

# Epidemiologic Research of Rare Cancers: Trends, Resources, and Challenges

Lisa Gallicchio, Danielle L. Daee, Melissa Rotunno, Rolando Barajas, Sarah Fagan, Danielle M. Carrick, Rao L. Divi, Kelly K. Filipinski, Andrew N. Freedman, Elizabeth M. Gillanders, Tram K. Lam, Damali N. Martin, Scott Rogers, Mukesh Verma, and Stefanie A. Nelson



## ABSTRACT

**Background:** The goals of this project were to assess the status of NCI's rare cancer-focused population science research managed by the Division of Cancer Control and Population Sciences (DCCPS), to develop a framework for evaluation of rare cancer research activities, and to review available resources to study rare cancers.

**Methods:** Cancer types with an overall age-adjusted incidence rate of less than 20 cases per 100,000 individuals were identified using NCI Surveillance, Epidemiology and End Results (SEER) Program data. SEER data were utilized to develop a framework based on statistical commonalities. A portfolio analysis of DCCPS-supported active grants and a review of three genomic databases were conducted.

**Results:** For the 45 rare cancer types included in the analysis, 123 active DCCPS-supported rare cancer-focused grants were identified, of which the highest percentage (18.7%) focused on ovarian

cancer. The developed framework revealed five clusters of rare cancer types. The cluster with the highest number of grants ( $n = 43$ ) and grants per cancer type (10.8) was the cluster that included cancer types of higher incidence, average to better survival, and high prevalence (in comparison with other rare cancers). Resource review revealed rare cancers are represented in available genomic resources, but to a lesser extent compared with more common cancers.

**Conclusions:** This article provides an overview of the rare cancer-focused population sciences research landscape as well as information on gaps and opportunities.

**Impact:** The findings of this article can be used to develop efficient and comprehensive strategies to accelerate rare cancer research.

*See related commentary by James V. Lacey Jr, p. 1300*

## Introduction

The NCI defines a rare cancer as one having an age-adjusted incidence of fewer than 15 cases per 100,000 per year (1). Despite their rarity, these cancers collectively account for approximately 25% of all new cases of adult cancers each year in the United States (1). For individuals diagnosed with a rare cancer, the 5-year relative survival rate, overall, is worse compared with those diagnosed with a more common cancer (2). In fact, many rare cancer types, such as pancreatic and liver cancer, are highly lethal, as no effective early detection methods have been identified and treatment options are limited. Conversely, the development of more effective therapies has led to decreasing mortality rates for other rare cancer types, such as testicular cancer (3), leading to an increase in the numbers of short- and longer-term survivors (i.e., any individual diagnosed with cancer). However, little is known about the survivorship issues experienced by these individuals.

Observational studies, such as prospective cohort and case-control studies, are critical to understanding factors associated with both the

risks and outcomes of rare cancers. In addition, beyond traditional treatment-focused randomized clinical trials, intervention trials can also be conducted to provide evidence on strategies to reduce the risk of rare cancers or to mitigate symptoms and side effects that occur as a result of a rare cancer diagnosis and its treatment. However, conducting observational studies and intervention trials focused on rare cancers is challenging. For example, recruiting a sufficient sample size of rare cancer patients or survivors to answer certain research questions with adequate statistical power is often difficult and cost prohibitive. These challenges hinder the generation of much needed evidence pertaining to rare cancers.

The Division of Cancer Control and Population Sciences (DCCPS) has the "lead responsibility at the NCI for supporting research in surveillance, epidemiology, health services, behavioral science, and cancer survivorship" (<https://cancercontrol.cancer.gov/od/history.html>). As such, the DCCPS manages NCI-funded observational studies and intervention trials that are not focused on cancer treatment across the cancer control continuum. With the goal of understanding the current status of and needs for population-based research focused on rare cancers, the following activities were carried out and are presented in this manuscript: (i) review of data from the NCI's Surveillance, Epidemiology and End Results (SEER) Program and development of a framework using these data to guide future rare cancer-related cancer control research efforts; (ii) a portfolio analysis of active DCCPS grants to determine research activity focused on rare cancer types and how this funded research fits within the framework developed with the SEER data; and (iii) an assessment of available population sciences resources, which include three genomic databases, for rare cancer research. Research in preclinical models, basic mechanistic studies, and cancer treatment-focused clinical trials are not included in the activities covered in this manuscript.

Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland.

