Center for Scientific Strategic Initiatives Briefing Book

Table of Contents

Center for Scientific Strategic Initiatives (CSSI) Summary

Table of Contents

1 CSSI Overview

Technological advances in molecular science have the potential to make quantum leaps in cancer research and care. Unfortunately, researchers in cutting-edge fields face many barriers to progress, including obtaining funding, accessing patient data and samples, identifying strategic collaborations, and preparing for regulatory challenges. Many fields of health research encounter similar roadblocks as they expand and mature. These are often related to a lack of infrastructure and prescribed, community-adopted standard operating procedures.

The Center for Strategic Scientific Initiatives (CSSI) lessens the burden of these obstacles by **developing resources and infrastructure** investigators need to traverse innovation bottlenecks. This **accelerates the translation of scientific discoveries** while allowing NCI to pilot new programs before incorporating them into the Institute. CSSI manages programs that promote or provide the following to support nascent or challenging fields of cancer research:

- Funding opportunities
- Shared reagent and/or database resources
- Assistance in the development of standards and protocols
- Partnerships among academic, industry, and government entities

2 CSSI Organization

CSSI's portfolio continually evolves to reflect the changing needs of the cancer research community and the maturity of relevant research fields. The three Offices and one Branch that comprised CSSI as of October 2017 are outlined in the table below and are detailed further in the following sections.

Table 1 The Offices and Branches of CSSI

Office/Branch	Component Programs/Initiatives	Program Goals
Office of the Director (OD)	Innovative Molecular Analysis Technologies (IMAT)	Support technology development research from proof-of-concept to rigorous validation
	Provocative Questions (PQ)	Stimulate research in understudied and difficult-to-address fields
	Pilot Projects	Test-run potential NCI programs and advance CSSI initiatives through early-stage research, training, and technology development

Table 1 The Offices and Branches of CSSI (continued)

Office/Branch	Component Programs/Initiatives	Program Goals
Office of Cancer Clinical Proteomics Research (OCCPR)	Clinical Proteomic Tumor Analysis Consortium (CPTAC)	Develop pipelines and standardized proteogenomics workflows to improve understanding of cancer biology and the mechanisms of drug response/resistance in clinical trials
	International Cancer Proteomics Consortium (ICPC)	Promote international collaboration and data sharing using proteogenomic data to understand cancer biology and predict treatment response
	Applied Proteogenomics OrganizationaL Learning and Outcomes (APOLLO)	Facilitate incorporation of proteogenomics into patient care using the nation's largest health systems at Department of Defense (DoD) and Veterans Affairs (VA)
Knowledge Management and Special Projects Branch (KMSPB)	Research, Condition, and Disease Categorization (RCDC)	Ensure the completeness and accuracy of the NCI-funded portfolio reports to Congress and the public
	NCI Funded Research Portfolio (NFRP)	Manage the NCI's official scientific portfolio reporting site and ensure the completeness, timeliness, and accuracy of scientific funding data for current and past fiscal years.

2.1 CSSI Office of the Director (OD)

The CSSI OD oversees the scientific and programmatic activities of all CSSI entities to effectively carry out the mission of the Center. This responsibility involves facilitating reviews and approvals from the NCI Scientific Program Leaders, NCI Board of Scientific Advisors, and the National Cancer Advisory Board. CSSI OD directly manages two trans-NCI grant programs, the *Innovative Molecular Analysis Technologies (IMAT)* program and the *Provocative Questions (PQ)* initiative.

2.1.1 Innovative Molecular Analysis Technologies (IMAT)

The IMAT program supports the development of potentially transformative technologies in cancer research using a strategy of phased grant support. IMAT is involved throughout the technology development timeline, from proof-of-concept demonstration to rigorous analytical validation. The goal of the program is to equip basic and clinical research communities with novel analytical capabilities through the development of next-generation, cutting-edge technologies.

2.1.2 Provocative Questions (PQ)

The goal of the PQ initiative is to stimulate research in understudied and difficult-to-address areas across the cancer research continuum. Twelve PQs are currently available to be addressed by the cancer research community. Proposals for these PQs are solicited through R01 and R21 mechanisms and more recently through competitive revisions to existing NCI grants. A companion program, *Pediatric PQ*, was launched in 2016.

2.1.3 Strategic Pilots Incubator and Data Coordinating Center

The Strategic Pilots Incubator (SPI) is a unit within Frederick National Laboratory for Cancer Research (FNLCR). The goal of SPI is to provide support to ongoing CSSI pilot projects at FNLCR and to launch projects that address research gap areas related to new technologies, standards, and preanalytical variables. To address a need identified by SPI to manage data from multiple projects, a pilot for a data coordinating center (DCC) was initiated in FY15. All data sets are publicly available, searchable, and accompanied by formatted metadata.

2.1.4 CSSI Pilot Projects

CSSI OD also oversees several pilot projects, many of which are housed within FNLCR. Pilot projects are an important component of CSSI, as they allow NCI to test-run exploratory programs in cancer research while helping CSSI to better execute its mission. The following CSSI-related pilot projects focus on the development of new technologies:

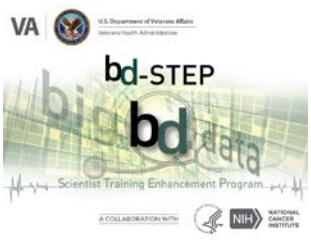
- Analytical Technologies to Objectively Measure Human Performance (ATOM-HP)
- High Content Tissue and Cellular Characterization Laboratories
- Tissue Imaging Laboratory
- High Content Single-Cell Analysis Laboratory

Several CSSI pilot projects include initiatives focused on preanalytical variables and standards:

- Optimizing Parameters and Techniques in Circulating Tumor Cell Collection (OPTICOLL)
- Thrombosis Preanalytical Variables
- HPV Serology Standards

One CSSI pilot project centers on training:

- Big Data Student Training Enhancement Program (BD-STEP)



BD-STEP is a pilot training program managed jointly by CSSI OD and the Veterans Health Administration (VHA). Program decisions are guided by entities across NCI, including the Center for Cancer Training (CCT) and Center for Biomedical Informatics and Information Technology (CBIIT). Recognizing the rapidly expanding volume of healthcare data, the program develops a diverse pool of specialists capable of employing data science in clinical cancer research. Postdoctoral fellows are matched to Veterans

Affairs (VA) medical centers, where they work with VHA clinicians over the course of one year to address clinically important questions in cancer research and care.

2.2 Office of Clinical Cancer Proteomics (OCCPR)

The mission of OCCPR is to advance proteomic science and technology development to gain a better understanding of the molecular mechanisms of cancer. This understanding has the potential to improve cancer prevention, diagnosis, and treatment. OCCPR executes its mission through programs like the *Clinical Proteomic Tumor Analysis Consortium (CPTAC)* and collaborative efforts like the *International Cancer Proteogenome Consortium (ICPC)* and the *Applied Proteogenomics OrganizationaL Learning and Outcomes (APOLLO)* network.

2.2.1 Clinical Proteomic Tumor Analysis Consortium (CPTAC)

The CPTAC network was created in 2006 to tackle barriers in the field of clinical proteomics, including a lack of standardized methodologies, comparable data, and well-characterized, quality reagents. CPTAC holds its members to an elevated level of analytical rigor and reproducibility while providing community resources like the CPTAC Assay, Data, and Antibody Portals. CPTAC has also been a pioneer in the field of proteogenomics, as exemplified by three seminal studies of colorectal, breast, and ovarian tumors characterized genomically by The Cancer Genome Atlas (TCGA) project. Using standardized proteomic workflows, CPTAC analyzed these tumors and elucidated functional consequences of cancer-associated mutations. CPTAC is extending its proteogenomic characterization efforts to additional cancer types and treatment applications.

2.2.2 International Cancer Proteomics Consortium (ICPC)

Inspired by CPTAC and catalyzed by the Beau Biden Cancer Moonshot℠, the ICPC investigates applications of proteogenomics in understanding cancer biology and predicting treatment response. The Consortium provides a forum for collaboration among the world's leading cancer research centers and unifies methodologies on a global scale. The ICPC supports public data sharing around the world to accelerate translation of results to patient care.

2.2.3 Applied Proteogenomics OrganizationaL Learning and Outcomes (APOLLO)

The APOLLO network, launched in 2016 in response to the Beau Biden Cancer Moonshot℠, is a collaboration between NCI, the Department of Defense (DoD), and the Department of Veterans Affairs (VA) to incorporate proteogenomics into patient care. Partnering with the nation's two largest health systems, DoD and VA, allows NCI to study more patients and obtain results more efficiently. The data will be made publicly available across various NCI data sharing platforms.

2.3 Knowledge Management and Special Projects Branch (KMSPB)

The role of KMSPB is to oversee the reporting of the NCI-funded portfolio of grants, intramural, and contract projects in official NCI and NIH reporting initiatives. KMSPB serves as the NCI lead for all official NIH Research, Condition, and Disease Categorization (RCDC) semi-automated categorization and reporting for NCI-funded research. This involves recruiting subject matter experts across NCI to help develop or update the definitions used to classify research into almost 300 research/disease reporting categories. This oversight of reporting role also requires the review of all projects categorized using the NIH RCDC methodology each fiscal year. Therefore, KMSPB collaborates closely with colleagues across the budget and scientific components within the NCI and is involved in the development of various reporting tools to ensure the completeness and accuracy of the NCI-funded portfolio reports to Congress and the public. The KMSPB also manages the NCI Funded Research Portfolio (NFRP) web site and database, to ensure the completeness, timeliness, and accuracy of NCI coded scientific funding data for current and past fiscal years.

CSSI Office of the Director (OD)

Table of Contents

1 CSSI Overview

CSSI, an operating entity of the NCI Office of the Director, offers strategic focus for innovative programs and cultivates cancer research in emerging fields. The Center serves research communities in these fields by developing infrastructure and national resources that help accelerate discoveries and their translation into the clinic. To stay abreast of new research areas, support collaboration across the NCI, and leverage resources, CSSI hosts workshops and think tanks among interdisciplinary experts. This includes the annual Science Day meeting, held since 2013, which brings together representatives from NCI Divisions, Offices, and Centers (DOCs), federal partners, and the extramural research community, to discuss current activities and future endeavors in cancer research. These discussions have spurred the development of multiple workshops and new scientific pilots.

2 CSSI Office of the Director (OD)

The Center is guided by the CSSI Office of the Director (CSSI OD), which provides oversight and coordination of scientific and programmatic activities for its offices and programs. This responsibility involves facilitating reviews and approvals from the NCI Scientific Program Leaders, NCI Board of Scientific Advisors, and the National Cancer Advisory Board. CSSI OD also facilitates collaboration among federal, academic, and industry entities to advance CSSI programs and initiatives.

CSSI OD oversees three trans-NCI grant programs, the PQ initiative, the Pediatric PQ initiative, and the Innovative Molecular Analysis Technologies (IMAT) program. Beyond these grant activities, CSSI OD participates in collaborations with federal and private partners, accomplishing objectives through contracts, interagency agreements, and the Frederick National Laboratory for Cancer Research (FNLCR). CSSI activities at FNLCR span a data coordinating center and pilots focused on development of new technologies and standards and assessment of preanalytical variables. Pilots across federal agencies include a study to evaluate fatigue in cancer patients and warfighters with the Department of Defense (DoD) and a data science training program in partnership with the Department of Veterans Affairs (VA), Veterans Health Administration (VHA). Key CSSI OD programs, responsibilities, and achievements will be described in the following sections.

3 CSSI OD Coordinated Programs

3.1 Innovative Molecular Analysis Technologies (IMAT)

3.1.1 Overarching Goals of IMAT

The goal of the IMAT program is to develop next-generation, cutting-edge technologies to equip basic and clinical research communities with novel analysis capabilities in the fight against cancer. Progress in detecting, monitoring, and targeting the biological mechanisms of cancer is accelerated when transformative technologies supplement or improve conventional technologies. The IMAT program is a key element of NCI's strategy to spur technology development research towards those ends. IMAT uses a strategy of phased grant support, as depicted in the figure below.

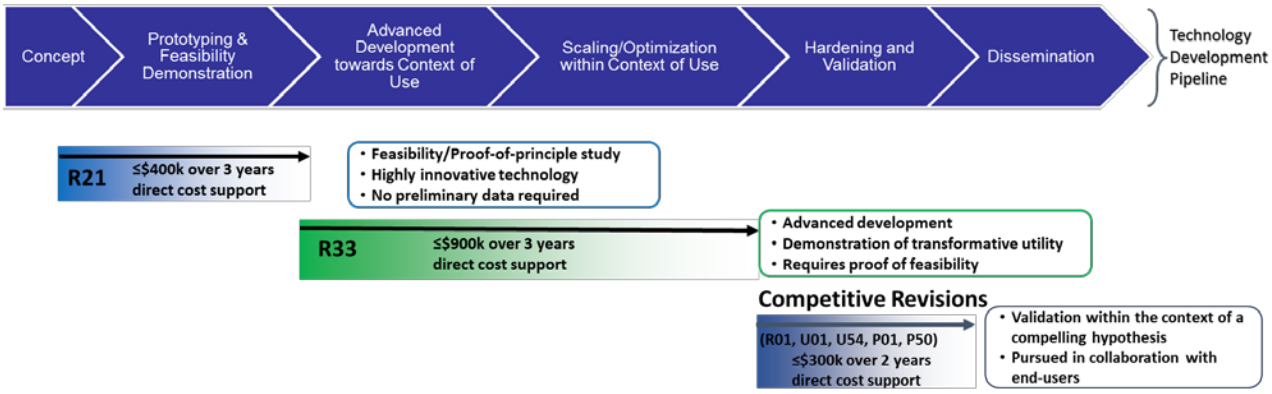


Figure 1 IMAT Grant Support Structure

To ensure that the technologies in IMAT's portfolio address the diverse needs of the cancer research continuum, the program includes participation from each of NCI's extramural divisions and is coordinated by CSSI. This trans-divisional management structure also minimizes duplication with other NCI programs or initiatives. Central coordination is required to help manage critical elements of the program, including development of solicitations, determination of application responsiveness, appropriate division referral based on potential applications of the technology (nearly all awards are held by division Program Directors), development of funding plans, and planning for the annual principal investigators' meeting. CSSI also works with the Small Business Innovation Research (SBIR) Development Center to offer parallel R43/R44 awards exclusively available to small business investigators. The SBIR-IMAT solicitation complements the IMAT R21/R33 programs, and other NCI technology development initiatives to fill a funding gap for small business concerns that need to perform further development and validation of next-generation technologies.

3.1.2 History

Since the inception of the IMAT program in 1998, the program has steadily evolved to meet the changing technology needs of cancer research and clinical care. The IMAT program has received nearly 5,000 applications and offered varying levels of support for roughly 500 unique technology platforms. A selection of successfully developed technology platforms supported by past IMAT awards is listed below:

- BeadChip [CA083398,1999] and BeadArray [CA081952,2000]
- Multidimensional Protein Identification Technology [CA081665, 1999]
- Isotope-Coded Affinity Tags [CA084698, 2000]
- Synchrotron Protein Footprinting [CA084713, 2000]
- Differential Methylation Hybridization [CA084701, 2000]
- Multi-Photon Intra-Vital Imaging [CA089829, 2001]
- Methylation-specific Amplification [CA089837, 2002]
- Nucleic Acid Programmable Protein Arrays [CA099191, 2003]
- ChIP-Seq [CA105829, 2004]
- Optical Mapping [CA111933, 2005]
- CryoXtract [CA114167, 2005]
- Proteolysis Targeting Chimera [CA118631, 2006]
- Activity-based protein profiling [CA118696, 2006]
- CellASIC [CA120619, 2006]
- RainDrop Digital PCR System [CA125693, 2007]
- Digital Transcriptome Subtraction [CA120726, 2007]
- Proximity Ligation Assays [CA126727, 2008]
- Exclusion-based Sample Prep [CA137673, 2009]
- Single Molecule Analysis (SiMoA) [CA133987, 2010]
- Oligo-selective sequencing [CA140089, 2010]
- NanoVelcro [CA151159, 2010]
- OncoPanel/Oncomap [CA155554, 2011]
- Single-molecule molecular inversion probes [CA160080, 2011]
- Crainbow Mouse [CA173245, 2012]
- NanoTrap™ [CA173359, 2012]
- Vortex Chip [CA177456, 2013]
- Protein Paint [CA177535, 2013]
- Duplex Sequencing [CA181771, 2014]
- Multiplexed Ion Beam Imaging [CA183654, 2014]

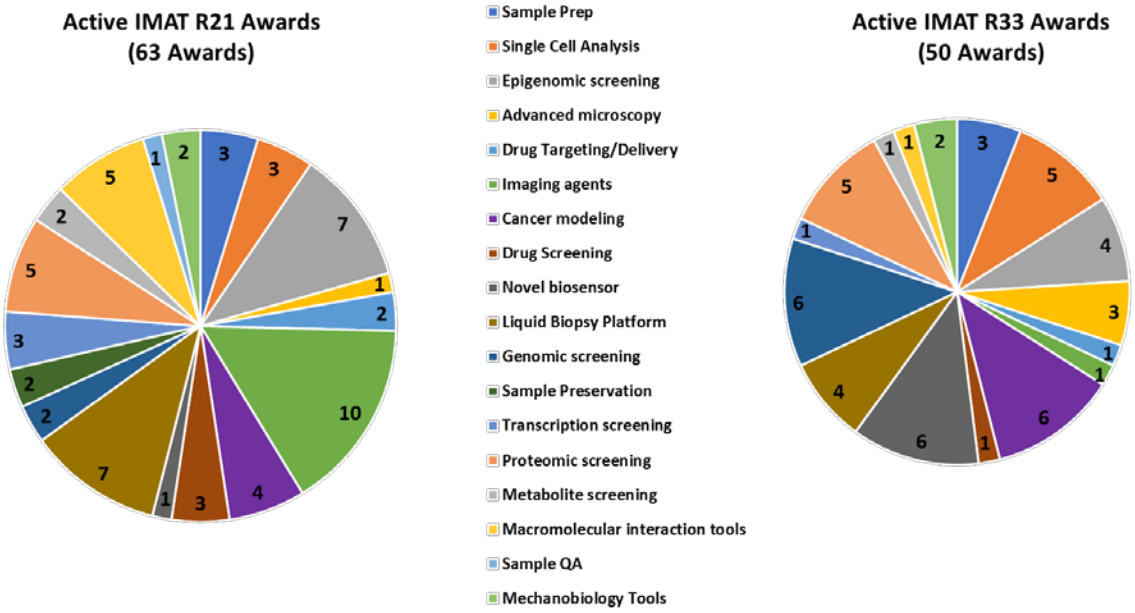
An extensive, external process and impact evaluation of the IMAT program was conducted during FY2015-2016 encompassing the full IMAT portfolio of projects receiving their first award before 2014. The evaluators collected archival data and conducted web-based surveys and telephone interviews with grantees and end-users unaffiliated with the IMAT-supported project to understand the outcomes and technology contributions of each project. Archival and web-based survey data were also collected from a comparison group of similarly-focused NIH grantees. The criteria used for this evaluation included those approved by both the NCI Scientific Program Leaders (SPL) committee and the NCI Board of Scientific Advisors (BSA) in the most recent reauthorization of the program. Highlights from the evaluation report include:

- As reported by the principal investigators (PIs), the IMAT program fills a very specific niche in cancer research that encourages cutting-edge, innovative research.

