

A Scalable Quality Assurance Process for Curating Oncology Electronic Health Records: The Project GENIE Biopharma Collaborative Approach

Jessica A. Lavery, MS¹; Eva M. Lepisto, MA, MSc²; Samantha Brown, MS¹; Hira Rizvi, BA¹; Caroline McCarthy, MPH¹; Michele LeNoue-Newton, PhD³; Celeste Yu, MS⁴; Jasme Lee, MS¹; Xindi Guo, BS⁵; Thomas Yu, BS⁵; Julia Rudolph, MPA¹; Shawn Sweeney, PhD⁶; AACR Project GENIE Consortium; Ben Ho Park, MD, PhD³; Jeremy L. Warner, MS³; Philippe L. Bedard, MD⁴; Gregory Riely, MD, PhD¹; Deborah Schrag, MD, MPH²; and Katherine S. Panageas, DrPH¹

abstract

PURPOSE The American Association for Cancer Research Project Genomics Evidence Neoplasia Information Exchange Biopharma Collaborative is a multi-institution effort to build a pan-cancer repository of genomic and clinical data curated from the electronic health record. For the research community to be confident that data extracted from electronic health record text are reliable, transparency of the approach used to ensure data quality is essential.

MATERIALS AND METHODS Four institutions participating in AACR's Project GENIE created an observational cohort of patients with cancer for whom tumor molecular profiling data, therapeutic exposures, and treatment outcomes are available and will be shared publicly with the research community. A comprehensive approach to quality assurance included assessments of (1) feasibility of the curation model through pressure test cases; (2) accuracy through programmatic queries and comparison with source data; and (3) reproducibility via double curation and code review.

RESULTS Assessments of feasibility resulted in critical modifications to the curation directives. Queries and comparison with source data identified errors that were rectified via data correction and curator retraining. Assessment of intercurator reliability indicated a reliable curation model.

CONCLUSION The transparent quality assurance processes for the GENIE BPC data ensure that the data can be used for analyses that support clinical decision making and advances in precision oncology.

JCO Clin Cancer Inform 6:e2100105. © 2022 by American Society of Clinical Oncology

Licensed under the Creative Commons Attribution 4.0 License 

INTRODUCTION

The future of precision medicine in oncology requires detailed patient data alongside molecular characterization of tumors to allow for discovery, risk stratification and ultimately, to inform the selection of optimal therapy.¹ Although efforts such as The Cancer Genome Atlas have molecularly characterized more than 200,000 primary cancer tumors and have led to insights regarding the genomic landscape of many cancers, phenomic (clinical) data that includes clinical characteristics, therapeutic exposures, and salient outcomes are limited in The Cancer Genome Atlas and similar data sources.² Historically, with limited treatment options and short survival for patients with advanced cancer, overall survival (OS) was viewed as the most relevant end point for clinical decision making. With more treatments now available and the need to make more rapid treatment decisions, intermediate end points such as progression-free survival (PFS) are

routinely used for patients treated as part of clinical trials. However, with the majority of patient care occurring outside of the clinical trial setting, there is a need for robust curation of intermediate end points such as real-world progression-free survival. Much of the critical data characterizing end points such as treatment duration, toxicity, progression, and recurrence are not captured using structured data fields in the electronic health record (EHR), posing challenges for the collection and synthesis of these data.³

The American Association for Cancer Research (AACR) Project Genomics Evidence Neoplasia Information Exchange Biopharma Collaborative (GENIE BPC) is a multi-institution effort to build a pan-cancer data repository of genomic, therapeutic, and phenomic data curated from the EHR. The GENIE BPC project builds on AACR Project GENIE, a publicly available registry of genomic data and limited clinical data from 19 cancer centers internationally.⁴⁻⁶ To

ASSOCIATED CONTENT

Appendix

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on January 6, 2022 and published at ascopubs.org/journal/cci on February 22, 2022; DOI <https://doi.org/10.1200/CCI.21.00105>

CONTEXT

Key Objective

To develop a scalable quality assurance process for a multi-institution initiative to curate disease characteristics, treatment, and outcomes from the electronic health record for patients with cancer.

Knowledge Generated

The transparent quality assurance processes and their findings ensure that the GENIE BPC data can be used for analyses that support clinical decision making and advances in precision oncology.

Relevance

These processes may also be adapted for other large-scale curation initiatives.

accomplish the goals of improving clinical decision making on the basis of real-world data, the existing Project GENIE genomic data in conjunction with the PRISMM framework^{7,8} to curate phenomic information from the EHR were leveraged to create a comprehensive data set characterizing the cancer treatment trajectories and outcomes of approximately 7,500 patients with cancer with six distinct cancer types.

The utility of the scientific research stemming from this open-access data source depends on having high-quality curated data that have undergone rigorous quality assurance (QA) processes to verify the completeness and accuracy of essential data elements. Even if performed at a single institution for a single cancer type, EHR data curation is complex and time-consuming; if not performed with appropriate quality control measures, the resulting data will be limited in its usability and generalizability. Harmonizing data curation is especially challenging when projects span multiple institutions that rely on distinct EHRs and informatics ecosystems. Furthermore, there are differences in provider care patterns and note structure, requiring coordination between institutions to ensure that data are consistently collected. For the research community to be confident that the data extracted from the EHR are reliable and reproducible, transparency of the QA approach is essential. This manuscript provides an overview of the GENIE BPC project and the approach used to ensure quality, rigor, and reproducibility so that the resulting data can support analyses that advance the goals of precision medicine.

MATERIALS AND METHODS

The AACR Project GENIE consortium has aggregated data from more than 120,000 tumors that have undergone high-throughput sequencing using targeted gene panels. Data from the sequencing reports and select clinical data elements, most notably age and OS, are publicly available, and have been previously described.⁹ The focus of the GENIE BPC project is to augment the genomic information with detailed phenomic data from the EHR and other institutional sources, such as the tumor registry, that characterize

baseline patient and tumor attributes, treatment exposures, and clinically relevant outcomes.

The initial phase of the GENIE BPC project included curation of six cancer cohorts: non-small-cell lung cancer (NSCLC; N = 1876), colorectal cancer (CRC; N = 1,501), breast cancer (N = 1,130), pancreatic cancer (N = 1,125), prostate cancer (N = 1,125), and bladder cancer (N = 750). Cases were randomly selected for phenomic data curation if the patient was age ≥ 18 years at the time of genomic sequencing, the sample met inclusion criteria (stage at diagnosis; eligible OncoTree code) and was performed during a specified time period, and eligible for 2 years of follow-up after sequencing. Each cancer-specific cohort is curated using the same data model with minor adaptations on the basis of features that are specific to the cancer type, such as human epidermal growth factor receptor 2 status for breast cancer. Data are curated locally at each institution and uploaded to a centralized data repository. There are four participating institutions: Memorial Sloan Kettering Cancer Center (MSKCC), Dana Farber Cancer Institute, Vanderbilt Ingram Cancer Center, and the University Health Network Princess Margaret Cancer Centre.

