National Cancer Institute

U.S. DEPARTMENT
OF HEALTH AND
HUMAN SERVICES

National Institutes
of Health

Division of Cancer Prevention

The Early Detection Research Network (EDRN)

Strategic Plan

October 2019

EXECUTIVE SUMMARY

Developing a collaborative network is a challenging, if not a daunting task. Identifying biomarkers for early cancer detection biomarkers involves a rigorous process that begins with discovery and leads to development, validation, and finally clinical application. The success of this process requires a dedicated infrastructure that facilitates coordination and collaboration among a variety of institutions and academic- and industry-based scientists and clinicians. The National Cancer Institute's Early Detection Research Network (EDRN) has fulfilled these expectations by establishing a process for biomarker development using a multidisciplinary and multi-institutional approach. This infrastructure serves as a model for the conduct of translational research and is well-aligned with the goals and objectives of the National Cancer Institute (NCI) and the National Institutes of Health (NIH) community.

The expectation is that validated biomarkers will result in an increase in early cancer detection, resulting in a decrease cancer morbidity and mortality. This expectation has resulted in an increase in the number of reported early detection biomarkers. Unfortunately, the vast majority of these biomarkers never progress beyond the initial discovery phase, usually because they have insufficient accuracy to be clinically useful or there is no process to translate them into clinically useful assays. The EDRN is designed to address both of these limitations by supporting high quality biomarker discovery using appropriate biospecimens and through incentives that facilitate the "hand-off" of biomarkers from the discovery laboratory to clinical validation centers. In addition to being dedicated to rigorous evaluation and validation of biomarkers, the EDRN has collaborated with NASA's Jet Propulsion Laboratory (JPL) to build a cutting edge, national, bioinformatics network for the capture, management, distribution and analysis of cancer research data and to develop a biomarker database that is available to all investigators.

The EDRN has made substantial progress over the past 19 years by utilizing its available resources and leveraging those of other organizations. Based on recommendations of NCI's Board of Scientific Advisors, the EDRN has expanded its goals during the past 5 years to include (1) the development and validation of imaging methods for early cancer detection, especially radiomics, (2) an increase in collection of biospecimens from high risk cohorts and minorities, (3) an increase in the number of biomarker trials that reflect the progress in biomarker development, and (4) an increase in interaction with implementation scientists. Successful biomarker translational research requires an understanding of the biomarker's potential clinical usage and what performance characteristics it needs to have. EDRN accomplishes this by involving basic scientists, clinical scientists, statisticians, information scientists, public health professionals, patient advocates, and clinical testing laboratories that are Clinical Laboratory Improvement Amendments (CLIA) approved.

EDRN's strategic goals are developed from discussions with both EDRN and non-EDRN investigators, from workshops convened by the EDRN, and program evaluation committees, such as the Network Consulting Team (NCT). This report outlines organ-specific research priorities to address pressing clinical needs in an efficient and timely manner. EDRN's strategic goals include the use of the latest technologies and systems biology approaches to develop and validate new biomarkers and imaging methods for early cancer detection and in the diagnosis of clinically significant diseases and to predict clinical outcomes.

Collaborative projects among EDRN and non-EDRN investigators facilitate biomarker research and its translation into clinical application in an efficient manner. Equally important, the EDRN has long-standing relationships with industry and the private sector. The EDRN also has strategic alliances with non-profit foundations, such as the Canary Foundation and the Lustgarten Foundation, federal agencies, such as the

National Institute of Standards and Technology (NIST) and the Food and Drug Administration (FDA), and professional organizations, such as American Society of Clinical Oncology (ASCO), to develop a transparent process for creating well-defined consensus standards and guidelines for biomarker development, validation and qualification to reduce the uncertainty in adoption of biomarkers for clinical use. In the future, the EDRN proposes to work with Centers for Medicare and Medicaid Services (CMS) to review new biomarkers and to reach consensus on coverage and pricing.

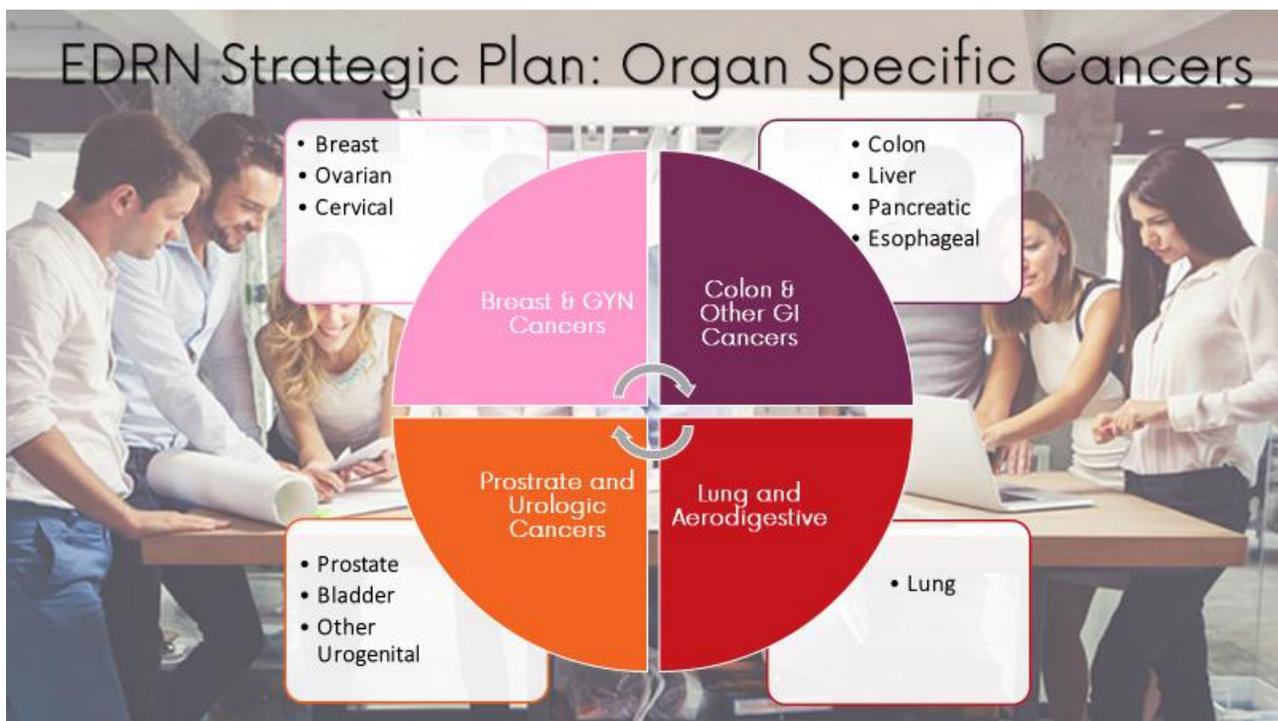In this document, we describe EDRN's short and long-term strategic goals.

TABLE OF CONTENTS

EDRN OBJECTIVES

The overarching goal of the Early Detection Research Network (EDRN) is to help reduce cancer morbidity and mortality by conducting research on biomarkers to detect early stage cancers and precancerous lesions and translating these biomarkers and imaging methods into clinical tests through:

- Strategic and systematic evidence-based discovery;
- Development and validation of biomarkers and imaging methods to detect cancer early and assess risk;
- Development and validation of biomarkers and imaging methods to distinguish indolent cancers from aggressive cancers to reduce overtreatment;
- Coordination of biomarker research in extramural community and with other NCI prevention and treatment programs to develop strategies to reduce cancer morbidity and mortality.

The EDRN fulfills its objectives by providing an infrastructure and environment that fosters translational research and by encouraging collaborations among basic scientists, population-based scientists, and physician scientists with expertise in clinical applications. These multidisciplinary investigators facilitate the integration of knowledge into evidence-based cancer biomarker research. The EDRN strategically focuses on cancers where there is a high potential for influencing cancer morbidity and mortality. To support this approach, the EDRN has established collaborative groups based on organ systems.

