Proteomics and cancer

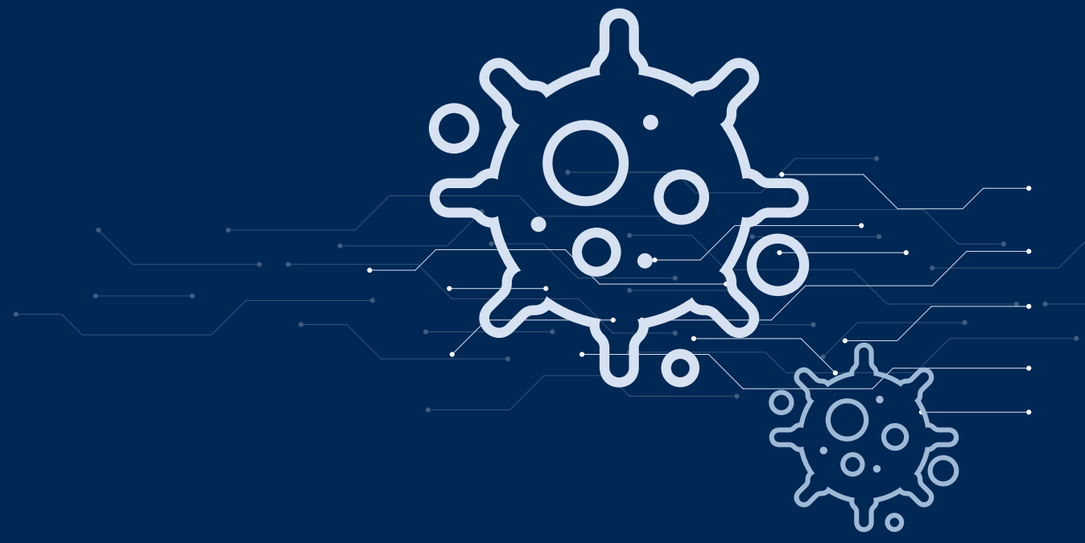
Demonstrating the potential of proteomics to complement genomics and advance cancer care

Use this QR code to share this eBook with friends and colleagues.

NAUTILUS™
BIOTECHNOLOGY

Table of Contents

Proteomics and cancer



INTRODUCTION

Demonstrating the potential of proteomics to complement genomics and advance cancer care

The widespread availability and affordability of genome sequencing has enabled researchers to thoroughly characterize the genetics of a wide array of cancers. They have identified many cancer-driving mutations and achieved inspiring successes in the field of precision medicine including:

- Antibodies targeting HER2 in breast cancer
- Small molecules targeting EGFR in lung cancer
- Small molecules targeting the BCR-ABL fusion protein in certain types of leukemia

These drugs can achieve high response rates in certain subsets of patients and have saved lives. Nonetheless, many cancer patients still don't have effective precision medicines or do not achieve durable responses (see Rodriguez et al. for an overview of successes and gaps in precision medicines for cancer).

Indeed, results from the recent NCI-MATCH trial, which paired patients with specific genetic alterations to targeted therapies, show an average response rate of about 10%. The patients in this study were heavily pre-treated and the finding that some did nonetheless respond is certainly promising, but the fact remains that simply knowing a patient has a driver mutation is not enough to direct consistently effective care.

Clear evidence for the power of proteomics in cancer research

We must look beyond the genome to understand the many ways biology goes awry in cancer. Given that proteins are the molecular machines that control most biological functions, and most drugs target proteins, it makes intuitive sense to look to the proteome to better understand the biology of cancer and find the appropriate pathways to drug in individual patients.

In this eBook, we distill results from a variety of recent proteomic investigations of cancer. These studies are often conducted through the Clinical Proteomic Tumor Analysis Consortium and are monumental efforts involving many skilled researchers. Their findings make it clear that proteomics is well-positioned to identify vulnerabilities in cancer and direct the development of novel therapeutics and diagnostics.

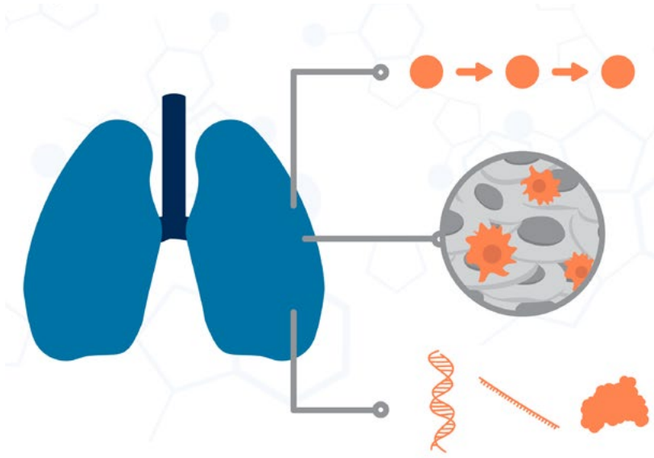
Designing a next-generation proteomics platform to accelerate cancer research and clinical development

We are developing the Nautilus™ Proteome Analysis Platform with the aim of making studies like those distilled here more accessible to more researchers. As we highlight throughout the eBook, there is exceptional potential for next-generation proteomics platforms to accelerate research, uncover the roles of new proteins and proteoforms in cancer, and facilitate the design and testing of novel therapeutics.

With the help of next-generation proteomics platforms, we are hopeful that researchers will make great strides against cancer in all its forms. We believe this will lead to vastly improved livelihoods for the many people who suffer from cancer around the world and invite you to dive into this eBook to be convinced of the power of proteomics in cancer research.

Insights from a multiomic analysis of lung adenocarcinoma

The American Cancer Society estimates that over 100,000 people will die from lung cancer in the US in 2023. According to the CDC, 90% of these deaths will be linked to smoking, but many will not. As with all cancers, lung cancer is a mixture of different diseases and parsing the molecular causes behind lung cancer cases is critical to diagnosing and treating them effectively.



- Identify active and druggable biological pathways
- Discover determinants of tumor invasion by immune cells
- Elucidate connections between omics levels

Toward this end, Gillette et al. recently conducted a multiomic analysis of over 100 lung adenocarcinoma samples isolated from tumors and matching adjacent tissues. They combined genomics, transcriptomics, proteomics, phosphoproteomics, and epigenomics to identify mutated genes, altered methylation, differential RNA and protein expression, as well as proteoforms associated with lung cancer. With this extensive dataset, they were able to:

- Divide adenocarcinoma into multiomics-defined subtypes
- Infer the activity of mutant kinases on various substrates
- Discover *cis* and *trans* effects from copy number alterations
- Assess the effects of cancer-associated mutations on various biological pathways
- Associate particular mutations with immune evasion by tumor cells
- Identify potential drug targets and biomarkers
- Characterize pathways associated with adenocarcinoma in smokers compared to non-smokers

This work resulted in a treasure trove of data that can be leveraged for many future studies of lung cancer biology, biomarkers, and treatments. While this work made use of mass spectrometry-based proteomics, it provides an excellent example of the highly actionable insights proteomics, in general, can provide.