The phenomic, therapeutic, and oncologic outcome data are curated from the EHR and other institutional data sources according to the patient- and cancer-centric PRISMM framework for determination of real-world outcomes. PRISMM uses a cancer-agnostic curation model that standardizes key components of outcome ascertainment in oncology on the basis of existing data standards. The PRISMM framework for defining cancer outcomes includes data collection with respect to pathology; radiology; imaging; signs and symptoms; tumor markers; and medical oncologist assessments.^{7,8,10} Specifically, all pathology reports, all radiology reports for imaging studies other than plain x-rays and ultrasounds, and one medical oncologist note per month are curated. Curation is partially automated; select variables (< 10%) that are recorded in the institution's tumor registry are transferred directly into the GENIE BPC database, whereas other variables are manually extracted from the EHR.

Curated data are stored in a Research Electronic Data Capture (REDCap) database of more than 700 data fields structured with a REDCap instrument corresponding to each of the PRISMM modules. Although the data dictionary is uniform across institutions, different EHR systems and versions of REDCap data management software across institutions add a layer of complexity to ensuring consistent quality.

Project Teams

Phenomic data curation across institutions requires a dedicated team at each institution with expertise in both informatics and clinical oncology, centralized project coordination and management, a centralized data repository and host, and a governance structure to facilitate streamlined decision making. The AACR Project GENIE Coordinating Center oversees management of the project and liaises between the biopharmaceutical partners and participating institutions that contribute data. Sage Bionetworks maintains the centralized data repository and is responsible for assuring data deidentification, versioning, provenance, access, and release internally and publicly. Each participating institution has a team of curators who abstract data from the medical record and QA manager(s) who are experts in clinical oncology research and

responsible for training curators and implementing and overseeing the QA processes. The Statistical Coordinating Center (SCC) is centralized at MSKCC and is responsible for supporting QA processes, providing feedback to institutions, stipulating the need for reabstraction of fields with unacceptably high rates of discordance, and preparing the final comprehensive data set for release, including the derivation of variables and their respective documentation.

QA Processes

Curation initiatives require assessment of different components of data quality throughout the curation process. Multipronged QA processes were developed at the start of the project to address feasibility, accuracy, and reproducibility (Table 1; Fig 1). We describe each process and its role in ensuring the quality of the data. At the time of this writing, the NSCLC, CRC, and breast cancer data have been curated; summaries of the results of the QA processes for these cancer sites are presented.

Assessing feasibility of the curation model. At the initiation of curation of every new cancer cohort, a pilot phase is conducted to assess the feasibility of the curation model. Although PRISMM is a pan-cancer curation model, there

TABLE 1. Summary of QA Processes

Timing	QA Process	Scale (per institution)	Responsible Party	Goal
Preproduction	Data dictionary pilot	Three-five cases	QA manager	Ensure that curators at each institution have a clear understanding of each data field and are able to identify the required information from the EHR Identify any modifications required to the REDCap data dictionary for each cancer site and develop training materials for curators
	Curation pressure test	10-15 cases	Curators and QA manager	Identify any cancer-specific curation challenges and identify any areas for curator training
Production (curation)	Data quality rules	100%	Development of data quality rules: QA Managers, SCC Implementation of rules: REDCap Database	Identify any impossible values in the data (eg, date of birth after date of death) in real time
	Double curation	5%	Identifying cases for double curation: SCC, Sage Bionetworks Performing double curation: curators Assessing double curation: SCC	Estimate the reproducibility of specific elements of the PRISMM framework
	Source data verification	20%	Identifying cases for source data verification: SCC, Sage Bionetworks Performing source data verification: QA manager Summarizing source data verification: SCC	Ensure a high degree of accuracy compared with the source data
Postproduction	Data quality queries	100%	Programming and providing query reports: SCC Resolving queries: QA managers	Identify any implausible or impossible values in the data that could not be identified through data quality rules

Abbreviations: EHR, electronic health record; QA, quality assurance; REDCap, Research Electronic Data Capture; SCC, Statistical Coordinating Center.

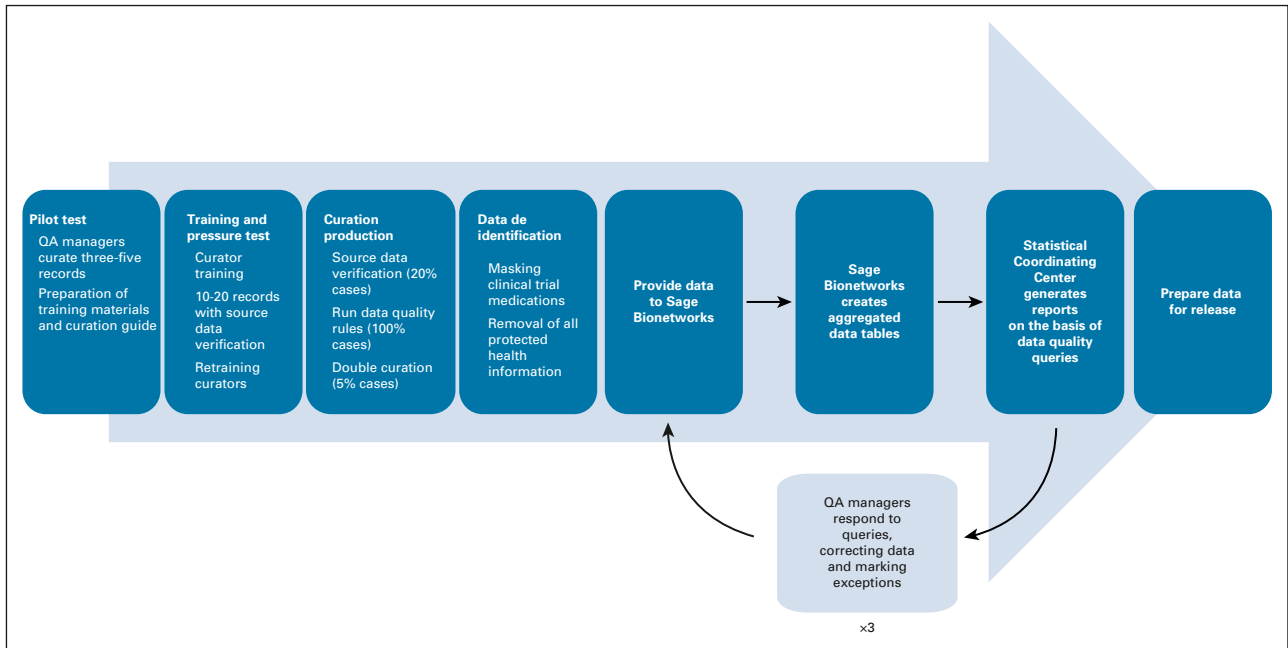


FIG 1. Flow diagram for QA processes and data release. QA, quality assurance.

are distinctions in the recording of data in the EHR across cancer sites. QA managers curate between three and five cases to assess the curation directives for each new cancer site. The goal of the pilot curation phase is to ensure that the data directives are comprehensible, cancer-specific modifications can be operationalized, and training materials for curators cover both common and ambiguous cases with specific examples. During this phase, QA managers from participating institutions contribute feedback on the content of the database, the training directives, and the feasibility of extracting specific elements in their local

environments. Modifications are made to the REDCap data dictionary, the accompanying curation guide, and the training materials, as needed.

After the pilot phase, curator training is performed by QA managers using the training directives with cancer-specific data elements and directives, which indicate the documents in the EHR that are in scope for review and abstraction under the PRISMM framework. Curators then review examples and a test case with QA managers. This process ensures that curators are abstracting relevant information from the EHR consistently.