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

**Corresponding Author:** Lisa Gallicchio, NCI/NIH, 9609 Medical Center Drive, Room 4E604, MSC 9763, Rockville, MD 20892. Phone: 240-276-5741; Fax: 240-276-7920; E-mail: [lisa.gallicchio@nih.gov](mailto:lisa.gallicchio@nih.gov)

Cancer Epidemiol Biomarkers Prev 2021;30:1305-11

doi: 10.1158/1055-9965.EPI-20-1796

©2021 American Association for Cancer Research.

Materials and Methods

For the work described in this article, we expanded the NCI definition of rare cancers (mentioned above) to include those with an overall age-adjusted incidence rate of less than 20 per 100,000 individuals. This decision to expand the definition of ‘rare’ was made *a priori*, and was done so to capture cancers with an age-adjusted incidence rate of less than 15 per 100,000 in either males, females, or individuals identified as black or white, but not when the sex and race-defined groups were combined.

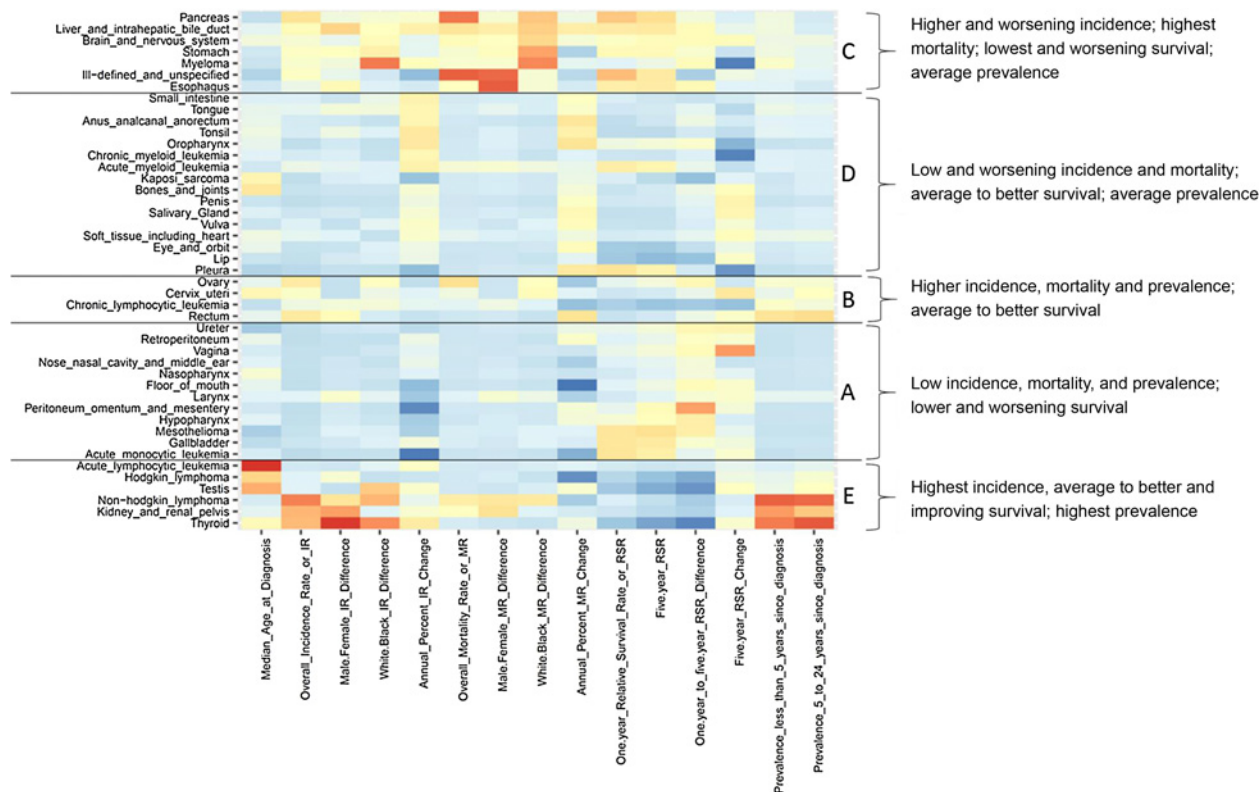
SEER statistics review and framework development

NCI’s SEER Program collects cancer incidence and survival data from population-based cancer registries covering approximately 34.6% of the U.S. population (<https://seer.cancer.gov/about/overview.html>). To better understand current incidence, mortality, and relative survival rates, as well as prevalence counts, for rare cancers in the United States, SEER data (primarily SEER21 and SEER18) were extracted from the SEER website (3; Supplementary Table S1) and reviewed. Specifically, we examined the highest and lowest age-adjusted incidence, mortality, and relative survival rates as well as prevalence counts, and the largest male–female and black–white disparities for incidence and mortality rates, for the cancers included.

