NIH NATIONAL CANCER INSTITUTE

Division of Cancer Epidemiology and Genetics

STRATEGIC PLAN

2020-2025

U.S. Department of Health & Human Services | National Institutes of Health

MISSION

To discover the causes of cancer and inform the means for prevention by conducting transdisciplinary epidemiological and genetic research.

STRATEGIC GOALS

DCEG has identified three primary goals and the following activities for each of the goals for achieving its mission:

Goal 1: Conduct world-class epidemiological and genetic research

- Improve process for prioritization of research
- Increase support and capacity for data analytics
- Improve data use and sharing following FAIR principles
- Increase the diversity of study populations
- Enhance research in health disparities
- Increase cross-branch research initiatives
- Efficiently use DCEG, NCI, and NIH resources
- Improve administrative processes

Goal 2: Maintain a highly skilled and diverse workforce and train the next generation of scientists to support public health

- Improve recruitment strategies to attract highly qualified and diverse candidates to DCEG
- Increase professional development and leadership opportunities
- Take advantage of existing training courses and create new training opportunities when needed
- Develop and implement strategies for workforce equity

Goal 3: Disseminate DCEG research on cancer risk and etiology

- Increase DCEG profile in the research community
- Describe and document policies for dissemination of research findings
- Develop strategies to help external investigators understand what we do
- Develop policy and provide guidance for investigators wishing to participate on policy and advisory boards

RESEARCH PORTFOLIO AND VISION

DCEG investigators conduct trans-disciplinary research on risk factors for cancer, involving extensive trans-divisional, trans-NIH, national and international collaborations. Predicated on a strong research portfolio and long-term investments, future studies will feature descriptive epidemiology and integrative analytical epidemiology with in-depth exposure assessments, novel biomarker measurements, germline genetic variation and extensive somatic characterization of target tissues and tumors. DCEG strives to conduct studies that reflect the racial, ethnic and geographical diversity of the US population. The intramural environment also facilitates rapid response to emerging public health or scientific issues. These research priorities are expected to lead to a better understanding of cancer etiology and its biological mechanisms, improve identification of individuals and populations at risk, and inform strategies for prevention and treatment of cancer, including approaches to reduce health disparities.

The DCEG research portfolio and vision can be captured in three broad areas:
(a) etiology of cancer,
(b) risk prediction, prevention, early detection and clinical outcomes, and
(c) methodological research on study design and data analytics.

Scientific excellence and potential impact on public health and clinical practice are the primary criteria used to evaluate and prioritize research in the Division. Distinctive high-risk, high-reward studies that build upon the strengths and expertise of DCEG investigators and infrastructure are of special consideration. Promoting innovation and initiative of individual investigators and trainees is a critical component of DCEG strategy to meet the full range of scientific opportunities, from large-scale research projects to small-scale efforts that could lead to new breakthroughs. Adopting FAIR (Findable, Accessible, Interoperable and Re-usable) data principles in the coming years is also critical to the DCEG strategy in order to facilitate data sharing and maximize return on investment.

a. Etiology of Cancer

DCEG studies the role of environmental and host factors in the development of a wide range of common and rare cancer types. Descriptive epidemiology utilizes large-scale databases to monitor exposures and cancers to identify areas of public health concern, while analytical epidemiology conducts in-depth investigations of established and novel risk factors for cancer. DCEG is committed to establishing and maintaining a wide range of resources, including study populations with rich data and biospecimen resources to address etiological questions. A major trans-divisional and trans-NIH initiative is the Connect for Cancer Prevention (Connect Study), a new long-term, prospective cohort of approximately 200,000 adults of diverse backgrounds, conducted in collaboration with integrated health care systems across the U.S., with longitudinally collected data and biospecimens.

Descriptive epidemiology

DCEG conducts surveillance and population-based studies to (i) evaluate the distribution of risk factors and cancers in different populations, including time trends and geographic, socio-economic, racial and ethnic distribution, (ii) quantify the population burden of cancer, (iii) estimate the fraction of the population burden that can be attributed to specific exposures, and (iv) forecast future trends and burden of cancer. Given the current explosion of Big Data resources at the population level available for linkage studies, and the development of statistical and data science methodology to analyze complex data, DCEG envisions a transformative period in this field, which has always been at the core of its epidemiological research. As an intramural program, DCEG can leverage internal access to government resources for descriptive studies working closely with scientists across NIH and other agencies. Studies that address important public health problems, cancer health disparities or emerging exposures are prioritized.