TABLE 2. Example Data Quality Rules and Data Quality Queries

PRISMM Module	Example Data Quality Rule in REDCap	Example Data Quality Query Implemented by SCC
Patient characteristics/vital status	Missing data (eg, birth year, vital status)	Patient recorded as alive, but date of death is populated
Cancer diagnosis	Missing data (eg, diagnosis date, birth year)	A cancer is marked as the cancer making the patient eligible for curation, but does not have associated genomic sequencing report indicated
Cancer-directed regimens	Drug end date before drug start date	Drug regimen started more than 30 days before its associated cancer diagnosis
Pathology	Missing specimen information	Source of cancer diagnosis is listed as pathology without a corresponding curated pathology report
Imaging	Image scan date after imaging report date	Sequential scans of the same anatomic site showed evidence of cancer changed from no to yes, but progression is not indicated
Medical oncologist assessment	Missing medical oncologist visit date	Curated medical oncologist assessment is more than 1 month before the first index cancer diagnosis
Cancer panel test	Missing link to corresponding pathology report	Cancer panel test report date is before its associated cancer diagnosis

Abbreviations: REDCap, Research Electronic Data Capture; SCC, Statistical Coordinating Center.

Following curator training, between 10 and 15 cases (one-two cases per curator) are curated at each institution, which are then fully reviewed by the QA Manager to ensure accuracy against the source data as part of a curation pressure test. The QA manager systematically records discrepancies. On the basis of the results of this source data verification, the data elements and/or directives may be modified; if data elements or directives are updated, curators at all institutions are retrained before production (Fig 1). If substantial modifications are required, then a second set of pressure test cases may be curated.

Assessing curated data accuracy. Data quality rules refer to programmatic checks implemented in the REDCap project that are executed in real time during curation. The data quality rules aim to identify missing data and any impossible values, for example, a date of death preceding the date of birth. There are more than 150 data quality rules implemented as part of GENIE BPC (Table 2; Appendix Fig A1). Identification of inconsistencies and errors during curation allows for corrections to be implemented in real time. The

number of records flagged by a rule and the number of exceptions recorded cannot currently be exported from REDCap and readily analyzed.

There are more than 75 data quality queries that address the consistency of data elements that cannot be compared using data quality rules because of REDCap limitations (Table 2). Data quality queries are implemented by the SCC after curation is complete on the data aggregated across institutions. For both data quality rules and queries, curators and QA managers determine whether the data require a correction; if not, the exception is documented. A minimum of three rounds of data quality queries are evaluated for each cancer cohort to ensure the data are of the highest integrity at the time of release.

The accuracy of the curated data is also assessed through source data verification, a manual process requiring familiarity with the curation directives, clinical content, and the nuances of phenomic data documentation in the EHR. QA Managers at each institution perform source data verification on 20% of records that are randomly selected

TABLE 3. Source Data Verification: Major and Minor Violations

PRISMM Module	Minor Issue	Major Issue
Patient characteristics/vital status	Missing source of death information Missing death indicator (alive/dead)	Incorrect/missed death date (> 1 month off) Incorrect/missed last alive date (> 1 month off)
Cancer diagnosis	Incorrect/missed sites of distant metastases at diagnosis Incorrect/missed cancer-specific questions (eg, smoking status)	Incorrect/missed cancer type Incorrect/missed dx date (> 1 month off) Incorrect/missed stage (early v stage IV) Missing cancer diagnosis
Cancer-directed regimens	Missing drug from regimen Incorrect end date (> 1 month off) Incorrect/missed discontinuation status Incorrect/missed clinical trial information Drugs grouped as regimen incorrectly	Incorrect overall start date (> 1 month off) Regimen associated with wrong cancer diagnosis Missing regimen
Pathology	Incorrect/missing histology Incorrect/missing specimens Incorrect report date (mm/yyyy)	Incorrect/missed procedure date (mm/yyyy) Incorrect/missed associated cancer Incorrectly identified/missed invasive cancer Missed biomarker test Incorrect PD-L1 test result Incorrect MSI test result Incorrect MMR test result Missed eligible path report
Imaging	Incorrect/missing sites Incorrect/missing report date Incorrect cancer status (other than major error re: progressing/responding)	Incorrect scan date (> 1 month) Incorrectly identified/missing invasive cancer Incorrectly identified or missed progression Incorrectly identified or missed response Missed eligible imaging report
Medical oncologist assessment	Incorrect cancer status (other than major error re: progressing/responding) Ineligible note included	Incorrect/missed visit date (mm/yyyy) Note associated with wrong cancer diagnosis Incorrectly identified or missed invasive cancer Incorrectly identified or missed progression Incorrectly identified or missed response Missed eligible medical oncologist note
Cancer panel test	NA	Incorrect pathology report CPT associated with wrong cancer diagnosis

Abbreviations: CPT, Cancer Panel Test; dx, diagnosis; MMR, mismatch repair; MSI, microsatellite instability; NA, not applicable; PD-L1, programmed death-ligand 1.

before beginning curation. Source data verification is performed after the case is fully curated, during production, so that any systematic issues identified across records can be rectified before completing curation of an entire cancer cohort. Discrepancies between the curated data and the EHR are classified into major, minor, and other issues (Table 3). The issues were categorized a priori by a team of clinical experts, taking into consideration the ramifications of erroneous data. Major issues are discrepancies that would affect data analysis of key outcome measures, such as incorrectly identified or missed progression on an imaging report. Minor issues, such as missing source of death information, represent discrepancies with lesser implications for analysis. Other issues refer to discrepancies that were not classified as major or minor, such as instances where the data are not incorrect but could be more specific. Although all discrepancies are corrected by QA managers, the classification into major and minor issues is used to systematically identify opportunities to revise curation directives and to initiate curator retraining. The SCC developed a QA application that updates automatically as data are centrally uploaded and summarizes the extent of issues within and across institutions (Appendix Fig A2). An acceptable benchmark regarding the number of violations is not easily defined; these data are regularly evaluated in a qualitative fashion to determine when and how to initiate corrections.

Assessing reproducibility. Double curation was performed to evaluate intercurator reliability. For a predetermined randomly selected subset of 5% of cases at each institution, two curators independently abstracted data for the same patient. Agreement was assessed using the kappa statistic for categorical variables, concordance correlation coefficient for continuous variables, and a correlation statistic on the basis of iterative multiple imputation for time-to-event data.¹¹

Statistical code review was performed by a second statistician to ensure reproducibility of the derivation of variables,

including outcomes such as PFS and OS provided in the data releases. Any discrepancies were discussed within the SCC and resolved accordingly before data release.

RESULTS

Over the course of curation, 34% (n = 647) NSCLC, 29% (n = 436) CRC, and 53% (n = 597) of patients with breast cancer had at least one data quality query. There were 1,074 queries identified for round one of NSCLC review, 404 for round two, and 65 for round three. For CRC, there were 699, 187, and 22 queries in rounds one, two, and three, respectively. For breast cancer, 846 queries were identified in round one, 350 in round two, and 154 in round three. Of the queries identified, 30% of NSCLC, 50% of CRC, and 49% of breast cancer queries did not result in a correction to the data (exception recorded). The high rates of queries designated as exceptions reflect an evolving process for evaluating data quality queries. If QA managers identify a common exception to a complex query, they provide feedback to the SCC so that the scenario is programmatically excluded from future query reports. Conversely, upon reviewing the data, the QA managers may recommend a new query be incorporated for subsequent rounds.