Below we cover a small fraction of these insights in more detail. We also highlight how we aim to make discoveries like these more accessible to people in many different research fields through the Nautilus Proteome Analysis Platform. It is our goal to create a next-generation proteomics platform that makes it easier to generate data like that discussed here. We hope researchers can use our platform to advance that data to new realms of scientific discovery as well as the clinic.

Inferring the activity of kinases and identifying druggable pathways

The genomics work carried out by Gillette et al. identified fusions between kinases and other proteins in lung adenocarcinoma. Some fusions inactivated the kinases while others activated them and some have been associated with lung cancer in the past. Examining just mRNA and protein expression would not capture how the fusions affect phosphorylation, so the researchers used phosphoproteomics to identify altered phosphorylation events indicative of kinase activity.

For example, fusions of the ALK kinase resulted in highly elevated phosphorylation of a variety of downstream targets. Not all such phosphorylation events necessarily drive cancer progression or will be druggable, but they point to promising avenues for further research that may lead to the identification of new drug targets.

Similar efforts identified copy number alterations as well as altered methylation patterns that impacted the expression of proteins both in *cis* and in *trans*. Some such mutations were in known cancer-associated proteins and, with their multiomic analysis, the researchers could trace the effects of these alterations to potentially druggable signaling pathways at the proteome and phosphoproteome levels.

Assessing immune activity in lung adenocarcinoma with multiomics

Tumors that have been invaded by immune cells are often associated with better cancer prognoses than those without invasion. The immune cells in these “immune hot” tumors have the potential to kill cancer cells and help prevent disease progression.

Gillette et al. used RNA-seq to group tumors into immune hot and immune cold subtypes. Then they used multiomics to find associations between immune invasion and particular mutations and potentially druggable signaling pathways. For instance, mutations in the STK11 protein were associated with low immune cell counts and high levels of proteins associated with an immune process called neutrophil degranulation. It is thus possible that manipulating this process in *STK11* mutant tumors may increase the ability of immune cells to infiltrate tumors. As tumors with high levels of immune cell infiltration are not necessarily easily killed by those immune cells, similar analyses can help researchers identify pathways to manipulate for more effective immune cell activity even in these “immune hot” tumors.

Using multiomics to identify biomarkers indicative of underlying mutations in lung cancer

With their tumor samples and matched normal adjacent tissues, Gillette et al. could identify proteins upregulated in tumor cells with druggable mutations. They found high differential expression of proteins in *TP53*, *EGFR*, *KRAS*, and *STK11* mutant tumors. Each of these mutant tumors had different sets of proteins that were differentially expressed and could be used as biomarkers. It was important to perform biomarker analysis at the protein level as opposed to the transcriptome level because mRNA expression did not correlate well with protein expression and may not predict the presence or activity of a protein.

Advancing lung cancer research with next-generation proteomics

The ground-breaking work of mass spectrometry experts has made it clear that proteomics is critically important to truly understand biological processes. Nonetheless, we need more accessible and higher throughput next-generation proteomics platforms that can quickly assess the proteomes of many samples and sample types across many more labs. This will make it possible to efficiently advance the work discussed here and will enable researchers to revolutionize our understanding of and ability to treat not just lung adenocarcinoma, but all types of cancer. It is Nautilus’ goal to make such studies possible through the Nautilus™ Proteome Analysis Platform. Learn more about our revolutionary platform here.

Proteomic analyses can reveal new breast cancer subtypes

In the U.S. alone, more than 260,000 people are diagnosed with and more than 40,000 people die of breast cancer every year according to the Centers for Disease Control and Prevention. Finding new ways to assess and treat the disease could potentially save thousands of lives every year. Toward this end, a recent proteomic analysis of breast cancer samples has revealed new subtypes of this common cancer. This work provides researchers with a more precise understanding of the disease that can hopefully lead to more effective diagnostics and treatments.

Protein Cluster-1
Fatty acid metabolism, catabolic, and oxidation-reduction associated processes

Protein Cluster-2
Stromal and extracellular matrix, also elevated DNA replication and repair functions

Protein Cluster-3
Immune response, including transporter proteins associated with antigen processing

Protein Cluster-4
Stromal and extracellular matrix, blood coagulation, humoral immune response, and hormone receptor binding

Published in *Nature Communications* in 2022, the work was conducted by a team of researchers from the Morin Lab at the University of British Columbia. They used proteomics to analyze 300 breast cancer specimens. Their analysis revealed the specimens could be grouped into distinct cancer subtypes based on their proteomes. Using information from a biobank, the researchers were able to see that patients with certain proteome-defined cancer subtypes had similar treatment outcomes.

Researchers can use these different breast cancer subtypes to help reveal what makes certain cancers deadly and why some patients respond to treatments. Proteomic analyses like this one demonstrate the mountain of actionable information in the proteome and showcase the value of bringing that information to light.

Finding more predictive breast cancer subtypes

Breast cancer arises when cells in the breast proliferate uncontrollably. Each specific instance of breast cancer can look different, depending on things like a person's genetics, and what mutations led to the cancer. For example, one way researchers classify breast cancers is a test known as the Prediction Analysis of Microarray 50 (PAM50) that looks at RNA expression from 50 genes known to be associated with the disease.

Knowing which breast cancer subtype a patient has helps show doctors which treatments may be best and also helps them deliver a prognosis. Subtypes based on RNA aren't perfect, though. Gene expression at the RNA level does not always correlate well with protein expression, and there are several processes, like mRNA splicing, as well as post-translational modifications that can cause proteins to look different from the genes that encode them. Ultimately, looking directly at proteins instead of transcripts may be a better way to differentiate one cancer from another.

Using proteomic analysis to understand breast cancer

For a better look at the diversity of breast cancer, the researchers turned to proteomic analysis. Using mass spectrometry, they analyzed the proteins in each breast cancer specimen, ending up with a list of 4,214 proteins that they were able to quantify in every sample.

The researchers' proteomic analysis enabled them to subdivide the samples into four unique groups that overlapped with canonical subtypes, but didn't mirror them. The researchers also used protein-level data to identify biological pathways associated with different prognoses. For example, they found that proteins indicative of a strong immune response were correlated with better odds of survival.

Looking specifically at triple-negative breast cancer, a variant of breast cancer that tends to grow faster and is harder to treat, the researchers were able to identify 85 proteins associated with better odds of avoiding relapse.