EDRN Strategic Plan: Organ Specific Cancers

- Breast
- Ovarian
- Cervical

Breast & GYN Cancers

- Colon
- Liver
- Pancreatic
- Esophageal

Colon & Other GI Cancers

Prostrate and Urologic Cancers

- Prostate
- Bladder
- Other Urogenital

Lung and Aerodigestive

- Lung

EDRN objectives include:

- Discover, develop, and validate biomarkers and technologies for risk assessment and detection of early stage cancers and precancerous lesions, and to distinguish indolent from aggressive cancers;
- Develop and validate imaging methods, especially radiomics to improve detection of early stage cancers and to distinguish indolent from aggressive cancers;
- Combine biomarkers with imaging to improve the detection of early stage cancer cancers;
- Develop and validate biomarkers and imaging methods to improve the detection of cancer progression in patients on active surveillance;
- Develop assays and evaluate advanced technologies to accelerate biomarker discovery;
- Facilitate the development of high-throughput, sensitive assay methods to identify and implement cancer biomarkers that are useful in assessing cancer risk and detecting early stage cancers;
- Support collaboration among academic and industrial leaders who have interests in molecular biology, clinical oncology, computer science, public health or related areas;
- Conduct clinical/epidemiological studies (e.g., cross-sectional, prospective, retrospective, etc.) in order to evaluate the predictive value of biomarkers;
- Expand the informatics infrastructure to facilitate pre-competitive data sharing on biomarker discovery, development, and validation; and
- Serve as a resource and infrastructure that can be leveraged by the NCI and the cancer community to facilitate translational cancer research.

Through its ongoing strategies to promote productive interactions among basic and clinical scientists, its expertise in epidemiology, clinical diagnostics, biostatistics and informatics, and its well-annotated biorepository, the EDRN represents a resource for NCI and the cancer research community to support and accelerate critical advances in cancer research and its translation into clinical cancer care. This role, as a core resource, positions the EDRN as a strong partner with other cancer programs, including clinical trials, cooperative groups, Specialized Programs of Research Excellence (SPOREs), and other NCI and community-designated cancer centers.

EDRN STRATEGIC GOALS

The goal of cancer screening is to detect either a preneoplastic lesion or a cancer at an early stage where treatment will change the outcome and prolong survival. The EDRN supports the development, validation and clinical application of biomarkers and imaging methods to improve the detection of early stage cancers. Areas of particular interest include the replacement of tissue-based assays with fluids-based assays, improvement of imaging techniques and radiomics, and development of a knowledge base for improving evidence-based screening of cancer. With the advent of new detection technologies, many candidate biomarkers are being identified. Unfortunately, the vast majority of these biomarkers lack sufficient sensitivity and specificity, and there is a need for improved study design in the discovery process. The translation of even the most promising of these biomarkers into clinically useful tests is limited due to the lack of reproducibility of the assays. EDRN is addressing these issues by developing quality specimens, robust study designs, standard operating procedures and collaborations among technology developers.

Many detectable lesions and cancers are not life threatening, and the current inability to discern which lesions will lead to clinically significant morbidity and mortality vs. benign or slow-growing cancers leads to excessive testing and treatment, a phenomenon known as overdiagnosis. About 25% of breast cancers detected on mammograms and about 60% of prostate cancers detected with PSA tests could represent overdiagnosis. There are ongoing debates regarding how to recognize and manage overdiagnosis. One proposed strategy is the development of disease-specific biomarkers or imaging methods that can distinguish aggressive from non-aggressive cancers, and one of the EDRN's strategic goals is to develop and validate these biomarkers and imaging methods.

Types of biomarkers relevant to the EDRN include genetic, genomic, epigenetic, gene expression, microRNA, exosomes, proteomic, glycomic, metabolomic, and other as yet uncharacterized novel categories of biomarkers. Genetic and genomic biomarkers include single-locus/gene, multiple-locus/gene, and genome-wide assays of gene copy number (including amplification or deletion), mutation, sequence variations/polymorphisms, linkage disequilibrium, chromosomal and sub-chromosomal translocations and other alterations or associations. Epigenomic biomarkers include DNA methylation, several types of histone modifications, and other changes to DNA that do not alter the DNA backbone. Expression biomarkers include, but are not limited to, messenger RNA and microRNA (additional species of small RNA, such as piwi RNA, and the large number of poorly characterized or as-yet completely uncharacterized noncoding RNAs also of potential interest). Proteomic biomarkers include not only the amino acid sequences of proteins or peptide fragments but also post-translational modifications of those proteins, including, but not limited to, phosphorylation, sulfation, myristoylation, farnesylation, glycosylation and many others. Glycomic markers include all measures of the sugar side chains found on many proteins, and these markers can also be detected by many different means.

Progress in early cancer detection and image-based diagnosis has been hampered by the lack of understanding regarding the natural history of the disease. Innovations in molecular biology, genomics, proteomics and immunology may provide insights.

EDRN is considering the following questions:

- o Why do some preneoplastic lesions progress rapidly and require intervention?
- o Therapeutic intervention is often triggered by assessment of a static picture of disease (histologic snapshot of observables assumed to be representative of the whole tumor), rather than knowledge of the inherently dynamic nature of the underlying disease. Current assessments are thus incomplete. What could be done to advance knowledge of cancer progression? Can information derived from the integration of imaging and molecular diagnostics further this understanding?
- o What types of molecular properties confer aggressiveness to some preneoplastic lesions? What could be done to study such behaviors to avoid overdiagnosis and overtreatment?
- o What could be done to define and characterize preneoplastic lesions in order to improve current standards in histology and cytology?

SHORT-TERM GOALS

- Work to augment the ability of commonly used screening tests to detect major epithelial cancers, such as lung, colon, breast, cervical, and prostate cancers. Also, facilitate the co-development of diagnostic tests for prevention or therapeutic interventions (theranostics).

- Employ cost-effectiveness measuring tools to evaluate biomarker discovery, development and validation, and to collaborate with the NCI's Cancer Intervention and Surveillance Modeling Network (CISNET) on integrating cost-effectiveness models in the discovery and development processes.

- Create well-defined consensus standards and guidelines for biomarker development, validation and qualification using the Translational Research Working Group (TRWG)-developed Device Pathway to reduce uncertainty in discovery and development of biomarkers.

LONG-TERM GOALS

- Develop new serum- and tissue-based methods for early cancer detection and diagnosis in order to identify clinically significant cancers and predict clinical outcomes.

- Develop and validate biomarkers and imaging methods for the early detection of cancers for which there are no currently available effective screening methods, e.g. pancreatic and ovarian cancer.

- Expand collaborative efforts and shared resources to enhance the capacity to conduct biomarker development and validation trials.

- Develop imaging methods, including radiomics, to improve the detection of early stage cancers and to distinguish indolent from aggressive cancers.

- Develop and validate multianalyte or multimodal biomarkers to improve the accuracy of early cancer detection. CancerSEEK, is an example of a multianalyte test. Five EDRN investigators participated in the development and testing of CancerSEEK, which

simultaneously determines the levels of eight proteins and the presence of cancer gene mutations in circulating DNA. The test is aimed at screening for eight common cancer types that account for more than 60 percent of cancer deaths in the U.S. Five of the cancers covered by the test currently have no screening test.

- EDRN's future directions include (1) development of strategies to study the natural history of cancer to aid in developing better tools for determining which cancers are clinically important, (2) integration of the genetic, cell signaling and biochemical pathways with biomarker discovery efforts to have a broader applicability across different tumor types, (3) determination of the potential of novel network- and pathway-based markers to detect and diagnose cancer, and (4) examination of whether pathway biomarkers would allow a systems biology approach to diagnosis, prevention and therapeutic strategies.

RESOURCES AND TECHNOLOGY

- Expand the EDRN biomarker database to capture and share methods and pre-competitive data on the validation and qualification of biomarkers.

- EDRN collects biospecimens that are well characterized, with comprehensive clinical, demographic, and epidemiologic information. EDRN will continue assembling prospectively collected specimen reference sets for each organ site to facilitate pre-validation and validation of biomarkers. EDRN is already collecting sample reference sets, a set of well characterized controls and cases, for rapid evaluation of technologies and biomarkers before initiating large, expensive validation trials.

- Standardize methods for tissue procurement and biospecimen banking that are not only indispensable for current studies but also for testing emerging technologies which require validation.

- Continue and expand partnerships with Human Proteome Organization (HUPO) and its US affiliates.

- Work with the American Association of Clinical Chemistry, the FDA, and the ASCO to set up clinical standards for diagnostic markers and assays.

- Collaborate with NCI cooperative groups to develop collaborative studies on prognostic and predictive markers.