The findings of source data verification yielded similar patterns across the three cohorts. The percentage of forms with major issues ranged from < 1% (cancer panel test instrument) to 8% (imaging and medical oncologist instruments), and the percentage with minor violations ranged from < 1% (vital status and cancer panel test instruments) to 20% (cancer diagnosis instrument; Table 4). All inconsistencies were corrected and triggered partial review of records that were not selected for source data verification to ensure high quality for data fields found to have high error rates.

Assessment of intercurator reliability for select variables on the cancer diagnosis, cancer-directed drug regimen, and cancer panel test instruments revealed adequate

TABLE 4. Summary of Source Data Verification

PRISSMM Module	NSCLC			CRC			Breast cancer		
	No. of Forms Reviewed	% Major Issue	% Minor Issue	No. of Forms Reviewed	% Major Issue	% Minor Issue	No. of Forms Reviewed	% Major Issue	% Minor Issue
Vital status	469	4	2	356	3	< 1	375	3	1
Cancer diagnosis	710	6	20	436	4	14	429	3	13
Cancer-directed regimens	1,066	7	13	1,280	4	8	2,106	4	10
Pathology	2,233	7	9	1,699	5	8	2,291	3	8
Imaging	9,573	8	10	6,342	4	7	8,180	6	7
Medical oncologist assessment	7,039	8	4	6,897	6	2	8,793	6	2
Cancer panel test	338	< 1	< 1	365	1	< 1	406	3	< 1

Abbreviations: CRC, colorectal cancer; NSCLC, non-small-cell lung cancer.

TABLE 5. Assessment of Reproducibility: Percent Agreement for Select Variables

Data Element	NSCLC		CRC		Breast Cancer	
	% Agreement	Agreement Statistic ^a	% Agreement	Agreement Statistic ^a	% Agreement	Agreement Statistic ^a
Cancer diagnosis						
Date of diagnosis, same date	93	1.0	89	1.0	95	1.0
Date of diagnosis (± 14 days)	95	—	97	—	95	—
Histology	95	0.91	100	1.0	98	0.85
Stage at diagnosis	86	0.82	93	0.90	85	0.80
Liver metastasis at diagnosis	97	0.60	98	0.94	98	0.79
Cancer-directed regimens						
Drugs in first regimen (up to 5)	85	0.80	86	0.84	75	0.71
Start date of first regimen	86	1.0	79	1.0	64	1.0
Start date of first regimen (± 14 days)	87	—	86	—	80	—
Cancer panel test						
Cancer panel test sequencing date ^b	91	1.0	87	1.0	98	1.0
Survival end points from diagnosis						
OS	—	0.98 (0.96-0.99)	—	1 (1-1)	—	1 (1-1)
OS status indicator	100	1.0	99	0.98	100	1.0
OS, days	69	0.99	63	1.0	78	1.0
OS, days (± 14 days)	79	—	79	—	83	—
PFS-I ^c	—	0.59 (0.35-0.75)	—	0.89 (0.83-0.93) ^d	—	—
PFS-I status indicator	99	0.97	98	0.95	—	—
PFS-I, days	91	0.63	85	0.91 ^e	—	—
PFS-I, days (± 14 days)	91	—	90	—	—	—
PFS-M ^c	—	0.79 (0.64-0.88)	—	0.80 (0.71-0.87)	—	—
PFS-M status indicator	98	0.94	100	1.0	—	—
PFS-M, days	92	0.61	87	0.93	—	—
PFS-M, days (± 14 days)	94	—	90	—	—	—

Abbreviations: CRC, colorectal cancer; NSCLC, non-small-cell lung cancer; OS, overall survival; PFS-I, progression-free survival-imaging; PFS-M, progression-free survival-medical oncologist.

^aThe measure of agreement was a kappa statistic for categorical variables, concordance correlation coefficient for continuous variables, and a correlation statistic on the basis of iterative multiple imputation for time-to-event data.¹¹

^bNo change in results for exact date versus date within 14 days.

^cPFS is only estimated among stage IV patients. For the breast cancer cohort, there were too few double-curated stage IV patients to assess intercurator reliability of PFS-I or PFS-M.

^dThe value presented on the table represents the agreement statistic after the removal of a single outlier. With the outlier, the agreement statistic is 0.67 (0.51-0.78).

^eThe value presented on the table represents the agreement statistic after the removal of a single outlier. With the outlier, the agreement statistic is 0.32.

agreement between curators (kappa ≥ 0.6; concordance correlation coefficient > 0.6). Estimates of the survival distributions were similar when analyzing the double-curated cases (Table 5; Appendix Figs A3 and A4).

DISCUSSION

The GENIE BPC Project represents a coordinated effort to generate a publicly released linked clinicogenomic data set that, when analyzed with methodologic rigor, can be used to further advance precision medicine in oncology. A comprehensive approach to QA was necessary to ensure high-quality data for curation of pan-cancer data from

EHRs across multiple institutions. As initiatives to curate EHR data are increasingly used to structure data for research purposes and further may inform regulatory decisions, it is important to highlight the need for transparent QA processes that are uniform across institutions. To enhance rigor and reproducibility, such initiatives require the means to implement a QA process that addresses the feasibility, accuracy, and reproducibility of data curation. An extensible and portable QA process and built-in data checks facilitate the scalability of the project. A rigorous and scalable QA process for GENIE BPC was accomplished by a multipronged process targeting a subset of records for the

most time-intensive QA processes (20% source data verification; 5% double curation), applying data quality rules and data quality queries to 100% of cases allowing for data corrections and the identification of systematic discrepancies.

Especially in a project of this scale, it is critical that QA processes involve harmonization of curation and QA achieved through frequent communication of project teams within and across institutions. The PRISSMM model was developed to standardize data collection across varied EHRs. Furthermore, the QA processes are performed on the harmonized data and are implemented uniformly. The QA processes described in this manuscript reflect a set of procedures that are optimal to this specific project in terms of balancing resources and data quality and have been refined based upon careful review of the curated data and a learning QA system. The QA processes use programmatic standardization and implementation when possible (uniform REDCap data dictionaries, data quality rules, and data quality queries), and rely on manual review for a subset of cases (source data verification) to systematically identify areas for improvement. Although the future of real-world data collection is advancing toward natural language processing for data abstraction, reducing the efforts and time required for human curation, rigorous QA processes that rely heavily on manual review would still be required to guarantee the accuracy of the data.¹²

AFFILIATIONS

¹Memorial Sloan Kettering Cancer Center, New York, NY

²Division of Population Sciences, Dana-Farber Cancer Institute Boston, MA

³Vanderbilt Ingram Cancer Center, Nashville, TN

⁴Princess Margaret Cancer Centre, University Health Network, Toronto, ON

⁵Sage Bionetworks, Seattle, WA

⁶American Association for Cancer Research, Philadelphia, PA

CORRESPONDING AUTHOR

Jessica A. Lavery, MS, Memorial Sloan Kettering Cancer Center, 485 Lexington Ave, 2nd Floor, New York, NY 10017; e-mail: laveryj@mskcc.org.

SUPPORT

Supported by The American Association for Cancer Research (all authors).

Support for this work was provided to Memorial Sloan Kettering Cancer Center by a core grant from the National Cancer Institute (P30 CA008748).