In addition to reinforcing the value of proteins for studying cancer, this work could point to new targets for cancer therapeutics, as well as identify new protein biomarkers that show how cancer therapies are proceeding.

Proteomics tools for cancer

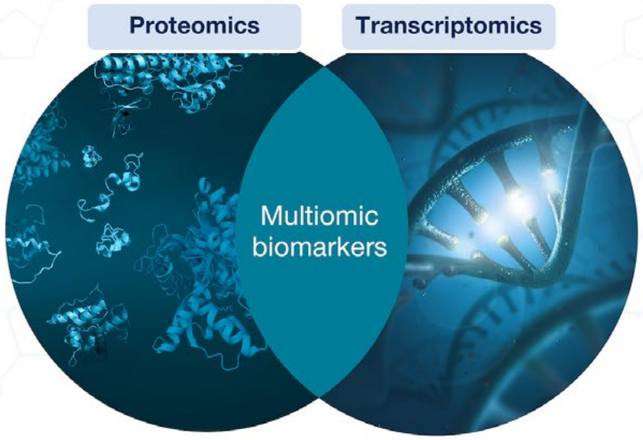
Scientists can use proteomic analyses to study the proteins that both aggravate and hinder cancer. At the moment, not every protein involved in cancer is visible to scientists. Current proteomics tools often miss low-abundance proteins, and don't reveal a sample's entire proteome.

With better proteomics tools, researchers may be able to confidently study more than the 4,214 proteins found in this study. That could reveal more than the 1,054 proteins used to subtype specimens here and may lead to better stratification of breast cancer types and risk. More accessible and rapid throughput technologies will help researchers efficiently perform similar studies across a wide range of cancers and other diseases.

New insights like these will hopefully be possible soon, as tools like the Nautilus™ Proteome Analysis Platform are expected to give scientists unprecedented views into the full proteome of cancer cells and other biological specimens.

Multiomic study highlights the value of the proteome for prostate cancer prognoses

In the United States, one in every eight men will be diagnosed with prostate cancer at some point in their life, according to the American Cancer Society. That number is even higher for Black men. Regular screenings can help diagnose the disease early and save lives, but some cases can be more aggressive than others. Additionally, existing cancer drugs don't work for every patient.



Combining information from proteomics, transcriptomics, and more can reveal biomarkers that may be better predictors of patient outcomes in prostate cancer.

Genetic alterations significantly impact prostate cancer risk and prognosis but are far from the end of the story. It's less clear how differences at the epigenomic, transcriptomic, and proteomic levels affect cancer outcomes. Studies in various types of cancer and other model systems have shown that genomic changes may not be strongly linked to changes at other levels. Thus, more effective biomarkers for prostate cancer progression could come from studying other omes or linking information from multiple omes.

In a study published in *Cancer Cell*, Sinha et al. applied genetic, epigenetic, transcriptomic, and proteomic profiling to see how changes at the genetic level flow through to other biology in prostate cancer. They found that genetic changes are often poorly linked to changes at the level of the transcriptome and proteome, and that transcriptomic and proteomic changes frequently do not align. Further, multiomic signals yielded significantly better biomarkers and predictions of patient outcomes than any single analysis alone.

Multiomic analysis reveals proteomic and transcriptomic biomarkers can complement one another

Sinha et al. performed a multiomic analysis of 76 samples from patients with localized, intermediate-risk prostate cancer who'd been treated with a prostatectomy. They analyzed their genomes, epigenomes, transcriptomes, and proteomes, and assessed how these datasets aligned with each patient's prognosis and treatment outcome.

Based on their data, the researchers grouped patients into four genetic subtypes and five proteomic subtypes. These subtypes aligned poorly with each other, indicating mutated genes weren't always causing changes to their associated proteins. That's a sign the proteome contains information about prostate cancer the genome doesn't, and studying it could reveal new prostate cancer insights. Indeed, several groups of proteins were associated with clinical phenotypes. For example, a group of 421 proteins correlated with percent genome altered (PGA), a biomarker of more aggressive prostate cancer, while a different group of eight proteins was associated with tumor size.

The research also highlighted the significance of fusion proteins containing the ETS-domain (this domain was originally named after the "**E t**wenty-**s**ix" oncogene). These gene fusions are common in prostate cancer patients and are thought to occur early in tumor development. With their multiomic analysis, the researchers were able to see that ETS fusions typically have stronger effects on the proteome than the transcriptome, making proteomic analysis a much more effective tool for assessing their impact. For example, ETS fusions left lysyl oxidase gene expression little changed at the RNA level but caused a 21,031 fold increase in protein abundance.

The researchers also used an information content analysis to see how genomic changes flowed through to proteomic changes and found significant variations between proteins. For example, less than 10 percent of PTEN abundance, but around 60 percent of NDRG3 abundance, was explained by genetic, epigenetic, and transcriptomic changes. Altogether, this data highlights the complexity of the relationship between genomic and proteomic changes and indicates the relationship is likely to be unique for each gene.

Combining prostate cancer biomarkers for improved prognoses

DNA and RNA-based biomarkers are commonly used for prostate cancer but it's currently difficult to adequately diagnose various kinds of prostate tumors. As a result, some patients receive treatment they don't need, while others may not receive it quickly enough.

To find better biomarkers, Sinha et al. generated 10 million sets of 100 randomly chosen genes and assessed the performance of these sets as biomarkers on genetic, epigenetic, transcriptomic, and proteomic levels. The researchers found RNA and protein biomarkers derived from these gene sets performed better than other types of biomarkers. Furthermore, pairing up sets of biomarkers, they found that combined methylation-protein biomarkers yielded the best performance overall. This work sets the stage for the development of novel biomarkers and highlights the benefits of leveraging multiomic data.

.