Analytical epidemiology

Environment

Environmental factors are integral to the etiology of most cancers, and thus studies on these factors using high-quality exposure assessment are central to DCEG research. DCEG investigators use state-of-the art exposure assessment methodology, including surveys, data linkages, wearable devices, chemical monitoring, geocoding and biomarkers in epidemiological studies to (i) identify new behavioral, contextual, infectious, medical, occupational and environmental risk factors for cancer, including the potential cancer risks from climate change, (ii) quantify exposure-response relationships, and (iii) identify co-factors and effect modifiers. This etiological research also provides the foundation for the translation of knowledge into guidelines, policies and

interventions to improve public health for all, for instance, leading to new regulations by state and federal agencies that reduce environmental and occupational carcinogenic exposures. These studies benefit from DCEG long-term investments that generate rich data and biospecimen resources including prospective cohort studies and case-control studies. The latter are particularly important for rare cancers, special populations, or where extensive and targeted exposure assessment is required.

Studies that are high-risk, require fast response due to public concern, or are of relevance for regulatory agencies are prioritized. Future studies will greatly benefit from advances in technologies to measure exposures, including the widespread use of wearable devices generating unprecedented amounts of high-quality data on individual exposures, as well as the increasing availability and standardization of electronic health records (EHR). The development of new studies of emerging risk factors and the application of novel methodologies for exposure assessment are paramount for advancing DCEG research.

Germline genetic variation

DCEG investigators evaluate how rare and common genetic variants influence susceptibility to cancer, and interact with the environment and other factors. Uncovering the genetic architecture of cancer susceptibility (i) provides important insights into cancer etiology and biology, (ii) enables cancer risk prediction based on polygenic risk scores and high-risk variants for precision prevention, and (iii) can lead to prediction of clinical outcomes and identification of treatment targets. This research requires both focused individual studies and large-scale collaborations through international consortia in order to conduct genetic studies in cancer families and the general population. DCEG also conducts functional laboratory studies to map and explore signals from association and rare variant studies to understand the underlying biology of susceptibility variants. Hallmark studies in DCEG include long-term, comprehensive studies of families and individuals at high risk of cancer, for instance families with Li-Fraumeni syndrome.

These studies benefit from the long-term commitment of DCEG to intensive collection, storage and analysis of biospecimen and clinical data, efficiently using resources at the NIH Clinical Center. DCEG investigators also conduct and play leadership roles in large-scale international consortia for genome-wide association studies that greatly benefit from previous investments in large epidemiological studies with DNA biospecimen collections. Multi-ancestry studies including adequate numbers of currently under-represented populations are a high priority. The DCEG Cancer Genomics Research (CGR) Laboratory provides critical support to all genetic studies across the Division.

Future genomic studies in DCEG will greatly benefit from the accelerated generation of shared genomic data facilitated by decreasing costs of technologies, which will enable

effective increases in sample sizes from studies of different ancestries, and linkage to EHR. Progress in data science to deliver Cloud-based platforms should enable secure federated data ecosystems and powerful analytics to accelerate the pace of research. Remaining at the cutting-edge of these developments will be critical for the success of future studies in the Division.

Biomarkers of exposure, early biological effects and host susceptibility

DCEG investigators use molecular measurements to (i) discover new biomarkers of cancer risk and evaluate mechanisms of action of known carcinogens in cancer biology, (ii) study the biological plausibility of suspected carcinogens, and (iii) conduct methodological studies for sample collection, processing and storage, as well as assay development and validation. In these studies, investigators use a wide range of biological specimens, such as blood, urine, buccal, hair, nails, cervical and stool samples, collected in epidemiological studies to provide measures of internal exposure, early biological effects and host susceptibility factors. Methodological research is critical for the conduct of high-quality biomarker studies and greatly benefits the wider research community. This area of research capitalizes on long-term investments in extensive collections of biological specimens drawn from high-quality epidemiological studies in the Division and elsewhere, sampled from the general population or special exposure populations. Studies use biological assays that can measure a small number of analytes of interest, or large numbers of analytes using targeted and untargeted "omic" technologies such as epigenomics, adductomics, transcriptomics, proteomics, metabolomics, and microbiomics.

