly, the pan-cancer, proteome-based subtypes the current study found cut across tumor lineages.

The compendium proteomic dataset came from 17 individual studies. Corresponding multi-omics data were available for most of these tumors, including mRNA levels, DNA somatic small mutations and insertions/deletions, and DNA somatic copy number alterations.

In general, the researchers found the protein expression of genes across tumors broadly correlated with corresponding mRNA levels or copy number alterations. However, there were some notable exceptions.

They identified 11 distinct proteome-based pan-cancer subtypes—named s1 through s11—that can provide insights into the deregulated pathways and processes in tumors that make them cancerous. Each subtype spanned multiple tissue-based cancer types, though subtype s11 was specific to brain tumors, spanning glioblastomas and pediatric brain tumors.

Each subtype expressed specific gene categories, some seen before in a previous, less comprehensive proteomic study. Three subtypes showed new gene categories: subtype s7 with "axon guidance" and "frizzled binding" genes, subtype s10 with "DNA repair" and "chromatin organization" genes, and subtype s11 with "synapse," "dendrite" and "axon" genes.

At the DNA level, the study detailed differences among the proteome-based subtypes in overall copy number alterations of genes, and somatic mutations in subtypes associated with higher pathway activity, as inferred by proteome or transcriptome data.

"Our study results provide a framework for understanding the molecular landscape of cancers at the proteome level to integrate and compare the data with other molecular correlates of cancers," Varambally said. "The associated datasets and gene-level associations represent a resource for the research community, including helping to identify gene candidates for functional studies and further develop candidates as diagnostic markers or therapeutic targets for specific subset of cancers.

"Furthermore, this study reinforces the notion that cancers should be comprehensively surveyed at the protein level, though expression

profiling on tumors has historically been mostly limited to the RNA transcript level. Many of the analyses in this ever-evolving cancer data analysis platform are based on user or expert requests, and the team is indebted to the support and encouragement from the researchers who use this platform to make discoveries that make a difference in cancer research."

Some of the large datasets for the UAB site are generated by consortiums like The Cancer Genome Atlas, or TCGA, and the Clinical Proteomic Tumor Analysis Consortium, or CPTAC, of the National Cancer Institute.

Precision targeting of cancer requires the identification of individual or subclass-specific genomic and molecular alterations. To help cancer researchers perform various data analyses for better understanding of these large datasets, Darshan Shimoga Chandrashekar, Ph.D., led the development of the UALCAN portal under the mentorship of Varambally. Updates to this continuously evolving portal were recently published in Neoplasia.

The UALCAN initiative and its continuous development involve contributions from a team of experts including bioinformaticians, computer scientists, statisticians, cancer biologists, pathologists and oncologists. "It is a team science approach to enable the global cancer research team to tackle cancer," Varambally said.