- 28% of PI interview participants stated that they would not have pursued additional alternative funding mechanisms if they had not received initial IMAT funding;
- Another 27% said they would have pursued other funding mechanisms, but would likely face challenges due to the innovative but unproven nature of the proposed research.

- IMAT awardees delivered more publications and with a higher average impact factor per dollar invested than the comparison group, and pursued patenting of their pursuits at a higher rate per dollar invested.

The IMAT program regularly makes 30-40 new R21 or R33 awards each year through Request for Applications (RFA) funding opportunities with multiple receipt dates per year. Awards are based on scientific merit, as determined by peer review scores and post-review program staff evaluation and discussion. Each award recommendation is presented to SPL for final approval. The active IMAT portfolio as of July 2017 consists of 113 projects covering a diversity of areas, as shown in the figure below.



3.2 Provocative Questions (PQ) Initiative

3.2.1 Overarching Goals of the NCI's PQ Initiative

The goal of the PQ Initiative is to stimulate areas of cancer research that are understudied, neglected, paradoxical, or historically difficult to address. Answers to these Provocative Questions (PQs) are solicited through R01 and R21 mechanisms and more recently through competitive revisions to existing NCI grants. One key aspect of the PQ Initiative is the solicitation of input from extramural investigators to identify key areas of cancer research that are challenging or understudied. Prior to each set of funding opportunities, the NCI organizes a series of workshops and gets input from dozens of investigators resulting in more than 100 potential questions. The questions are then evaluated and refined by a group of NCI program directors before being finalized by senior NCI leadership. This process

helps ensure that exciting questions cover a broad range of topics, and disciplines are identified for each set of funding opportunities.

Another key component of the PQ Initiative is the trans-divisional management of the program. The current program is coordinated by CSSI and involves more than 75 program directors from across each of NCI's extramural divisions. Like the IMAT program, this trans-divisional coordination ensures sufficient breadth of expertise and knowledge regarding the wide-ranging topics covered by the PQ Initiative.

3.2.2 History of the PQ Initiative

Evolution of the PQ Initiative

The PQ Initiative was launched in 2011 as one of the signature initiatives of former NCI Director, Dr. Harold Varmus. The initiative aimed to "…engage a diverse range of scientists in a challenging intellectual exercise to define then solve the major unsolved or neglected problems in oncology" [1]. The initial phase of the PQ initiative ran from 2011-2014 and included the issuance of three separate sets of Funding Opportunity Announcements (FOAs). In total, 43 distinct questions covering a broad spectrum of understudied, neglected, or difficult-to-address areas of cancer research were explored through these FOAs.

In June 2014, the NCI's BSA approved the PQ Initiative for three additional issuances. Based on lessons from the first phase of the program, a handful of changes were made to the program to attract the most relevant and high-quality applications. Changes included the addition of competitive revision applications, the development of an 'Intent Statement' for each PQ, and the formation of Question Teams. The revision applications allow investigators to address new ideas or interesting observations that arise in NCI-supported research projects by adding a PQ component to their grant. Intent Statements are used to assess the scientific responsiveness of proposed projects to ensure only relevant applications are considered. Question Teams were established to manage and implement programmatic and scientific aspects of the initiative and ensure that each PQ is assigned to NCI program directors with appropriate expertise. The Question Teams also help the long-term success of their respective PQ by holding workshops to increase awareness within specific research communities and planning new FOAs if a question is retired from the PQ RFAs. There are over 75 NCI program directors currently serving on the 12 PQ Question Teams (4–9 people/team).

The first FOAs for the second phase of the program were issued in 2015 and 2016, and the final grants from these solicitations were funded at the end of Fiscal Year 2017. In the Fall of 2016, a series of workshops were convened with extramural investigators to identify new potential PQs, and more than 100 questions were submitted by the community. A group of Program Directors from across the NCI (the PQ "Executive Committee") evaluated existing and potential new questions and proposed a set of questions to NCI senior leadership. Based on feedback from NCI leadership, the questions were refined, and a final set of 12 was established. This includes seven new PQs and five PQs that were reused or rewritten from previous FOAs. Applications from the first of four receipt dates for these 2017 FOAs were received in June of 2017 and are being reviewed in October and November.

1. Varmus H and Harlow E. Provocative Questions in Cancer Research. *Nature* 2012, *481*, 486–487.

Pediatric Provocative Questions

In February 2016, NCI senior leadership approved a companion to the PQ Program focused on understudied questions in pediatric cancer and simultaneously approved nine Pediatric PQs. Like the standard PQs, the Pediatric PQs were selected based on recommendations from Pediatric PQ workshops, input from NCI program directors, and review by the NCI senior leadership. The Pediatric PQs come from a broad range of cancer research fields and are framed to inspire the extramural research community to conceive feasible new approaches to challenging issues in pediatric cancer research. The goal of the initiative is to promote creativity and originality combined with scientific rigor to expand innovation and eventually solve problems in childhood or adolescent cancer that have been identified within the Pediatric PQs. The Pediatric PQ Program is managed by a trans-NCI team and coordinated by CSSI. The final receipt date for the Pediatric PQ FOAs is November 2017 and there are currently no plans to reissue the program.

3.2.3 PQ Portfolio

2011–2014 PQ Funding

For the 2011, 2012, and 2013 FOAs, the NCI funded 188 new awards seeking to address 38 of the 43 Provocative Questions.

2015–2017 PQ Funding

For the 2015/2016 FOAs, the NCI funded 95 awards covering all 12 questions (see Table 1).

Table 1 PQs and Funded Applications 2015–2016

Provocative Question	Number of Funded Applications
1. For tumors that arise from a premalignant field, what properties of cells in this field can be used to design strategies to inhibit the development of future tumors?	13 (8 R01s, 4 R21s, 1 P50rev)
2. What molecular mechanisms influence disease penetrance in individuals who inherit a cancer susceptibility gene?	7 (6 R01s, 1 R21)
3. How do variations in tumor-associated immune responses contribute to differences in cancer risk, incidence, or progression?	15 (10 R01s, 4 R21s, 1 P50rev)
4. Why do some closely related tissues exhibit dramatically different cancer incidence?	2 (2 R21s)
5. How does mitochondrial heterogeneity influence tumorigenesis or progression?	14 (9 R01s, 3 R21s, 2 R01revs)
6. What are the underlying molecular mechanisms that are responsible for the functional differences between benign proliferative diseases and premalignant states?	6 (2 R01, 3 R21s, 1 R01rev)

Provocative Question	Number of Funded Applications
7. What in vivo imaging methods can be developed to determine and record the identity, quantity, and location of each of the different cell types that contribute to the heterogeneity of a tumor and its microenvironment?	5 (4 R01s, 1 R21)
8. What cancer models or other approaches can be developed to study clinically stable disease and the subsequent transition to progressive disease?	3 (3 R21s)
9. What are the molecular and/or cellular mechanisms that underlie the development of cancer therapy-induced severe adverse sequelae?	20 (16 R01s, 4 R21s)
10. How do microbiota affect the response to cancer therapies?	6 (3 R01s, 3 R21s)
11. What mechanisms of action of standard-of-care cytotoxic, radiologic, or targeted therapies affect the efficacy of immunotherapy?	3 (2 R01s, 1 R21)
12. What methods and approaches induce physicians and health systems to abandon ineffective interventions or discourage adoption of unproven interventions?	1 (1 R01)

Pediatric PQ Funding

For the first two receipt dates for the 2016 Pediatric PQ FOAs, the NCI has funded seven projects covering five of the nine Pediatric PQs (see Table 2).

Table 2 Pediatric PQ and Funded Application 2015–2016

Pediatric PQ	Number of Funded Applications
1. What are the processes of normal development that are -permissive to the development of specific pediatric cancers?	3 (2 R01s, 1 R21)
2. What is the functional mechanism by which gains and losses of large segments of chromosomes contribute to childhood cancer development and response to treatment?	1 (1 R01)
3. What role do alterations in noncoding sequences play in the development of pediatric cancers?	0
4. What are the molecular vulnerabilities for tumors of childhood identified through CRISPR or equivalent screens?	0
5. What molecular and cellular mechanisms allow reactivation or bypassing of specific silenced tumor suppressor genes in pediatric cancers?	0
6. How can mouse or other preclinical models be used to study how standard of care and investigational therapies affect normal tissue and lead to adverse events later in life?	1 (1 R21)
7. How can prediction models be developed and used to identify patients at highest risk of treatment-related complications?	1 (1 R01)

Pediatric PQ	Number of Funded Applications
8. What are the molecular mechanisms that define how pediatric solid tumors (both tumor cell and stroma) evolve in response to standard pediatric cancer therapy?	0
9. What are the underlying molecular mechanisms that cause accelerated aging seen in some pediatric cancer survivors?	1 (1 R01)

4 CSSI Pilot Projects

4.1 Overview of the Strategic Pilots Incubator and Data Coordinating Center

The Strategic Pilots Incubator (SPI) is a unit within FNLCR responsible for marshaling and coordinating CSSI-related FNLCR resources and projects to facilitate the advancement of CSSI initiatives. The goal of SPI is to: 1) design, plan, implement, and launch new inter- and intra-CSSI pilot projects to address research gap areas and capitalize on strengths of CSSI initiatives, and 2) to provide readily deployable support to ongoing CSSI projects at FNLCR. To assess needs and identify potential projects, SPI obtains input from all stakeholders, including government employees, contractors, and extramural investigators.

A major need identified by SPI was to house and manage disparate data from multiple cell or tissue characterization projects associated with CSSI and other projects across NIH. In response to this need, a pilot was initiated to create an integrated, open data store as a public resource to increase data sharing and secondary analyses by the cancer research community. A data coordinating center (DCC) was designed in FY2015 and has been developed, launched, and refined across FY2016 and 2017. The DCC's searchable data are displayed with visualization tools, and all data sets are accompanied by formatted metadata describing how they were obtained and processed. As new projects are completed, the center is prepared to bring in new data sets via direct investigator upload.

The ability to readily share, view, and download data in a consistent format has the potential to improve data sharing and quality and to expand the value of CSSI-sponsored projects. Moving forward, the DCC development team will continue to engage with leaders of the NCI cancer research data ecosystem to ensure the DCC is compatible with the new infrastructure and continues to fill a unique need.

4.2 New Technologies

4.2.1 Analytical Technologies to Objectively Measure Human Performance (ATOM-HP)

The Analytical Tools for Objective Measurement of Human Performance (ATOM-HP) project was spurred by the jointly-managed NIH/DoD Human Performance Optimization Working Group (HPO-WG), which identifies opportunities for collaboration between defense and medical research communities. The

objective of ATOM-HP is to develop a wearable technology platform that can objectively monitor fatigue and provide insights about human performance in the face of fatigue. Such devices may have applications for both cancer patients and warfighters alike.

The monitoring device will capture a range of physiological data that will be algorithmically correlated with performance on objectively scored tests of physical readiness. The objective tracking of fatigue in real time will equip physicians and commanders with accurate information about the physical capacity of an individual or group. This is particularly important in the context of cancer treatment, because agreement between patient-reported outcomes (PRO) and physician-reported ECOG scores is often low. These scores, which categorize the progression of a patient's disease based on his or her activity level, help physicians determine the most appropriate treatments for their patients.

Following extensive review of commercially-available wearable monitoring devices, the ATOM-HP team selected the Microsoft Band and Microsoft Kinect to be used for the study. The Microsoft (MS) Band records measurements of step count, heart rate (mean, peak, min) and calories burned, while the MS Kinect senses movement of the human body and face. The devices were distributed to cancer patients and warfighters to begin data collection for parallel clinical and military studies. The monitoring scheme for these trials is summarized in Figures 2 and 3 below. The clinical study would focus on the effect of chemotherapy administration, gender, and age on the activity of cancer patients, while the military study would monitor marksmanship, physical performance (via the fatigue-inducing Marine Corps' Combat Fitness Test (CFT)), and cognitive performance (via a trail-making test).

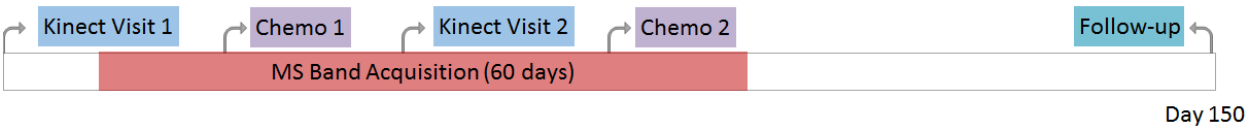


Figure 2 **Band and Kinect Data Acquisition for Clinical Trials**

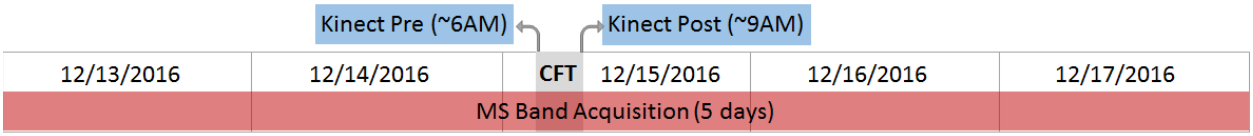


Figure 3 **Band and Kinect Data Acquisition for Military Trials**

Conclusions and Findings

Preliminary findings from the clinical study suggest that objectively measuring markers of fatigue might better predict patient outcomes than physician assessment. Fatigue scores from PRO survey reports were found to be correlated with step count data from the MS Band, while physician assessments differed significantly. Physicians on average over-estimated the health and activity of their patients, reporting lower (healthier) ECOG scores for groups determined to be less active using monitoring data. Importantly, patients' self-assessments were found to be much more predictive of clinical events.

The military study showed no relationship between physical performance, as measured by the MS Band and Kinect, and marksmanship or cognitive decision-making skills. However, there were significant

differences between the pre- and post-CFT trail-making tests, indicating that the Marines performed better on the cognitive testing when subjected to fatigue. The Marines were also found on average to have more accurate marksmanship after the CFT. Data analysis for both studies is ongoing, but preliminary results show promise for future applications of objective fatigue monitoring in both military and clinical settings.

4.2.3 High-Content Tissue and Cellular Characterization Laboratories

Precision medicine initiatives in oncology focus on delivering targeted therapeutics to specific patient populations where they will be most effective. One of the challenges in achieving the promise of precision oncology is that single measured parameters fail to provide a complete picture of the complex biology within a patient's tumor composition, thus limiting the ability to predict therapeutic responsiveness. Two capabilities are being established at FNLCR to contribute to the ongoing development of highly-multiplexed assays capable of comprehensive analysis of heterogeneous tumor cell populations within solid or liquid tissue samples. High content technologies are designed to capture as much image information from a tissue specimen as possible and have the potential to increase the capacity for better disease diagnosis, drug development, and biological research. These laboratories, described below, will include capabilities for single-cell and tissue proteomics and morphology characterization to allow for analysis and comparisons of rare cells and tumor tissue.

Tissue Imaging Laboratory

FNLCR is establishing a laboratory to explore the suitability of highly-multiplexed imaging for use in translational and clinical research. Initially, the laboratory will evaluate imaging mass cytometry (IMC) for reproducible high-content analysis of clinical and preclinical tissue samples. IMC is an expansion of mass cytometry, using a laser to ablate tissues or cells labeled with antibodies carrying high-mass metal tags. The particles are carried to the mass cytometer, with the ions from each ablated spot measured by time of flight (TOF) mass spectrometry. The use of rare earth metals allows for multiplexing up to 40 antibodies per sample and facilitates the quantitation and subcellular localization of proteins in both tissue and single cell samples.

The Helios CyTOF IMC platform was acquired through Fluidigm's early access program, and the initial establishment of the laboratory will focus on two demonstration projects. The first project is a collaboration with the NCI's Division of Cancer Treatment and Diagnosis (DCTD) and will evaluate IMC with validated clinical biomarker assays from the Clinical Pharmacodynamics Program and the Molecular Characterization Laboratory (MoCha). The second project will use the same panel to examine epithelial-mesenchymal transition at the tumor margins in triple negative breast cancer. This will involve collaboration with a team in the Center for Cancer Research (CCR). These projects will also address the impact of the CyTOF's one micron pixel size on image quality and explore the compatibility of CyTOF image files with downstream image analysis algorithms currently in use for quantifying and visualizing histopathological diagnostic grade images.

FNLCR is establishing a scientific advisory group made up of NCI and Leidos stake holders to evaluate the outcomes of the demonstration projects and prioritize potential future projects.

High Content Single-Cell Analysis Laboratory

FNLCR will establish a laboratory to explore technologies for the isolation and analysis of rare cells obtained from liquid biopsies. The first step is to establish a High Definition-Single-Cell Analysis (HD-SCA) platform for the analysis of liquid biopsies. The HD-SCA platform combines immunofluorescent staining and automated digital microscopy to identify individual circulating tumor cells (CTCs), circulating tumor microemboli (CTC clusters), and other rare cell populations to generate liquid biopsy profiles. This assay has potential advantages over other CTC capture technologies, because it analyzes all nucleated cells in blood samples, ensuring a non-biased approach to CTC isolation and allowing a comprehensive analysis of all cell populations.

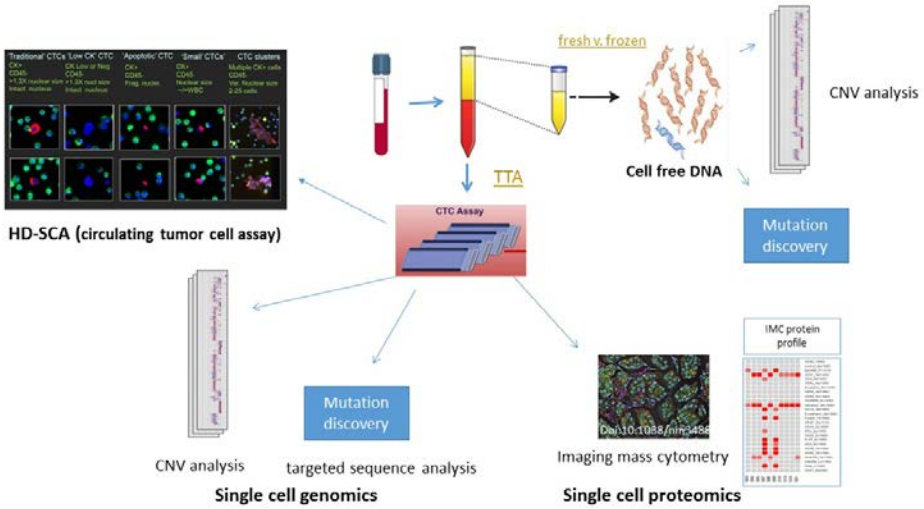
Initially, the laboratory will focus on validation experiments done in collaboration with extramural teams and will focus on the technical reproducibility of sample preparation and analysis between multiple sites. Concordance studies of control samples prepared at each site will be carried out, followed by clinical samples from cancer patients to extend the validation of the HD-SCA platform. The comparison of multiple parameters will be assessed, including enumeration of nucleated cells and CTC events detected, morphometric characteristics such as cell and nuclear size/ratio, and relative intensity of antibody staining. Future investigations will include coupling the HD-SCA platform with the multiplexing capabilities of IMC, allowing more comprehensive CTC characterization, including functional analysis of tumor cell-immune cell interactions in the circulation.

As with the IMC laboratory, future application of the HD-SCA platform to biological and clinical questions of interest will be identified and prioritized by Leidos, NCI, and extramural stakeholders.