In addition to the SEER variables described above, data on the following SEER variables were also extracted, and synthesized with the incidence, mortality, relative survival, and prevalence data, to identify

statistical commonalities between the rare cancer types: median age at diagnosis, annual percent change in incidence rate, annual percent change in mortality rate, difference between the most recent 1- to 5-year relative survival rates, and 5-year change in 5-year relative survival rate (see Fig. 1; Supplementary Table S1). While not truly independent of the other included statistics, prevalence count data were kept in as part of the analytic framework because these data are more of a reflection of past, and not current, incidence, mortality, and relative survival rates. Data for all of the *a priori* identified SEER variables were analyzed using the “heatmap” function in the R statistical package. This approach allowed for the ordering of the cancer types based on a computed correlation between SEER statistical values. Data were coded (or recoded) so that the data points indicating “worse” (or “better”) individual or public health cancer burden had a higher (or lower) absolute value and were colored in red (or blue) on the heatmap. For example, data for 5-year relative survival rates were entered as “1-the 5-year relative survival rate” so that lower 5-year relative survival rates would have higher values, and, therefore, be colored red on the heat map. Boundaries of the clusters were then identified using visualization of the heatmap (i.e., not by using the “clusters” function in R).

To describe each cluster, the statistics within each cluster were labeled with descriptors such as “low,” “average,” and “high” (e.g., “high” incidence, “low” mortality); these descriptors were utilized to generally compare the clusters to each other; there were no specific data-derived cut-points for the descriptors. Furthermore, the



**Figure 1.** Clusters of rare cancer types based on statistical commonalities. Heat map was generated utilizing the R statistical package. Data were coded so that the data points indicating “worse” individual or public health cancer burden had a higher absolute value and are colored in red on the heatmap. Data points indicating “better” individual or public health cancer burden are colored in blue. Boundaries of the clusters were identified using visualization of the heatmap.

Downloaded from <http://aacrjournals.org/cebp/article-pdf/30/7/1305/3100913/1305.pdf> by guest on 25 August 2022

descriptors of “low,” “average,” and “high” pertain to the cancer types included in this analysis and, thus, the descriptors are relative to the values of the other included cancer types. For example, overall, the cancers types are all lower incidence cancers (age-adjusted incidence rate of less than 20 per 100,000 individuals), thus, a cluster labeled as a “high” incidence cluster generally includes cancer types on the high end of the age-adjusted incidence rate range of 0.1 to 20 per 100,000 individuals. Similarly, a cluster labeled as “high” prevalence generally includes cancer types on the high end of the prevalence case count range for the cancer types included in this analysis. As a comparison, the highest prevalence case count for the cancer types included in this analysis was 644,761 for non-Hodgkin lymphoma (NHL; January 1, 2016, 24-year limited duration, first per type in previous 24 years) while the highest prevalence case count overall in SEER in 2016 was 3,112,731 for breast cancer. It should be noted that due to the nature of the statistical clustering technique utilized, the assigned cluster descriptors did not necessarily apply to all of the individual cancer types within that cluster.

### Portfolio analysis of DCCPS-supported grants

Research project grants (which include cooperative agreements, program project grants, and research career awards) related to rare cancers that were supported by the DCCPS and active on December 10, 2018 were included in the portfolio analysis, with the terms searched being the specific SEER cancer types identified as being rare (see Supplementary Table S2). Non-research grant mechanisms (R13) were excluded. The search of the Information for Management, Planning, Analysis, and Coordination (IMPACII) records (NIH's proprietary system) using NCI's proprietary Portfolio Management Application (PMA) software version 13.4, identified 126 active grants at the time of the search. These grants were manually reviewed in duplicate by the authors of this manuscript for inclusion into the analytic dataset. A grant was included in the portfolio analysis if a rare cancer was mentioned as a focus of the study in the abstract or the specific aims page. Review and all coding (including that which is described below) were blinded; discrepancies in coding were resolved by an independent third reviewer. Of the 126 active grants identified, three were excluded because a rare cancer was not a focus of the study, resulting in 123 grants in the analytic data set.