BREAST AND GYNECOLOGIC CANCERS

Breast • Ovarian

Breast Cancer

Strategic Goals

The increase in incidence of breast cancer observed over the past 20 years is almost entirely attributable to the detection of ductal carcinoma in situ (DCIS) and stage I cancer by imaging. The large majority of these lesions remain indolent. At the same time, there are many cancers that are being missed by the current screening modalities, many of which tend to be aggressive disease, such as "interval" and hormone-receptor negative or triple-negative breast cancers (TNBC). The incidence of the latter is significantly higher in premenopausal women where imaging screening modalities are significantly less effective. The ultimate goal is to develop non-invasive methods for detecting and characterizing pre-cancerous and cancerous breast lesions with poor prognosis with certainty when they are small and more easily treatable. Specifically, biomarkers are needed that can either augment mammography in the short term or replace mammography in the long term. Biomarkers are also sought for the assessment of risk of progression from benign breast disease (BBD) or DCIS to invasive breast cancer (IBC), and for the detection of aggressive cancers, such as TNBC, the majority of which are not detected by routine imaging.

- Improve the performance of screening mammography. Can biomarkers further improve the interpretation of conventional mammography or other computer-aided technologies?

- Distinguish benign from malignant breast lesions. Can biomarkers detect characteristics of specific types of benign and malignant breast lesions and stratify benign disease into high and low risk for progression?

- Improve early detection of different molecular subtypes of breast cancer

- Can tumor-specific biomarkers be identified and used s contrast agents to improve the performance of any imaging modality?

The Plan

Identification and validation of:
- Biomarkers to further improve the interpretation of conventional mammography or other computer-aided technologies;

- Biomarkers that detect characteristics of benign and malignant breast lesions and stratify benign disease into high and low risk for progression;

- Biomarkers which, in conjunction with mammography, can distinguish malignant from benign lesions in order to reduce or eliminate unnecessary biopsies;

- Biomarkers to detect highly proliferative early malignant lesions associated with increased mortality;

- Tumor-specific biomarkers that could be used as contrast agents to improve the performance of existing imaging modalities;

- Combination of blood- based biomarkers and radiomics to distinguish benign from malignant breast lesion

Ovarian Cancer

Strategic Goals

The absence of accurate screening biomarkers, coupled with the typical late stage diagnosis of ovarian cancer, contributes to the significant lethality of the disease. Thus, early detection is important as currently there are no reliable biomarkers available for screening for ovarian cancer. Transvaginal ultrasound and the serum tumor marker CA-125 have been explored as a strategy for the early detection of ovarian cancer, but the sensitivity, specificity, and lead time (earliness of detection) are not optimal. For example, increased CA-125 levels are found in about three percent of post-menopausal women, resulting in false positives for this biomarker. Recent morphologic and molecular genetic studies have resulted in a paradigm shift with regard to the origin of ovarian cancer and its pathogenesis. This has important implications for research and for radically changing our approaches to early detection, prevention, and treatment of the disease. The development of new circulating biomarkers to be used as a first-tier screening modality for the general or high risk population is a strategic goal that would improve the early detection of the most lethal, high-grade serous ovarian cancers is of utmost importance. Hence, there are two key questions:

- Can biomarkers further improve the interpretation of conventional transvaginal ultrasound or other computer-aided technologies?

- Could a strategy involving the use of risk stratification, accurate biomarkers, and secondary diagnostic imaging tests be a cost-effective model for ovarian cancer screening in a high risk or even general population?

The Plan

- Develop a blood test for early detection of ovarian cancer.

- Identification and validation of biomarkers that can further improve the interpretation of conventional transvaginal ultrasonography or other computer-aided imaging technologies.

- Identification and validation of biomarkers that can identify and stratify early ovarian lesions as benign and at high risk of progression to ovarian cancer.

- Develop a high sensitivity uterine sampling test for high risk patients.

- Develop a strategy combining risk stratification, biomarkers, and secondary diagnostic imaging as a cost-effective screen in high risk or general population.
- Build a consolidated pre-diagnostic specimen repository (PLCO, WHI, CARET, Nurses' Health Study, UKCTOCS), statistically powered for discovery and rapid pre-validation of candidate markers;

- Utilize the NCI TCGA data on ovarian cancer somatic genetics to inform early detection biomarker development; and

- Biomarker discovery efforts on a better understanding of the natural history of the disease and better characterization of putative pre-malignant ovarian lesions.

COLORECTAL AND OTHER GASTROINTESTINAL CANCERS

Colon • Esophagus • Liver • Pancreas

Colon Cancer

Strategic Goals

Colon cancer is both the third most frequently diagnosed cancer and the third cause of cancer deaths in the United States. Successful prevention of colon cancer depends on early detection. Overall, the lifetime risk of developing colorectal cancer is about 4%. Current screening technologies include fecal occult blood test, fecal immunochemical test, stool DNA test, sigmoidoscopy, and colonoscopy. Although screening has been shown to reduce cancer deaths, all of the current screening methods have limitations that reduce their effectiveness. Fecal occult blood test and fecal immunochemical test fail to detect a significant fraction of colon cancers and advanced adenomas and have high rate of false positive rates. Although the stool DNA test (Cologuard) has significantly better performance, (stool-based testing is rejected by 40% of the population. Sigmoidoscopy and colonoscopy are invasive, expensive, and cause patient discomfort. Consequently, many individuals who should be screened are not. Thus, there is a need to develop blood or urine-based biomarkers that accurately identify individuals that are at risk of having colon cancer or advanced adenomas and that need further testing by colonoscopy. Genetic, epigenetic, and proteomic methods are being used to identify potential colon cancer biomarkers.

- Can biomarkers be used to accurately determine which patients are at risk and in need of further testing (i.e., colonoscopy)?

- Can biomarkers detect characteristics of benign and malignant lesions and stratify benign disease into high and low risk for progression?

The Plan

- Develop and validate blood or urine-based biomarkers with accuracy comparable to FIT to increase the number of people being screened (stool-based testing is rejected by 40% of the population).

- Develop and validate ctDNA as a biomarker to monitor for early recurrence of colorectal cancer.

- Develop colon cancer and advanced adenoma reference sets comprised of serum, plasma, urine, DNA from WBC (white blood cells), and paraffin embedded tissues from normal colon, adenomas, inflammatory bowel disease, and colorectal cancer.

Liver Cancer

Strategic Goals

Hepatocellular carcinoma (HCC) accounts for approximately 85-90% of all primary liver cancer. The five-year survival rate for patients detected with early stage HCC is greater than 70% with transplant or resection, but for patients with advanced HCC, the 5-year survival is less than 5%. The etiological/risk factors for liver cancer include viral hepatitis (HBV and HCV), non-alcoholic fatty liver disease (NAFLD), non-alcoholic steatohepatitis (NASH), obesity, metabolic syndrome, alcohol abuse, and environmental factors (e.g. smoking and aflatoxin B1). Approximately 80-90% of HCC occurs in patients with liver cirrhosis. In patients with cirrhosis, the five-year cumulative risk of liver cancer ranges from 5-30%, depending on the etiology (4), and patients with advanced fibrosis or cirrhosis represent a high-risk group for liver cancer and are recommended for surveillance.

Surveillance of patients with cirrhosis is an important goal for early detection of hepatocellular carcinoma. AFP (alpha-fetoprotein) level in the blood and ultrasound are the current standard methods used to detect liver cancer. However, AFP has a high false positive rate and can miss many early stage cancers. Better biomarkers need to be developed for hepatocellular carcinoma for early detection and diagnosis, which will reduce the mortality of this cancer.

- Combine biomarkers and clinical data to develop better methods to screen cirrhotic patients. Examine blood-based biomarkers, longitudinal changes in biomarker, and clinical data (i.e. etiology, age, gender) to predict the risk of developing liver cancer.

- Use circulating tumor DNA to detect HCC based on cancer-specific genetic and epigenetic aberrations found in the liver.

- Improve imaging methods for HCC. Increase detectability of HCC using tagged HCC-specific peptides. Leverage machine learning and novel imaging and biomarker data to differentiate benign from malignant liver nodules. Test accuracy of abbreviated MRI and the benefit of adding serum biomarkers, and clinical data for early detection of HCC.

- Assess biomarkers, imaging and algorithms to determine if they have a better sensitivity and specificity than AFP or use in combination with AFP.

The Plan

- The HEDS (Hepatocellular carcinoma Early Detection Strategy) study serially collected serum and plasma from patients with cirrhosis. These patients were also monitored with imaging (ultrasound) and followed until the development of HCC. The serially collected reference set is will be used to validate of biomarkers for the early detection of HCC.

The goal is to use this prospective liver cancer reference set to test the ability of previously developed biomarkers (e.g., AFP, DCP and AFPL3%) and newly discovered biomarkers to detect early stage cancer in cirrhotic patients and determine which patients with cirrhosis are likely to progress to cancer.