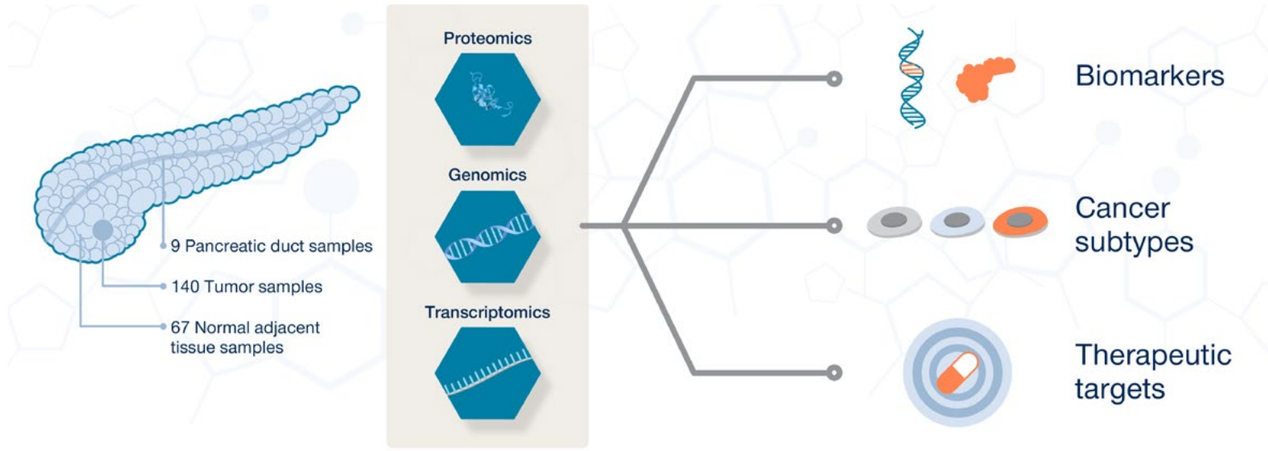
An important role for next-generation proteomics: developing and applying biomarkers for prostate cancer

Today, technologies for proteomics cannot fully analyze a sample, meaning potential biomarkers could be missed. Additionally, high-throughput clinical tests based on proteomics are not yet part of routine practice, making it challenging to utilize new biomarkers clinically.

Next-generation proteomics technologies, like the Nautilus™ Proteome Analysis Platform could begin to change that. The Nautilus Platform is designed to be more sensitive, have a higher dynamic range, and higher throughput than current technologies. It may make analyzing the full proteome possible for more scientists. With such platforms, researchers may discover new multiomic biomarkers like those described above. Doctors may one day use next-generation platforms to apply these biomarkers in the clinic and improve patient lives.

Identifying potential biomarkers and therapeutic targets in pancreatic ductal adenocarcinoma

Pancreatic ductal adenocarcinoma (PDAC) is projected to become the second leading cause of cancer deaths by 2030 and is currently the third leading cause of cancer deaths in the US. The disease has a five-year survival rate below 10% and has been difficult to treat because of a lack of early symptoms, reliable screening methods, and early detection tools.



Past genomic studies on PDAC have identified somatic mutations in genes such as *KRAS*, *TP53*, *CDKN2A*, and *SMAD4*. *KRAS* activating mutations are the most prevalent genetic alteration in PDAC. However, despite recent successes with the G12C *KRAS* mutation, *KRAS* is generally considered difficult to target. Other studies have identified tumor-specific therapeutics that target only a small subset of pancreatic cancers leaving many people without treatment options. While studies that use genomics or transcriptomics can identify signaling pathways involved in PDAC, they alone fall short in fully revealing the connections between genes, transcripts, and proteins.

To overcome these challenges, Cao et al. examined PDAC using a multiomic approach. They identified potential biomarkers and therapeutic targets for PDAC that may have gone undetected in more limited studies of individual omes. More studies like this one will hopefully become possible if next-generation proteomics technologies are broadly available and accessible. We're designing the Nautilus™ Proteome Analysis Platform to become one such game-changing technology and hope it will drive impactful insights into cancer like those discussed below.

Unraveling the proteogenomic landscape of PDAC

Cao et al. analyzed pancreatic tumor samples from patients, paired NATs (normal adjacent tissues), and normal pancreatic duct tissues. These samples were of particularly high quality because they were all frozen in liquid nitrogen within, on average, 20 minutes of collection to ensure post-translational modifications could be analyzed accurately. This is incredibly important in a disease like cancer where disruptions to all sorts of signaling pathways involving various proteoforms could result in different prognoses.

This study amassed an immense amount of data including tens of thousands of transcripts, proteins, phosphosites, and glycopeptides. This is a lot of data, but by leveraging information from previously identified mutations (ex: *KRAS*), the scientists were able to begin homing in on molecular details that only a multiomic approach could capture.

Below, we highlight a few important findings that illustrate the benefits of this multiomic approach

Linking genomic mutations to functional impact

Genome content doesn't necessarily correlate with transcript abundance or protein expression, and it doesn't reveal much about post-translational modifications. By merging several omics techniques together, the researchers were able to gain additional insights that connect the genome to function. For example, they linked certain *KRAS* mutations to the upregulation of specific glycoproteins. These glycoproteins mediate cell migration, invasion, and adhesion. They also appear to protect neoplastic cells from a type of programmed cell death that occurs when cells detach from the extracellular matrix and that plays a role preventing cancer metastasis.

The connection between KRAS and these glycoproteins provides a potential point of therapeutic intervention.

While KRAS itself is difficult to drug, it may be possible to develop monoclonal antibodies against these glycoproteins. In patients with *KRAS* mutations that upregulate these glycoproteins, such monoclonal antibodies could work in conjunction with first-line chemotherapies.

Addressing difficulties in treating PDAC: Immune evasion and biomarkers

One reason it is difficult to treat PDAC with immunotherapies is that immune cells typically have difficulty penetrating pancreatic tumors. Using a combination of transcriptomic and phosphoproteomic approaches, Cao et al. discovered that "immune-cold" tumors had reduced expression of endothelial adhesion proteins and upregulation of a variety of pathways that might make the tumors impermeable to small molecules and immune cells. Such pathways included:

- VEGF – Involved in endothelial cell remodeling during tumorigenesis
- Hypoxia – Involved in endothelial cell remodeling during tumorigenesis
- Glycolysis – Involved in the generation of ATP (most cancer cells rely on glycolysis for proliferation)
- Cell junctions – Involved in regulating the permeability of endothelial cells to small molecules and immune cells

Targeting the above processes pharmacologically may make it easier for immune cells to invade pancreatic tumors and prevent tumor growth.

Another reason that PDAC is difficult to treat is that it is often diagnosed at an advanced stage. Therefore, new biomarkers could enable earlier diagnoses. To work towards this goal, Cao et al. identified 12 secreted proteins that are over 2-fold upregulated in PDAC. They also found thousands of protein phosphorylation sites and N-linked glycosites that were increased in abundance in PDACs.

These may one day make great biomarkers for the early identification of the disease and may therefore result in many saved lives.