To better understand in more detail what types of rare cancer research were currently being funded, data on study design and broad area of scientific interest were reviewed. Specific aims and abstracts were reviewed and coded for study design. A study was coded as an intervention trial if the investigator proposed an intervention in some or all members of a group of participants; otherwise, the study was coded as observational. In situations where the reviewed grant included both intervention and observational components, the grant was coded as an intervention trial.

Common Scientific Outline (CSO) codes were used to identify the broad scientific area(s) of interest for each grant (<https://www.icrpartnership.org/cso>). The CSO is a coding system used by public and private organizations in the United States and other countries to describe research projects, making it possible to compare research portfolios across public, nonprofit, and government agencies. Because survivorship is a specific area of interest for DCCPS, the overarching CSO category of “cancer control, survivorship, and outcomes research” was separated into two distinct categories: (i) survivorship and (ii) cancer control and outcomes. A grant could be assigned multiple CSO codes, in fractions of applicability, with the sum of all of the CSO fractions for a grant equaling 1.

### Resource assessment

The existence of databases of population-level genomic and epidemiologic data greatly facilitates research on rare cancer types. To assess existing resources that may be leveraged for rare cancer research, three widely used NIH genomic databases —The Cancer Genome Atlas (TCGA), the database of Genotypes and Phenotypes (dbGaP), and the Genome-Wide Association Studies (GWAS) catalog— were reviewed. The number of rare cancer cases in TCGA (<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>, accessed July 2018) was determined by using an advanced search query on the international classification of diseases codes (ICD-10) that matched the codes for selected rare cancer types. Studies within dbGaP (<https://www.ncbi.nlm.nih.gov/gap/>, accessed May 2020) were assigned to rare cancer groups by searching the “study disease/focus” and “study description” fields through the advanced search function. Number of cases, total studies, and studies with available germline data were extracted using information from the main study page information and the relevant tabs. The GWAS catalog (<https://www.ebi.ac.uk/gwas/ancestry>, accessed May 2020) was searched for all studies that pertained to cancer, which were then grouped by rare cancer types via keywords search of the fields “trait,” “reported traits,” and “ontology traits synonyms.” Number of studies, genetic associations, and ancestry were found by searching the “reported traits” field.

## Results

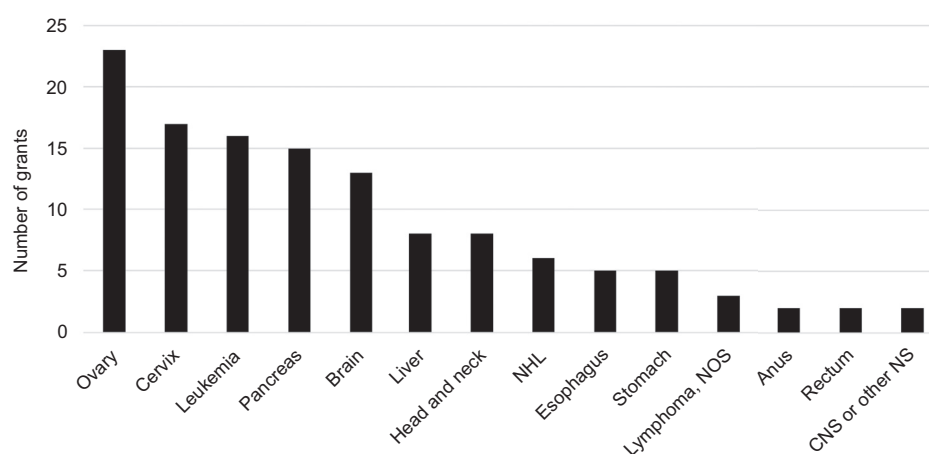
### SEER statistics review

Age-adjusted incidence rates for the 45 cancers that met our criterion ranged from 19.6 per 100,000 for NHL to less than 0.1 per 100,000 for pleural cancer. Thyroid cancer showed the largest male–female disparity, while myeloma showed the largest black–white disparity. More than half (57.4%) of the rare cancers showed a trend indicating decreasing (i.e., improving) incidence rates from 2007 to 2016, although not all decreases were statistically significant. The rare cancer types with the highest increases (i.e., worsening) in incidence rate were oropharynx and tonsil.