Pancreatic Cancer

Strategic Goals

Pancreatic cancer has a very high mortality rate, with the mean survival time of less than six months, largely due to the late diagnosis. The current standard biomarker for the diagnosis of pancreatic cancer is the serum marker, CA 19-9. In an asymptomatic population, this biomarker has a positive predictive value below one percent. Currently, universal screening for pancreatic cancer is not recommended. Identification of high-risk populations and better biomarkers are needed for the early detection and diagnosis of pancreatic cancer. There are currently three known high-risk groups; individuals with a family history of pancreatic cancer, patients with pancreatic cysts and patients with new onset diabetes. Approximately 0.5 - 1% of adult new onset diabetes is due otherwise asymptomatic pancreatic cancer, and 25-50% of pancreatic cancer patients may have new onset diabetes as their first symptom.

Precursors of invasive ductal adenocarcinoma of the pancreas include pancreatic intraepithelial neoplasias (PanINs), intraductal papillary mucinous neoplasms (IPMNs) and mucinous cystic neoplasms. PanIN's are classified from PanIN1-3, with PanIN-3 being clinically significant because of its potential to progress to invasive cancer. Commonly used imaging methods include endoscopic ultrasound, abdominal CT scan, or MRI. PanINs are difficult to detect using current imaging modalities. However, these methods are increasingly detecting mucinous cystic lesions (IPMNs) in the pancreas. Clinically, the study of IPMNs that have the potential to progress to pancreatic cancer is important and has potential in early detection in identifying asymptomatic patients.

- What is the link of other diseases in increasing the risk for pancreatic cancer, especially Type 3c diabetes?

- Can imaging be used to identify or stratify populations at risk for the development of pancreatic cancer (e.g. IPMNs)? Furthermore, can biomarkers be used to improve imaging techniques and better identify cysts with potential to progress toward pancreatic cancer?

- Can biomarkers be identified and used with CA 19-9 to better identify pancreatic precursor lesions with the potential to progress toward pancreatic cancer?

The Plan

- Develop and test biomarkers and/or clinical factors that can identify those patients with new on et diabetes that have pancreatic cancer.

- Identify patients at high risk for developing pancreatic cancer and develop methods (*in vitro* diagnostic and/or imaging) to detect early stage pancreatic cancer and PanINs in these patients.

- Develop biomarkers to distinguish IPMNs with high risk of developing pancreatic cancer from those with low risk of developing cancer.

- Use panels of markers to identify profiles for precursor lesions associated with pancreatic cancer (IPMNs or PanINs).

- Develop imaging and biomarkers to stratify populations at risk for developing pancreatic cancer.

- Establish and maintain biorepositories with specimens from patients with early stages of pancreatic cancer (Stage IA, IB, and Stage IIA and IIB), with pancreatic cancer precursor lesions (IPMNs, PanINs), and with benign pancreatic conditions (acute and chronic pancreatitis, biliary obstruction).

Esophageal Cancer

Strategic Goals

The incidence of esophageal adenocarcinoma (EAC) in the United States has been increasing for the past four decades. The prognosis for most patients diagnosed with EAC is poor with the 5-year relative survival rate of 17%. Barrett esophagus (BE) is the only established precursor for EAC. BE is typically diagnosed in individuals with chronic acid reflux symptoms and affects an estimated 1%–5% of the general U.S. population. Detection of BE currently requires performing esophagogastroduodenoscopy (EGD); however, because of the high cost of EGD and the lack of a randomized controlled trial demonstrating cost-effective reduction in EAC, endoscopy screening for BE has not been routinely recommended. Only a minority of patients who develop EAC have a prior diagnosis of BE and of those with BE, less than 1% will progress to EAC per year.

The presence of the antecedent BE remains undetected and unknown in about 95% of cases of EAC. There is a need for alternative methods for BE detection that are less expensive than EGD and can be readily implemented in an at-risk population. For example, EDRN investigators used an experimental, swallowable, balloon-like sampling device (EsoCheck) to check esophageal tissue for changes in DNA methylation in two genes, CCNA1 and VIM (EsoGuard), each of which they have previously shown are biomarkers for BE. When combined, tests of methylated CCNA1 plus methylated VIM DNA detected BE metaplasia with 90.3% sensitivity and 91.7% specificity. They have proposed that this approach could be a cost-effective, sensitive, and well-tolerated way of screening for BE in at-risk individuals. They are currently developing biomarkers than can distinguish dysplastic from non-dysplastic BE.

- Can biomarkers be used to accurately determine which patients may have BE and need further evaluation ?

- Can non-endoscopic screening biomarkers be used to distinguish dysplastic from non-dysplastic BE and the need for further evaluation?

- Can non-endoscopic biomarkers be used to detect EAC?

The Plan

- Develop and validate non-endoscopic methods to detect esophageal adenocarcinoma.

- Develop and validate a non-endoscopic method to distinguish dysplastic from non-dysplastic Barrett's esophagus.

- Validate Esoguard (licensed to PAVmed), a non-endoscopic method to detect Barrett's esophagus using EsoCheck (licensed to PAVmed), a swallowable balloon-like sampling device.

LUNG CANCER

Strategic Goals

Lung cancer continues to be the most lethal cancer in the US with over 160,000 deaths per year. The incidence of lung cancer is driven predominantly by smoking with a prevalence of lung cancer in the smoking and former smoker populations in the range of 10–15%. About 90% of patients with lung cancer have a significant smoking history. In the EDRN, a variety of lung cancer markers have been pursued including panels of gene methylation markers, mitochondrial DNA mutations, mitochondrial number, chromosomal abnormalities, proteomic profiling, autoantibodies to defined proteins or glycans, miRNAs, and gene expression profiles. Within a high risk population of smokers, many markers were not able to distinguish non-diseased smokers from lung cancer patients; however, they clearly differentiate smokers from nonsmokers. These molecular alterations persist later in life as evidenced in cohorts who ceased smoking five years previously. These findings highlight how smoking induces profound molecular alterations in the epithelial linings of the lungs setting them on a path towards neoplasia. Recently, interest has arisen to characterize the sequence of driver mutations that occur during oncogenesis from the earliest stages of atypical adenomatous hyperplasia (AAH) to invasive cancer. Understanding how somatic gene defects accumulate over time and determine which critical genes are responsible for progression to invasive cancer will yield greater insights into predicting and stratifying those smokers that may be at greater risk for developing lung cancer.

Results from the National Lung Screening Trial (NLST) found that screening by low-dose CT decreased lung cancer mortality by 20% highlighting the benefits that can be achieved with more frequent detection of early stage lung cancers. It is noteworthy that 25% of the subjects presented with nodules of which 96% were benign. Current management of abnormalities found by chest CT can lead to expensive and invasive diagnostic follow-up. The utilization of non-invasive biomarkers to discriminate between early stage malignancies and benign nodules could greatly reduce the need and costs associated with clinical follow-up of subjects with indeterminate nodules.

- Can biomarkers be used in combination with CT imaging in patients presenting with small pulmonary nodules to stratify those that will need to undergo further work up for diagnosis of lung cancer from those with benign disease? This test could spare follow-up on many subjects who do not have an early stage malignancy.

Can gene mutations occurring in early stages of hyperplasia through successive stages to invasive adenocarcinoma be used to understand the sequence of genetic insults that drive lung cancer progression?

The Plan

- Develop and validate biomarkers and imaging methods to detect lung cancer among smokers with indeterminate nodules that are detected by low dose CT.
- Conduct studies using semantic and radiomic image-based features to reduce false positives and distinguish indolent from aggressive pulmonary nodules.

- Discover and validate biomarkers for the early detection and prognostication of pleural mesothelioma.

- Validate all promising lung cancer biomarkers on a prospectively collected cohort of 200 subjects with indeterminate pulmonary nodules. Approximately 30 lung cancer cases are estimated to be found in this cohort after 2 year follow-up. Biomarkers to be tested among EDRN BDLs are gene methylation markers, gene expression profiles, a host of serum protein markers, anti-glycan autoantibodies, and miRNAs.

- Recruit 200 subjects for a prospective study of current or former smokers greater than 45 years of age and over 25 pack-years of smoking. Subjects must have noncalcified pulmonary nodules on chest CT in the range of 7–20 mm in diameter. Specimens to be obtained from each subject are serum/plasma, bronchial brushings during bronchoscopy, and nasal brushings. All subjects will be followed up for two years until a diagnosis of lung cancer is determined. All samples will be sent to one site where relabeling and blinding will occur before being sent out to all biomarker testing labs.