AUTHOR CONTRIBUTIONS

Conception and design: Jessica A. Lavery, Eva M. Lepisto, Samantha Brown, Caroline McCarthy, Michele LeNoue-Newton, Shawn Sweeney, Gregory Riely, Deborah Schrag, Katherine S. Panageas

Administrative support: Shawn Sweeney, Ben Ho Park, Gregory Riely, Deborah Schrag

Provision of study material or patients: Eva M. Lepisto, Philippe L. Bedard, Deborah Schrag

Collection and assembly of data: Jessica A. Lavery, Eva M. Lepisto, Samantha Brown, Hira Rizvi, Caroline McCarthy, Michele LeNoue-

Newton, Celeste Yu, Xindi Guo, Thomas Yu, Julia Rudolph, Philippe L. Bedard, Gregory Riely, Deborah Schrag, Katherine S. Panageas
Data analysis and interpretation: Jessica A. Lavery, Samantha Brown, Jasme Lee, Ben Ho Park, Jeremy L. Warner, Philippe L. Bedard, Deborah Schrag, Katherine S. Panageas
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

Although the curation and QA processes can easily be extended to all solid tumors, adaptations to curator training, data abstraction, and QA methods may be required to maintain high-quality data for hematologic malignancies, which are distinct from solid tumors in terms of their characterization, treatment, and outcome ascertainment. For example, radiologic evidence of disease is a key end point in PRISSMM, but is less important for hematologic cancers, whereas assessment of residual disease according to blood-based assays is important for hematologic cancers but is not applicable for solid tumors. As such, the curation model would require revision and validation to appropriately account for these differences. Given the complexity of hematologic malignancies and related treatments, we anticipate that additional curator training may be needed. The other QA processes would remain largely the same, modified in terms of the addition of hematologic-specific logic checks and queries, with a potentially increased percentage of cases that undergo source data verification.

In conclusion, the GENIE BPC data contribute a standardized assessment on the basis of the PRISSMM framework of structured information from the radiology reports, pathology reports, and medical oncologist assessments, alongside important outcomes in oncology, namely treatment duration and PFS. In the setting of this pan-cancer multi-institution curation project, a comprehensive approach to assessing feasibility, accuracy, and reproducibility is necessary to ensure the highest data quality.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Jessica A. Lavery

Research Funding: AACR Project GENIE (Inst)

Eva M. Lepisto

Employment: Kiniksa Pharmaceuticals (I)

Stock and Other Ownership Interests: Kiniksa Pharmaceuticals (I)

Samantha Brown

Research Funding: AACR

Michele LeNoue-Newton**Employment:** DaVita (I)**Stock and Other Ownership Interests:** DaVita (I)**Research Funding:** GE Healthcare (Inst)**Shawn Sweeney****Employment:** ConcertAI (I)**Research Funding:** Amgen, AstraZeneca, Bristol Myers Squibb, Genentech, Bayer, Boehringer Ingelheim, Janssen, Merck, Novartis, Analysis Group**Ben Ho Park****Leadership:** Loxo**Stock and Other Ownership Interests:** Loxo, Celcuity**Consulting or Advisory Role:** Horizon Discovery, Loxo, Casdin Capital, Jackson Laboratory for Genomic Medicine, Celcuity, Sermonix Pharmaceuticals, Hologic, EQRx**Research Funding:** AbbVie, Pfizer, GE Healthcare, Lilly**Patents, Royalties, Other Intellectual Property:** Royalties paid through inventions at Johns Hopkins University by Horizon Discovery Ltd**Travel, Accommodations, Expenses:** Lilly, Loxo**Uncompensated Relationships:** Tempus**Jeremy L. Warner**

This author is an Associate Editor for *JCO Clinical Cancer Informatics*. Journal policy recused the author from having any role in the peer review of this manuscript.

Stock and Other Ownership Interests: HemOnc.org**Consulting or Advisory Role:** Westat, IBM, Roche, Flatiron Health**Travel, Accommodations, Expenses:** IBM**Philippe L. Bedard****Consulting or Advisory Role:** BMS (Inst), Pfizer (Inst), Seattle Genetics, Lilly, Amgen, Merck**Research Funding:** Bristol Myers Squibb (Inst), Sanofi (Inst), AstraZeneca (Inst), Genentech/Roche (Inst), Servier (Inst), GlaxoSmithKline (Inst), Novartis (Inst), PTC Therapeutics (Inst), Nektar (Inst), Merck (Inst), Seattle Genetics (Inst), Mersana (Inst), Immunomedics (Inst), Lilly (Inst), Amgen (Inst), Bicara Therapeutics (Inst)**Gregory Riely****Research Funding:** Novartis (Inst), Roche/Genentech (Inst), GlaxoSmithKline (Inst), Pfizer (Inst), Infinity Pharmaceuticals (Inst), Mirati Therapeutics (Inst), Merck (Inst), Takeda (Inst)**Patents, Royalties, Other Intellectual Property:** Patent application submitted covering pulsatile use of erlotinib to treat or prevent brain metastases (Inst)**Travel, Accommodations, Expenses:** Merck Sharp & Dohme**Other Relationship:** Pfizer, Roche/Genentech, Takeda**Deborah Schrag****Stock and Other Ownership Interests:** Merck (I)**Honoraria:** Pfizer**Consulting or Advisory Role:** JAMA-Journal of the American Medical Association**Research Funding:** AACR (Inst), GRAIL (Inst)**Patents, Royalties, Other Intellectual Property:** PRISMM model is trademarked and curation tools are available to academic medical centers and government under creative commons license**Other Relationship:** JAMA-Journal of the American Medical Association**Katherine S. Panageas****Stock and Other Ownership Interests:** AstraZeneca, Pfizer, Sunesis Pharmaceuticals

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

The authors would like to acknowledge the American Association for Cancer Research and its financial and material support in the development of the AACR Project GENIE registry as well as members of the AACR Project GENIE consortium for their commitment to data sharing. Interpretations are the responsibility of study authors. Members of the AACR Project GENIE consortium are listed in the [Appendix 1](#).

REFERENCES

1. Dolley S: Big data's role in precision public health. *Front Public Health* 6:68, 2018
2. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al: The Cancer Genome Atlas pan-cancer analysis project. *Nat Genet* 45:1113-1120, 2013
3. Kim E, Rubinstein SM, Nead KT, et al: The evolving use of electronic health records (EHR) for research. *Semin Radiat Oncol* 29:354-361, 2019
4. AACR Project GENIE: Powering precision medicine through an international consortium. *Cancer Discov* 7:818-831, 2017
5. Litchfield K, Turajlic S, Swanton C: The GENIE is out of the bottle: Landmark cancer genomics dataset released. *Cancer Discov* 7:796-798, 2017
6. Micheel CM, Sweeney SM, LeNoue-Newton ML, et al: American Association for Cancer Research Project Genomics Evidence Neoplasia Information Exchange: From inception to first data release and beyond—Lessons learned and member institutions' perspectives. *JCO Clin Cancer Inform* 2:1-14, 2018
7. Schrag D: GENIE: Real-World Application. Chicago, IL, American Society for Clinical Oncology Annual Meeting, 2018
8. Real-World Applications of GENIE and a Taxonomy for Defining Cancer Outcomes. 2018. <https://www.oncnct.com/news/immunotherapy-beyond-checkpoints-cars-trucks-and-more>
9. Consortium APG: AACR Project GENIE: Powering precision medicine through an international consortium. *Cancer Discov* 7:818-831, 2017
10. Conley RB, Dickson D, Zenklusen JC, et al: Core clinical data elements for cancer genomic repositories: A multi-stakeholder consensus. *Cell* 171:982-986, 2017
11. Ploner M, Kaider A, Heinze G: SurvCorr: Correlation of Bivariate Survival Times. R package version 1.0, 2015. <https://CRAN.R-project.org/package=SurvCorr>
12. Kehl KL, Elmarakeby H, Nishino M, et al: Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports. *JAMA Oncol* 5:1421-1429, 2019



APPENDIX

A

Cancer-Directed Drug 1

Name of Cancer-Directed Drug 1 Anastrozole(Arimidex,ICID1033,ZD1033)

Start Date for Cancer-Directed Drug 1 02-05-2016 M-D-Y
Record the first date that the patient received the drug for this combination drug therapy. Identified data that will not be shared.