The importance of a multiomic approach in understanding cancer

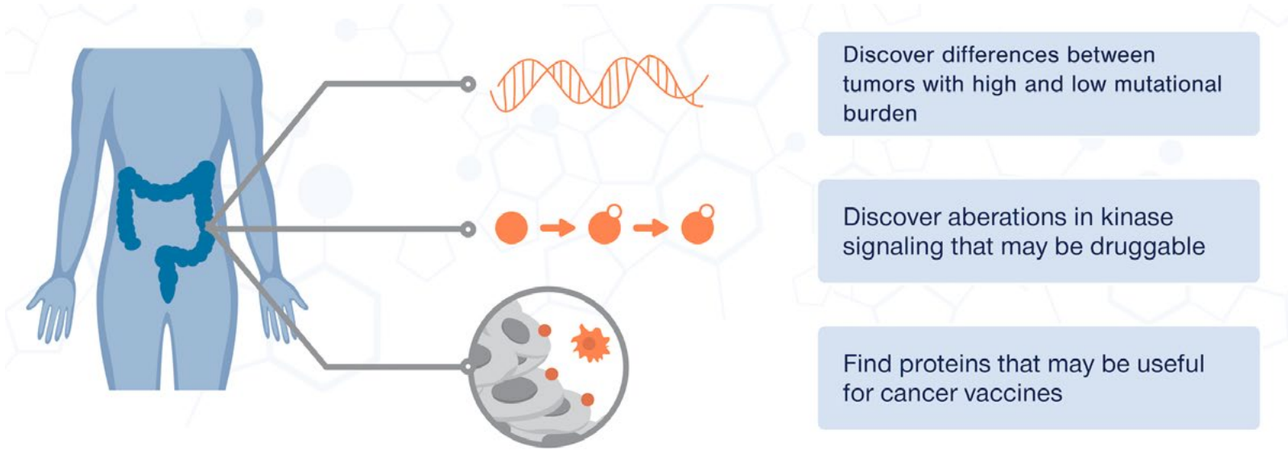
This study highlights the importance of a multiomic approach in understanding disease. Many disease biomarkers and affected pathways cannot be identified based on genomics or transcriptomics alone and this study strongly reinforced that point. With their multiomic approach, the researchers identified new links between genomic alterations and their impact on the proteome, revealed various pathways involved in PDAC immune evasion, and discovered potential biomarkers for the early detection of PDAC.

Next-generation proteomics technologies, such as the Nautilus™ Proteome Analysis Platform, are designed to reveal even more biological links in cancer. They aim to give researchers the ability to rapidly improve their understanding of cancer and other diseases by identifying up to 95% of proteins in a sample over a wide abundance range. The wide accessibility of such technologies will hopefully reveal new biomarkers and drug targets that will vastly improve our ability to diagnose and treat pancreatic cancer and many other diseases.

Discover additional ways next-generation proteomics can fuel biology research in our Applications of proteomics blog posts.

Insights from a multiomic analysis of colorectal cancer

Colon cancer is one of the most deadly cancers in the US causing roughly 8.6% of cancer deaths according to the National Cancer Institute. It's predicted that more than 153,020 people in the US will be diagnosed with this devastating disease in 2023. If caught early, colon cancer can be treated by tumor surgical resection and new targeted therapies for advanced disease are in development. Nonetheless, given the incredible burden of this disease, it's clear that we need a better understanding of its molecular underpinnings.



Researchers associated with the Clinical Proteomic Tumor Analysis Consortium have carried out monumental efforts to characterize a variety of tumors using multiomics analyses. Vasaikar et al. conducted such an analysis on colon cancer tumors in 2019 and compared their genomic, transcriptomic, proteomic, and phosphoproteomic profiles to those of matched adjacent tissues. We cover some of their key findings below and highlight how their work elucidates the biology of various colon cancer subtypes while also identifying points for targeted therapeutic intervention and the development of cancer vaccines. Future work in this area may be accelerated with next-generation proteomics technologies like the Nautilus™ Proteome Analysis Platform.

Characterising colon cancer by genomic mutations and their effects on the transcriptome and proteome

Vasaikar et al. first divided the tumor samples into two groups – those that were hypermutated and those that were not. The hypermutated group generally contained samples with microsatellite instability – a large number of microsatellite length polymorphisms - and a higher number of mutations overall including single nucleotide variants and indels. In many cases, microsatellite polymorphisms likely resulted from mutations in mismatch repair genes which are known to cause microsatellite instability.

Overall, the hypermutated and non-hypermutated groups had frequent mutations in different sets of genes indicating that they had different etiologies. Investigating the frequently mutated genes provided clues as to the mechanisms underlying tumor growth.

Hypermutated colon tumor samples often had mutations in DNA repair enzymes as well as high frequency mutations in 9 additional genes. These genes included *CASP5*, which encodes a protein involved in

programmed cell death and *RNF43*, which encodes a protein that regulates cell growth pathways.

Non-hypermutated tumors made up most samples and had a high frequency of mutations in a different set of genes that included:

- *APC* - a protein that plays a variety of roles in regulating cellular proliferation and is a well-known tumor suppressor.
- *TP53* - also a known tumor suppressor.
- *SOX9* - *SOX9* was particularly interesting because it was both frequently truncated at the gene-level and over-expressed at the protein level. Gene truncation often leads to non-functional proteins and a high level of truncation in tumor cells would usually indicate that a gene encodes a tumor suppressor. Yet, the high levels of SOX9 protein observed here indicate that the truncated protein acts to promote tumor cell proliferation.

Analysis of copy number alterations across colon cancer genomes identified correlations between genomic, transcriptomic, and proteomic alterations. Genes with correlated changes across all these levels were prioritized as drivers of colon cancer and were found to be associated with particular biological pathways. For instance, 6 of the 90 genes prioritized as potential drivers through this multiomic analysis were involved in endocytosis. This process can impact, among other pathways, growth signaling and metastasis and is a promising therapeutic target.

The role of phosphorylation in colon cancer signaling pathways

Kinase signaling pathways can drive tumor growth and the multiomic work here identified some of the kinase pathways at work in colon cancer. For example, phosphoproteomic analysis revealed that the retinoblastoma (RB) protein, which is normally considered a tumor suppressor, was surprisingly upregulated in both gene copy number and in protein expression. Phosphosite analysis revealed that, in many cases, the up-regulated RB protein was more highly phosphorylated in tumors than normal adjacent tissues, and the researchers identified CDK2 as the likely kinase responsible. Further analysis revealed pathways through which phosphorylated RB could both activate tumor cell proliferation and inhibit apoptosis. This work suggests CDK2 inhibition as a potential means of treating colon cancer.

Beyond RB, Vasaikar et al. identified many additional proteins and phosphosites that were either upregulated or downregulated in tumor tissues. While more research is necessary to parse out the many ways these proteins may be involved in disease, some of them had up-regulated expression that was relatively restricted to tumors. This is an exciting finding because cancer vaccines, which prime the immune system to recognize cancer cells, require such tumor specific proteins, and researchers may be able to use these proteins in future colon cancer vaccines.