4.3 Preanalytical Variables and Standards

4.3.1 High-Content Screening of Physical-Based Properties in Biospecimens Phase II

Blood-based biomarkers like CTCs and cell-free DNA (cfDNA) are of great importance in cancer diagnostics due to their correlation with tumor measures and the relative ease of liquid biopsy compared to invasive tissue biopsy. However, differences in the way that patient blood samples are handled prior to analysis can lead to variable results. To examine the effect of preanalytic variables on the detection and characterization of CTCs and cfDNA in liquid biopsies (Figure 4) from breast cancer patients, the High-Content Analysis of Physical-Based Properties in Biospecimens Phase II pilot program was executed in 2016-2017. Managed by FNLCR, this program was an effort between CSSI, the Division of Cancer Biology's Office of Physical Sciences–Oncology (OPSO), and the Division of Cancer Treatment and Diagnosis' Biorepository and Biospecimen Research Branch (BBRB).



Abbreviations: CNV, copy-number variation; CTC = circulating tumor cell; HD-SCA, high-definition single cell assay; TTA = time-to-assay.

Figure 4 Flow of Specimen Preparation and Assays for the High-Content Screening of Physical-Based Properties in Biospecimens Phase II

In phase I of this effort, experiments were conducted to compare the detection of CTCs in samples analyzed 24 and 72 hours post blood draw. Using the HD-SCA platform described under the *High Content Single-Cell Analysis Laboratory*, it was determined that the 24-hour "time-to-assay" period produced superior results both in terms of overall number and quality of identified CTCs. Building upon these results, Phase II compared 24- and 48-hour time-to-assay periods and investigated differences in the analysis of cfDNA in fresh versus frozen plasma. Using blood samples from breast cancer patients: 1) CTCs were isolated and analyzed by single-cell genomics and proteomics, or 2) plasma was prepared and cfDNA was isolated and analyzed.

Overall results from the time-to-assay studies indicated that the 48-hour period was equivalent to the 24-hour period in the detection and characterization of CTCs. This is important for clinics performing these studies, because it means that shipping samples overnight is not necessary. The results of single-cell genomics approaches (both copy-number variation and targeted sequencing) were also compared for the different time-to-assay periods. The DNA quantity and quality and the sequencing quality were found to be comparable for both time periods. As a pilot for this study, a single-cell proteomics assay was performed using imaging mass cytometry and provided a readout of over 25 biomarkers from each single cell examined.

In the study comparing fresh versus frozen plasma, cfDNA was characterized by measuring copy-number variations and targeted sequencing. Although the overall yield was low in either condition, the fresh and frozen samples yielded comparable results. The data from this project are publicly available through the CSSI DCC and are expected to help the cancer research and diagnostic communities by providing evidence-based best practices for the collection and storage of CTCs and cfDNA.

4.3.2 Thrombosis Preanalytical Variables

Thrombosis, particularly venous thromboembolism (VTE), is a major source of increased morbidity and mortality in cancer patients. Cancer patients have a four- to seven- fold increased risk of venous thrombosis, resulting in a decreased quality of life and reduced overall rate of survival. The increased risk of VTE in cancer patients can also confound the choice of effective cancer treatment modalities and require additional healthcare resources. Motivated by the scientific and clinical significance, NCI and the National, Heart, Lung, and Blood Institute (NHLBI) convened a working group to identify gaps and priority research areas [2]. Among recommendations ranging from mechanistic studies to intervention trials, the need to identify and validate improved, actionable biomarkers and risk factors for thrombosis in cancer patients was highlighted. As a first step towards biomarker improvement and development, an NCI-sponsored pilot study, Thrombosis in Cancer Patients, was launched in FY2015 with the goal of assessing the effects of preanalytical variables on thrombosis biomarkers in cancer patients.

The study is managed by FNLCR, with high level guidance provided by an external scientific advisory committee with expertise across clinical pathology, coagulation/hemostasis testing, preanalytical variables, biospecimen research, and thrombosis in cancer patients. Blood is being collected from cancer patients and healthy donors under tightly controlled standard operating procedures. Preanalytical variables experienced in the clinical setting, including delay to blood processing, delay to assay, and freeze-thaw cycles, will be measured on markers of coagulation, fibrinolysis, cellular injury, and inflammation. Participants will be monitored over six months for cancer and thrombosis outcomes, as outlined in Figure 5. Overall, the study aims to advance standardization and provide guidance for the measurement of thrombosis biomarkers in cancer patients. This includes identifying steps during biospecimen procurement, handling, and processing which are critical for optimal specimen preservation and accurate marker detection.

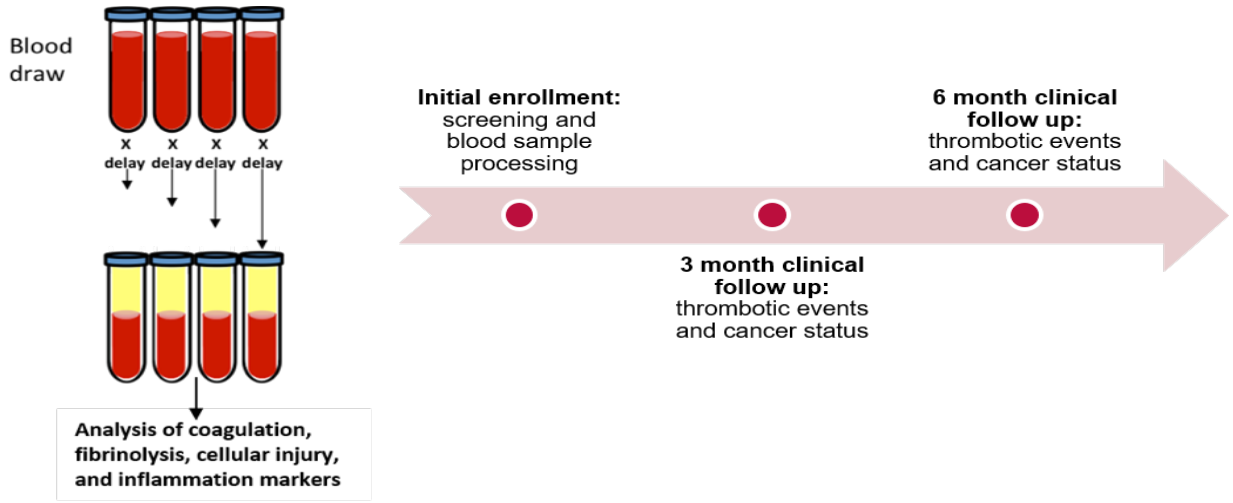


Figure 5 **Blood Sample Flow for Delay to Blood Processing Variable at Initial Enrollment (Left) and Data Capture over the Study (Right)**

2. Key et al. Thrombosis in Cancer: Research Priorities Identified by a National Cancer Institute/National Heart, Lung, and Blood Institute Strategic Working Group. *Cancer Res*. 2016 Jul 1;76(13):3671–5.

The Thrombosis in Cancer Patients projects will finish in the last quarter of FY2018 and will provide best practices for biomarker assessment, a potential biospecimen source to enhance future NIH funding opportunities, and a foundation for future biomarker studies to assess thrombotic risk and predict efficacy of anti-thrombotics. The data generated will be widely shared with the research community through publications and a public data repository, the CSSI DCC.

4.3.3 HPV Serology Standards

The Human Papillomavirus (HPV) Serology Laboratory (HSL) was established in January 2017, as part of an initiative to standardize and harmonize serological assays for HPV antibody testing in the context of vaccine trials.

HPV serology is an essential tool for the measurement of vaccine immunogenicity. Serology standardization is particularly important as new HPV prophylactic vaccine trials are proposing to use serology data as endpoints for licensure of new vaccine indications or new vaccines. However, there is a lack of uniform, standardized assays, procedures, and reagents accessible to the scientific community for assessment of immune responses to HPV prophylactic vaccines. In addition, there is an incomplete understanding of the correlates of efficacy and minimal titers of vaccine-induced antibodies required for protection against HPV infection. To address these gaps, the HPV Serology Laboratory, co-funded by the Bill and Melinda Gates Foundation and the NCI, is working in partnership with several members of the HPV scientific community. The main goals of the Laboratory include the development of qualified secondary assay standards, critical reagents (HPV Virus-Like Particles), and assays that will be made available to the scientific community.

The HSL includes personnel with expertise in HPV, assay development, molecular biology, and Quality Assurance/Control leading this initiative. As assays and standards are developed and made available for clinical trial use, expansion to immune monitoring activities in support of HPV vaccine trials with various partners from government, industry and/or academia will be developed as needed and made available at FNLCR through additional sources of funding.

Overall, this initiative will enable comparisons of data across different vaccines and studies. Thus, it will facilitate vaccine development and implementation of new vaccine indications and candidates. In addition, the tools developed may drive new discoveries such as novel surrogate markers of protection, with the potential to contribute to a reduction of HPV and its associated cancer burden in the world.

4.4 Training

4.4.1 Overview of the Big Data Scientist Training Enhancement Program

The Big Data Scientist Training Enhancement Program (BD-STEP) was launched in 2015 as a partnership between NCI and the VA/VHA. Recognizing the rapidly expanding volume of healthcare data, the program aims to develop a diverse pool of scientists and engineers capable of employing data science in clinical cancer research. The long-term goal is to improve the treatment and care of cancer patients by increasing capacity for manipulation and analysis of large-scale patient data sets and construction of new algorithms that advance patient-centered outcomes research.

Competitively selected postdoctoral fellows from physical science, engineering, or computer science backgrounds with demonstrated expertise in bioinformatics, modeling, or management of large data sets are matched to VA medical centers for one year fellowships. With joint mentorship from academic investigators and VA clinical researchers and care providers, fellows are immersed in collaborative training and research environments at the medical centers. Combining exposure to the healthcare setting with access to rich VA data resources, fellows are primed to apply their quantitative skills to address clinically important questions in cancer research and care. These resources include: clinical data from the VA's integrated, national health care system; genetic data from the Million Veteran Program (MVP); research and care data from a Precision Oncology Program; and diagnosis and treatment information from the VA Central Cancer Registry.

The BD-STEP VA medical center network consists of six locations that were selected through an open competition request for proposal in 2015 based on strength in mentoring and training, research capabilities, utilization of VA informatics systems, and leadership support. The six awarded sites, as indicated in Figure 5, are: 1) VA Western New York Health Care System, Buffalo, NY; 2) VA Boston Healthcare System, Boston, MA; 3) Durham VA Medical Center, Durham, NC; 4) Michael E. DeBakey VA Medical Center, Houston, TX; 5) VA Palo Alto Health Care System, Palo Alto, CA; and 6) VA Puget Sound Health Care System, Seattle, WA. The sites are coordinated into a national network through VHA to facilitate collaborative research and curriculum development. Overall network guidance is provided by an Advisory Council with membership from VHA and NCI, including the NCI's Center for Cancer Training (CCT) and Center for Biomedical Informatics and Information Technology (CBIIT). VHA supports fellowship salaries and NCI provides support for program administration, curriculum development and implementation, and program and scientific meeting travel.

BD-STEP has supported 19 fellows across the FY2016 and FY2017 cohorts, and launched a third cohort of 15 fellows in October 2017 for FY2018 (highlighted in Figure 5). Fellows come from strong research institutions, including the University at Buffalo, Harvard, Duke, University of Southern California, Stanford, Oregon Health & Science University, and University of Washington. These institutions also bring academic mentorship and technical expertise to collaborative research partnerships with VA. Fellows have initiated diverse studies such as predicting hepatocellular carcinoma in hepatitis C patients

using a cohort of more than 180,000 veterans, investigating the role of platelet counts on the outcomes of lung cancer patients, and developing and validating surgery- and pain-related quality metrics. A panel of fellows presented research outcomes at a national health data meeting in May 2017 and first research publications are expected in the last quarter of FY2017. Tracking early career outcomes, three fellows have obtained permanent VA positions, two have obtained new outside faculty and research scientist positions, six fellows have continued the program for second year, and the remainder are continuing their research at a variety of academic institutions.

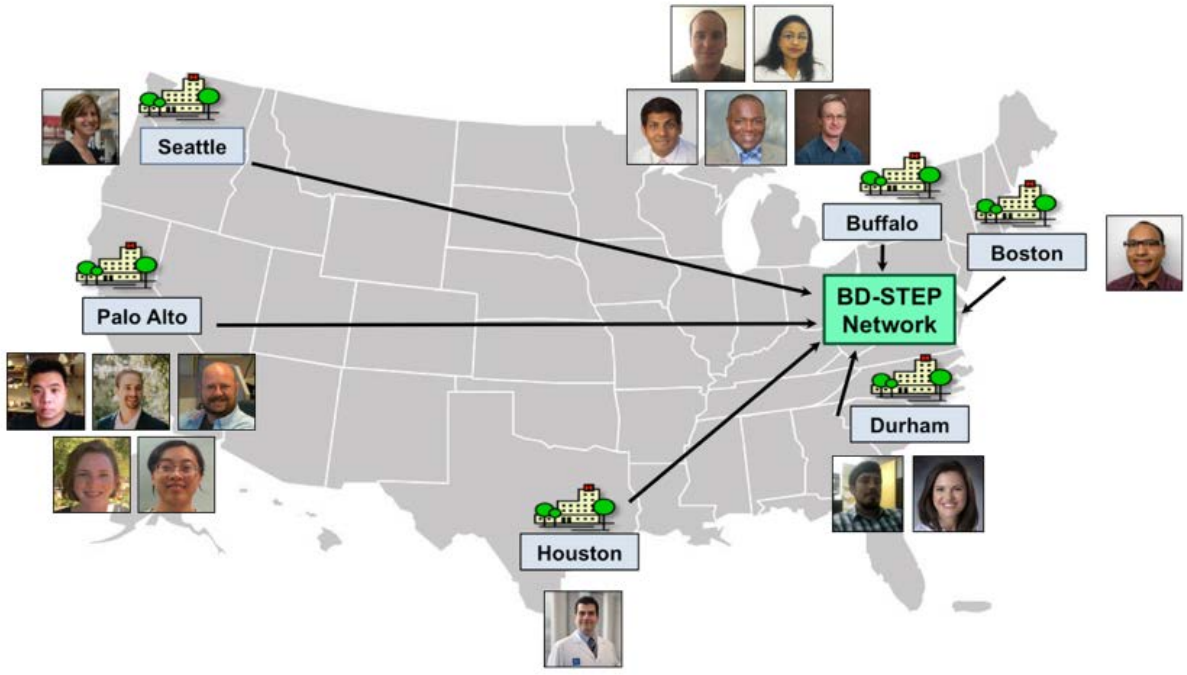


Figure 5 BD-STEP VA Sites and Matched Fellows for FY18

Beyond individual projects and outcomes, new cross-site collaborations have formed to take advantage of scientific, clinical, and technical expertise across the BD-STEP network. These collaborations are bolstered by monthly meetings of the fellow network and of the VA site director network. Activities across the network and with academic institutions will strengthen with future projects and cohorts. Based upon the early engagement, training opportunities, research development, and partnership establishment, BD-STEP was recognized as an exemplar program in the Beau Biden Cancer Moonshot^{SM} task force report for developing the biomedical data science workforce. The program will continue to support development of data scientists in the healthcare space, with potential to expand the model to other federal partners in the health sector in future years.

Appendix 1
Current Staffing

Douglas Lowy, MD

Dr. Lowy serves as acting director of CSSI. In addition to this role, he serves as chief of the Laboratory of Cellular Oncology in the Center for Cancer Research, deputy director of the NCI, and deputy director of the Center for Cancer Research. Doug is a member of the National Academy of Sciences (NAS) and the Institute of Medicine of the NAS.

Dr. Lowy has directed a research laboratory at NCI since 1975. For his research with John Schiller on the technology that enabled preventive HPV vaccines, he and Dr. Schiller have jointly received numerous honors, including the 2007 Federal Employee of the Year Service to America Medal from the Partnership for Public Service, the 2011 Albert B. Sabin Gold Medal Award, the 2012 National Medal of Technology & Innovation (awarded in 2014), and the Lasker-DeBakey Clinical Medical Research Award, the country's most prestigious honor for biomedical research. Dr. Lowy has also received the National Medal of Honor for Basic Research from the American Cancer Society.

Jerry S. H. Lee, PhD

Dr. Lee serves as deputy director of CSSI, where he is responsible for strategic scientific initiatives that focus on the integration of advanced technologies and transdisciplinary approaches to accelerate the creation of publicly available, broadly accessible, multidimensional data, knowledge, and tools to empower the entire cancer research continuum for patient benefit. In 2016, he was assigned to Office of the Vice President to serve as the deputy director for Cancer Research and Technology for the White House Cancer Moonshot Task Force. A few key efforts he helped coordinate include the Applied Proteogenomics OrganizationaL Learning and Outcomes Network (APOLLO), international collaborations to share molecular characterization datasets, the Blood Profiling Atlas in Cancer pilot, as well as cochairing an interagency group focused on cancer data and technology policy issues.

Dr. Lee continues research as an adjunct associate professor at Johns Hopkins University, where he also earned his bachelor's degree in biomedical engineering and PhD degree in chemical and biomolecular engineering. Dr. Lee is a member of the Innovation Policy Forum of the National Academies Board on Science, Technology, and Economic Policy; the Foundation for the NIH's Biomarkers' Consortium Cancer Steering Committee; and the Health and Environmental Sciences Institute's Board of Trustees.

Sean E. Hanlon, PhD

Dr. Sean E. Hanlon is an associate director of the Center for Strategic Scientific Initiatives (CSSI), where he contributes to the vision and strategic plans of the Center, provides leadership in the analysis and evaluation of emerging fields, and develops and implements new initiatives. Additionally, Dr. Hanlon serves as a CSSI/NCI representative on NCI, NIH, and interagency working groups and committees, including the trans-NCI Data Sharing working group. He also facilitates collaborations and provides strategic and scientific leadership to collaborative transdisciplinary programs, including the NIH Common Fund's 4D Nucleome program. Dr. Hanlon is interested in promoting data sharing and collaborative team science, improving scientific research evaluation, and enhancing science education.

Prior to joining CSSI, Dr. Hanlon served as program director within the NCI Division of Cancer Biology where he served as director of the Physical Sciences-Oncology Network (PS-ON). In this role, he led the scientific management and oversight of the PS-ON and worked to identify synergistic opportunities and foster new collaborations. He also managed the PS-ON Trans-Network Projects program and the PS-ON Data Coordinating Center, which together leveraged the expertise of multiple teams to test new physical sciences–based cancer questions and fostered the sharing of PS-ON generated data.

Michelle A. Berny-Lang, PhD

Dr. Michelle Berny-Lang serves as a program director for CSSI. In this capacity, she participates in the organization, development, and evaluation of CSSI initiatives, managing a portfolio of grants and contracts, and working to facilitate multidisciplinary collaborations among researchers, NCI and NIH colleagues, and external partners.

As one example, Dr. Berny-Lang collaborated with the Veterans Health Administration (VHA) to launch the Big Data Scientist Training Enhancement Program (BD-STEP), designed to bring new expertise in physical sciences into Veterans Affairs (VA) clinical research teams with the goal of enhancing utilization and application of VA informatics systems to positively impact patient-centered care. She currently serves on the BD-STEP Advisory Council, is a member of the Health and Environmental Sciences Institute (HESI) Cardiac Biomarkers Working Group, and is a founding member of the National Heart, Lung, and Blood Institute (NHLBI) and NCI Cancer and Thrombosis Working Group.

Tony Dickherber, PhD

Dr. Tony Dickherber is director of the NCI Innovative Molecular Analysis Technologies (IMAT) program and a program director for CSSI, co-directing the Cancer Moonshot Team for New Technologies. He holds a master's degree in electrical engineering and PhD in bioengineering, both from the Georgia Institute of Technology. In addition to his duties of directing the IMAT program team and grant portfolio, Tony also assists with general strategic planning for initiatives involving emerging cancer technologies and in development of strategic partnerships.