Interval in days between the Date of Birth and the Start Date of Drug 1 19258 View equation
Calculated Field Calculated field. Do not enter data.

End Date for Cancer-Directed Drug 1 02-20-2015 M-D-Y
Record the last date that the patient received the drug. If there is no indication that the patient has stopped taking this drug, then leave blank. Identified data that will not be shared.

Interval in days between the Date of Birth and the End Date of Drug 1 18908 View equation
Calculated Field Calculated field. Do not enter data.

B

⚠ WARNING: Data Quality rules were violated! ✕

The Data Quality rules listed below were found to have discrepancies for this record. Review the table below to see which rules were violated so that the data values for the fields involved can be corrected, if necessary. If you have some data query privileges, you may click the button in the 'Resolve issue' column to open the Data Resolution Workflow popup for a given rule. After that data query is set to 'closed', that rule will no longer be displayed for this record whenever the record is saved.

	Rule(s) violated	Fields involved	Resolve issue
	Rule #31: Start Date of Drug 1 after End Date Drug 1 [drugs_startdt_drug_1]<>" and [drugs_enddt_drug_1]<>" and [drugs_startdt_drug_1]>[drugs_enddt_drug_1]	drugs_startdt_drug_1: 02-05-2016 drugs_enddt_drug_1: 02-20-2015	0 comments

FIG A1. Example of (A) incorrectly curated data and (B) corresponding REDCap data quality rule alert. Example REDCap data quality rule for incorrectly curated data on the basis of simulated data for an example patient. REDCap, Research Electronic Data Capture.

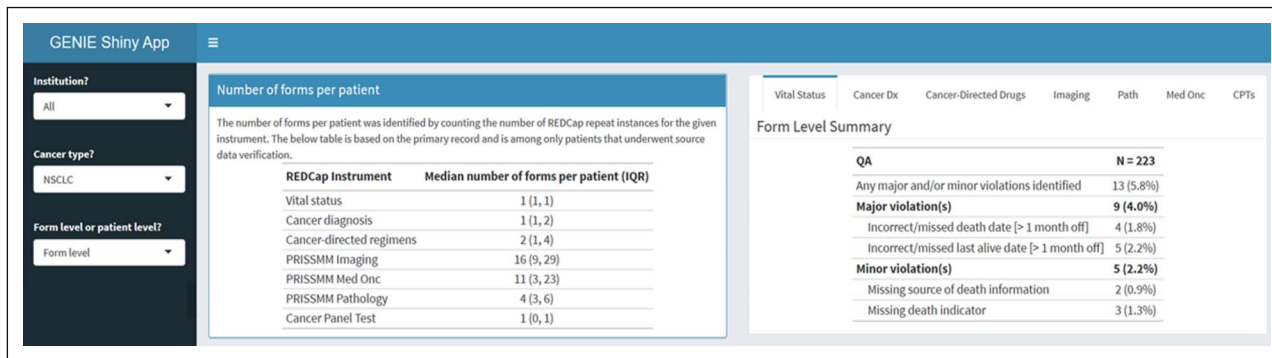


FIG A2. Snapshot of quality assurance application summarizing source data verification findings. The screenshot indicates a dropdown on the left-hand side where the user can specify a particular institution, cancer diagnosis, and summary level for the source data verification findings. The first table on the left shows the number of forms that were reviewed per patient. The table to the right has tabs for each PRISMM module and shows the number of forms that were compared with the electronic health record and the extent and type of major and minor issues.

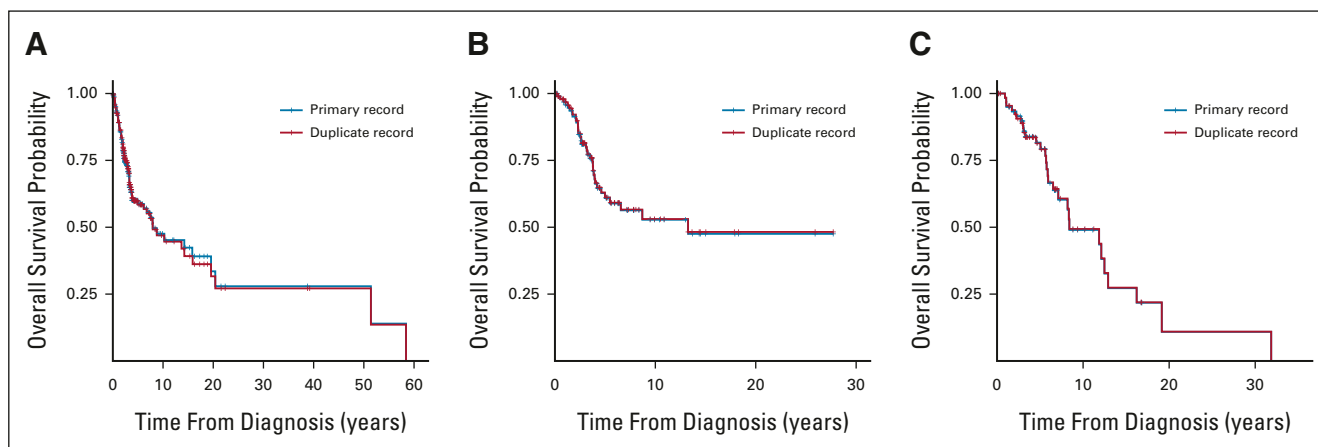


FIG A3. Kaplan-Meier estimates of OS stratified by curator for (A) NSCLC, (B) CRC, and (C) BrCa. Kaplan-Meier estimates of OS stratified by the primary curator and secondary curator among records that underwent double curation for the purposes of assessing reproducibility. Note that curves are not intended to describe estimates of time to event end points, but to demonstrate the assessment of reproducibility of curation of time to event end points across curators. BrCa, breast cancer; CRC, colorectal cancer; NSCLC, non-small-cell lung cancer; OS, overall survival.