Colon cancer subtypes based on genomics, transcriptomics, and proteomics

Vasaikar et al. combined genomic, transcriptomic, and proteomic data to delineate three "Unified multiomics subtypes." These subtypes were broadly associated with:

- Microsatellite instability (MSI) – These tumor samples had microsatellite instability, were associated with a previously identified proteomic colorectal cancer subtype (ProS-B, Zhang et al. 2014), and were associated with a previously identified transcriptomic colorectal cancer subtype (CMS1, Guinney et al. 2015).
- Chromosome instability (CIN) – These tumors had high chromosome instability which could lead to alterations in chromosome number, chromosomal rearrangements, and gene copy number changes. They were also associated with a previously identified proteomic colorectal cancer subtype (ProS-E, Zhang et al. 2014) and a previously identified transcriptomic colorectal cancer subtype (CMS2, Guinney et al. 2015).
- Epithelial to mesenchymal transition – These tumor samples had multiomic signatures indicative of the epithelial to mesenchymal transition, a process whereby tumor cells gain the ability to invade other tissues. They were also associated with a previously identified proteomic colorectal cancer subtype (ProS-C, Zhang et al. 2014) and a previously identified transcriptomic colorectal cancer subtype (CMS4, Guinney et al. 2015).

Importantly, tumors from different subtypes are potentially susceptible to different kinds of treatment. For example, tumors in the MSI subgroup have a high level of immune cell infiltration. This indicates that MSI tumors may be susceptible to treatments that enhance the ability of immune cells to kill tumor cells.

Applying findings from the proteogenomic analysis of colon cancer to future studies with next-generation proteomics

The findings from this combined genomic, transcriptomic, and proteomic analysis are highly actionable. They are likely to spur further research both into the roles of the up and down-regulated pathways in colon cancer progression and the potential therapeutic strategies identified. Importantly, this additional work will require higher sample numbers and throughput to validate biological and therapeutic hypotheses. Accordingly, these efforts will be greatly aided by next-generation proteomics technologies, like the Nautilus™ Proteome Analysis Platform, that are designed to be more accessible to more researchers in more labs. We hope researchers can use our platform to advance studies like this one and truly have an impact on devastating diseases like colon cancer.

Proteomics reveals inner workings of head and neck squamous cell carcinoma

Head and neck squamous cell carcinoma (HNSCC) is the sixth-most-common epithelial cancer, and it accounts for around four percent of all cancers in the United States. It includes tumors in the mouth, nose, larynx, throat, and tongue. One common risk factor for HNSCC is infection with the human papillomavirus (HPV); for this reason, HNSCC is broadly divided into HPV-positive and HPV-negative subtypes. The HPV-negative subtype typically has a much poorer prognosis, making it a high priority for new treatments.

Even within the HPV-negative classification, there can be significant differences in HNSCC cancer types involving different mutations, pathways, and more. A better understanding of HNSCC will include not only the genes that predispose people to this cancer, but the transcriptomic and proteomic factors salient to each subtype, allowing much more precise treatments.

In a 2021 paper in *Cancer Cell*, Huang et al. used a multiomic analysis to study HPV-negative HNSCC patient samples. Their analysis uncovered a wealth of information that could make HNSCC treatments more effective. Some of their high-level findings include:

1. The connections between DNA alterations, mRNA expression, protein expression, and pathway activity in HNSCC are not obvious and require multiomic analyses to understand.
2. Protein levels and data on phosphorylation status can identify potential pathways to target with precision medicines, including existing immunotherapy drugs.
3. Combining multiomic data enables researchers to separate cancers into robust subgroups with actionable biomarkers.

HNSCC is a cancer with diverse drivers and biological mechanisms that make it difficult to study and treat. The multiomics research discussed here significantly advances our understanding of how various pathways are altered in this disease and could lead to new precision medicines for patients.

Proteomics applications in HNSCC

To better understand HPV-negative HNSCC, Huang et al. conducted a multiomic analysis encompassing genomics, epigenetics, proteomics, transcriptomics, and phosphoproteomics of 108 samples taken from patients with the cancer. Using mass-spectrometry-based proteomics, they compared the proteomes of tumor cells to the proteomes of matched samples from healthy tissues nearby. This discovery proteomics analysis revealed 3,355 proteins that were significantly increased and 3,163 proteins that were significantly decreased in tumors compared to healthy tissues.

Biomarkers indicative of HNSCC could help researchers better diagnose the disease, identify pathways to target therapeutically, and update prognoses. Looking at the proteins that were increased the most in tumors, the team uncovered 22 potential protein biomarkers for HNSCC, including seven currently targeted by FDA-approved drugs. There were also 162 proteins associated with progression-free survival that could be useful prognostic biomarkers.

Linking genomics and proteomics also helped reveal new insights into the mechanisms behind HNSCC. For example, the researchers looked at two common types of genetic alteration in HNSCC:

- Truncations to the *FAT1* gene
- Amplification of the 11q13.3 region

They found that both types of mutation decrease actin protein levels with variable affects on actin mRNA. Amplification of 11.q13.13 was additionally associated with increased phosphorylation of an actin binding protein. Together, these results indicate that disrupted actin dynamics are an important component of HNSCC.

Their multiomic analysis also provided valuable insights into the regulation of the cell cycle in HNSCC through the cyclin D-CDK4/6-RB pathway. It is often assumed that alterations to upstream genes *CDKN2A* or *CCND1* will ultimately change the phosphorylation status of the RB tumor-suppressor, and thereby impact the cell cycle. However, the researchers found aberrations in these genes did not always impact RB. Thus, one key conclusion is that RB protein status, and not just gene expression is an effective and necessary indicator for CDK4/6-dependent cell-cycle activity. Only by looking at this pathway through a multiomic lens could the researchers accurately connect changes at the gene level to the output of the pathway.

Multiomic insights into HNSCC treatment

Huang et al. also uncovered new insights into why leading treatments for HNSCC fail in some patients. Mutated forms of the EGFR protein can drive runaway cell proliferation and are found in many kinds of cancer. Drugs based on monoclonal antibodies can inhibit EGFR, potentially helping control tumor growth, but the antibodies don't work in every patient. From their proteogenomic analysis, Huang et al. saw that EGFR ligand abundance is predictive of successful monoclonal antibody therapy in HNSCC patents. Levels of *EGFR* gene amplification or overexpression on the other hand, were less predictive, a finding that may lead to better approaches to treating the cancer.

The team also identified three subtypes of HNSCC based on proteomic, transcriptomic, and other data, each characterized by the overexpression of different genes, and potential treatment options.

- Chromosome instability (CIN): characterized by mutations to the *CCND1* and *CDKN2A* genes as well as high CDK4/6 activity. Cyclin-dependent kinase 4 and 6 (CDK4 and CDK6) inhibitors could be more effective in this group.
- Basal: characterized by elevated EGFR ligand expression and high EGFR pathway activity. Monoclonal antibodies targeting EGFR could be more effective in this group.
- Immune: characterized by elevated expression of immune checkpoint proteins. Immunotherapy drugs targeting checkpoints may be more effective.