- Perform deep genomic sequencing of heterogeneous sites within a single subject to uncover the sequence of genetic mutations that lead to progression of invasive adenocarcinoma. Two strategies are used in this paradigm. The first is to sample adenocarcinoma tissue, premalignant lesions, and adjacent normal lung tissue from individual lung cancer patients. The second approach is to sample various histological zones of progression from subjects presenting with lesions characterized as atypical adenomatous hyperplasia, adenocarcinoma in situ, or minimally invasive adenocarcinoma. DNA obtained by laser capture microdissection will be analyzed by deep sequencing methods where comparisons of the similarities or heterogeneity in the mutational landscape within each lesion should reveal the important cascade of mutational events leading to invasive adenocarcinoma.

PROSTATE AND OTHER UROLOGIC CANCERS

Prostate • Bladder and Other Urogenital Cancers

Prostate Cancer

Strategic Goals

Prostate-specific antigen (PSA) had a remarkable effect on the detection diagnosis and monitoring of prostate cancer. Although clinically localized prostate cancer has become highly curable, the overall death toll remains high due to recurrence and progression to hormone-refractory and metastatic disease, which remains incurable. PSA tests resulted in detection of a large number of false positive prostate cancers, leading to the phenomena known as "overdiagnosis" and repeated biopsies.

There is an urgent need for predictive markers for early detection of aggressive prostate cancer, which would be distinguished from the less aggressive non-lethal forms of prostate cancer. Markers should be developed based on recent advances in cancer biology with application of Omics related approaches (genomics, epigenomics, metabolomics, and proteomics).

The recently discovered fusion transcripts (TMPRSS2-ETS), which are frequently (~50%) expressed in prostate cancers, are promising markers that were or are being validated for early detection; as a prognostic markers for the development of aggressive cancer; or as risk markers. Those include TMPRRS2-ERG, TMPRRS2-ETV1, TMPRSS2-ETV4 as well as genes, which are exclusively expressed in fusion negative prostate cancers such as the mutant SPOP and the overexpressed SPINK1. In addition, a recently discovered marker, SChLAP-1, a long non-coding RNA, that is highly correlated with ~20% of aggressive prostate cancers.

In light of unmet needs of minimizing false positives, differentiating between aggressive and non-aggressive forms of prostate cancer, and estimating risk, EDRN investigators are trying to address the following areas:

- An overarching goal is to improve early detection of prostate cancer using molecular markers (RNA, DNA, protein, metabolites) in blood or urine to predict the presence of aggressive prostate cancer on histopathology of subsequent biopsy or prostatectomy.

- Further development and validation of promising markers for early detection of significant cancers; for risk assessment and for prostate cancer stratification (e.g., fusion genes positive as compared to fusion genes negative cancers).

- Development and validation of markers based on recent advances in cancer biology (e.g., markers associated with prostate cancer stem cells; markers associated with reactive

stroma adjacent to a cancer focus; and application of integrative approaches to biomarker development for early and aggressive forms of prostate cancer.

- Development and validation of biomarkers for prostate cancer risk assessment based on recent results of whole genome association studies; and provide guidelines for integration of risk markers into clinical practice.

- Improve the performance of existing markers (e.g., PSA by evaluating its molecular forms, such as [-2]proPSA or combining it with a panel of tumor specific markers such as TMPRSS2-ERG and mutant SPOP.

- Combining biomarkers with imaging to improve the diagnostic performance of imaging – reducing the false positive rate; improving the detection of significant cancers; and to improve testing of same foci for patients on active surveillance.

- Reduce the number of unnecessary biopsies (primary and repeat biopsies).

- Develop and validate biomarkers associated with grade and stage upgrading.

- Continue to assemble prostate cancer "reference sets" for biomarkers discovery and validation studies. Composition of each collection of "reference set" should be tailored to answer specific clinical question.
  - Body fluids to include, plasma, serum, urine, EPS (Expressed Prostatic Secretions)
  - Circulating tumor cells, WBC
  - Tissue

The Plan

- Develop non-invasive tests to discern indolent cancers (i.e. histopathological Grade Group I prostate cancers, which do not need treatment) from aggressive cancers (i.e. histopathological Grade Groups II, III, IV or V) treatment of which reduces cancer death.
- Determining whether MRI prostate imaging and biomarkers can improve the prediction of cancer extent and aggressiveness to determine suitability for active surveillance or treatment.

- Develop and identify differential expressed markers for detection of prostates cancer perpetuating cells – prostate cancer stem cells.
  - Detection of cancer stem cells among exfoliated cells (for cancers of the urological system such as prostate and bladder).
  - *In vivo* early detection of early stage prostate cancer by detection of prostate cancer stem cells. This approach will be based on combination of affinity reagents, specific for the cancer stem markers, with imaging technologies. The

proof of principle will be established in animal models. Similar approaches should be conducted for other cancers as well.

- Development of biomarkers based on cancer stroma cells associate markers.
  - Identification of prostate cancer stoma markers in tissue culture, 3-D cultures, cancer tissue, and in animal models.
  - Detection of cancer reactive stroma markers in prostatectomy specimens.
  - Detection of cancer stroma markers in body fluids

- Application of Systems Biology approach to integrate variety of cancerous processes as detected by Omic technologies [e.g., differential expression of genes, non-coding RNAs including miRNAs, fusion transcripts (TMPSS2-ERG), mutant or amplified oncogenes, inactivation of tumor suppressor genes; mutations / inactivation of DNA repair mechanisms; differential proteomics; and abnormal PTMs; differential expression of epigenomic markers (signatures for abnormal DNA methylation; signatures of abnormal histone methylation and acetylation; differential expression of metabolites, etc] to identify perturbed biochemical pathways and processes that could be used for early detection and early prediction of aggressive cancers. .

- Initiate validation studies based on new generation of biomarkers, such as gene fusion product (TMPRSS2-ETS); mutant genes and differentially overexpressed coding and non-coding genes and or metabolites in tissue and body fluids (biopsy and prostatectomy specimens as well as in urine, EPS and blood).

- Circulating Tumor Cells (CTCs) to estimate cancer progression and early recurrence. CTCs could be early indicators for the development of an aggressive cancer.

- To develop surrogate markers, the GU (Genitourinary) collaborative group will assemble a collection of WBC from cancer and control patients, which could be used to develop surrogate markers, such as functional biochemical tests, and polymorphism, which correlate well with risk; e.g., the capacity to repair damaged DNA by certain DNA repair enzymes like OGG1 was recently correlated with risk of developing lung cancer due to smoking. The activity of this enzyme was identical in the surrogate tissue, WBC, and lung epithelia cells in the same individuals.

- Continue to establish collection of body fluids (plasma, serum, urine, EPS) as "reference sets" for rapid evaluation of biomarkers before entering the validation trials; and for discovery purpose using well annotated and well represented collection of specimens to minimize the presence of confounders and bias.

- Reference Sets and Cohorts:
  - EDRN will assemble reference sets based on tissue and body fluids to discover and test candidate biomarker/s that may assist in answering specific clinical needs.

- Biopsy positive reference sets will be necessary for:
  - o For general population screening with marker/s adding or replacing PSA, all cases and controls should have a biopsy. This collection should not be triggered by elevated PSA.
  - o For identification and testing markers that will be used to assist in clinical decisions (i.e., whether a patient needs a radical prostatectomy). One need to keep in mind that the current practice is based on clinical predictors (e.g., Gleason score).
  - o For prediction of cancer progression, including the development of metastasis and mortality, there is a need for specimens from cohorts with many years of follow up. For this purpose we will collaborate with the appropriate programs (i.e., collaborative groups, PLCO).
  - o Expand the collection biopsy from high risk individuals, men with elevated PSA, or abnormal DRE (Digital Rectal Exam).
  - o For Biopsy negative population. There is a need to develop a tissue resource and combine tissue-based marker with body fluid markers to increase negative predictive value.

Bladder Cancer

Strategic Goals

Bladder cancer is the fifth most common cancer in the Western world, affecting about 4% of all cancer patients and is the cause of about 3% of all cancer-related deaths. The estimated life probability of developing bladder cancer in the United States is 1 in 28 for men and 1 in 87 for women. Bladder cancer occurs in two clinically significant forms: Superficial (TNM: Ta, TIS, T1) and Invasive (TNM: >T2). Seventy-five percent of the patients are diagnosed with superficial disease, and only a minority (about 15%) is at risk for progression. Approximately 70% of these patients will experience recurrence of the disease within 10 years. The majority of recurrences occur within the first two years after diagnosis. The vast majority of invasive bladder cancers occur in patients without a prior history of papillary tumors. Although urine cytology and cystoscopy are considered standards of care, these are less than optimal in detecting all forms of bladder cancer. The sensitivity and specificity of urinary cytology are 25-50% and 90-100%, respectively. The sensitivity and specificity of cystoscopy is 90-100% and 75%, respectively. In recent years, several new biomarkers and tests for detection of bladder cancer gained acceptance and FDA approval (BTA™, BTA stat™ FDP™ NMP22™ and the UroVysion). Most of these FDA-approved tests can augment, but not replace, the cystoscopy for diagnosis of bladder cancer. Consequently, there is a need to improve the current practice of bladder cancer detection and surveillance. Strategic goals are to:

- Develop non-invasive diagnostic tests for early detection of superficial bladder cancer (to minimize the number of unnecessary cystoscopies) and for early recurrence of superficial bladder cancer.