Dr. Dickherber joined NCI as an AAAS Science & Technology Policy Fellow in 2009 to assist with the design and development of the Cancer Human Biobank (caHUB) with the Office of Biorepositories and Biospecimen Research. Prior to joining the NCI, Tony's research focused on innovative biosensor platforms for early detection of cancer and arrayable ion-trapping structures for quantum-bit computing at the Georgia Tech Microelectronics Research Center in Atlanta, GA. He also spent four years as a research engineer at the Georgia Tech Research Institute working on military-related telecommunications projects.

Frankie Cozzens Philips

Ms. Frankie Cozzens Philips serves as a program analyst for CSSI, where she supports the organization and management of CSSI programs, facilitating collaborations among researchers, NIH/ NCI staff, and industry colleagues. She also works to establish new projects and initiatives within CSSI.

One such effort she leads is a new program evaluation project on comparative assessment of research initiatives. The goal of this project is to assess methods to compare the effectiveness and scientific impact of initiatives developed within CSSI. She currently serves on the NCI Technology Refresh Program Committee, the International Travel and IT Policy Task Force, and represents CSSI as the NCI software coordinator.

Steven Cole

Mr. Steven Cole serves as a program analyst for CSSI. In this capacity, he provides analytics reporting, input, planning, and development of CSSI initiatives, working to further promote the mission of finding new and exciting technologies.

Mr. Cole's first role with the NCI was in 2007, where he served as an administrative assistant to the NCI Deputy Director. In 2009, he transitioned to program analyst and assisted with the inaugural funding of NCI's Provocative Questions Initiative. In this role, he assisted with the analysis of contracts, grants, and cooperative agreements for CSSI Programs and Initiatives. Using his experience at the NCI and Office of the Director, he hopes to play an important role in finding and implementing advanced technologies to support CSSI's mission to empower the entire cancer research community.

Norbert Tavares, PhD

Dr. Norbert Tavares is a 2017–18 AAAS Science & Technology Policy Fellow in CSSI. In this role, he supports the goals of CSSI to identify and promote emerging and innovative solutions to cancer research problems. He is involved in ongoing trans-NCI programs, including the Provocative Questions Initiative and initiatives related to the Beau Biden Cancer Moonshot^{SM}, such as the Human Tumor Atlas Network.

Dr. Tavares is a microbiologist with a background in bacterial physiology. Before coming to NCI, Dr. Tavares worked as a research associate at CuraGen Corp., where he was involved in the process validation of large-scale bacterial protein drug production. He also has four years of procurement experience from his previous work at ESPN/Disney Corp. Dr. Tavares has strong interests in problem solving and advancing science by supporting basic research, and the advancement of women and underrepresented individuals in science and other areas.

Sarah G. Elder, MS

Sarah Elder is a health communications fellow for CSSI. In this role, she spearheads the development of a uniform communication plan across multiple platforms and writes pieces about the scientific and programmatic activities of CSSI.

Sarah received her MS in medicinal chemistry and molecular pharmacology from Purdue University, where her research focused on the development of an analytical platform for disease biomarker discovery. Before coming to NCI, Sarah worked as a science writer for the Purdue student newspaper, *The Exponent,* in which she detailed scientific achievements and discoveries on campus. She also worked as a market research consultant for a medical device startup company in Oslo, Norway.

Office of Cancer Clinical Proteomics Research (OCCPR)

Table of Contents

1 Mission

The mission of the National Cancer Institute (NCI)'s Office of Cancer Clinical Proteomics Research (OCCPR) is to use proteomic data to improve the prevention, early detection, diagnosis, and treatment of cancer. The Office works to enhance understanding of the molecular mechanisms of cancer and streamline the translation of these findings into the clinic. The OCCPR also supports the development of promising proteomic and proteogenomic technologies as well as provides publicly accessible data, assays, and reagents to the research community.

2 Office/Program Evolution

Creation of NCI's Clinical Proteomic Tumor Analysis Consortium (CPTAC) was recommended to the NCI by the National Cancer Advisory Board's Working Group on Biomedical Technology, as a part of the Clinical Proteomic Technologies for Cancer (CPTC) initiative. This working group, commissioned in 2003 and chaired by Drs. Eric Lander and Lee Hartwell, provided the framework for a molecular diagnostic development program that utilized recent technological advances for detecting proteins in patient samples. The working group emphasized the importance of team science, streamlined sample collection, data standards, robust informatics platforms, and the availability of well-characterized reagents. In response to this recommendation, NCI held a series of workshops that led to the development of the CPTAC program.

Subsequently, the NCI Executive Committee approved this original concept encompassing biomarker discovery and technology development. In interactive reviews with the NCI Board of Scientific Advisors (BSA), it was recommended that the program focus primarily on proteomic technology assessment and development. These recommendations were incorporated into the concept, and in late June 2005 the BSA unanimously approved the CPTC initiative to address a lack of reproducibility and transferability of measurement technologies across laboratories and a lack of quality reagents for the cancer research community. To achieve the mission of OCCPR and with input from the extramural scientific community, CPTAC was established in 2006.

CPTAC 1.0 (2006) sought to address these challenges through analytical rigor and reproducibility between laboratories, standardization of sample collection, proper sample storage and processing, experimental design, data analysis and reporting, and community resources to the public. Program highlights include the standardization of mass spectrometry (MS) methodologies for untargeted protein analyses (discovery proteomics) and targeted protein analyses (confirmatory proteomics), including their reproducibility and transferability among labs; adoption of selective/multiple reaction monitoring mass spectrometry (S/MRM-MS) assays by clinical reference laboratories; development of an open-source computational tool for designing MRM assays supported by all major instrument vendors (Skyline software); development of mock 510(k) device clearance documents on how to design studies to address regulatory approval requirements on multiplexed protein-based In Vitro Diagnostics (IVD) assays/platforms (mass spectrometry and protein array) done in coordination with the US Food and Drug Administration (FDA) and the American Association for Clinical Chemistry (AACC); and development of proteomic data-sharing policies (Amsterdam Principles) that are supported by peer-reviewed journals.

To begin to apply these technological outputs, the second phase of CPTAC was initiated in 2011 (CPTAC 2.0; (unanimously approved by BSA in March 2010)). The goal of this pilot program was to address biological questions using analytical outputs from CPTAC 1.0. A coordinated team effort was necessary to apply CPTAC's standardized workflows to three genomically characterized tumors (colorectal, breast, and ovarian; from The Cancer Genome Atlas – TCGA). The aim was to systematically identify biologically relevant proteins derived from alterations in cancer genomes, providing insight into the molecular basis of cancer beyond what can be elucidated using genomics alone.

More recently, CPTAC's "proteogenomics" approach has integrated comprehensive proteomics with genomics to produce a more unified understanding of cancer biology. CPTAC has demonstrated the scientific benefits of this approach on a large number of clinical samples. These efforts have paved the way for possible therapeutic interventions for patients, while also creating resources that are widely used by the global cancer community.

CPTAC was reissued in late 2015 (unanimously approved by the BSA in June 2015) to expand its efforts on the comprehensive proteogenomic characterization of additional cancer types with data and assays released to the public. This also enabled CPTAC to support clinically-relevant research projects geared at elucidating the biological mechanisms of therapeutic response, resistance, and toxicity. CPTAC 3.0 (2016) will leverage the achievements in proteogenomics from CPTAC 2.0, with all investigators collaborating and sharing data and expertise across the consortium, to address two major goals:

- Comprehensively characterize the proteome and genome of tumors from up to five new cancer types to better understand the interplay between genes and proteins.

- Address questions of biology in the context of a clinical trial (NCI-sponsored) using a proteogenomics-based research approach. Focus on mechanisms of drug response and resistance.

2.1 CPTAC: Clinical Proteomic Tumor Analysis Consortium (CPTAC) 1.0

2.1.1 Goals

In the early 2000s, clinical proteomics research lacked standardized technologies and methodologies as well as well-characterized, quality reagents, critically hampering this emerging scientific field. CPTAC 1.0, comprised of five Clinical Proteomic Technology Assessment Centers, was created by the NCI in 2006 to address these issues.

2.1.2 Key Accomplishments

Rigor and Reproducibility

CPTAC 1.0 pioneered an efficient translational pipeline to winnow down a large number of protein candidates from discovery proteomics (untargeted MS-based workflow) [1]. This pipeline utilizes a confirmatory step (multiplexed targeted MS-based workflow [2]) prior to large-scale downstream clinical validation studies. CPTAC teams successfully addressed measurement variability issues in proteomics resulting from analytical platforms and biospecimen handling by implementing metrics at every step of the proteomics pipeline. In collaboration with the National Institute of Standards and Technology (NIST), the variability of discovery proteomics due to the stochastic nature of MS was addressed through the use of a yeast reference material [3] and MassQC software [4] for monitoring instrument performance. At the confirmatory stage, the variability of targeted MS-based quantitative assays was addressed using quality controlled peptide standards, standard operating procedures (SOPs), and Skyline software [5] to reproducibly verify discovered targets.

By enhancing analytical capabilities to measure proteins accurately and reproducibly, the CPTAC program demonstrated the effectiveness of a multi-disciplinary, multi-institutional approach in addressing long standing problems of analytical variability in proteomics. This has paved the way to overcome the inherent variability of specific analytical platforms to uncover and quantify real biological differences.

Proteomics Open Data-Sharing Policy (Amsterdam Principles)

To emulate the data-sharing path taken by the genomic community, development of open data-sharing policies in proteomics was critical to maximize the benefit of high-quality data for the greater research community. In 2008, the OCCPR supported the "International Summit on Proteomics Data Release and Sharing Policy" workshop, which explored various approaches for data release and sharing principles that would make large-scale proteomic data widely available on a pre-competitive basis. As a result, the "Amsterdam Principles" were born to help identify and address potential roadblocks to rapid and open access data [6]. Two years later, the OCCPR convened the "International Workshop on Proteomic Data Quality Metrics" as a follow-up, addressing issues in the development and use of methods for open access proteomic data. The workshop enumerated the key principles underlying a framework for data quality assessment in MS data that will meet the needs of the research community, journals, funding agencies and data repositories. Subsequently*, leading proteomics journals (e.g.,* Molecular and Cellular Proteomics*) and international research programs have adopted the Amsterdam Principles to ensure open access to proteomics data* [7].

1. J Proteome Res. 2010; Feb 5;9(2):761-76. PMID: 19921851
2. Nat Biotechnol. 2009; Jul;27(7):633-41. PMID: 19561596
3. Mol Cell Proteomics. 2010; Feb;9(2):242-54. PMID: 19858499
4. Mol Cell Proteomics. 2010; Feb;9(2):225-41. PMID: 19837981
5. Bioinformatics. 2010; Apr 1;26(7):966-8. PMID: 20147306
6. J Proteome Res. 2009; Jul;8(7):3689-92. PMID: 19344107
7. J Proteome Res. 2012; Feb 3;11(2):1412-9. PMID: 22053864

Regulatory Science and Clinical Chemistry Initiatives (Partnerships with FDA and AACC)

To educate the community about designing studies that meet the clinical and analytical standards set by the FDA for multiplex, protein-based assays, CPTAC investigators and the FDA's Office of In Vitro Diagnostics developed mock 510(k) pre-submissions on platforms being assessed through the CPTAC network. These first-of-their-kind analytical validation review documents, done in coordination with the FDA and the AACC, illustrate the details involved in regulatory pre-submissions. The documents serve to benefit the global proteomics community and to streamline the regulatory process by providing examples of submission formatting. These mock pre-submissions, along with the comments from the FDA review staff, were published in a special issue of *Clinical Chemistry* [8,9]. These efforts in regulatory science earned a *Leveraging/Collaboration Award* from the FDA in 2010.

2.2 CPTAC 2.0

2.2.1 Goals

CPTAC 2.0 applied state-of-the-art, standardized proteomic workflows developed from CPTAC 1.0 on genomically characterized tumors (from TCGA), adding an additional layer of functional biology to understand cancer biology. The goal was to determine if additional biological insights would be identified, including details that are difficult or impossible to obtain solely through genomic approaches. CPTAC 2.0 was comprised of five Proteome Characterization Centers (PCCs) with expertise in proteomics, genomics, cancer biology, oncology, and clinical chemistry that performed coordinated research to comprehensively characterize and analyze the cancer specimens selected for study. CPTAC's unique proteogenomic approaches (comprehensive proteomics combined with comprehensive genomics) successfully demonstrated that integrating these two scientific disciplines can produce a more complete understanding of cancer biology and identify potential therapeutic targets. This research has led to the development of some of the world's largest public repositories (open community resources) of proteogenomic sequence data and targeted proteomic fit-for-purpose assays.

2.2.2 Key Accomplishments

Comprehensive Proteomic Characterization of TCGA Genomically Characterized Tumors (Biology Studies)

Colorectal Cancer

This first large-scale proteogenomic study produced several key findings. First, measurements of mRNA abundance did not reliably predict protein abundance. This discordance was not surprising, because many regulatory controls lie between RNA and protein expression. Second, most of the focal amplifications (increased amounts of certain chromosome segments) observed in the earlier genomic analyses of the same tumors did not result in corresponding elevations in protein level. Proteomic analyses identified a few amplifications that add dramatic effects on protein levels and may represent

8. Clin Chem. 2010; Feb;56(2):165-71. PMID: 20007858
9. Clin Chem. 2010; Feb;56(2):237-43. PMID: 20007859

potentially important targets for diagnosis or therapeutic intervention. Third, proteomics identified five colon cancer subtypes, including classifications that could not be derived from genomic data. Interestingly, protein expression signatures for one of the subtypes not differentiated by the corresponding transcriptomic subtype [10] indicated molecular characteristics associated with highly aggressive tumors and poor clinical outcome.

Ovarian Cancer

Deep proteogenomic characterization of TCGA high-grade serous ovarian tumors yielded several insights, such as the influence of copy number changes on the proteome, the proteins associated with chromosomal instability, the sets of signaling pathways that diverse genome rearrangements converge on, and the ones most associated with short overall survival. Specific protein acetylations associated with homologous recombination deficiency suggested a potential means for stratifying patients for therapy. In addition to providing a valuable resource, these findings generated a view of how the somatic genome drives the cancer proteome and associations between protein and posttranslational modification (PTM) levels (phosphorylation) and clinical outcomes in high-grade serous carcinomas [11].

Breast Cancer

This study produced a broad overview of the proteomic and the phosphoproteomic landscape across a set of breast cancer tumors that had been genomically characterized by the TCGA project. Although TCGA produced an extensive catalog of somatic mutations found in cancer, the effects of many of those mutations on cellular functions or patients' outcomes remain unknown. Furthermore, not all mutated genes are true "drivers" of cancer; some are merely "passenger" mutations that have little functional consequence. Thus, winnowing the list of candidate genes by studying the activity of their protein products can help identify useful therapeutic targets.

This analysis uncovered new protein markers and signaling pathways for breast cancer subtypes and tumors carrying frequent genomic alterations such as PIK3CA and TP53 mutations. It also correlated copy number alterations in some genes with protein levels, identifying ten new candidate regulators. Two of these candidate genes, SKP1 and CETN3, are connected to the oncogene EGFR, a marker for a particularly aggressive breast cancer subtype characterized by "basal-like" tumors [12].

Understanding Tumor Preanalytical Variables

The adage of "garbage in, garbage out" in proteomics research refers to non-biologically relevant observations that can result from preanalytical variables. Improper biospecimen collection, storage, and processing are some of the culprits that can create an array of artifacts prior to labor-intensive analytical endeavor in proteomics workflows. For example, TCGA retrospectively collected treatment-naïve tumors that involved a time-lapse from excision to freezing (cold ischemia) of up to 60 minutes, but there was a lack of actual recording of "excision-to-freeze" delay time. Furthermore, TCGA tumor samples were embedded in optimal cutting temperature compound (OCT) commonly used for histological evaluation, which raised the possibility of signal interference when performing ultra-deep proteomic analysis. Due to the lack of understanding of the effect of these preanalytical variables on

10. Nature. 2014; Sep 18;513(7518):382-7. PMID: 25043054
11. Cell. 2016; Jul 28;166(3):755-765. PMID: 27372738
12. Nature. 2016; Jun 2;534(7605):55-62. PMID: 27251275

sample quality, the CPTAC program became involved in the evaluation of the effect of cold ischemia and OCT on the stability of proteome and/or phosphoproteome as well as the *N*-glycoproteome prior to sample analysis. To account for possible analytical drift over time among the platforms used in CPTAC laboratories, development of a common reference standard for establishing comparable platform performance metrics across centers was necessary.

Preanalytical Variable 1: Excision-to-Freeze Delay Time (Tumor Ischemia)

The effect of cold ischemia time on the stability of proteome, the phosphoproteome and *N*-glycoproteome of several human ovarian tumors and patient-derived breast cancer xenograft (PDX) tissues frozen at defined ischemic intervals (0, 5, 30, 60 min) was studied. Results revealed *that after 60 minutes of cold ischemia, there were no significant changes detected in the global proteome,* N-*glycoproteome, or most of the >25,000 phosphosites (pSer, pThr, pTyr) of each tumor type. However, fluctuations in protein phosphorylation at specific phosphosites, were observed in up to ~24% of the phosphoproteome starting as early as 5 minutes after tumor excision, as demonstrated by different clusters of dynamic changes in phosphorylation* (Fig. 1). Such effects were biologically coherent (not random), with pathways affected reflective of responses to ischemic insult including stress response, transcriptional regulation, and cell death, among others [13].

These results have been used to help the research community avoid erroneous interpretation of preanalytical variability as meaningful biology. CPTAC used the results to (1) develop an ischemic signature phosphoproteomics database and (2) develop standardized best practice procedures for prospective biospecimen procurement tailored for proteomics, including prospective tumor collection for verification purposes. These studies provide useful resources for the research community and suggest caution in interpreting activation of stress pathways in retrospectively collected tumor samples. CPTAC's deep characterization of the impact of cold ischemia time on tumor samples, especially with respect to biologically relevant phosphosite regulations, has informed recommendations for proteomic sample collection/processing that will be distributed and enforced by the College of American Pathologists.

Preanalytical Variable 2: Optimal Cutting Temperature Compound

Most TCGA tumor samples contain Optimal Cutting Temperature Compound (OCT), a compound commonly used to embed and stabilize specimens for tissue section cutting. OCT contains polyethylene glycol (PEG) polymers, which carry through proteomic sample preparation and suppress ion signals in MS proteomic analyses. CPTAC investigators developed a variety of OCT removal protocols to minimize the interference of PEG polymers in a variety of tissue types [14] in proteome analyses. SOPs for OCT removal developed by the CPTAC program are publicly available for the research community. CPTAC also demonstrated that variation in OCT content between TCGA tumor specimens does not introduce protein abundance differences after OCT removal, ensuring confident identification and quantification of proteins from subsequent TCGA sample analyses.