Next-generation proteomics tools for cancer research

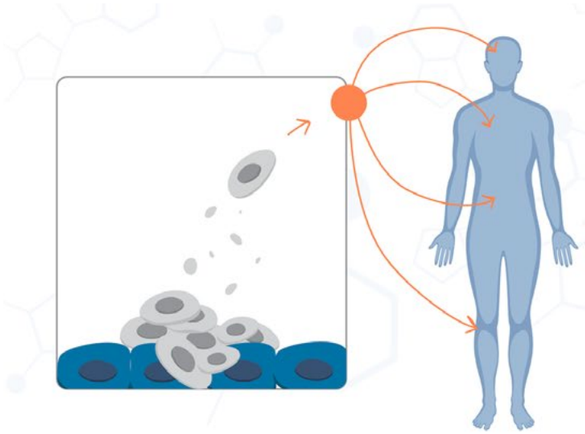
By integrating proteomic insights with those from genomics and other disciplines in a multiomic framework, Huang et al. were able to glean insights into the fundamental biology of HNSCC, as well as identify potential biomarkers and avenues for targeted therapeutics. For instance, proteomics helped identify nuances in the way the EGFR protein behaves in different cancer patients, with clear clinical relevance for monoclonal antibody treatments.

Proteomic insights can be invaluable for studying disease biology, and, as the proteomics revolution progresses, the power of the proteome will only continue to grow. Including proteomics in a multiomic analysis is a powerful way to advance discovery and increase our knowledge of the molecular mechanisms behind various diseases and their subtypes.

Next-generation proteomics tools like the Nautilus Proteome™ Analysis Platform aim to make such analyses routine for researchers studying a wide variety of topics and reveal the vast potential of personalized treatments currently hidden in the depths of the proteome. We are designing our platform to make more studies like this one accessible to more labs and thereby spur the creation of more effective treatments for a wide variety of cancers.

Revealing molecular mechanisms of the epithelial to mesenchymal transition

The epithelial to mesenchymal transition (EMT) is an incredibly important process in both healthy development and cancer. In development, this transition enables cells to migrate to various locations in a developing organism and seed the creation of tissues and organs. Tumor cells that undergo EMT, on the other hand, gain the ability to travel throughout the body and seed the growth of metastases.



The epithelial to mesenchymal transition enables cells to migrate across tissue boundaries during development and during cancer metastasis.

Multiomic analyses empower researchers to uncover the intracate molecular mechanisms underlying this dynamic and consequential process.

There have been many studies into the processes underlying EMT, but omics-scale technologies are enabling researchers to move beyond studies of single proteins or pathways. These technologies empower scientists to get a holistic understanding of the many coordinated pathways underlying this complex and dynamic process and drive the identification of key points where the process can be inhibited.

In this blog post, we cover fantastic work by Paul et al. demonstrating the ways proteomics, transcriptomics, metabolomics, and more can be combined to create detailed, mechanistic multiomic models of complex and incredibly consequential processes like EMT.

Uncovering EMT mechanisms through multiomic analyses

Paul et al. performed extensive studies on the MCF10A human mammary epithelial cell line *in vitro*. They treated the cells with the TGF-beta cytokine over a period of 12 days to induce them to undergo EMT and performed multiomic analyses including, proteomics, phosphoproteomics, transcriptomics, metabolomics, and single cell RNA sequencing at various time points. They even measured a variety of subcellular proteomes. Then, they used bioinformatics techniques to identify relationships between the various omic data sets and progression through EMT.

There was often poor correlation across the omes. For example, the coefficient of variation between mRNA and any of the other omes was no higher than 0.2 across all time points. This highlights that it is essential to do multiomic analyses to get a mechanistic understanding of complex cellular processes like EMT.

Even though the various omics layers were often poorly correlated across time points, the researchers could

cluster data across the time points and partition distinct stages of EMT:

- Epithelial (Day 1)
- Transition state 1 (Days 2-3)
- Transition state 2 (Day 4)
- Mesenchymal (Days 5-12)

The multiomic data showed that distinct molecular pathways were active at these different stages. For example, pathways enriched in the mesenchymal stage included smooth muscle contraction, regulation of TNFR1, NF-kB signaling, transferrin endocytosis and recycling, and insulin receptor signaling. These results show that cells can be manipulated in different ways depending on what stage they are in and what pathways are active in that stage.

Phosphoproteomics and metabolomics serve as important indicators of protein activity

Digging deeper, the researchers used phosphoproteomics to identify highly phosphorylated peptides and metabolomics to discover metabolic pathways active at different stages of EMT. Phosphorylation was often poorly correlated and sometimes anti-correlated with protein abundance but could be used to hypothesize roles for particular kinases and phosphorylation events. For example, the researchers hypothesized that phosphorylation of MICAL3 during EMT regulated its nuclear localization and went on to show that knock down of MICAL3 inhibited TGF-beta induced EMT.

With metabolomics, they discovered the arachidonic acid metabolism pathway was upregulated during EMT. siRNA mediated knock down of a key enzyme in this pathway prevented EMT. This is an exciting finding given that this pathway is not well-studied in EMT.

Single cell RNA sequencing enabled the researchers to identify individual cells at various stages along the EMT trajectory. Such analyses pointed to transcription factors important at steps along the transition and helped the researchers infer intercellular communication pathways active in cells undergoing EMT.

Combining data from these omic layers enabled the researchers to generate a mechanistic model of EMT. This model identified a cascade of pathways and controller proteins driving EMT, many of which had not been identified before. It helped the researchers identify two drugs capable of blocking EMT through the inhibition of the SMO and CAMK-II proteins. Applying these inhibitors in a model of mammary cell invasiveness showed that the drugs effectively decreased invasiveness *in vitro*. These and similar inhibitors may have therapeutic potential in cancer.

Further insights into biological processes enabled by next-generation proteomics

The research efforts described above show just how important it is to move beyond single omic measurements and instead develop multiomic perspectives on biological processes. Had these researchers stuck to one type of measurement, they may have been misled about the activities of various pathways and would not have been able to create such a rich model to identify new drug targets.

We hope to make studies like this one accessible to a wide variety of researchers through easy-to-use, next-generation proteomics technologies. In addition, we plan to enable researchers to achieve more comprehensive assessments of the proteome through a platform that is designed to provide in-depth views of proteins across the wide dynamic range of the proteome.

We hope to give researchers the ability to develop in-depth models of broad biological processes and thereby identify means of manipulating these processes to treat disease and more.