The Plan

- Develop biomarkers associated with the four major subtypes of bladder cancer: (1) transitional cell carcinoma; (2) squamous cell carcinoma; (3) adenocarcinoma; and (4) small cell carcinoma. Also, indentify biomarkers associated with bladder cancer stem cells, bladder stroma cells, and others.

- Validate (analytical and clinical validation) promising biomarkers for the various subtypes of bladder cancer [i.e., methylated DNA sequences, genetic alterations (mutations, amplifications, and deletions) in candidate oncogenes and tumor suppressor genes, and alterations in mtDNA, etc.].

- Assemble bladder cancer "reference sets" and appropriate controls for biomarker discovery and validation studies. The composition of each collection of "reference sets" should be tailored to answer specific clinical questions.

- For prediction of cancer progression, including the development of metastasis and mortality, there is a need for specimens from cohorts with many years of follow-up. For this purpose, EDRN will collaborate with the appropriate programs (i.e., collaborative groups). Collected specimens will include body fluids (urine, plasma, and serum), tissues, circulating tumor cells, and WBCs.

- Develop biomarkers including biomarkers derived from stromal and stem cells for molecular classification of bladder cancer subtypes.
    - Use circulating tumor cells (CTCs) to estimate cancer progression and early recurrence. CTCs could be early indicators for the development of an aggressive cancer.

Other Urogenital Cancers

At present, there is no established serum or urinary biomarker for the diagnosis or management of kidney cancer as well as a lack of specific symptoms in people with early stage disease. Furthermore, an increasingly larger subgroup of patients with small renal masses are not treated but are instead monitored for disease progression by CT or MRI.

The Plan

- Priorities for kidney cancer are the development of biomarkers for non-invasive early detection and as prognostic indicators of aggressiveness of disease.

OTHER CANCERS

Although the focus of the EDRN is on the early detection of the major epithelial cancers and those cancers with very high rates of mortality, e.g. pancreatic and ovarian cancers, it does support pilot projects on other cancers and on novel discovery technologies through an Associate Member Program (https://edrn.nci.nih.gov/colops/assoc). The EDRN also supports biomarker validation trials on other types of cancers. Requests for support of validation trial, which can be from both EDRN and non-EDRN investigators, are reviewed by the EDRN Executive Committee and require strong preliminary data demonstrating the biomarker, panel of biomarkers and/or imaging method has the sensitivity and specificity to be clinically useful (https://edrn.nci.nih.gov/colops/vsp).

- Scientific merit
- Study design: e.g., prospective versus cross-sectional
- Technical parameters: reproducibility, sensitivity, specificity, throughput, automation, and cost
- Clinical or Translational impact: e.g., more common cancers or a significant impact for rare cancers; tests geared toward screening of the general or high-risk populations for early detection of cancer versus risk assessment or prediction of disease progression, etc.
- Portfolio balance within EDRN and NCI's needs
- Practicality, Portability, and Feasibility: e.g., required sample size or amount of biospecimen; non-invasively obtained samples; bias in study population.
- Collaborative strength, including contribution of resources and technology. Collaboration is a central mission of EDRN.

Other cancers of potential interest include

- Head and neck cancer is the sixth leading cancer in the world and more than 65,000 Americans are diagnosed with HNC per year. Head and neck squamous cell carcinoma (HNSCC) is the most common type, accounting for 70-90% of HNC. A subset of HNSCC is caused by human papillomavirus (HPV), where HPV-related HNC have increased significantly in recent years

  o Develop biomarkers to distinguish benign from aggressive lesions

  o Develop biomarkers to combine with currently available methods for increasing test sensitivity and specificity in medical diagnosis.

- Mesothelioma: biomarkers and/or imaging methods for the early detection of mesothelioma in asbestos exposed individuals

- Cervical: improve the effectiveness of cervical cancer screening in the United States; improve on risk-stratification provided by HPV testing to allow screening intervals to be increased in order to reach unscreened populations by using a more culturally acceptable sampling method (self-sampling, urine, blood).

INTERNATIONAL COLLABORATIONS

According to the World Health Organization in 2008, 36 million deaths occurred due to noncommunicable diseases, including cardiovascular diseases, cancer and chronic respiratory diseases. Cancer contributed to 7.6 million deaths. Diabetes, a risk factor for a number of cancers, caused an additional 1.3 million deaths. Behavioral risk factors, including tobacco use, physical inactivity, and unhealthy diet, contributed significantly to most of the deaths due to cancers. More than two thirds of all cancer deaths occur in low- and middle-income countries. Lung, breast, colorectal, stomach and liver cancers cause most cancer deaths. In high-income countries, the leading causes of cancer deaths are lung cancer among men and breast cancer among women. In low- and middle-income countries cancer levels vary according to the prevailing underlying risks and exposures. In Sub-Saharan Africa cervical cancer, due to HPV infection, is the leading cause of cancer death among women despite proven screening tests and effective vaccines. Risk factors for many cancers include four shared behavioral factors (tobacco use, unhealthy diet, insufficient physical activity and the harmful use of alcohol), but infections such as hepatitis B, hepatitis C (liver cancer), human papillomavirus (HPV; cervical cancer) and Helicobacter pylori (stomach cancer) cause up to 18% of cancer burden. In addition, radiation and a variety of environmental and occupational exposures are of varying importance.

A recent report by the American Cancer Society examined the growing burden of cervical cancer in low- and middle-income countries. The report, titled "Saving Women's Lives: Accelerating Action to Eliminate Cervical Cancer Globally," calls on Congress to dedicate a portion of U.S. global health funding to improve access to preventive vaccinations, screening and treatment through international health initiatives. Research to optimize these strategies should be a priority. Today cervical cancer is the leading cause of cancer-related death for women in 42 countries. These high rates of cervical cancer around the globe clearly illustrate the need for research and timely action. Working with international partners to ensure that girls and women in LMICs have access to the HPV vaccine, along with services to screen and treat cervical cancer will go a long way toward the goal of eliminating cervical cancer.

EDRN Collaboration with the NCI Center for Global Health: The EDRN has worked closely with NCI's Center for Global Health on cost-effective cancer technologies that can be beneficial in global low- and middle-income countries settings and may also be beneficial in underserved US populations. Although prevention, diagnosis, and treatment approaches exist in the US for most cancers, there are many examples of disparities in cancer outcomes for certain underserved populations. This effort has focused on technologies for the detection that show promise to deliver improved cancer outcomes and has addressed cancers of the cervix, colon/rectum, esophagus, and oral cavity. Three Funding Opportunity Announcements have been published which encourage grant applications to develop/adapt, apply, and validate existing or emerging technologies into user-friendly cancer prevention, diagnosis, or treatment to be used in low resource settings.

China EDRN: (C-EDRN): The Chinese National Cancer Center has developed a C-EDRN, which is modeled after US EDRN and consists of Biomarker Discovery Laboratories, Epidemiology Laboratories, and a Data Analysis Center. As part of this collaboration, there are regular conference calls, in person meetings, and participation in each other's scientific conferences. A joint publication on next gen sequencing of prostate cancer in African Americans and Chinese Population was published in collaboration with the Shanghai Biotechnology Center, a premier gene sequencing center (*Science Reports* 8:12868,2018).

EDRN-Japan's Agency for Medical Research and Development: Since 2012, Japan AMED and EDRN have conducted joint meetings, alternating between US and Japan. The 7th Annual Workshop is scheduled for Tokyo in January/February 2020. This collaboration has yielded research publications and the validation of biomarkers from Japan using EDRN reference samples and has led to the approval of Apolipoprotein A2 as potential biomarkers for pancreatic cancer. A joint publication (EDRN and AMED investigators) on Apo-2 has been published in Science Reports

EDRN-Cancer Research United Kingdom (CR-UK): CR-UK program managers regularly consult with EDRN program officials to ensure that there is no duplication and overlaps with CR-UK funded projects. In the case of overlaps, we bring together investigators from both sides to collaborate and exchange information. A CR-UK program manager spoke at a recent EDRN Scientific Workshop.