DCEG prioritizes distinct studies in understudied or special populations (e.g. highly exposed populations due to occupation, underrepresented minorities, geographical location, or accidents), and well-powered studies that leverage existing resources. New technologies to measure thousands of analytes on minimal amounts of specimens coupled with advances in methods and Cloud infrastructure for integrative analytics of Big Data provide exciting opportunities. However, progress requires decreasing assay costs and continuing investments in collection of biospecimens from epidemiological studies, particularly serial biospecimens collected prior to disease diagnosis to study changing biomarkers in relation to risk, and the use of modern collection and storage protocols. The new DCEG prospective cohort, Connect Study, has been designed to address these challenges and to become the workhorse for future investigations integrating biomarkers and exposures in the Division and elsewhere.

Natural history of cancer and tumor heterogeneity

Epidemiological and clinical studies in DCEG use organ and tissue imaging, as well as morphological and multi-omic analyses of normal, precursor and tumor tissues to (i)

understand etiology and cancer pathogenesis, (ii) identify high-risk populations who can benefit from tailored screening and prevention strategies, and (iii) identify differences in risk profiles for tumors growing in the same organ site that are morphologically, molecularly, and clinically heterogenous. These studies elucidate underlying cancer biology to inform strategies for cancer prevention and early detection, including development of drug targets for prevention and treatment. These studies leverage the long-term commitment of DCEG resources to collect well-annotated tissues, many from epidemiological studies and genetic studies of high-risk families with high-quality exposure assessment and clinical follow-up. This represents valuable and sometimes unique resources that, together with the breadth and depth of transdisciplinary expertise and collaborative environment in DCEG, provide remarkable research opportunities. These are often high-risk studies that benefit from stable resources in DCEG as well as close collaborations with extramural partners. DCEG also develops new analytical methods and tools (e.g. digital image analysis, bioinformatic pipelines) that can benefit the whole research community.

The scientific opportunities available to DCEG are at a critical moment with the convergence of technological advances and decreasing costs for image and genomic analyses of tissues and increasing availability of large databases of biological features, Cloud infrastructure, and Big Data analytics. Because of its resources and stable funding, DCEG is in a distinct position to carry out high-risk, high-reward studies that are often resource-intensive. Our focus is on populations of special interest (e.g. highly exposed to carcinogens, rare cancers, families with unusual patterns of cancer, or cancers with mostly unknown etiology) and in areas of strength in the Division because of its expertise. Consequently, DCEG will need to increase investment in imaging and tissue collections, investments in technology and Cloud infrastructure and computing, development of stable and reproducible bioinformatics pipelines, professional development, and form strong strategic partnerships with other expert leaders in the field.

b. Risk prediction, prevention, early detection and survivorship

DCEG investigators integrate information on multiple cancer risk factors derived from etiological studies to predict the risk of developing cancer over a given period of time. This research includes (i) development of methods and approaches for accurate and comprehensive risk assessment, including artificial intelligence (AI) predicted analytics, and (ii) evaluation of the levels of risk stratification that can be attained in the population for precision prevention strategies. The emphasis is on cancers with actionable public heath interventions, such as screening for breast, cervix, melanoma, and lung cancers, which could be tailored according to an individual risk profile. As a general principle, the Division focuses on research that informs implementation strategies that ultimately are

developed by other institutions into public health interventions or treatments. Translational research on prevention, early detection and prognosis of cancer is conducted in selected areas where (i) DCEG has high-level expertise, (ii) it is closely related to etiological research, or (iii) there is an urgent public health need. This research benefits from the strong etiologic research program in DCEG, long-term investments and commitments on longitudinal studies, breadth and depth of expertise, and collaborations across NIH and with other agencies and clinical groups. Methods and tools developed in DCEG, such as cancer risk assessment tools, can be broadly used by researchers, in clinical practice and can inform public health guidelines. As the population of cancer survivors increases in the US due to improved treatment and early detection, we increasingly leverage our etiological studies to evaluate survivorship. This includes evaluation of treatment-related subsequent neoplasms, which is an area of special expertise in the division as well the role of genetics and environmental factors in survival.