The rare cancer types with the highest age-adjusted mortality rates were those categorized as ill-defined and unspecified (11.8 per 100,000) and pancreatic cancer (11.0 per 100,000). Esophageal cancer showed the largest male-female disparity in mortality rate, and, similar to the disparity in incidence rate, myeloma showed the largest black-white mortality rate disparity. Less than half (45.6%) of the rare cancers identified showed a trend indicating decreasing mortality rates from 2007 to 2016. The rare cancer types with the highest mortality rate increases were oropharynx and anus.

Both the 1-year and 5-year relative survival rates were highest for thyroid; the lowest 1-year and 5-year relative survival rates were for cancers in the ill-defined and unspecified group. Five-year relative survival rates have been increasing (i.e., improving) for most (76.5%) of the rare cancers identified, with the greatest improvements observed for certain blood cancers (myeloma, chronic myeloid leukemia), and pleural cancer. Vaginal cancer had the greatest decrease in 5-year relative survival.

Finally, prevalence counts for the total number of survivors and the number of longer-term survivors (living 5 or more years after diagnosis) were examined. For both of these statistics, the largest numbers of cancer survivors among the cancer types included in this analysis are those with a history of NHL, thyroid cancer, or kidney and renal pelvis cancers. These higher prevalence counts relative to the other rare cancers reflect the fact that the incidence rates and relative survival rates for these cancers are among the highest for the rare cancers.

**Figure 2.**

Number of active rare cancer-focused grants managed by the NCI's Division of Cancer Control and Population Sciences, by cancer type. A grant may include several rare cancer types. CNS or other NS, central nervous system or other nervous system; NOS, not otherwise specified; NHL, non-Hodgkin lymphoma.

### Framework

Five clusters containing rare cancer types with similar SEER statistical characteristics were identified (Fig. 1). The resulting clusters are those including cancer types with the following broad characteristics:

- Low incidence, low mortality; lower and worsening survival; low prevalence
- Higher incidence, higher mortality; average to better survival; higher prevalence
- Higher and worsening incidence, highest mortality, and lowest and worsening survival; average prevalence
- Low and worsening incidence and mortality; average to better survival; average prevalence
- Highest incidence; average to better and improving survival; highest prevalence

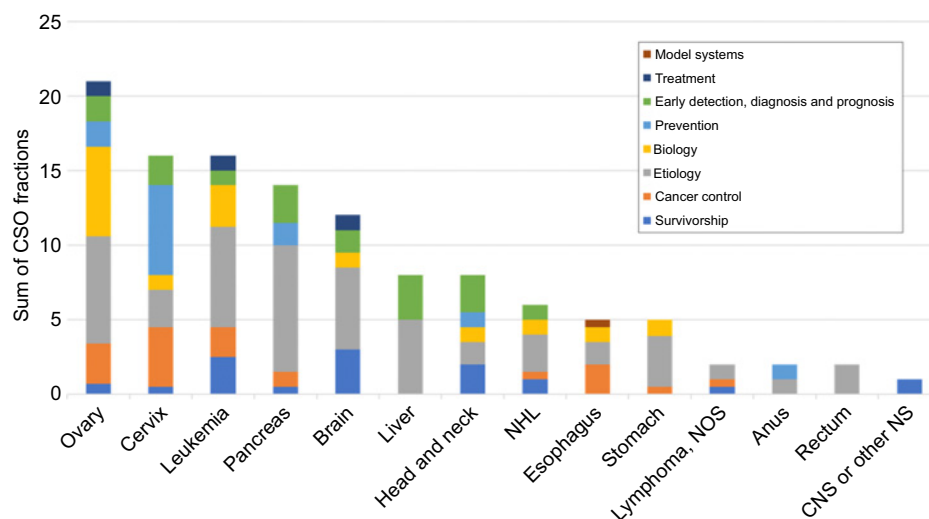
The clusters identified comprise a framework for evaluating the DCCPS-funded grant portfolio and reviewed resources.