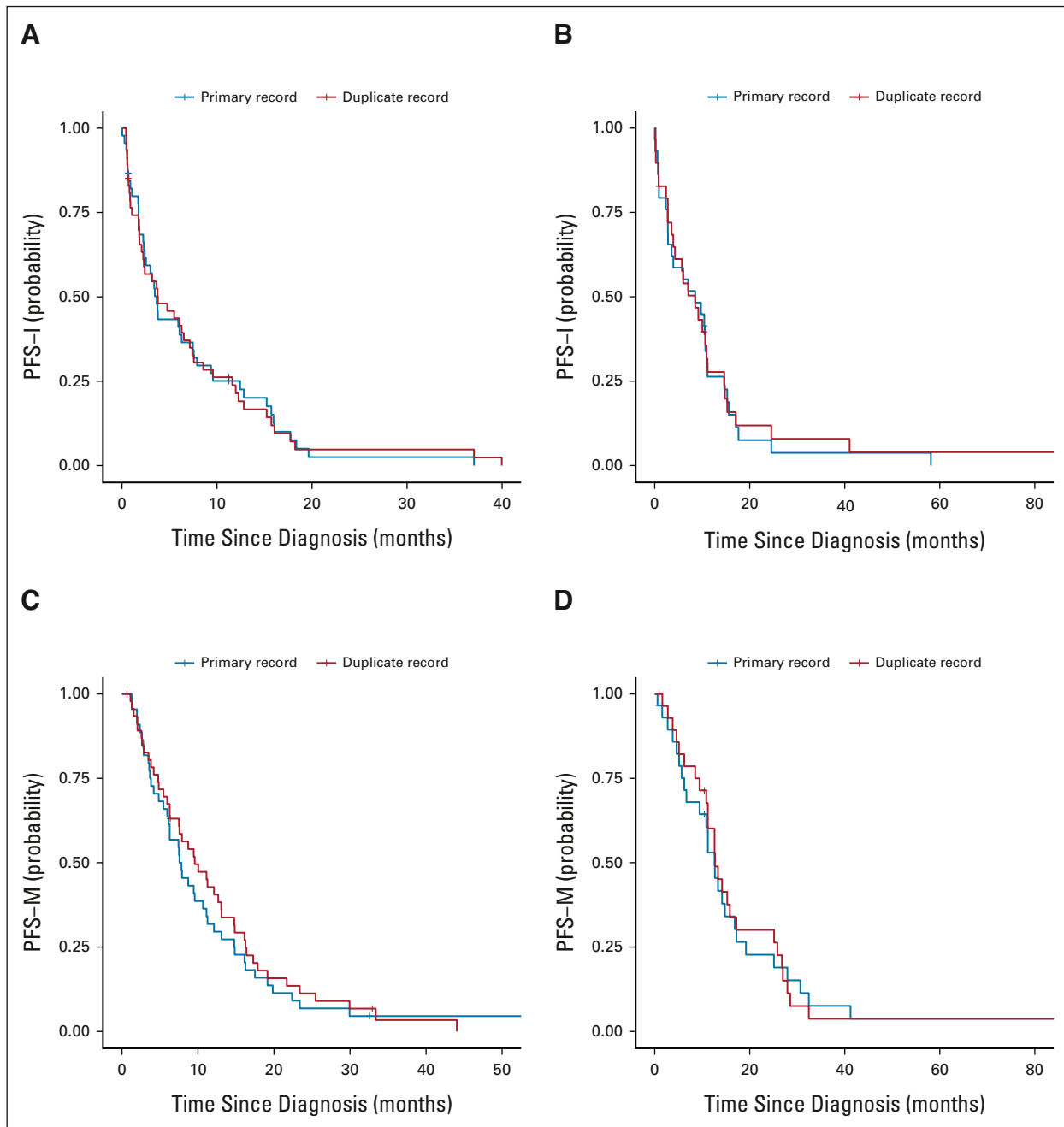


FIG A4. Kaplan-Meier estimates of PFS-I and PFS-M stratified by curator for NSCLC and CRC: (A) PFS-I NSCLC, (B) PFS-I CRC, (C) PFS-M NSCLC, and (D) PFS-M CRC. Survival curves are stratified by the primary curator and secondary curator among records that underwent double curation for the purposes of assessing reproducibility. Note that curves are not intended to describe estimates of time to event end points, but to demonstrate the assessment of reproducibility of curation of time to event end points across curators. CRC, colorectal cancer; NSCLC, non-small-cell lung cancer; PFS-I, progression-free survival according to imaging; PFS-M, progression-free survival according to medical oncologist.

APPENDIX 1. AACR PROJECT GENIE CONSORTIUM

First Name and Middle Initial(s)	Last Name	Institution	Location (city, state/providence, country)
Michael	Fiandalo	American Association for Cancer Research	Philadelphia, PA
Margaret	Foti	American Association for Cancer Research	Philadelphia, PA
Yekaterina	Khotskaya	American Association for Cancer Research	Philadelphia, PA
Shawn	Sweeney	American Association for Cancer Research	Philadelphia, PA
Jean	Abraham	Cancer Research UK Cambridge Centre	Cambridge, UK
Chris	Bournell	Cancer Research UK Cambridge Centre	Cambridge, UK
James	Brenton	Cancer Research UK Cambridge Centre	Cambridge, UK
Carlos	Caldas	Cancer Research UK Cambridge Centre	Cambridge, UK
Raquel	Garcia	Cancer Research UK Cambridge Centre	Cambridge, UK
Birgit	Nimmervoll	Cancer Research UK Cambridge Centre	Cambridge, UK
Ezequiel	Rodriguez	Cancer Research UK Cambridge Centre	Cambridge, UK
Oscar	Rueda	Cancer Research UK Cambridge Centre	Cambridge, UK
Dilrini	Silva	Cancer Research UK Cambridge Centre	Cambridge, UK
Mikel	Valgañón	Cancer Research UK Cambridge Centre	Cambridge, UK
Marilyn	Li	Children's Hospital of Philadelphia Cancer Center	Philadelphia, PA
Jena	Lilly	Children's Hospital of Philadelphia Cancer Center	Philadelphia, PA
Adam	Resnick	Children's Hospital of Philadelphia Cancer Center	Philadelphia, PA
Mahdi	Sarmardy	Children's Hospital of Philadelphia Cancer Center	Philadelphia, PA
Angela	Waanders	Children's Hospital of Philadelphia Cancer Center	Philadelphia, PA
Richard	Carvajal	Columbia	New York, NY
Susan	Hsaio	Columbia	New York, NY
Matthew	Ingham	Columbia	New York, NY
Raul	Rabadan	Columbia	New York, NY
Simon	Arango Baquero	Dana-Farber Cancer Institute	Boston, MA
Simon Arango	Baquero	Dana-Farber Cancer Institute	Boston, MA
Ethan	Cerami	Dana-Farber Cancer Institute	Boston, MA
Oya	Cushing	Dana-Farber Cancer Institute	Boston, MA
Matthew	Ducar	Dana-Farber Cancer Institute	Boston, MA
Alexander	Gusev	Dana-Farber Cancer Institute	Boston, MA
Bill	Hahn	Dana-Farber Cancer Institute	Boston, MA
Kevin	Haigis	Dana-Farber Cancer Institute	Boston, MA
Michael	Hassett	Dana-Farber Cancer Institute	Boston, MA
Katherine	Janeway	Dana-Farber Cancer Institute	Boston, MA
Pasi	Jänne	Dana-Farber Cancer Institute	Boston, MA
Jason	Johnson	Dana-Farber Cancer Institute	Boston, MA
Kenneth	Kehl	Dana-Farber Cancer Institute	Boston, MA
Priti	Kumari	Dana-Farber Cancer Institute	Boston, MA
Eva	Lepisto	Dana-Farber Cancer Institute	Boston, MA
Neal	Lindeman	Dana-Farber Cancer Institute	Boston, MA
Laura	MacConaill	Dana-Farber Cancer Institute	Boston, MA
Matthew	Meyerson	Dana-Farber Cancer Institute	Boston, MA
Diana	Miller	Dana-Farber Cancer Institute	Boston, MA