New approaches for risk prediction, precision prevention and survivorship can be enhanced by easier access to EHR data from health care systems to conduct research that informs individual health care and future research. This model where research moves closer to care delivery operations requires access to large dynamic datasets and biospecimen collections from clinical trials and observational studies, in order to take advantage of developments in data science. The Connect Study, designed within integrated health care systems in the U.S., will be an invaluable resource for advancing a learning health care system for precision prevention.

c. Methodological research on study design and data analysis

The use of rigorous and state-of-the art methodologies for study design and data analysis of epidemiological studies are central to the success of DCEG research. Statisticians, biostatisticians, and data scientists in DCEG are fully integrated within research projects in which they also develop and apply novel methodology to address challenges faced in epidemiological studies. This form of team science ensures a rigorous use of existing and new methods by providing technical expertise, and at the same time, increases training opportunities. Methodological research in DCEG includes developing efficient study designs and sampling strategies, and integrative analyses of high-dimensional, time-dependent data such as data derived from EHR, image analyses, geographical/spatial analyses, biosensors, or "omic" technologies to measure biomarkers. Causal inference and designing validation studies and methods to evaluate and correct for bias such as measurement error in exposures and clinical outcomes are also important areas of research in the Division.

The NIH-wide movement towards Data Commons platforms, increasingly in interoperation with other federal and state health related systems, represents a new landscape for epidemiology methods development. In particular, the development of Cloud-based technologies and AI constitutes a critical shift in research that is transforming many aspects of how we conduct and communicate science. AI offers unprecedented opportunities to analyze complex, fast growing data generated in epidemiological studies including genomic analyses, digital pathology, radiological imaging, EHR, and other sources. The ubiquitous nature of modern computational infrastructure is creating new opportunities to engage the systems where these data are being generated and analyzed directly. To realize this, it is paramount that these advances are well integrated into epidemiological research using time-tested study design principles that take into account potential biases inherent in observational studies. Consideration of the appropriateness of different analytical tools in statistics, bioinformatics, and AI to address specific questions, rather than using the method of the moment, will continue to be a core principle. DCEG is committed to remain at the trans-disciplinary methodological forefront with the adoption of FAIR data principles through interoperable platforms offering trusted data governance, easy access to data/meta-data and powerful analytics, and through the distribution of open-source tools.

STRATEGIC PLANNING PARTICIPANTS

Chair

Montserrat García-Closas
Division Deputy Director
Trans-divisional Research Program Director

Co-Chair

Amy Berrington
Senior Advisor for Strategic Activities
Radiation Epidemiology Branch Chief

Participants

Christian Abnet
Metabolic Epidemiology Branch Chief

Paul Albert
Biostatistics Branch Chief

Laura Beane Freeman
Occupational and Environmental Epidemiology Branch

Amanda Black
Associate Director for Biological Resources

Stephen Chanock
Division Director

Jonas De Almeida
Chief Data Scientist

Eric Engels
Infections and Immunoepidemiology Branch Chief

Gretchen Gierach
Integrative Tumor Epidemiology Branch Chief

Belynda Hicks
Cancer Genomics Research Laboratory Deputy Director

Robert Hoover
Former Epidemiology and Biostatistics Program Director

Maria Teresa Landi
Integrative Tumor Epidemiology Branch, Lead of Somatic Genomic Studies

Jackie Lavigne
Office of Education Chief

Jennifer Loukissas
Communications Teams Chief

Ludmila Prokunina
Laboratory of Translational Genomics Chief

Mark Purdue
Occupational and Environmental Epidemiology Branch

Nathaniel Rothman
Occupational and Environmental Epidemiology Branch, Head of Molecular Studies

Sharon Savage
Division Clinical Director
Clinical Genetics Branch Chief

Debra Silverman
Occupational and Environmental Epidemiology Branch Chief

Nicolas Wentzensen
Clinical Genetics Branch Deputy Chief

Hannah Yang
Associate Director for Scientific Operations

Meredith Yeager
Cancer Genomics Research Laboratory Director

NATIONAL CANCER INSTITUTE

April 2021