### Portfolio analysis

Of the 123 active DCCPS-funded grants identified in the portfolio analysis, 23 focused on ovarian cancer, 17 on cervical cancer, 16 on

leukemia, 15 on pancreatic cancer, and 13 on brain cancer (Fig. 2). In general, more active rare cancer grants were observational studies ( $n = 105$ ) versus intervention trials ( $n = 18$ ). Intervention trials included a web-based physical activity program for children with acute lymphoblastic leukemia, a telehealth intervention targeting distress among rural cancer survivors, and a smoking cessation intervention for cervical cancer survivors. In terms of the scientific area of interest, although the fractions of grants assigned to each of the CSO scientific areas of interest differed by cancer type, etiology, and biology were, in general, the most highly represented areas in DCCPS active grants (Fig. 3). Higher fractions of the leukemia, brain, and head and neck grants were coded as pertaining to survivorship, while a higher fraction of the cervical cancer grants was coded as relevant to prevention. Among the biology or etiology grants, noted broad areas of science were identification of genetic risk variants; analyses of lifestyle risk factors; and the role of human papillomavirus (HPV) in carcinogenesis. Among the survivorship grants, noted broad areas of science were physical, psychological, and social adverse sequelae; caregiving; behavioral, education, or complementary alternative medicine interventions; and economic costs of cancer and cancer treatment.

The SEER-based framework was used to determine the number of active grants with shared statistical characteristics, identified by the five

**Figure 3.**

Common Scientific Outline broad scientific area code fractions for active rare cancer-focused grants, by cancer type. CNS or other NS, central nervous system or other nervous system; NOS, not otherwise specified; NHL, non-Hodgkin lymphoma.

**Table 1.** Description of active grants in rare cancer clusters identified on the basis of SEER statistical commonalities.

	<b>Cluster A: Low incidence, mortality, and prevalence; lower and worsening survival</b>	<b>Cluster B: Higher incidence, mortality and prevalence; average to better survival</b>	<b>Cluster C: Higher and worsening incidence, highest mortality, lowest and worsening survival; average prevalence</b>	<b>Cluster D: Low and worsening incidence and mortality; average to better survival; average prevalence</b>	<b>Cluster E: Highest incidence, average to better and improving survival; highest prevalence</b>
Number of cancers in cluster	12	4	7	16	6
Cancer types in cluster	Ureter, retroperitoneum, vagina, nose/nasal cavity/middle ear, nasopharynx, floor of mouth, larynx, peritoneum, hypopharynx, mesothelioma, gallbladder, acute monocytic leukemia	Ovary, cervix, chronic lymphocytic leukemia, rectum	Pancreas, liver/bile duct, brain/nervous system, stomach, myeloma, ill-defined and unspecified, esophagus	Small intestine, tongue, anus/anal canal/anorectum, tonsil, oropharynx, chronic myeloid leukemia, acute myeloid leukemia, Kaposi sarcoma, bones/joints, penis, salivary gland, vulva, soft tissue, eye, lip, pleura	Acute lymphocytic leukemia, Hodgkin lymphoma, testis, non-Hodgkin lymphoma, kidney/renal pelvis, thyroid
Number of active grants (based on cancer type) <sup>a</sup>	2	43	33	17	18
Number of active grants <sup>a</sup> per cancer type	0.17	10.8	4.7	1.1	3.0
Number of interventional study active grants	1	10	4	1	3
Number of observational study active grants	1	33	29	16	15

<sup>a</sup>Grant could be counted in more than one cluster if it included more than one cancer type as a focus.

clusters. The number of grants in each cluster ranged from 2 to 43 (Table 1). Cluster A ( $n = 2$  grants), comprised of cancer types characterized by low incidence and mortality rates as well as lower and worsening survival rates, had the lowest number of grants per cancer type (0.17 grants/cancer type). Conversely, Cluster B, comprised of cancers with higher incidence and relative survival rates, had the highest number of grants per cancer type (10.8 grants/cancer type). The two clusters containing cancer types with the highest prevalence (which reflects higher incidence and higher relative survival rates), Clusters B and E, had the highest and third highest grants per cancer type. Cluster C and Cluster E both included cancer types with greater black-white disparities.

### Resource assessment

#### TCGA

TCGA is a valuable resource for tumor molecular data (e.g., genotype and methylation arrays, DNA and RNA sequencing). We found that most of the rare cancer types in our framework were represented in TCGA, but case numbers were generally smaller than for more common types, as expected (Supplementary Table S3). The exceptions to this are ovarian cancer, for which there were over 500 cases, and brain and nervous system cancers, with over 1,000 cases. In comparison, TCGA included 1,024 lung cancer cases.