13. Mol Cell Proteomics. 2014; 13(7):1690-1704. PMID: 24719451
14. Anal Biochem. 2015; 469:27-33. PMID: 25283129

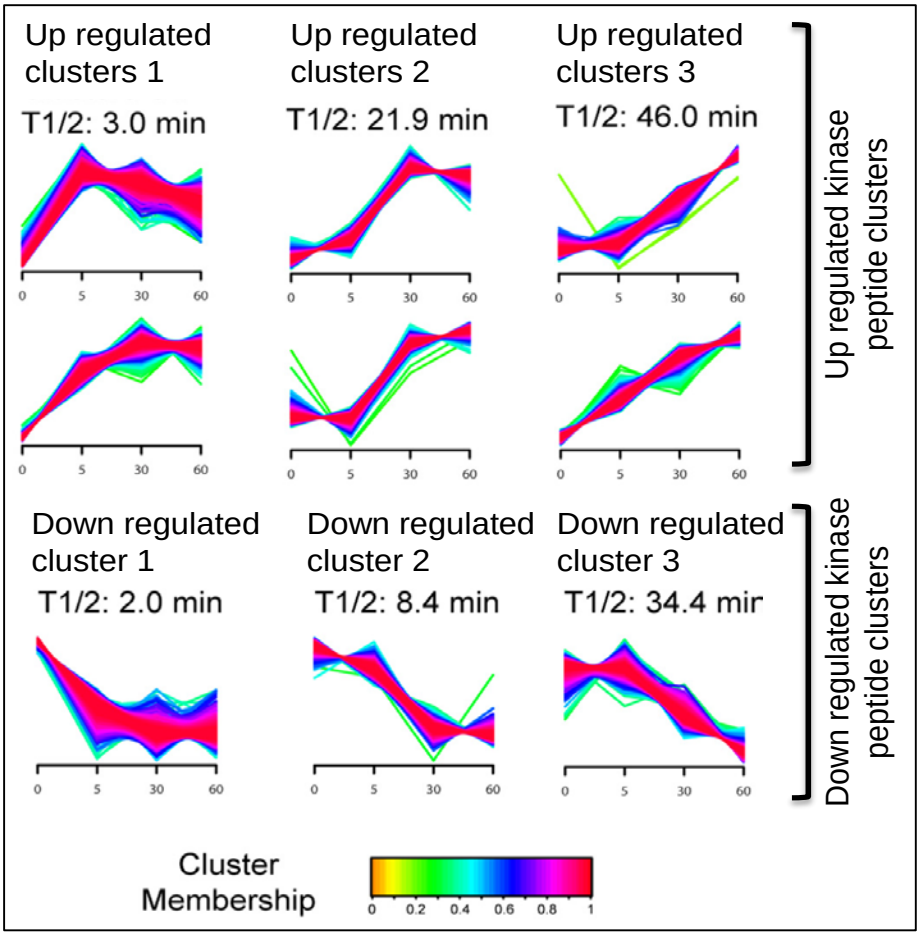


Figure 1 **Temporal Dynamics of Phosphorylation Changes Resulting from Cold Ischemia During Surgical Procedures**

Generation and Dissemination of Proteomic Community Resources

CPTAC is focused on advancing clinical cancer research through proteomic measurement capabilities. Its reputation of quality is internationally recognized. The CPTAC program currently has three dedicated proteomic public resources for disseminating its research findings. They are a Data Portal, Assay Portal, and Antibody Reagents Portal – each with SOPs that are downloadable.

Data Portal

The ability to share and re-use data across the biomedical research community is vital to accelerating scientific discovery and clinical translation. The CPTAC Data Portal is a centralized repository for the public dissemination of proteomic sequence datasets collected by CPTAC, along with corresponding genomic sequence datasets. It also houses the analyses of CPTAC's raw MS-based data files (mapping of spectra to peptide sequences and protein identification) by individual investigators from CPTAC as well as a Proteomics Common Data Analysis Pipeline (P-CDAP). A core principle of CPTAC is the sharing and re-use of data across the biomedical research community, which is vital to accelerating scientific discovery and clinical translation to patient care. The *NCI CPTAC Data Portal represents the largest public*

repository of proteogenomic comprehensive sequence datasets, essentially making it a Proteogenomic Cancer Atlas. Proteomic data and related data files are organized into datasets by study, sub-proteome, and analysis site. All data are freely available to the public and subject to data use guidelines. Reference mass spectral peptide libraries resulting from these studies may also be downloaded freely from the NIST peptide library.

Assay Portal

The CPTAC Assay Portal serves as a centralized public repository of "fit-for-purpose" multiplexed quantitative MS-based assays [15]. *Unlike databases or libraries currently in use, the CPTAC Assay Portal contains analytically validated assays and SOPs that can be used to compare results across the board. This provides a common ground for clinicians, systems biologists, and analytical chemists to facilitate widespread adoption of targeted MS assays by disseminating SOPs, reagents, and assay characterization data.* Because targeted proteomic assays use MS to get direct readout of a signal, they eliminate issues associated with conventional protein detection systems like ELISA, which can only provide a surrogate readout of a signal.

The goal of the portal is to widely disseminate highly characterized proteomic assays to the global research community, with access to SOPs, reagents, and assay characterization/validation data in support of the National Institutes of Health (NIH)'s Rigor and Reproducibility Principles and Guidelines. Targeted MS assays are characterized per CPTAC's "fit-for-purpose" assay characterization guidelines. The *NCI CPTAC Assay Portal represents the largest public repository of fit-for-purpose proteomic targeted MS assays*. The portal is also designed to bring together biologists seeking to ask hypothesis-driven questions about the proteome they study and analytical chemists equipped to perform targeted proteomic assays. Furthermore, a trademark seal of "CPTAC Characterized Assay" helps users evaluate assay performance before investing time, money, and energy in adopting and deploying the assays in their own laboratories. Assays in the portal adhere to validation guidelines developed in coordination with FDA and AACC. Assays are searchable through NIH's PubMed databases (via LinkOut), with analytical parameters downloadable into a laboratory's mass spectrometer via Skyline (a freely available open-source quantitative proteomics software tool developed in CPTAC 1.0).

15. Nat Methods. 2014 Jul;11(7):703-4. PMID: 24972168

Antibody Reagents Portal

Antibodies are among the most commonly used tools in biological sciences, put to work in many experiments to identify and isolate other molecules. The CPTAC Antibody Portal serves as a community resource of unbiased antibody validation in a centralized location for a large variety of renewable antibodies (monoclonal), with all associated SOPs and characterization data made publicly available. Antibodies and hybridomas are branded with the "CPTC" prefixes, allowing them to be tracked and identified in searchable databases that include synchronization with the NIH PubMed databases (via LinkOut) to ensure broad accessibility from the global scientific community.

Rigorous antibody characterization is performed at the Antibody Characterization Laboratory (ACL), an intramural laboratory located at the Frederick National Laboratory for Cancer Research (FNLCR) in Frederick, Maryland. While initially created as a resource for CPTAC investigators, the program has effectively transitioned to one that now solicits the entire cancer research community for requests on cancer-related protein or peptides to be used for antibody generation.

The Antibody Characterization Program opens the reagent target request to the research community approximately once a year. The Antibody Scientific Committee evaluates proposed antibody targets based on their relatedness to cancer, the availability of commercial antibodies/affinity binders for the target, and justification and contribution to existing NCI-funded projects. Once approved, up to three affinity binders per protein/peptide target are generated and characterized using standardized assays including but not limited to: SDS-PAGE, Western Blot, ELISA, immunohistochemistry, immuno-mass spectroscopy, and Surface Plasmon Resonance. All assays are done in alignment with the NIH's Rigor and Reproducibility Principles and Guidelines. Affinity binders, such as monoclonal antibodies, are subsequently made available to the research community through the Developmental Studies Hybridoma Bank at the University of Iowa and/or other third-party vendors (public-private partnerships). The *NCI CPTAC Antibody Portal represents one of the largest public repositories of extensively-characterized monoclonal antibodies, with reagents and SOPs accessible to the public*.

Rigor and Reproducibility

Analytical Benchmarking of Laboratories (Development of a Proteomics Reference Material)

For discovery proteomics, CPTAC teams employed a pair of reference patient-derived xenograft proteomes (CompRef) for initial platform qualification to determine intra- and interlaboratory reproducibility. Daily analytical monitoring of data collection was also carried out for hundreds of TCGA tumors, allowing for laboratories to monitor analytical drift over the course of an analysis. These two xenografts, representing basal and luminal B human breast cancer, were extensively evaluated by genomics and proteomics. These data represent a unique opportunity to evaluate the stability of proteomic differentiation by MS instruments over many months through individual instruments or across instruments running dissimilar workflows [16].

16. J Proteome Res. 2016; Mar 4;15(3):691-706. PMID: 26653538

International Targeted Proteomics Assay Benchmarking Study (Demonstrating Reproducibility and Transferability of Targeted Proteomic Assays)

Multiplexed MRM-MS assays were conducted in three laboratories (two in the USA and one in South Korea) on breast cancer cell lines using standardized procedures. This interlaboratory effort sought to extend the effort of the seminal CPTAC 1.0 multi-site study by increasing the multiplexity of targeted assays and including a site outside the USA. The laboratories generated comparable results that discriminated among molecular subtypes of breast cancer, identified genome-driven changes in the cancer proteome, and provided additional information about the cell lines beyond genomic profiles. These results established the feasibility of a large-scale effort to develop and deploy MRM-MS assay resources on an international level [17].

Fit-for-Purpose Assay Community Guidelines (Development of Community Guidelines That Define Targeted Proteomic MS Assays)

NCI and National Heart, Lung and Blood Institute (NHLBI), in coordination with FDA and AACC, held a workshop with representatives from research communities to address the wide range of criteria applied to claim successful assay development. The group discussed analytical goals and the experimental evidence required to confirm that assays work as intended and achieve the required level of performance. The group collectively defined best practices for MS-based assay development using a "fit-for-purpose" approach.

Assays were classified into three tiers based on their performance and the extent of analytical characterization. Importantly, these *CPTAC assay guidelines have been adopted by the journal Molecular and Cellular Proteomics*. This scheme was designed to help users evaluate assay performance before investing time, money, and energy into adopting and deploying these assays in their own laboratories. To comply with such criteria, assays currently residing in the CPTAC Assay Portal are comprised of a minimum of "Tier 2" characterization [18]. To help populate this high-quality resource, a document detailing this scheme provides instructions to help investigators submit their own targeted assays to the portal.

3 CPTAC 3.0

3.1 Program Goals

Building upon CPTAC 2.0's achievements stemming from the integration of genomics and proteomics, CPTAC 3.0 seeks to expand the understanding of biological processes in additional cancer types. The goals of CPTAC 3.0 are to 1) systematically identify proteins derived from alterations in cancer genomes to gain insights into the molecular basis of cancer that cannot be fully elucidated through genomics alone, and 2) to accelerate the translation of such molecular findings into the clinic. This project will use a proteogenomic approach that prioritizes driver genes, enhances understanding of pathogenesis through proteomic subtyping, illuminates dynamic alterations in PTMs responsible for the dysregulation

17. Nat Methods. 2014; Feb;11(2):149-55. PMID: 24317253
18. Mol Cell Proteomics. 2014; Mar;13(3):907-17. PMID: 24443746

of cancer signaling networks and pathways, and improves understanding of drug response and resistance to therapies in a clinical setting.

3.2 Program Structure

CPTAC 3.0 operates through three components that coordinate research activities. The three components are:

3.2.1 Proteome Characterization Centers (PCCs; RFA-CA-15-021)

The PCCs work as a group to perform comprehensive proteomic characterizations on genomically characterized biospecimens provided by NCI, and quantitative measurements of protein targets of biological/clinical relevance. The PCCs are:

Broad Institute, Cambridge, MA
Principal investigators: Steven A. Carr, PhD; Michael Gillette, MD, PhD

Johns Hopkins University, Baltimore, MD
Principal investigators: Daniel W. Chan, PhD; Hui Zhang, PhD; Zhen Zhang, PhD

Pacific Northwest National Laboratory, Richland, WA
Principal investigators: Tao Liu, PhD; Richard D. Smith, PhD

3.2.2 Proteogenomic Data Analysis Centers (PGDACs; RFA-CA-15-023)

The PGDACs integrate, visualize, and analyze genomics, transcriptomics, proteomics, imaging, and clinical datasets to improve the understanding of genome-proteome relationships and the interplay/regulation of signaling pathways involved in cancer. The PGDACs are:

Baylor College of Medicine, Houston, TX
Principal investigator: Bing Zhang, PhD
Additional area of focus: pathway and molecular network visualization

Broad Institute, Cambridge, MA
Principal investigators: Denkanikota R. Mani, PhD; Gad Getz, PhD
Additional area of focus: patient-centric protein databases and proteogenomic data analysis pipeline and visualization portal (CPTAC Firehose)

Icahn School of Medicine at Mount Sinai, New York, NY
Principal investigators: Pei Wang, PhD; Eric Schadt, PhD
Additional area of focus: missing proteomics data, batch effects, and global regulatory networks

New York University School of Medicine, New York, NY; Washington University in St. Louis, St. Louis, MO; Pacific Northwest National Laboratory, Richland, WA
Principal investigators: David Fenyo, PhD; Li Ding, PhD; Samuel Payne, PhD

Additional area of focus: quantitative trait loci-centered approach, outlier analysis using PTMs data, and druggability of kinase inhibitors using protein structural information

University of Michigan, Ann Arbor, MI
Principal investigators: Alexey I. Nesvizhskii, PhD; Arul Chinnaiyan, MD, PhD
Additional area focus: quality control scoring measures to identify genomic/transcriptomic variants expressed at protein level

3.2.3 Proteogenomic Translational Research Centers (PTRCs; RFA-CA-15-022)

Three PTRCs collaborate with NCI to address questions of biology in a clinical trial context through integrated proteomics and genomics. The PTRCs focus is on understanding and predicting drug response and resistance to therapies. The PTRCs are:

Baylor College of Medicine, Houston, TX; Broad Institute, Cambridge, MA
Principal investigators: Matthew Ellis, MB, BChir, BSc., PhD; Steven A. Carr, PhD
Cancer focus: Breast cancer

Fred Hutchinson Cancer Research Center, Seattle, WA; Massachusetts General Hospital, Boston, MA
Principal Investigators: Amanda Paulovich, MD, PhD; Michael Birrer, MD, PhD
Cancer focus: Epithelial ovarian cancer

Pacific Northwest National Laboratory, Richland, WA, Oregon Health & Science University, Portland, OR
Principal Investigators: Karin Rodland, PhD; Brian J. Druker, MD
Cancer focus: Acute myeloid leukemia

All CPTAC investigators collaborate, share data and expertise across the consortium, and participate in consortium activities. Data (genomics [NCI Genomic Data Commons], proteomics [CPTAC Data Portal], The Cancer Imaging Archive [TCIA]), assays, and reagents are made available to the public as a community resource to accelerate cancer research and advance patient care.

3.3 Consortium Governance

3.3.1 Steering Committee (SC)

The CPTAC SC serves as the primary governing body of the CPTAC program. The committee is jointly established by all the awarded project directors/principal investigators of the PCCs, the PGDACs, the PTRCs, and the NCI program staff members. The CPTAC SC provides strategic coordination for the activities of the PCC network and the CPTAC program overall.

3.3.2 Working Groups (WGs)

Working groups are organized as subcommittees of the SC, with members composed of one or more subject matter experts designated by the team lead, from each of the funded CPTAC components. The WGs independently develop recommendations according to their specific missions, which are then presented to the SC for approval and subsequent implementation.

3.3.3 Disease Working Groups (DWGs)

DWGs are interdisciplinary and consist of international researchers and physicians. They are recruited and designated by the CPTAC SC to identify important biological and clinical questions for a designated cancer type. Questions identified by a DWG help guide which types of molecular data should be generated and which CPTAC platforms should be used. They also help guide the analysis of the data produced.

3.3.4 External Scientific Panel (ESP)

The ESP provides independent assessment of research directions and progress of the CPTAC awardees. The ESP annually evaluates research conducted by CPTAC members. This includes the review of the overall program metrics, progress of individual awardees, strategic plans, etc. The panel may also recommend to the SC new research opportunities to explore, adjustments in priorities, and/or actions to advance the overall CPTAC goals. ESP members are not principal investigators for or funded by the CPTAC program.

3.4 Scientific Approaches

CPTAC employs two complementary scientific approaches, a "Targeting Genome to Proteome" (Targeting G2P) approach and a "Mapping Proteome to Genome" (Mapping P2G) approach, to address biological questions from data generated from a sample. In a "Targeting G2P" approach, a genome dataset defines the protein sequences (candidates) to be targeted in proteomic measurements. With this approach, laboratories detect and quantify protein products that correspond to genomic abnormalities including splice variants, mutations, insertions, deletions, rearrangements, copy number aberrations, or epigenomic changes detected at the genome level.

In a "Mapping P2G" approach, the integration of the genomic and proteomic datasets is postponed until after the completion of both types of measurements. An advantage of this approach is that it allows a broader inventory of the detectable proteins in a tumor, including up- or down-regulation of protein abundance, and identification of PTMs that may be critical to cell signaling pathways and networks. In addition, this integrative approach can be used to improve the quality of genome annotations, as it uses proteomic information to confirm protein-coding genes. The combination of these two approaches is anticipated to produce a more comprehensive inventory of the detectable proteins in a tumor and advance our understanding of cancer biology. The targets identified are subsequently configured into multiplexed targeted assays that can be tested in relevant cohorts of biospecimens.

3.5 Accomplishments to Date

3.5.1 Completion of Harmonization and Benchmarking Studies

The PCCs have collaborated extensively to evaluate proteomic workflows for global proteome and phosphoproteome analyses on CompRef benchmarking reference materials using state-of-the-art MS and multiplexed labeling strategies. To demonstrate the reproducibility and high quality of untargeted MS data regardless of sites, the PCCs standardized the workflows starting from protein extraction from tissue samples to mass spectrometric runs. This interlaboratory demonstration study prior to tumor analyses has generated highly correlated data between sites in terms of identification and quantitation of tens of thousands of proteins and phosphopeptides (manuscript in preparation). Currently, proteogenomic analyses of renal clear cell carcinoma and endometrial cancer samples (100 cases per cancer type) are underway.