APPLICATIONS OF PROTEOMICS IN CANCER

Assessing the potential of PROTACs

Researchers have discovered many proteins that enable cancer cells to proliferate. Unfortunately, these proteins often have important roles to play in healthy cells. Thus, targeting them with drugs that inhibit their activity or knock down their expression can lead to detrimental side effects for patients Recognizing this issue, researchers have begun developing means of selectively targeting therapeutics or therapeutic activity to cancer cells. One elegant way they do so is through protein degraders such as PROTACs.

PROTACs or “proteolysis-targeting chimeras” recruit ubiquitin ligases to a protein of interest. They consist of three parts:

- A target-binding piece
- A ubiquitin ligase-binding piece
- A linker that ties the two pieces together

When target and ligase are in the same cell, the PROTAC brings them together and the ligase ubiquitinates the target. This marks the target for proteolysis by the ubiquitin-proteasome system ultimately leading to the target’s degradation.

Critically, different ubiquitin ligases are expressed in different cells. Thus, with knowledge of protein expression across cancer cells and healthy cells, researchers can design a PROTAC that degrades a target protein selectively in cancer cells.

There are many protein degraders in clinical trials, and below we cover just one example of PROTAC development. As you’ll learn, next-generation proteomics technologies have a powerful role to play in the development of these promising therapeutics.

Selective PROTACs in action

Khan et al. demonstrated the potential for tumor selective PROTACs in a 2019 publication in Nature Medicine and are currently conducting a clinical trial based on their work. They designed a PROTAC that selectively kills leukemia cells by tethering the anti-apoptotic protein BCL-X_L to the Von Hippel Lindau (VHL) E3 ligase. BCL-X_L has long been a drug target in leukemia, but small molecule drugs against it also inhibit its activity in healthy platelets leading to thrombocytopenia, a dangerously low platelet count.

As the VHL protein is not abundant in platelets, the BCL-X_L-VHL PROTAC has minimal effects on them but effectively kills leukemia cells. The researchers demonstrated this in both cultured cells and mouse xenograft models. The authors go on to show that their PROTAC can be combined with other antitumor drugs and chemotherapy regimens to treat cancer more effectively in mouse xenograft models of leukemia.

Although these findings alone are promising, the authors also leveraged mass spectrometry-based proteomics to show proof of the designed mechanism of action by demonstrating ubiquitination of a specific lysine residue of BCL-X_L . They also leveraged proteomics to show that their PROTAC only reduced BCL-X_L levels and not those of any other proteins. This is a promising finding because it indicates the PROTAC should have minimal off-target effects.

Interestingly, this specificity was not demonstrated *in vitro* where the PROTAC was able to bind to other BCL proteins. These results show that it is crucial to assess functional activity in cells and not just binding of proteins in simple, defined systems. Proteomics techniques help researchers get a comprehensive view of drug activity in live cells where the milieu is complex and dynamic. All in all, these results reveal the incredible potential of PROTACs as cancer therapeutics as well as the need to evaluate their mechanism of action carefully.

PROTAC-based treatments require a comprehensive understanding of the proteome

Khan et al.'s work shows how knowledge of gene expression and the proteome can guide the creative development of new and promising therapeutics. To create an effective PROTAC, these researchers needed to know where their target was expressed, where the ligase was expressed, whether their PROTAC impacted other proteins, and how to design a PROTAC to interact with the appropriate target and ligase in the first place.

Next-generation proteomics technologies can help future researchers rapidly assess all these questions comprehensively and efficiently without the need for expertise in complex mass spectrometry workflows. This will hopefully lead to more effective treatments. Indeed, we hope that researchers will be able to use the Nautilus™ Proteome Analysis Platform in similar ways to impact not just cancer but a wide range of diseases.

For more on the potential for proteomics in drug development, check out our blog post titled, "Using proteomics to improve the drug development process."

Proteomics and cancer

CONCLUSION

Putting the power of proteomics into the hands of more cancer researchers

Despite decades of research and progress, cancer is still a devastating set of diseases that kills millions of people every year. There is, however, good reason to believe that the application of next-generation proteomics can stem the impacts of cancer and bring hope to many cancer patients and their loved ones.

Combining genomics and proteomics for clinical impact

Essential work in genomics has laid the foundation for thinking of cancer as a set of molecular diseases propelled by driver mutations in particular genes. This thinking has led to the development of therapeutics that effectively target driver mutations in a subset of cancer patients. While these therapeutics have saved many lives, we now realize that driver mutations are far from the end of the story. Some driver mutations are difficult to drug, targeting others leads to the rapid development of resistance, and many late-stage cancers have additional mutations or downstream effects that persist once a driver is targeted.

Far from throwing their hands in the air frustrated by this complexity, researchers have sought ways to dissect it and exploit its vulnerabilities through new and often combinatorial therapies. Proteomics technologies have risen to prominence as essential tools in understanding and utilizing cancer's complexity.

Indeed, as you've discovered in this eBook, proteomics tools can:

- Predict the activity of kinases and other enzymes essential for cancer growth
- Identify druggable pathways downstream of driver mutations
- Define cancer subtypes with characteristic biomarkers and functions that can be drugged and used to develop more accurate prognoses
- Determine whether a drug acts through a proposed mechanism or is likely to have off-target effects

All of these proteomic insights are highly actionable and can lead to the development of more effective treatments.

Moreover, the examples here provide just a small peek at the many ways scientists can leverage proteomics in their cancer research. Additional examples of the power of proteomics include:

- Identifying proteins with consistently correlated abundance. Such proteins are likely to interact and aberrations in their correlations may point to disease mechanisms or even a mutation's direct impacts on protein-protein interactions.
- Identifying proteoforms that drive cancer.
- Discovering whether novel or repurposed drugs have their desired effects in clinical trials.

It is clear that proteomics combined with other omics-scale analyses can generate a massive amount of actionable information that has broad-reaching clinical implications. With the power of proteomics thus demonstrated in cancer, it is time for the development of next-generation proteomics platforms that make this information accessible to researchers who don't have expertise in mass spectrometry workflows.

Powering a healthcare revolution with the Nautilus™ Proteome Analysis Platform

We're creating the Nautilus™ Proteome Analysis Platform precisely to make proteomics accessible to more researchers. Our platform is designed to have the sensitivity and dynamic range necessary to cover substantively the entire proteome while still being easy to use thanks to integrated workflows and rapid run times. We aim to make it possible for any researcher to use our platform to perform proteomic analyses on their samples of interest and achieve reproducible insights from easy-to-understand data. We truly believe that, by getting our platform into the hands of researchers in a wide variety of fields, we will revolutionize not just cancer research but healthcare in general.

The power of proteomics in cancer is proven. Now we must get that power to the researchers who know how to use it to develop treatments and save lives.

NAUTILUS™

BIOTECHNOLOGY