#### dbGaP

dbGaP is a primary resource for genomic studies of many phenotypes and diseases. Because we could only access publicly available

study pages and not the data directly, it was difficult to determine the exact counts of all rare cancer studies and cases. We found studies for most rare cancer types, however, the number of studies for each type was small (range: 1 to ~10) compared with more common types. For example, the rare cancer type found to have the most studies registered in dbGaP was brain and nervous system (~10 studies); lung cancer was found to have approximately 25 registered studies.

#### GWAS catalog

The GWAS Catalog provides information and summary statistics for genome-wide association studies. Most rare cancer types were represented, but again, there were fewer studies than for common cancer types. Ovarian cancer had the most studies ( $n = 20$ ) and liver and pancreatic were second with 17 and 16 studies, respectively. Most of the studies were conducted in European ancestry populations, although for 16 of the 28 rare cancer types for which data were located, studies in Asian populations were reported. As a comparison, there were 74 studies reported for lung cancer, and these included populations with European, African American, Chinese, East Asian, Japanese, Han-Chinese, and/or Korean ancestry.

### Discussion

Cancer epidemiology research on rare cancer types is challenging. The low incidence rates of these cancers, often combined with high mortality rates, demand innovative strategies and approaches to work with the small sample sizes available for research. To understand

research opportunities and challenges related to population sciences rare cancer research, we developed a framework to better visualize and understand shared characteristics and potential research needs for rare cancers. In addition, we conducted a portfolio analysis of DCCPS-funded studies to understand how these studies fit into the framework.

An analysis of SEER data for rare cancer types was used to create a framework and identify clusters of cancer types with shared statistical characteristics. This framework provided a novel way of viewing rare cancers, by creating groups based on characteristics other than anatomic site, which may point to both unique and shared research needs. For example, we noted clusters of cancer types with worsening incidence (C and D); these cancers might be the focus of research to discover or understand emerging or changing exposures that could be leading to the increase in incidence. Similarly, we identified clusters of cancer types for which the number of prevalent cases was high (B and E); the cancer types within these clusters represent populations of patients in which to study the needs and challenges faced by cancer survivors. Cancer types with improving survival rates might also be assessed to determine what is the underlying cause for this improvement, for example, whether screening/early detection or treatment options have improved. For many research questions, it may be possible to group patients with different types of cancer based on other shared characteristics, thereby overcoming the problem of small population sizes that often hinders research on many rare cancer types. This could mirror current strategies to assess cancer treatment options based on the genomic characteristics of the tumor, rather than on anatomic site.

The currently funded DCCPS rare cancer grants focus on a broad range of cancer types and cover a variety of research domains, as defined by CSO codes. To assess our portfolio in the context of the SEER-based framework, we assigned grants to clusters based on the cancer type studied in the grant. The clusters with the largest numbers of grants (B and C) were those with cancer types that have higher incidence rates (relative to other rare types), and in the case of Cluster C, worsening, or increasing, incidence rates. Although higher incidence rates likely aid in recruiting larger sample sizes, other factors that may contribute to the higher number of studies of particular cancer types include increased interest in cancers with a more substantial public health impact due to higher lethality, easier recruitment due to higher prevalence, or successful coalition building to achieve necessary sample sizes, such as is the case for ovarian and pancreatic cancer. By pooling their data, consortia such as the Ovarian Cancer Association Consortium have enabled well-powered studies of environmental and genetic risk factors that might not otherwise be possible.

We found the fewest numbers of grants in Cluster A. The cancer types in this cluster had lower incidence rates compared to other clusters (e.g., B and C), which limits the number of cases available to conduct population-based studies. Cancer types in Cluster A also have low and worsening survival, which would make survivorship studies more difficult. Cluster D includes cancers with low and worsening incidence rates and mortality; similar to the cancers in Cluster A, the small numbers may be a barrier to conducting risk studies. Cluster E includes only six cancer types (testis, CLL, Hodgkin lymphoma, NHL, kidney, and thyroid) and is characterized by increasing (i.e., better) survival rates, despite worsening incidence. Increasing survival rates is suggestive of improvements in the detection and treatment for these cancers, although our work here did not specifically include assessment of screening and treatment options, nor did we assess the potential for overdiagnosis. Research on very rare, and often more lethal, cancer types could benefit from novel approaches to recruitment to more quickly identify patients. Efforts to connect researchers

working on these cancer types (consortium building) and to connect patients and patient advocacy groups to opportunities to participate in research would be helpful for studies of rare cancers.