3.5.2 Development/Refinement of Proteogenomic Data Analysis Pipeline

The PGDACs have completed a coordination plan that summarizes proposed analysis approaches from each PGDAC (developed and/or refined), and establishes a core analysis plan (similar to P-CDAP for proteomics data) for all cancer types studied by all PGDACs. In addition, the PGDACs have established requirements for genomic raw data harmonization by the Genome Data Commons, evaluated options for proteomic raw data processing in collaboration with the P-CDAP, and developed a tool catalog summarizing algorithms available for proteogenomic analyses.
STOP HERE AT 4:57

4 Extension Programs of CPTAC (Cancer Moonshot)

4.1 Applied Proteogenomics OrganizationaL Learning and Outcomes (APOLLO) Network

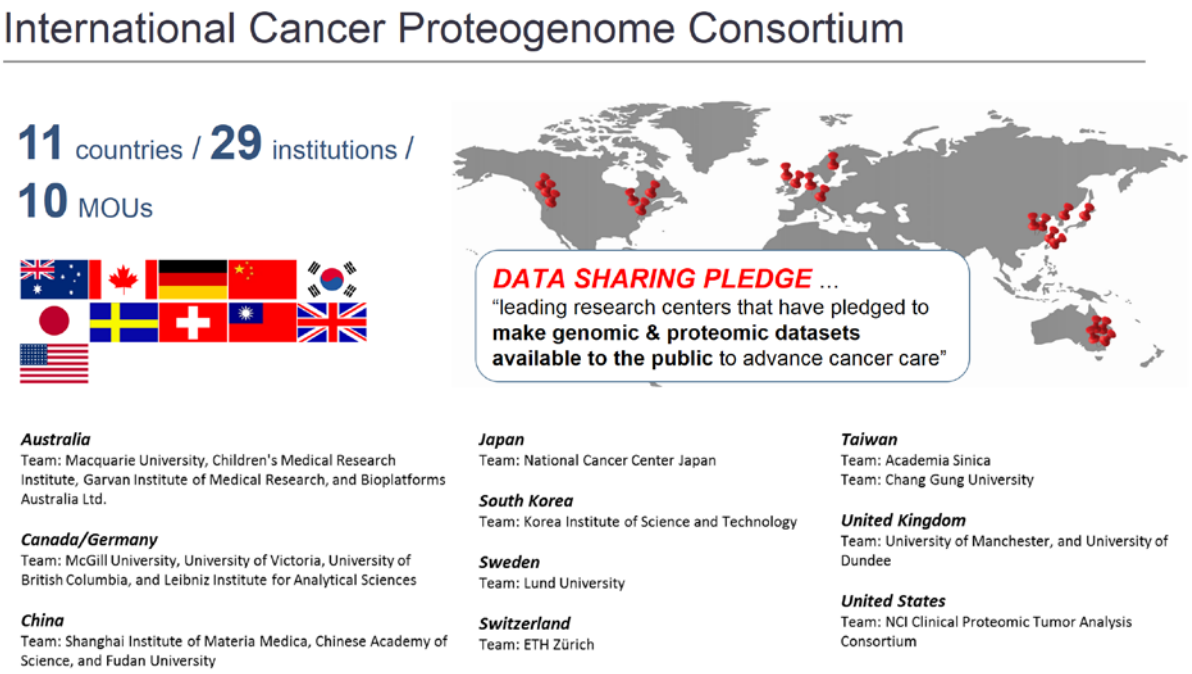
The APOLLO network is a collaboration between NCI, the Department of Defense (DoD), and the Department of Veterans Affairs (VA) to incorporate proteogenomics into patient care. It seeks look beyond the genome to the activity and expression of the proteins that the genome encodes. The APOLLO network was launched in 2016 in response to the Beau Biden Cancer Moonshot^{SM} challenge for federal agencies to work together to hasten the progress of cancer research [19]. The network increases the resources devoted to proteogenomics research and accelerates its progress. Partnering with the nation's two largest health systems – DoD and VA – allows NCI to study a larger number of patients and obtain results more efficiently.

The data will be curated and made available publicly through the NCI Genomic Data Commons, CPTAC Proteomic Data Portal, and the NCI Division of Cancer Treatment and Diagnosis (DCTD) Cancer Imaging Archive. Using all the data available (analytical, invasive, noninvasive, and clinical) will enable researchers to study the relationships among these data, validate results, and develop predictive and prognostic models to improve patient care. The DoD agency leading the APOLLO network is the Murtha Cancer Center (MCC) at the Uniformed Services University of the Health Sciences (USUHS) in Bethesda, Maryland. DNA sequencing and RNA sequencing will be performed by The American Genome Center (TAGC) at USUHS. Proteomic workflows will be analyzed on multiple platforms by NCI's CPTAC, led by the Fred Hutchinson Cancer Research Center in Seattle, Washington, and the MCC's Clinical Proteomics Platform (CPP), led by the Gynecologic Cancer Center of Excellence (GYN-COE) in Fairfax, Virginia.

19. Clin Pharmacol Ther. 2017 May;101(5):619-621. PMID: 28187513

4.2 International Cancer Proteogenome Consortium (ICPC)

The ICPC is a voluntary scientific organization that provides a forum for collaboration among some of the world's leading cancer and proteogenomic research centers. ICPC spans 11 countries, encompassing 29 institutions connected through 10 MOUs. Catalyzed by new efforts in precision medicine and the Beau Biden Cancer Moonshot^{SM} initiative to encourage international cooperation and investments in cancer research and care, ICPC was launched in late 2016. The ICPC brings together more than a dozen countries to investigate applications of proteogenomics in predicting cancer treatment success. The ICPC supports the public sharing of cancer-associated proteogenomic data for use by cancer researchers and physicians around the world to accelerate translation of results to patient care. Genomic, proteomic, and imaging data will be shared through NCI's Genomic Data Commons, CPTAC Proteomic Data Portal, and the NCI DCTD Cancer Imaging Archive.



Current institutions include (alphabetical order):

Australia

Team: Macquarie University, Children's Medical Research Institute, Garvan Institute of Medical Research, and Bioplatforms Australia Ltd.

Canada/Germany

Team: McGill University, University of Victoria, University of British Columbia, and Leibniz Institute for Analytical Sciences

China

Team: Shanghai Institute of Materia Medica, Chinese Academy of Science, and Fudan University

Japan

Team: National Cancer Center Japan

South Korea

Team: Korea Institute of Science and Technology
Team: Korea University

Sweden

Team: Lund University

Switzerland

Team: ETH Zürich

Taiwan

Team: Academia Sinica
Team: Chang Gung University

United Kingdom

Team: University of Manchester and University of Dundee

United States

Team: NCI Clinical Proteomic Tumor Analysis Consortium

5 Future Vision of OCCPR

With input from the research community, OCCPR will continue to develop programs that proteogenomically characterize new cancer types, while providing outputs and knowledge (data, assays, and reagents) to the public as a global community resource– creating a Human Cancer Proteogenome Atlas. This will significantly advance our knowledge of human cancer genomes and their functional impact, while bringing scientists closer to understanding the workings of a cancerous cell.

Moving forward, OCCPR (mainly through CPTAC) will continue to investigate the anticipated value of proteogenomics in a clinical setting. In the past decade, genomic studies have helped make significant progress in stratifying patients. This has been critical to the discovery and development of targeted therapies and predictive models. For example, previous studies in colon cancer indicated that KRAS mutations in codons 12 or 13 resulted in a lack of therapeutic response to anti-EGFR antibody treatment and an increased risk of toxicity. The identification of similar drug response/toxicity predictions associated with targeted therapeutics is highly critical to the success of precision medicine. However, the relatively rapid acquisition of resistance to such treatments significantly limit their utility. It is anticipated that complementing genomic analysis with comprehensive proteomic analysis would provide new insights to tumor resistance predicting clinical response to therapeutic agents.

To ensure a continuance of state-of-the-art proteomic technologies entering the CPTAC program (and proteomics community) and subsequent clinical environment, OCCPR will continue to support research in technology development/optimization in proteomics and bioinformatics, both paramount to proteogenomic research. In addition, OCCPR will expand its interaction/coordination with the FDA and AACC, respectively. Development of high-sensitivity, high-content, multidimensional technologies will be an integral part of comprehensive molecular characterization of cells required to further our understanding of the complex biology of cancer. Approaches to increase analytical throughput and resolution of analyses, sample fewer materials, streamline sample preparation with automation, and simplify bioinformatic analyses of multidimensional data will be pursued.

Appendix 1
Current Staffing

Henry Rodriguez, PhD, MS, MBA, Office Director

Dr. Rodriguez is director of the OCCPR at the NCI, NIH. Prior to the NCI, he was director of the Cell & Tissue Measurements Group, director of the Tissue Engineering program, principal scientist in the DNA Damage and Repair program, and program analyst (Office of the Director), at the NIST. Dr. Rodriguez's research has focused on understanding mechanisms of cancer and age-related diseases, including the development of molecular-based technologies in basic, translational, and clinical science.

Dr. Rodriguez has led the development of NCI's clinical proteomic and proteogenomic research programs, which today includes the world's largest public repository of proteogenomic sequence data and targeted fit-for-purpose assays. These efforts led to the formation of two Cancer Moonshot initiatives – the ICPC and the APOLLO network, of which he developed and co-developed.
Dr. Rodriguez's honors include Presidential Citation, AACC; NIH Director's Award, National Institutes of Health; NCI Director's Award, National Cancer Institute, NIH; Wertheim Global Medical Leadership Award, Herbert Wertheim College of Medicine at Florida International University; and Leveraging Collaboration Award, US FDA. He has authored more than 250 scientific articles, which include 122 original research papers in peer-reviewed journals, including co-editing a best-selling book on oxidative stress and aging. Dr. Rodriguez received his BS in biology/chemistry and MS in biology/toxicology from Florida International University, PhD in cell and molecular biology from Boston University, and MBA in finance and management from Johns Hopkins University Carey Business School. Research fellowships were conducted at The Scripps Research Institute (Department of Immunology) and at City of Hope National Medical Center (Department of Medical Oncology).

Emily Boja, PhD, Program Director

Dr. Boja provides leadership in managing the scientific portfolio of the CPTAC program involving complex proteogenomic analyses of proteins and PTMs in biological and clinical samples to ensure scientific and programmatic goals are met. She also oversees the analytical and regulatory aspects of molecular diagnostics technologies and platforms via interactions with the US FDA and the AACC. Prior to the NCI, Dr. Boja served as a staff scientist and lead of proteomics at the Laboratory of Cardiac Energetics and the Laboratory of Biophysical Chemistry, National Heart, Lung and Blood Institute, NIH. Her expertise originates from her research on structural/functional studies of key enzymes and pathways involved in one-carbon metabolism critical for energy metabolism in many diseases including cancer, and later proteomic analyses of complex systems via the integration of genomics and proteomics data, and models of systems biology. Dr. Boja holds a PhD in biochemistry and molecular biology (1999) from the Medical College of Virginia, Virginia Commonwealth University.

Tara Hiltke, PhD, Program Director

Dr. Hiltke provides leadership and oversight to the Monoclonal Antibody Characterization program. She also works in developing other methods of antigen generation. Previously, Dr. Hiltke served as a senior scientist/project manager in assay development at both Wellstat Diagnostics and BioVeris Corporation, where she developed clinical assays for diagnostic markers using electrochemiluminescence platform and magnetic beads. Dr. Hiltke holds a PhD (1999) in biology from the University of Buffalo.

Chris Kinsinger, PhD, Program Director

Dr. Kinsinger focuses on the expansion and coordination of open data access and programmatic goals involving MS, informatics, and biospecimens. In this role, he works with NCI staff and investigators to optimize proteomics technology, establish policies for sharing data and biospecimens, and generally improve the quality and reliability of proteomic measurements. Dr. Kinsinger completed postdoctoral training at NIST, where he researched fragmentation pathways of peptide ions in MS. He holds a PhD in chemistry (2004) from the University of Minnesota.

Mehdi Mesri, M.Med.Sci., PhD, Program Director

Dr. Mesri provides leadership in integrating emerging technologies for the development of cancer protein diagnostics and therapeutics. He coordinates activities with NCI's Small Business Innovation Research (SBIR) Office and manages OCCPR proteomics investigator grants. Prior to the NCI, Dr. Mesri served as a principal scientist/projects leader in the Department of Protein Therapeutics at Celera. There, he used MS technologies to discover and validate biologic antibody targets in oncology, including prostate and lung cancers and angiogenesis. Prior to that, Dr. Mesri was a project leader at CuraGen Corporation, where he investigated genomically derived high value protein drugs and fully human monoclonal antibody targets for cancer and immunology/inflammation programs, among other responsibilities. Dr. Mesri holds a BSc. in Biomedical Sciences (1990) from the University of Bradford (UK), a M.Med.Sci. in clinical pathology (1991) from the University of Sheffield (UK), and a PhD in immunology (1995) from the University of Aberdeen (UK).

Etaria Omekwe, M.Sc., Scientific Program Analyst

Ms. Etaria Omekwe serves as liaison between the various Tissue Source Sites, Biospecimen Core Resource, and the Program Office for the CPTAC Program. Ms. Omekwe mainly focuses on managing the CPTAC Biospecimen Collection/Accrual and works with CPTAC components to implement program-wide quality metrics. Ms. Omekwe holds a B.Sc. in biology from Hampton University and a M.Sc. in physiology and biophysics from Georgetown University.

Alexis Carter, PhD, Cancer Research Training Award Fellow

Dr. Alexis Carter provides scientific and technical writing support as a health communications intern. She develops and manages website content and information materials, and handles news releases and email updates for CPTAC, ICPC, and APOLLO programs. She is also responsible for creating and disseminating internal communication documents such as meeting strategy notes. Dr. Carter holds a PhD in neuroscience (2016) from The University of Pittsburgh.

Gordon Whiteley, PhD, Director, CPTAC Antibody Characterization Laboratory, Frederick National Laboratory for Cancer Research

Dr. Whiteley spent more than 30 years in the medical diagnostics industry directing conception, development, and FDA filing of more than 75 diagnostic tests and accompanying instrumentation. This included 510(k), PMA, and licensed biological filings for a variety of infectious diseases, cancer biomarkers, steroid and hormone, therapeutic drugs, and cardiac markers. Prior to that, Dr. Whiteley directed the clinical laboratory service in microbiology for MDS Health Group Ltd. In Toronto. Dr. Whiteley joined Leidos Biomedical Research (formerly SAIC) in 2001 to evaluate MS as a possible diagnostic test platform. Since 2008, Dr. Whiteley has directed the Antibody Characterization Laboratory for the OCCPR. Dr. Whiteley graduated from the University of Toronto with a PhD in microbiology.

William Bocik, PhD, Scientific Project Manager, CPTAC Antibody Characterization Laboratory, Frederick National Laboratory for Cancer Research

Dr. Bocik assists in coordinating both proteogenomic informatics research strategies and assay development efforts within the CPTAC program. Prior to joining Leidos Biomedical Research, Dr. Bocik served in Public Services at the National Center for Biotechnology Information as a liaison to the scientific community with a focus on business intelligence. Dr. Bocik received his PhD in molecular and computational biophysics from Johns Hopkins University with a focus on biomolecular NMR spectroscopy.

Linda I. Hannick, PhD, Project Manager, Frederick National Laboratory for Cancer Research

Dr. Hannick provides leadership to a team of specialists with expertise in biospecimen acquisition and management, pathology review, genomics and bioinformatics, and Quality and Ethical, Legal and Regulatory Affairs within the CPTAC program. She has 30 years of experience in research that includes macromolecular crystallography at the Naval Research Laboratory and National Institute of Allergy and Infectious Diseases, followed by genomics and bioinformatics at The Institute for Genomic Research (J. Craig Venter Institute). At Leidos Biomedical Research, Dr. Hannick has experience with NCI's Physical Sciences-Oncology Network (PS-ON), TCGA program, and Molecular Analysis for Therapy Choice (MATCH) project. Dr. Hannick received her MS and PhD, both in physical chemistry from the University of New Orleans, post-doctored in molecular biophysics at the Naval Research Laboratory, and has authored over 30 scientific papers in peer-reviewed journals. Dr. Hannick is a PMI-certified project management professional (PMP).

Mathangi Thiagarajan, MfmS, Technical Project Manager, Frederick National Laboratory for Cancer Research

Mathangi Thiagarajan manages the Genomic Characterization Center (GCC) contracts for CPTAC Phase III and NCI-MATCH programs for the Frederick National Laboratory for Cancer Research. Prior to this role, she managed the CPTAC phase II and Cancer Genomics HUB (a genomic data repository for the TCGA program). She also contributed to the transition of the legacy system to the Genomic Data Commons (GDC). Ms. Thiagarajan received her MS in biomedical informatics with a specialization in genomics and bioinformatics from the University of Medicine and Dentistry of New Jersey (UMDNJ), and M.Sc. in biochemistry from India. Ms. Thiagarajan has been a scientific and technical lead for over a decade at the J. Craig Venter Institute. She has authored over 20 scientific research papers and has presented at numerous scientific conferences. She is a PMI-certified PMP.

Knowledge Management and Special Projects Branch (KMSPB)

Table of Contents

1 KMSPB Background and Mission

The Knowledge Management and Special Projects Branch (KMSPB) was created in 2009 as an office within the National Cancer Institute (NCI) Center for Strategic Scientific Initiatives (CSSI) to recognize its responsibility for Research, Condition, and Disease Categorization (RCDC) activities, beginning years earlier while a team in NCI Research Analysis and Evaluation Branch, in NCI's Division of Extramural Activities. Its role is to oversee and monitor the reporting of the entire NCI-funded portfolio of grants, intramural, and contract projects. KMSPB serves as the NCI lead for all official National Institutes of Health (NIH) RCDC reporting for NCI-funded research. KMSPB collaborates closely with colleagues across the budget and scientific components within the NCI and is involved in the development of various reporting tools. Its goal is to ensure the completeness and accuracy of the NCI-funded portfolio in NIH reports to Congress and the public.

Since 2008, the NIH RCDC coding program, which is managed within the NIH Office of Extramural Research (OER), has been used to create the annual budget reports to Congress and the Biennial Report of the Director. The RCDC program represents the NIH research categorization and reporting program that was re-engineered as authorized by Section 402B of the NIH 2006 Reform Act: *"The Secretary, acting through the Director of NIH, shall establish an electronic system to uniformly code research grants and activities of the Office of the Director and of all the national research institutes and national centers."* Additionally, Section 403 requires Biennial Reports of the Director of the NIH: *"(4) A catalog of all the research activities of the agencies including numerous metrics spelled out in the law."*

RCDC uses a computerized reporting process at the end of each fiscal year to categorize NIH funding in medical research into over 290 categories. A text-mining computer application (Collexis) assigns NIH-funded grants and contracts to the various categories. In addition to using trans-NIH definitions that take the form of a list of scientific terms and concepts referred to as 'FingerPrints' (FP), the automated categorization process relies solely on the clarity and specificity of the research stated in the grant application's title, abstract, and specific aims, which is extracted for each awarded project. The tool attempts to match biomedical concepts/terms in the FP definition to those in the extracted text using a thesaurus that is regularly curated to add or remove concepts and synonyms based on current scientific need.

The categories included in RCDC are those that the NIH has historically reported to Congress and the public. New categories are considered and added annually to accommodate requests from Congress and the public. From 2008 to 2017 the number of reported categories increased 40% from 210 to 291. NCI reports in ~70% of the NIH RCDC categories, as shown in Fig. 1. The RCDC computer-based process sorts NIH-funded projects into categories of research area, disease, or condition. The main steps in the RCDC categorization process can be viewed at https://report.nih.gov/rcdc/process.aspx.