DATA SCIENCE

The opportunities and challenges of data-driven computing are enabling a major shift for scientific research. The National Research Council, in its report "Frontiers in the Analysis of Massive Data" (2013), identified many of the challenges in scientific analysis for data-driven disciplines occurring in fields such as astronomy and biology. NASA/JPL and Caltech, through their joint initiative for Data Science and Technology, have been active in working to advance research, capabilities, and education across these fields with two of the National Cancer Institute's major research programs, the Early Detection Research Network (EDRN) and the Molecular and Cellular Characterization of Screen-Detected Lesions (MCL) by enabling data-driven discovery for cancer biomarker research. In particular, the team pioneered a national data ecosystem for EDRN's to systematically capture, process, manage, share, and analyze data across multiple research centers. A critical focus is the advancement of a scalable and systematic approach for reproducibility of derived data analysis results. As the data across such areas as radiomics, genomics, and pathomics substantially increase, the need to systematically capture, manage and analyze data through an advanced knowledge system is critical for integrated data use across Laboratories.

To establish the EDRN knowledge environment, JPL has researched and developed an innovative informatics infrastructure of data science services tailored to support biomarker research. By collaborating on software and data-driven methods developed by JPL and Caltech for space and earth science research, the EDRN has been able to heavily leverage cloud-based capabilities that positions the EDRN to scale to support the data and computational demands of the program. The architecture of the knowledge environment is based on supporting and linking diverse data from clinical phenotypes to imaging to genomics through the EDRN's efforts to standardize data models and common data elements for cancer biomarkers. The infrastructure captures and links data from across the EDRN using nearly a thousand annotations of cancer biomarkers to terabytes of analysis results in the EDRN data commons, known as "LabCAS". LabCAS provides support to capture data from validation studies linking data from laboratory tracking tools (e.g., VSIMS) at the DMCC to the analysis of data captured in the EDRN. Capabilities are also in place to run repeatable analysis pipelines for genomics, image analysis, and other complex data types. These can be customized for each study as shown in figure 1. Several validation study teams are working with JPL to capture their data (e.g., Prostate MRI, Breast Ref Set Imaging Project, etc.). Other efforts such as NCI's pre-cancer atlas are developing central data pipelines that shepherd annotated data into the data commons. Similar capabilities are being developed for alignment of massive imaging data for pancreatic cancer. As EDRN is amassing more data, there is greater opportunity for applying advanced data mining to and across different repositories of data (e.g., lung, pancreas, pathology, radiology, genomics, etc.). Integrated tools provide access to and visualization of the data in the data commons. Several of the tools are open source tools (e.g., QuPath, 3D Slicer, OHIF, caMicroscope, etc.) and are developed through collaborations with NCI's Information Technology for Cancer Research (ITCR) program. The entire knowledge environment is integrated with the EDRN portal, providing secure, multi-layer access to data for EDRN, NCI, research, and public communities. Approximately 2500 unique visitors come to the portal each month.

Beyond providing the core data science infrastructure, the EDRN is active in the research of next generation capabilities for crowd sourcing, machine learning, computational analysis, and visualization. JPL and Caltech have been working to bring in tools such as Zoonverse which have been successfully used in fields such as astronomy of crowdsourcing. These tools, in particular, have been useful for generating large labeled sets required to train machine learning algorithms in the classification of features in images. JPL and Caltech have been working with different PIs, in particular with lung imaging, to explore the use of these tools to improve the capture, annotation and construction of databases using crowd sourcing and collaborative methods in data analysis with a goal of looking at how data-driven approaches (e.g. deep learning) can be applied. The capture of EDRN's data in LabCAS provides a foundation for opening up new possibilities in these areas, and enabling new analysis approaches for consortiums like the EDRN which are highly distributed and diverse.

Additionally, Caltech, JPL, and NCI have explored the use of nascent virtual reality (VR) capabilities for analysis of multi-dimensional data. A recent prototype exploring lung cancer was presented at SigGraph 2019, one of the foremost research conferences in visualization and graphics. The VR prototype demonstrated multidimensional 3D data visualization of EDRN's radiology data, integrating with LabCAS. The prototype demonstrated opportunities for entirely new approaches for data exploration as data increases in size and complexity.

The following provides an overview of EDRN's progress in data science:

- Standards and a process for capturing highly annotated cancer biomarker data.
- Development of the LabCAS data commons infrastructure enabling EDRN to capture data, run pipelines, and link analytics including 77 collections and 23, 200 files.
- Deployment and instantiation of LabCAS on Amazon Web Services to support massive scalability and computation as well as collaborative analysis tools.
- Integration of analytical tools including OHIF, caMicroscope, QuPath, 3D Slicer..
- Development of cancer biomarker ontology and common data elements for annotating data in the data commons, and sharing of these standards with the research community by deposition in caDSR.
- Collaboration with the ITCR program investigators to integrate tools
- Integration of EDRN's data commons LabCAS into validation studies/reference sets including:
  - Breast Reference Set - BRSI (Jeffery Marks and John Heine)
  - Prostate MRI - P-MRI (John Wei)
- Development of a genomics pipeline for Boston University (Mark Lenburg)
- Development of a Secretome tool/pipeline (Michael Birrer, Steven Skates and Levi Waldron)
- Development of a pipeline for miRNA Measurements (Scott Pine)
- Capture and curation of 940 EDRN Biomarkers Curated in the EDRN biomarker database
  - Linking of EDRN biomarker studies to researched and discovered biomarkers

- o Linking of publications and data
- o Linking of external data sources

- Development and collaboration on OncoMX Portal through ITCR with George Washington University and integration with EDRN to link biomarkers and gene mutations.
- Development of crowd sourcing techniques based on Zoonverse for analysis of lung imaging data with UCLA and Moffitt.
- Development of the EDRN Portal to integrate sites, protocols, biomarkers, and data into a searchable knowledge environment
  - o Collaboration and coordination with the Data Management and Coordinating Center (DMCC) to maintain portal information
  - o Daily integration of databases to support an integrate biomarker data environment
  - o Collaboration with NCI's Center for Bioinformatics and Information Technology (CVBITT) to provide operational support for running the portal.
- Research and development of image alignment and an automated data pipeline with University of Nebraska (Tony Hollingsworth) to support automated alignment of 3D imaging for biomarker discovery in pancreatic cancer.
- Presentations to the HTAN on the biomarker knowledge environment and model-drive architecture for large scale data integration, sharing and analytics.
- Pilot of an experimental VR platform, spun out from Caltech, for exploring lung images as a future concept for multi-dimension data analysis. Presentation at Association of Computer Machinery (ACM) SigGraph (2019), a highly competitive conference.

Data Science in Improving Diagnostic Accuracy

EDRN has built a number of data repositories on biomarkers, imaging and analytical tools. These resources will be employed to help build a specific, interoperable data platforms that will allow investigators to mine and analyze data using Artificial Intelligence and other Machine Learning Languages. Some example are provided below:

Building Cancer Biomarker Data Aggregator (CBDA): Despite unprecedented technological advancements, only a handful of cancer biomarkers get through regulatory approval or clearance every year. Inadequate data infrastructure and statistical or bioinformatics rigor have been identified as key contributing factors this lag in clinical translation of promising candidate biomarkers. Specific challenges in translational biomarker research include missing clinical meta/data, low statistical power, custom data schemas and analytical pipelines, over-fitting of multi-biomarker models, and inflated evaluations of model performance. These practices have led to a buildup of siloed, underutilized data collections, delays in algorithm development, and inefficient regulatory-grade evidence generation needed for biomarker clearance or approval.

EDRN plans to build a platform that will integrate data from NCI data commons (genomics, proteomics, imaging) and serve as a repository for less common data types such as lab findings, chemicals, and cell types. CBDA will function as an interoperable component of an overarching

Cancer Data Ecosystem and leverage the existing NCI Cancer Research Data Commons infrastructure.

Medical Imaging Analytics, Imaging Biomarkers, and Artificial Intelligence. Medical images are routinely used to detect the presence of cancer in symptomatic individuals or at-risk groups and monitor patient responses to therapy. Recent developments in imaging and data technologies provide unprecedented opportunities for new or improved risk stratification, early detection, and precision prevention strategies, particularly in patients with ambiguous symptoms or at high risk for the disease. Clinicians have to evaluate large amounts of imaging data and combine it with clinical information in order to diagnose patients, recognize tumor types, or make any follow-up care and treatment decisions. Despite much progress in imaging research, it is extremely difficult to accurately identify early-stage aggressive neoplasms from indolent or benign lesions for many cancers and make predictions about their aggressiveness.