SEER data on male/female and black/white disparities were included in our framework exercise. Two clusters (C and E) contained cancers for which the magnitude of the racial/ethnic disparities in incidence and mortality is larger than for the cancer types in the other clusters. DCCPS evaluates the specific aims, design, and analysis sections of each of its supported grants to determine whether the grant addresses health disparities or includes medically underserved populations (<https://maps.cancer.gov/overview/DCCPSGrants/grantlist.jsp?method=dynamic&division=dccps&menu=division&raId=2&codeCategoryId=2>). In a review of this publicly available database, we found that 10 of the grants included in our portfolio analysis focused on underserved populations and 20 addressed health disparities. Three grants each from Clusters C and E included underserved populations, and eight grants from Cluster C and five from Cluster E focused on health disparities. Addressing health disparities and improving inclusion is an area of interest across NCI and NIH, and our framework analysis may point to cancer types particularly in need of research to better understand the underlying factors contributing to disparities.

Resource availability is key to conducting rare cancer research. The TCGA initiative has generated a range of genomic data types for several rare cancers; indeed, ovarian cancer was one of the first cancers to be molecularly characterized by TCGA. dbGaP, which contains data for a wide variety of phenotypes, and both germline and somatic data for cancer, also includes at least one study/dataset for most of the cancer types included in our framework. However, sample sizes for the rare cancer types were much smaller than for common types. The GWAS Catalog includes fewer studies of rare cancer types, which is not surprising given the need for large sample sizes for adequate power to examine rare variants in association with cancer risk. Most GWAS of rare cancers were conducted in European ancestry populations, underscoring the need to increase representation of diverse populations in genomic research. The cancer epidemiology cohorts supported by DCCPS are also a resource for rare cancer studies, not only because of the wealth of epidemiological data that they contain but also because of the numerous types of biospecimens that are collected (<https://cedcd.nci.nih.gov>).

Given the generally poor survival trends and lack of effective prevention strategies and treatment options for many rare cancers, additional efforts to understand these cancers clearly are needed. Our framework development, portfolio analysis, and resources review provide an overview of the rare cancer population sciences research landscape as well as information on gaps and opportunities that can be used to develop efficient, comprehensive strategies and to improve resources with the goal of accelerating rare cancer research.

## Authors' Disclosures

No disclosures were reported.

## Authors' Contributions

**L. Gallicchio:** Conceptualization, resources, formal analysis, supervision, methodology, writing—original draft. **D.L. Dae:** Conceptualization, resources, data curation, formal analysis, methodology, writing—original draft, writing—review and editing. **M. Rotunno:** Conceptualization, resources, data curation, formal analysis, writing—original draft, project administration, writing—review and editing. **R. Barajas:** Resources, formal analysis, writing—review and editing. **S. Fagan:** Resources, data curation, writing—review and editing. **D.M. Carrick:** Conceptualization, resources, methodology, writing—review and editing. **R.L. Divi:** Conceptualization, resources,

methodology, writing–review and editing. **K.K. Filipski:** Conceptualization, methodology, writing–review and editing. **A.N. Freedman:** Conceptualization, resources, methodology, writing–original draft, writing–review and editing. **E.M. Gillanders:** Conceptualization, resources, methodology, writing–original draft, writing–review and editing. **T.K. Lam:** Conceptualization, resources, investigation, methodology, writing–review and editing. **D.N. Martin:** Conceptualization, methodology, writing–review and editing. **S. Rogers:** Resources, data curation, formal analysis, writing–

review and editing. **M. Verma:** conceptualization, resources, methodology, writing–review and editing. **S.A. Nelson:** Conceptualization, resources, methodology, writing–original draft, writing–review and editing.

Received December 19, 2020; revised February 15, 2021; accepted March 10, 2021; published first April 1, 2021.

## References

1. Greenlee RT, Goodman MT, Lynch CF, Platz CE, Havener LA, Howe HL. The occurrence of rare cancers in U.S. adults, 1995–2004. *Public Health Rep* 2010;125:28–43.
2. DeSantis CE, Kramer JL, Jemal A. The burden of rare cancers in the United States. *CA Cancer J Clin* 2017;67:261–72.
3. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, et al. (eds). *SEER Cancer Statistics Review, 1975–2016*. Bethesda, MD: National Cancer Institute. Available from: [https://seer.cancer.gov/csr/1975\\_2016/](https://seer.cancer.gov/csr/1975_2016/).