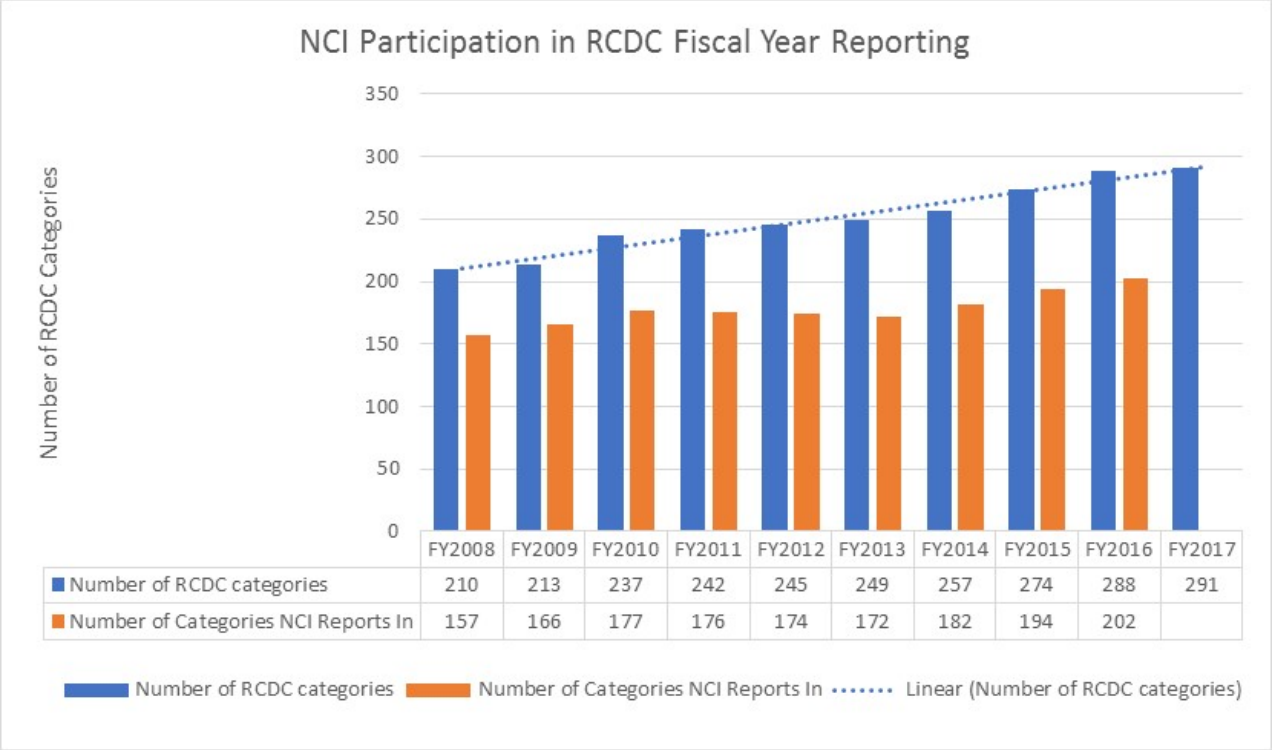


Figure 1 Fiscal Year Trend Analysis of the Publicly Reported RCDC Categories

KMSPB workload increased annually from 2008-2017 (Fig. 1) and can vary depending on the RCDC program activities (Table 1). Table 1 illustrates KMSPB participation in different types of category maintenance activities related to the RCDC program during this period. The level of IC participation required is based on the type of category maintenance activity. Creating new categories and re-visiting existing categories involve a significant amount of work, consisting of trans-NIH FP sessions to reach consensus on category definitions, inclusion, and exclusion criteria. General maintenance categories involve similar work to review the new and revisit prior or existing categories, but do not require any trans-NIH meetings. Fundamental maintenance categories require minimal FP refinements and as such require minimal participation from the ICs. These distinctions are used to cycle categories through revisit maintenance periodically to accommodate emerging areas of science.

Table 1 RCDC Category Maintenance Schedule for FY2008–2017

FY	New Categories	Revisit Categories	Categories in General Maintenance	Categories in Fundamental Maintenance	No FP Required Categories	Total
2017[1]	3	19	1	265	3	**291**
2016	14	14	79	178	3	**288**
2015	18	16	48	191	1	**274**
2014	8	6	103	139	1	**257**
2013[2]	4	8	108	128	1	**249**
2012	4	13	74	153	1	**245**

FY	New Categories	Revisit Categories	Categories in General Maintenance	Categories in Fundamental Maintenance	No FP Required Categories	Total
2011[3]	5	24	148	64	1	**242**
2010	24	6	206	0	1	**237**
2009	3	52	64	93	1	**213**
2008	14	0	195	0	1	**210**

[1]The ICs are being asked to help evaluate the impact on the categories of switching to a new RCDC system indexer in 2017, in addition to routine category validation activities.

[2]ICs assisted in the evaluation of a text-mining software as a comparison to the current technology, Collexis. This was in addition to routine validation activities.

[3]ICs assisted in evaluating the impact of the conversion of RCDC indexing software from Collexis version 6 to version 7, in addition to routine category validation activities.

As part of its role in providing oversight, coordination, and interface with the NIH about NCI's participation in RCDC, KMSPB is charged with:

- Recruiting subject matter experts from NCI Divisions, Offices, and Centers to help develop or update definitions (search term lists called FPs) that are used for classifying NIH-funded research into the almost 300 research/disease reporting categories made public each fiscal year.
- Participating in trans-NIH category FP sessions
- Evaluating and validating the draft and final category project lists annually (20,000+ projects each year)
- Representing the NCI in all RCDC working groups
- Representing the NCI in the RCDC Points of Contact (POCs) committee and ensuring that information about the RCDC program is disseminated to appropriate staff within the NCI, and that information from the NCI is relayed back to designated staff within RCDC
- Assisting the NIH OER, as needed, in periodic evaluations and assessments of the RCDC system categorization, when NIH/OER/ORIS (Office of Research Information Systems) makes periodic system upgrades or changes to a different indexing engine
- Collaborating with NCI Divisions, Offices, and Centers to assist with portfolio analyses to facilitate strategic planning activities
- The KMSPB also manages the NCI Funded Research Portfolio (NFRP) web site and database, to ensure the completeness, timeliness, and accuracy of scientific funding data for current and past fiscal years

2 KMSPB Program Activities

2.1 Official NIH RCDC Scientific Categorization and Reporting Activities

KMSPB participates, year-round, in the NIH reporting program through the following RCDC-related activities:

January to April

- KMSPB analyzes prior year fiscal data and makes recommendations to RCDC for category FP revisit sessions, which require an in-depth review and analysis of the category FP and participation by subject matter experts to address emerging areas of science and other critical issues.

April to September

- KMSPB recruits subject matter experts as needed to participate in FP sessions with KMSPB staff, and/or provide feedback for refinements in inclusion/exclusion criteria for category FP refinements.

April to October

- KMSPB performs quality control checks for RCDC reported R&D Contracts and Inter/Intra Agency Agreements (IAAs).

April to November

- KMSPB staff review and validate NCI projects pulled into the RCDC categories in the RCDC Project Review System (PRS) module. This involves providing project-level feedback to RCDC with regard to the project's scientific relevance to a category. This feedback is used by RCDC to perform category FP refinements.

June to September

- KMSPB staff perform quality control checks for RCDC reported grant subprojects.

August to October

- KMSPB participates in 'User Assisted Categorization' (UAC). The UAC module was created to accommodate categories that have unique or additive reporting requirements, such that a project captured by the category may only be reported into one of the sub-categories. KMSPB performs the manual sub-categorization of the Networking and Information Technology R&D, Nanotechnology, and Pediatric Research Initiative categories.

September to November

- KMSPB, consulting with NCI Office of Budget and Finance (OBF), reconciles NCI Cancer report data and all funding streams prior to OER/ORIS data freeze.

November to December

- KMSPB review of NIH draft frozen data (these are likely to be the Final Official Reports).

2.2 NCI Guidelines and Standard Operating Procedure Development

The importance of the RCDC program is not only based on the use of the categories for official reporting to the Department of Health and Human Services (DHHS) and Congress. The reports are also significant because they are featured on the NIH Office of Budget Categorical Spending site and the NIH RePORTER site, which are public NIH websites that were created to facilitate the query and analysis of the RCDC data. The RePORTER website is increasingly used internally and by the extramural community to perform portfolio analyses. For example, in 2013 (using search criteria of NIH, RePORTER, portfolio) there were five articles published in PubMed referencing the use of RePORTER data to perform analyses on NIH funding levels in certain areas of interest. As of July, there are 17 articles published in 2017 using the same search criteria, indicating an increased level of interest and use of publicly available RePORTER data.

This increasing demand for access to the RCDC data, for both official congressional reporting and portfolio analyses using RePORTER, was an incentive for the KMSPB office to develop a comprehensive quality management program for the NCI participation in the RCDC program. Some examples of the quality assurance/quality control measures that have been implemented are as follows:

1) A KMSPB Standard Operating Procedures (SOP) training document has been created for new personnel. The KMSPB validity review process involves the evaluation of the scientific relevance of the NCI-funded research portfolio that is reported by the NIH RCDC system for delivery to Congress. The SOP document provides background information on the RCDC system and defines important terminology pertinent to the KMSPB validity review process. The document also contains detailed descriptions of existing guidelines and policies regarding the KMSPB validity process.

2) A suite of SQL-based user tools has been initiated and developed to enhance the KMSPB validity review process. Each SQL query has been designed to utilize either one or more user defined data sets to generate a data rich document. Each query has been tailored for specific KMSPB related activities, such as FP analysis, False Negative searches, RCDC validity process, etc. These queries are time saving tools that are intended to improve the overall efficiency and accuracy of the validity review process.

3) Summaries of each of the RCDC categories that have NCI-funded projects were developed. These summaries contain the NIH RCDC category parameters and any pertinent additional information from the NIH RCDC category FP session meeting minutes. They contain the NCI KMSPB guidelines that are developed to create inclusion and exclusion criteria for the NCI portfolio in each category. The KMSPB guidelines align with the general NIH category parameters but also include specific feedback KMSPB solicits from NCI program staff as needed.

Not all the RCDC and other analysis and reporting activities are amenable to written procedures. Changes can be made to the RCDC system and program that are unexpected and require innovative adaptations by the ICs to accommodate these changes.

2.3 Collaboration with NCI Scientific Subject Matter Experts

The NIH RCDC reporting program relies on IC scientific expertise in the development of category FP definitions. KMSPB recruits the NCI program subject matter experts that participate in this activity. Often, there are multiple NCI experts for a category, from more than one NCI Division, Office or Center (DOC); see Figure 2.

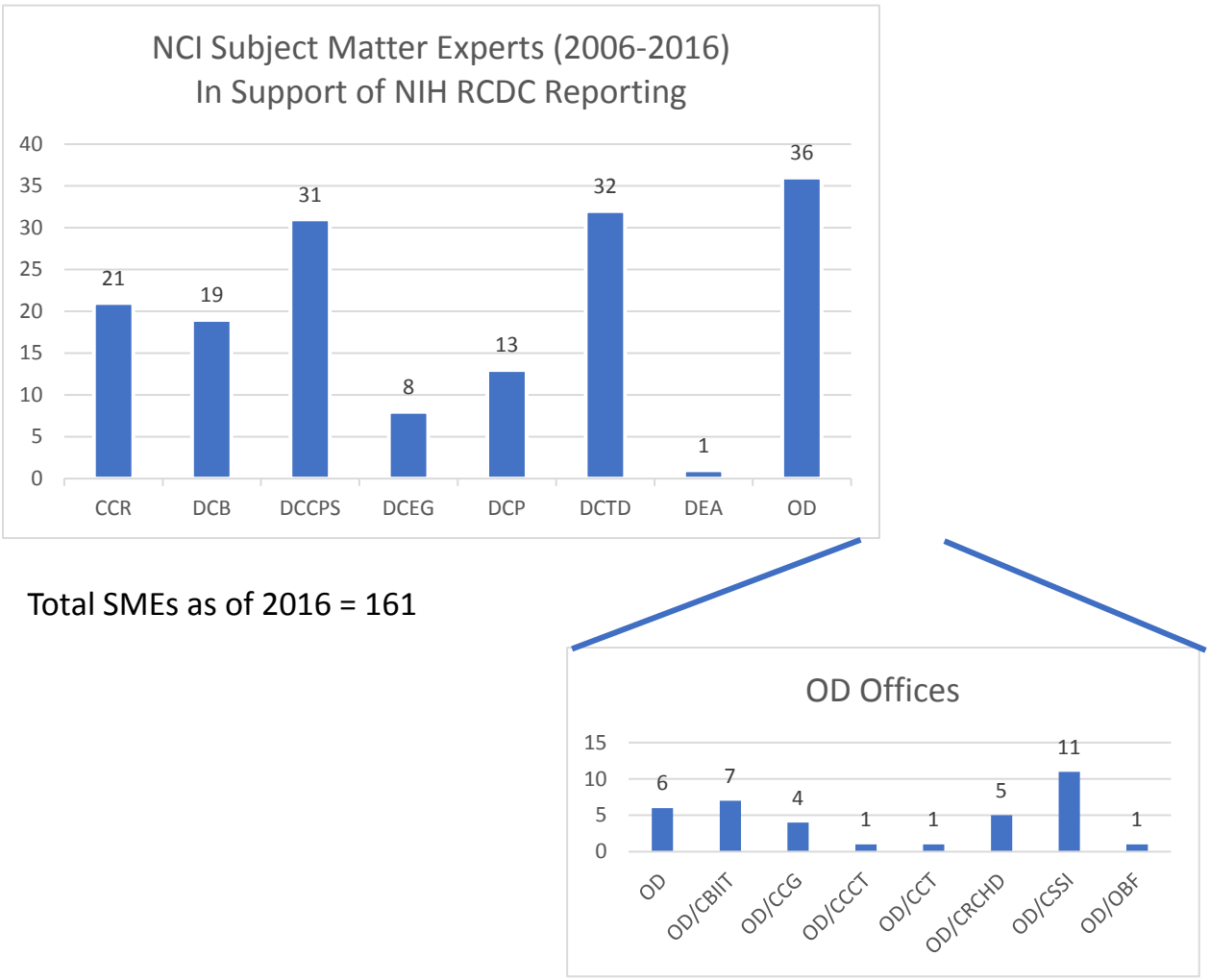


Figure 2 **Recruitment of NCI Subject Matter Experts by Division**

2.4 Validation of NCI Portfolio in Official NIH Budget Reports to Congress

KMSPB staff evaluate NCI projects in each of the RCDC reporting categories and enter comments into the RCDC system regarding each project's scientific relevance; this constitutes IC validity feedback. RCDC categories are assigned based on the scientific background and expertise of each KMSPB analyst. KMSPB also identifies false negatives for each of the categories and enters them into the system. NIH refers to this feedback when they make FP refinements to remove false positives or capture false negatives. These changes to the FP sometimes require updates to the RCDC thesaurus.

The NIH ICs are required to officially report using the RCDC program but the level of participation by the ICs in the validity feedback process varies. NCI has validated an average of 91% of total projects over the 2012-2016 timespan.

2.5 Support of Other NCI Offices' RCDC Data Entry Requirements

KMSPB supports the RCDC data entry requirements of several NCI Office of Management (OM) offices, including the NCI Office of Acquisitions, the Office of Grants Administration and the Administrative Resource Center.

- The Office of Acquisitions has the ultimate responsibility for ensuring that all the required contract data entry into the RCDC Contract and IAA Management System is completed by the appropriate deadlines. KMSPB supports the Contracting Officers (CO) and the Contracting Officer Representatives (COR) by running reports for them to track their progress with data entry. We also offer training customized for NCI staff and help COs and CORs resolve issues and errors they find in the system, and work with NIH RCDC staff to get the problems resolved.

- The Office of Grants Administration has the ultimate responsibility for ensuring that all the required subproject data entry is completed and correct, since RCDC reports out multi-component grants at the subproject level. KMSPB supports them in that effort by running reports that allow us to identify issues, such as grants with incomplete subproject data or duplicate subprojects. We send them any examples of these that we find throughout the year so that they can fix it in the subproject module.

3 Development of Knowledge Management Toolkit to Facilitate Evaluation and Validation of the NCI Portfolio and Official NIH Reports

KMSPB reviews over 20,000 NCI projects annually distributed in over 290 RCDC categories. Therefore we have developed of a suite of tools to facilitate the review and evaluation process by KMSPB staff.

Scientific Managers Analysis and Reporting Tool (SMART)

The KMSPB SMART application (Fig. 3) is an analysis and reporting tool that runs on RCDC and NCI coded data that is downloaded from both NIH and NCI databases. The tool compares the NCI scientific coding applied to grants with the RCDC categorization. The SMART comparison tool enables KMSPB analysts to perform scientific validation of projects categorized into the ~290 categories publicly reported by NIH Office of Budget annually. SQL statements query and retrieve data from the resource databases. Retrieved data is stored in local tables in an Oracle database. The SMART application runs on an original Java-based technology. Additional functionality added to SMART would involve moving to a newer version of Java and Tomcat which would mean significant cost and longer development time. We are in the process of transitioning to a newer technology, the Scientific Coding System (SCS), which allows much more flexibility in functionality and performance than a Java-based application.

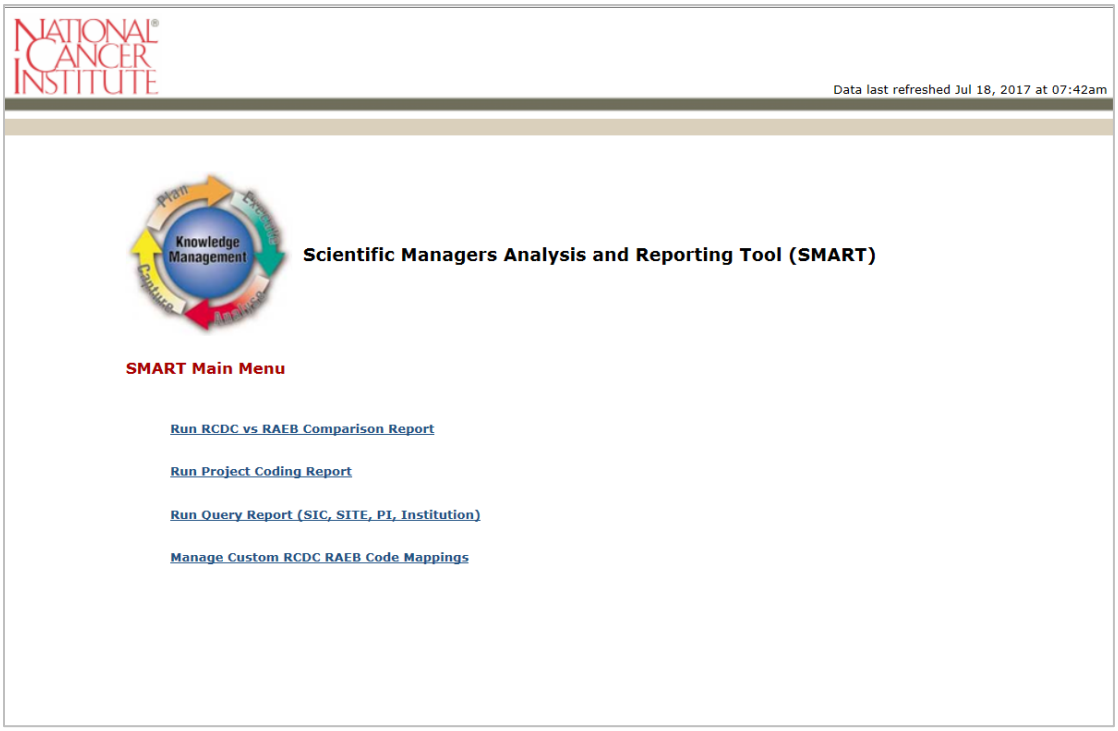


Figure 3 SMART User Interface

Scientific Coding System (SCS) OnDemand

SCS OnDemand (Fig. 4) is a web-based scientific coding solution managed by NIH Center for Information Technology (CIT). It uses an Oracle database platform which CIT hosts and provides database services. Currently, six NIH ICs are using this application to code grants, contracts, intramural projects, and IAA based on each project's scientific research focus. This system has been customized to incorporate the specific coding practices, vocabulary, and reporting requirements of each IC. Each IC develops its own unique coding structure and process to reflect reporting and portfolio analysis needs and requirements which are then integrated into and supported by the SCS application. The system also allows participating ICs to categorize and report their portfolio as a formatted data feed to various NIH-wide reports. The KMSPB use of this tool is primarily to compare NCI and RCDC coding. This comparison functionality of SCS supports and streamlines our validation process.