This process of qualitative interpretation of cancer imaging and related patient information by expert clinicians is highly subjective and laborious, which leads to wide variations in the assessment of the aggressiveness of the lesions as well as to under- or overdiagnosis. The process, however, can be greatly augmented by artificial intelligence (AI), a powerful computational approach that allows computers to solve problems they were not explicitly programmed to address. One of the largest areas of AI, machine learning (ML) provides computers the ability to learn patterns from data (e.g., digitized images) and make predictions (generalizations) based on prior examples. By automating analytical model building through "training", AI extract meaningful patterns from digitized images (i.e., specific features corresponding to a particular type

In digitized tumor images, ML algorithms can automatically identify minute imaging features ("imaging biomarkers") frequently present in neoplastic lesions but missed by or invisible to clinicians. The algorithms can find thousands of novel biomarkers in addition to the traditional ones such as the presence of a nodule, nodule size, location, or rate of growth. ML has been used successfully to streamline and optimize imaging analytics by reducing inefficiencies and decreasing error rates in diagnostic and risk assessment workflows. The power of ML can be also applied to omics data as well as to combined omics and imaging datasets. In personalized cancer prevention, identification of imaging features present in preneoplastic lesions may provide biomarkers for risk assessment, early detection of cancer, and identification of new targets for chemopreventive treatment in at risk individuals for this disease.
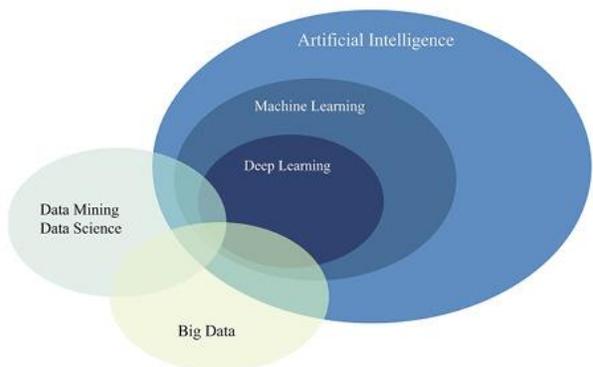
Until recently, ML algorithms required model features to be provided manually through a painstaking process of feature engineering. A new subfield of ML, called **deep learning (DL)**, has emerged recently, which has far surpassed these traditional methods in terms of performance, scalability, and speed of the software development, because DL algorithms (also known as deep neural network algorithms) can learn directly from data. DL often uses many layers of computation or a deep neural network, which enables more powerful analyses of the input data. For example, in image classification tasks, DL can automatically discover recurrent features within digitized images ("imaging biomarkers"), obviating the need for feature engineering. DL is also more scalable and efficient, has better analytics and accuracy, can handle imbalanced datasets, changing conditions, and better suited for knowledge integration than traditional ML algorithms.
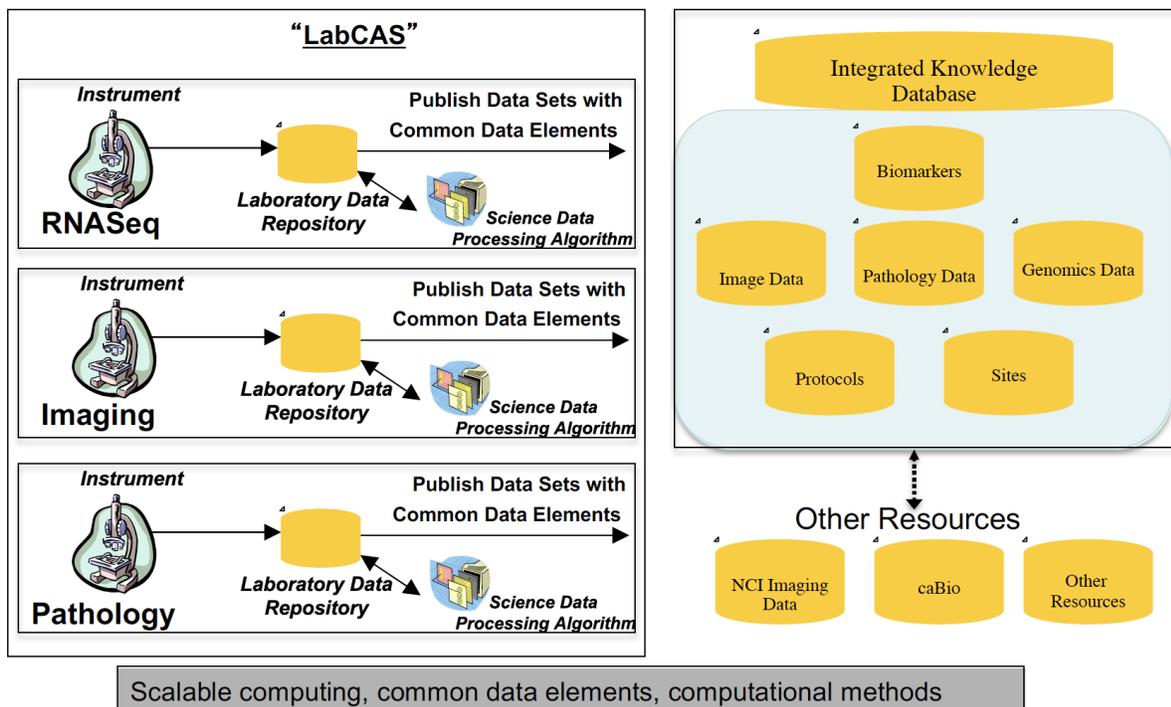

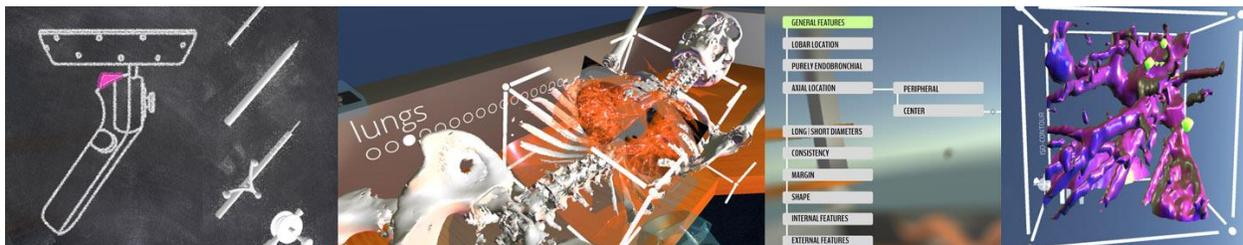**FIGURE 1: SCALABLE ANALYTICAL PIPELINES TO CAPTURE AND SHARE DATA ACROSS LABS**

FIGURE 2:EDRN VIRTUAL REALITY PROTOTYPE FOR EXPLORING MULTI-DIMENSIONAL IMAGING DATA

CONCLUSION

Translation of biomarker discovery into early detection and diagnostic tests requires a broad spectrum of expertise. A network, such as the EDRN, facilitates the interaction of scientists and physicians with these expertise and ensures that the most promising biomarkers coming from discovery laboratories are moved forward into clinical validation. The EDRN has established a unique group of multidisciplinary investigators dedicated to the EDRN Strategic Plan. They include basic scientists, epidemiologic and population researchers, and physician scientists, who are on the front line of patient care. Collaborative projects amongst these EDRN investigators facilitate biomarker research and its translation into clinical application. Equally important, EDRN has long-standing relationships with industry and the private sector, which are changing their business model to support biomarker discovery for early detection and diagnostics.

EDRN also has strategic alliances with non-profit foundations, such as the Canary Foundation and the Lustgarten Foundation, Federal agencies, such as NIST (National Institute of Standards and Technology) and the FDA (Food and Drug Administration); and professional organizations, such as ASCO (American Society of Clinical Oncology) to develop a transparent process for creating well-defined consensus standards and guidelines for biomarker development, validation and qualification to reduce the uncertainty in adoption of biomarkers for clinical use. In the future, the EDRN proposes to work with CMS (Centers for Medicare and Medicaid Services) to review new biomarkers and to reach consensus on coverage and pricing.
This Strategic Plan includes both EDRN's short and long-term goals. While the discovery, validation and implementation of biomarkers for population-level early detection and screening remain part of both the short-term long-term goals, the improvement of current detection and diagnostic modalities are achievable short-term goals. In response to developments in the field and to recommendations by NCI's Board of Scientific Advisors, the EDRN has incorporated imaging into its research projects, mostly to be in combination with biomarkers, to improve both accuracy and to distinguish clinically significant cancers from those that do not require treatment.