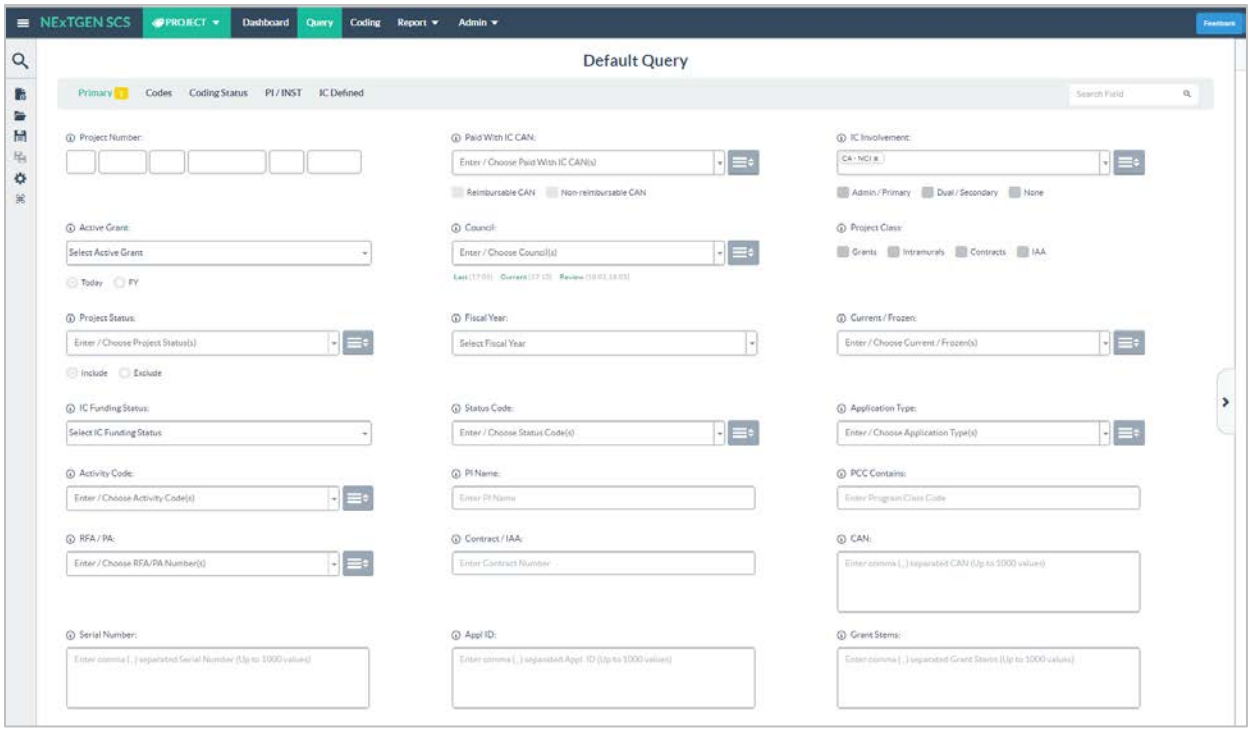


Figure 4 SCS OnDemand User Interface

SCS OnDemand's use of a database structure allows the application to perform:

- comparative analyses of RCDC and NCI coding, at project and category level
- scientific coding of grants, intramural projects, contracts/IAAs
- generation of detailed reports
- portfolio analyses
- configuration of SQL scripts to download IC specific grant data
- uploading of NIH data such as contracts, intramural projects, and IAAs
- application customizations such as downloads of NCI-specific data and data elements
- advanced searches

NCI-CIT Partnership to Develop Additional Knowledge Management Tools

This project is a collaboration between NCI and CIT to develop tools to assist with portfolio analysis. The overarching goal is to create opportunities for users to interact with sophisticated information retrieval tools, supported by visualization software to steer the machine learning process as well as refine the classification output. Success will conserve users' time as well as maximize the efficient integration of users' judgment and expertise. The focus is on the evaluation and optimization of currently available tools and resources.

4 Data Manager and Administrator of the NCI Funded Research Portfolio

The NCI Funded Research Portfolio (NFRP) web site and database (Fig. 5) contains information about research grants, contract awards, and intramural research projects funded by the NCI. It provides the ability to search the database for NCI-specific data elements and cancer-relevant scientific codes, as well as the ability to perform keyword searches of the project abstracts. The NFRP was launched as a public website in 1998 and in 2015 the management of the website and database was transferred to the CSSI KMSPB.

The NCI employs a sophisticated system of scientific coding in which trained professionals and/or scientific staff analyze grant applications, contracts, and intramural projects to classify each project for its degree of relevance to Special Interest Category (SIC) and Organ Site (SITE) codes. NCI staff apply codes using percent relevance and prorated dollars, for precise budget forecasting. The primary NCI Divisions that supply scientific coding data for the NFRP are: the Division of Extramural Activities (Extramural Grants and Research Contracts); Division of Cancer Epidemiology and Genetics and the Center for Cancer Research (Intramural projects); and the OBF (Research Management and Support and Leidos projects). The KMSPB and OBF work closely during each new fiscal year's data consolidation and validation process, prior to posting to the NFRP.

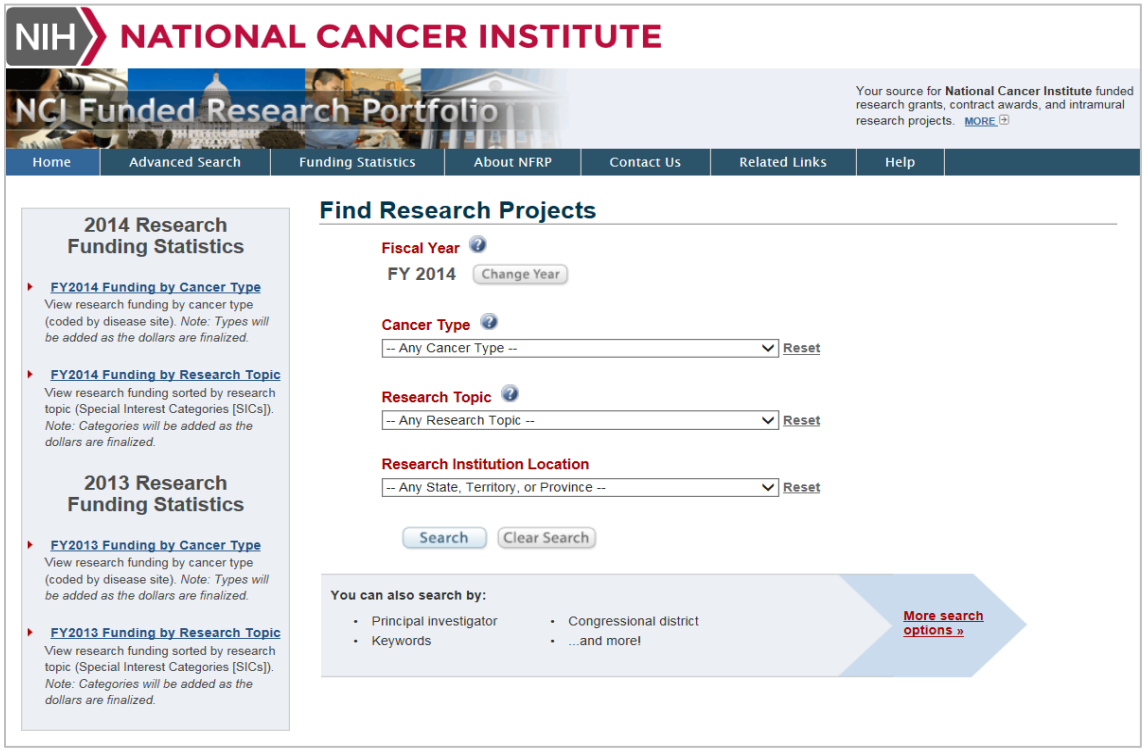


Figure 5 NFRP User Interface

5 Future Vision

5.1 Categorization and Reporting Activities

Our efforts to create SOPs and archive each year's data and summary of activities is motivated by our desire to identify real-time coding and reporting issues so that we can propose new categories for trans-NIH review and re-visit prior categorization issues as soon as possible. We will continue to do that and try to find new processes and tools that will help us be even more efficient and timely in our feedback.

Another goal is to build on the NCI relationships we have already established with NCI program staff and perform more outreach to them, especially with regard to the NIH RCDC categorization and reporting program. For example, we have started to provide training through the NCI Division of Extramural Activities Program and Review Extramural Staff Training Office (PRESTO). Initially, we are focusing on the data entry requirements in support of the RCDC process, but in the future we would like to add more training sessions or brown bag meetings to update our colleagues on the RCDC categories and the available tools that might be of interest, especially for portfolio analyses and strategic planning.

5.2 Additional Tool Development

Each year we will progress in our customization of the SCS application to improve our assessment and evaluation of RCDC's categorization of the NCI research portfolio. Further development of this tool will aid our ability to perform project-level annotation for specialized areas of research, emerging fields of science, and new technologies. In the future we hope to enhance the application to use a definition-based controlled vocabulary to code NCI research projects for portfolio analyses and evaluation.

5.3 NFRP

Now that the NFRP is managed by the KMSPB, we would like to create new and update existing SOPs as needed for data collection and data quality control, as well as the annual data update itself. The SOPs will include timelines so that the NCI partners in this process can plan for the work and time this effort requires. Our goal is to publish the new fiscal data within six to eight months of the end of each fiscal year.

We have recently partnered with the NCI OBF and the Office of Communications and Public Liaison, in planning new enhancements and functionality for the NFRP, which will be based on user testing feedback. Hand in hand with updated functionality, we would like to add new Research Topic reports to the site, that are timely and of interest to our users.

Appendix 1
Current Staffing List

Lisa Krueger, MS, Branch Chief, KMSPB

Ms. Krueger is the chief of the KMSPB in the CSSI in the Office of the Director of the NCI. She has been a staff member of NCI since 1997, when she was recruited to NCI from the University of Minnesota Hospital and Clinics by Dr. Diane Arthur to help establish and then manage the Clinical Cytogenetics Section in the Laboratory of Pathology in the Center for Cancer Research. She moved from the intramural to the extramural side of NCI in 2001 when she joined the Research Analysis and Evaluation Branch (RAEB) in the Division of Extramural Activities, where she became the lead of the section in RAEB responsible for participation in new NIH scientific coding and reporting efforts, such as the NIH RCDC program, which NIH now uses for official congressional reporting. The team was transitioned to the KMSPB in CSSI in 2009. The KMSP branch is charged with the oversight, coordination, and interface with the NIH as regards all aspects of the NCI participation in RCDC. Ms. Krueger has responsibility for all aspects of planning and implementation of NCI Knowledge Management (KM) reporting initiatives, serve as the NCI representative on trans-NIH KM initiatives such as the NIH RCDC project, create collaborative relationships across organizational lines and between ICs to further KM efforts within the NCI and throughout the NIH, and support development of KM tool-based search strategies to perform NCI portfolio analyses, prepare reports or summaries as needed. She supervises four full-time employees (FTEs) and acts as the NCI Point of Contact for four contractors assigned to KMSPB. Her BS degree in biology is from George Mason University and her MS degree in genetics is from the University of Minnesota, where her research focused on molecular and cytogenetic factors influencing the expression of chromosomal fragility at FRAXA in Fragile X syndrome.

Michele Vos, MS, Deputy Branch Chief, KMSPB

As NCI's secondary point of contact for the NIH RCDC program Ms. Vos helps manage the institute's commitments to the annual RCDC reporting activities. She is the system administrator for several IT knowledge management applications developed and maintained within KMSPB. These systems support administrative processes that help ensure all NCI awarded projects are accurately categorized in the ~290 research and disease areas NIH publicly reports each fiscal year. Ms. Vos received her BS degree in biology from the University of Maryland and an MS degree in biomedical science from Hood College, where her thesis project focused on the role of amidating enzymes and how they potentiate the growth response of tumor cells in the presence of neural and gastrointestinal peptides. As a research biologist for over 20 years in NCI's Intramural Program, she conducted and published on several independent research projects while in the Division of Clinical Sciences, Medicine Branch's Molecular Signaling and Oncogenesis Section, the Division of Clinical Sciences, Biomarkers and Prevention Research Branch, and the Division of Cancer Biology and Diagnosis, Experimental Immunology Branch.

Luciana B. Crotti Espinoza, PhD, Health Scientist Administrator, KMSPB

Dr. Crotti Espinoza manages a scientific portfolio of cancer-relevant topics on basic, clinical, and translational-focused research. She provides technical review on a variety of issues to capture government-funded research projects for congressional reporting. She earned a bachelor's degree in

pharmaceutical science, a master's degree in chemistry and biology, and a PhD degree in biochemistry from the University of Sao Paulo, Brazil. She continued her scientific training in biochemistry, genetics, and molecular biology as a postdoctoral fellow at the NIH. Her research at the NIH started by studying cell cycle control, cell growth, and differentiation in *Saccharomyces cerevisiae.* She then shifted her focus to defining determinants of chromosome segregation in *S. cerevisiae* and extrapolating these findings to causes and consequences of genome instability in humans. Before joining the KMSP branch in 2015, she was involved in a research program to study alternative splicing defects and the development of genetic diseases at the Uniformed Services University of the Health Sciences in Bethesda, MD, as a senior research associate.

Linnia H. Mayeenuddin, PhD, Health Scientist Administrator, KMSPB

Dr. Mayeenuddin completed undergraduate and graduate studies in pharmacology at the University of Toronto (Ontario, Canada). Her graduate studies involved training in protein biochemistry, molecular/cellular biology, and signal transduction. As a research associate/instructor at the University of Virginia (Charlottesville, VA), she conducted research studies focused on functional genomics and signal transduction. She has been with the NCI for the last ten years. She began her career at NCI as a postdoctoral fellow at the Molecular Targets Core housed within the Genetics Branch (Bethesda, MD). As a postdoctoral fellow, her research focused on oncology biomarkers and signal transduction. In August of 2009, she transitioned to a policy analyst role at the NCI, at the KMSP branch within the CSSI. She is currently a health scientist administrator at KMSP, where she participates in a trans-NIH process to facilitate accurate congressional reporting of NCI's research funding dollars, as determined by NIH's automated RCDC tool.

Tina Branscombe Miranda, PhD, Health Scientist Administrator, KMSPB

Dr. Miranda completed her PhD in biochemistry at the University of California, Los Angeles. Her graduate thesis work focused on identifying novel enzymes (protein methyltransferases), important for modifying histones within chromatin and affecting gene expression, while her postdoctoral studies at the USC Cancer Center focused on DNA methylation and chromatin biology. Dr. Miranda then did a fellowship with NCI Center for Cancer Research (CCR) in Dr. Gordon Hager's laboratory looking at how chromatin structure affects transcription factor binding. She maintains working knowledge of NCI scientific research programs and their policies and practices in order to oversee a portfolio of reporting categories including Cancer Genomics, Precision Medicine, Human Genome, Orphan Drug, and Rare Diseases.

We also are supported by four contractors:

Julie Beaudet, PhD, Science Program and Policy Analyst, KMSPB

Dr. Beaudet earned a bachelor of science in chemistry and her PhD in biochemistry and biophysics from Renesselaer Polytechnic Institute (Troy, NY). Her graduate studies focused on protein-glycosaminoglycan interactions and metabolic engineering of heparin. She recently completed a postdoctoral fellowship in the Laboratory of Bacterial Polysaccharides at the FDA. Her postdoctoral research examined the binding behavior of meningococcal bacteria proteins for next-generation vaccine production. She primarily supports the Branch by performing complex analyses and validations of categories with a drug and/or bioengineering component.

Pawel Sulima, PhD, Science Program and Policy Analyst, KMSPB

Dr. Sulima worked for several years as a scientific data analyst in discovery logic at Thomson Reuters where he analyzed big data sets and provided reports and visualizations to support portfolio planning and management. He obtained his PhD at K. Marcinkowski University of Medical Sciences in Poznan, Poland and completed postdoctoral training at NIDCR/NIH. As a postdoctoral fellow, his research was focused on the molecular basis of hereditary diseases, oral microbiome, and bioinformatics. He primarily supports the Branch performing analyses and validations of categories that have an infectious disease and/or immunology component.

Cathy Rowe, BS, Scientific Program Manager, KMSPB

Ms. Rowe performs administrative functions and portfolio analyses. She gained a bachelor of science in biology from Howard University in Washington, DC, translating her scientific background through various careers. First, performing viral clearance studies required by the U.S. Food and Drug Administration (FDA) for Invitrogen BioServices, she then participated in the coordination of four simultaneous research studies at the Walter Reed Army Medical Center/Vaccine Healthcare Centers, which involved developing, maintaining, and analyzing research data, planning and implementing research protocols, instituting new contracts and business processes, and managing the processing of biological samples. At the NIH, she first worked for the NIH National Institute of Allergy and Infectious Diseases (NIAID), where she supported clinical research in the Division of AIDS (DAIDS), performing protocol and clinicaltrials.gov data management. At NCI, Ms. Rowe administratively supports the KMSP branch, as well as performs advanced analytics on NCI-funded extramural research, assisting the Branch in its support of the NIH RCDC program.

Elaine Taylor, BS Data Manager, KMSPB

Ms. Taylor graduated from Old Dominion University in Norfolk, Virginia with a BSCS in computer science. She has spent the last 30 years as a software developer and database designer, with the last 18 of those years at NCI. This experience resulted in her acquisition of extensive knowledge of NCI business data, processes, and systems from project work experience and interaction with NCI staff from various groups. Her subject matter focus at NCI has been NCI scientific coding data as it relates to NCI financial data. She worked on and then managed the Fiscal Linked Analysis of Research Emphasis (FLARE) system in the Division of Extramural Activities from 1999 to 2014.

Dr. Sulima worked for several years as a scientific data analyst in discovery logic at Thomson Reuters where he analyzed big data sets and provided reports and visualizations to support portfolio planning and management. He obtained his PhD at K. Marcinkowski University of Medical Sciences in Poznan, Poland and completed postdoctoral training at NIDCR/NIH. As a postdoctoral fellow, his research was focused on the molecular basis of hereditary diseases, oral microbiome, and bioinformatics. He primarily supports the Branch performing analyses and validations of categories that have an infectious disease and/or immunology component.

Cathy Rowe, BS, Scientific Program Manager, KMSPB

Ms. Rowe performs administrative functions and portfolio analyses. She gained a bachelor of science in biology from Howard University in Washington, DC, translating her scientific background through various careers. First, performing viral clearance studies required by the U.S. Food and Drug Administration (FDA) for Invitrogen BioServices, she then participated in the coordination of four simultaneous research studies at the Walter Reed Army Medical Center/Vaccine Healthcare Centers, which involved developing, maintaining, and analyzing research data, planning and implementing research protocols, instituting new contracts and business processes, and managing the processing of biological samples. At the NIH, she first worked for the NIH National Institute of Allergy and Infectious Diseases (NIAID), where she supported clinical research in the Division of AIDS (DAIDS), performing protocol and clinicaltrials.gov data management. At NCI, Ms. Rowe administratively supports the KMSP branch, as well as performs advanced analytics on NCI-funded extramural research, assisting the Branch in its support of the NIH RCDC program.

Elaine Taylor, BS Data Manager, KMSPB

Ms. Taylor graduated from Old Dominion University in Norfolk, Virginia with a BSCS in computer science. She has spent the last 30 years as a software developer and database designer, with the last 18 of those years at NCI. This experience resulted in her acquisition of extensive knowledge of NCI business data, processes, and systems from project work experience and interaction with NCI staff from various groups. Her subject matter focus at NCI has been NCI scientific coding data as it relates to NCI financial data. She worked on and then managed the Fiscal Linked Analysis of Research Emphasis (FLARE) system in the Division of Extramural Activities from 1999 to 2014.