(Continued on following page)

(Continued)

First Name and Middle Initial(s)	Last Name	Institution	Location (city, state/providence, country)
John	Orechia	Dana-Farber Cancer Institute	Boston, MA
Sindy	Pimentel	Dana-Farber Cancer Institute	Boston, MA
Daniel	Quinn	Dana-Farber Cancer Institute	Boston, MA
Deborah	Schrag	Dana-Farber Cancer Institute	Boston, MA
Priyanka	Shivdasani	Dana-Farber Cancer Institute	Boston, MA
Parin	Sripakdeevong	Dana-Farber Cancer Institute	Boston, MA
Angela	Tramontano	Dana-Farber Cancer Institute	Boston, MA
Eliezer	Van Allen	Dana-Farber Cancer Institute	Boston, MA
Jonathan	Bell	Duke University Cancer Institute	Durham, NC
Michael	Datto	Duke University Cancer Institute	Durham, NC
Michelle	Green	Duke University Cancer Institute	Durham, NC
Shannon	McCall	Duke University Cancer Institute	Durham, NC
Fabrice	Andre	Gustave Roussy Cancer Campus	Villejuif, France
Monica	Ardenos	Gustave Roussy Cancer Campus	Villejuif, France
Fabien	Calvo	Gustave Roussy Cancer Campus	Villejuif, France
Semih	Dogan	Gustave Roussy Cancer Campus	Villejuif, France
Meurice	Guillaume	Gustave Roussy Cancer Campus	Villejuif, France
Lacroix	Ludovic	Gustave Roussy Cancer Campus	Villejuif, France
Jean	Scoazec	Gustave Roussy Cancer Campus	Villejuif, France
Gilles	Vassal	Gustave Roussy Cancer Campus	Villejuif, France
Stefan	Michels	Gustave Roussy Cancer Campus	Villejuif, France
Raymond	DuBois	Medical University of South Carolina	Charleston, SC
Mike	Berger	Memorial Sloan Kettering Cancer Center	New York, NY
Marufur	Bhuiya	Memorial Sloan Kettering Cancer Center	New York, NY
Samantha	Brown	Memorial Sloan Kettering Cancer Center	New York, NY
Ari	Caroline	Memorial Sloan Kettering Cancer Center	New York, NY
Cynthia	Chu	Memorial Sloan Kettering Cancer Center	New York, NY
Stu	Gardos	Memorial Sloan Kettering Cancer Center	New York, NY
Benjamin	Gross	Memorial Sloan Kettering Cancer Center	New York, NY
Alexia	Iasonosa	Memorial Sloan Kettering Cancer Center	New York, NY
Ritika	Kundra	Memorial Sloan Kettering Cancer Center	New York, NY
Andrew	Kung	Memorial Sloan Kettering Cancer Center	New York, NY
Marc	Ladanyi	Memorial Sloan Kettering Cancer Center	New York, NY
Jessica	Lavery	Memorial Sloan Kettering Cancer Center	New York, NY
Axel	Martin	Memorial Sloan Kettering Cancer Center	New York, NY
Brooke	Mastrogiacomo	Memorial Sloan Kettering Cancer Center	New York, NY
Caroline	McCarthy	Memorial Sloan Kettering Cancer Center	New York, NY
Kathy	Panageas	Memorial Sloan Kettering Cancer Center	New York, NY
John	Philip	Memorial Sloan Kettering Cancer Center	New York, NY
Greg	Reily	Memorial Sloan Kettering Cancer Center	New York, NY
Hira	Rivzi	Memorial Sloan Kettering Cancer Center	New York, NY
Julia	Rudolph	Memorial Sloan Kettering Cancer Center	New York, NY
Charles	Sawyers	Memorial Sloan Kettering Cancer Center	New York, NY

(Continued on following page)

(Continued)

First Name and Middle Initial(s)	Last Name	Institution	Location (city, state/providence, country)
Nikolaus	Schultz	Memorial Sloan Kettering Cancer Center	New York, NY
Julian	Schwartz	Memorial Sloan Kettering Cancer Center	New York, NY
Joseph	Sirintrapun	Memorial Sloan Kettering Cancer Center	New York, NY
David	Solit	Memorial Sloan Kettering Cancer Center	New York, NY
Stacy	Thomas	Memorial Sloan Kettering Cancer Center	New York, NY
Andrew	Zarski	Memorial Sloan Kettering Cancer Center	New York, NY
Ahmet	Zehir	Memorial Sloan Kettering Cancer Center	New York, NY
Lailah	Ahmed	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Philippe	Bedard	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Helen	Chow	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Samantha	Del Rossi	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Sevan	Hakgor	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Suzanne	KamelReid	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Geeta	Krishna	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Natasha	Leighl	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Zhibin	Lu	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Alisha	Nguyen	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Demi	Plagianakos	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Trevor	Pugh	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Nazish	Qazi	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Elizabeth	Shah	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Nitthusha	Singaravelan	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Lillian	Siu	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Natalie	Stickle	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Carlos	Virtaenen	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Stuart	Watt	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Celeste	Yu	Princess Margaret Cancer Centre	Toronto, Ontario, Canada
Brady	Bernard	Providence Cancer Institute	Portland, Oregon
Carlo	Bifulco	Providence Cancer Institute	Portland, Oregon
Julie	Cramer	Providence Cancer Institute	Portland, Oregon
Brian	Piening	Providence Cancer Institute	Portland, Oregon
Paul	Tittel	Providence Cancer Institute	Portland, Oregon
Walter	Urba	Providence Cancer Institute	Portland, Oregon
Alyssa	Acebedo	SAGE Bionetworks	Seattle, WA
Xindi	Guo	SAGE Bionetworks	Seattle, WA
Haley	Hunter-Zinck	SAGE Bionetworks	Seattle, WA
Thomas	Yu	SAGE Bionetworks	Seattle, WA
Alexander	Baras	Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University	Baltimore, MD
Julie	Brahmer	Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University	Baltimore, MD
Christopher	Gocke	Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University	Baltimore, MD

(Continued on following page)

(Continued)

First Name and Middle Initial(s)	Last Name	Institution	Location (city, state/providence, country)
Victor	Velculescu	Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University	Baltimore, MD
Neil	Bailey	Swedish Cancer Institute	Seattle, WA
Philip	Gold	Swedish Cancer Institute	Seattle, WA
Mariska	Bierkens	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Jacques	Craenmehr	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Annemieke	Hiemstra	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Jelle	Hoeve	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Hugo	Horlings	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Jan	Hudecek	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Martijn	Lolkema	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Geritt	Meijer	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Kim	Monkhorst	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Les	Nijman	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Gabe	Sonke	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Tony	van de Velde	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Harm	van Tinteren	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Emilie	Voest	The Netherlands Cancer Institute, Amsterdam, on behalf of the Center for Personalized Cancer Treatment	Utrecht, The Netherlands
Steinhardt	George	The University of Chicago Comprehensive Cancer Center	Chicago, IL
Sabah	Kadri	The University of Chicago Comprehensive Cancer Center	Chicago, IL
Wanjari	Pankhuri	The University of Chicago Comprehensive Cancer Center	Chicago, IL
Jeremy	Segal	The University of Chicago Comprehensive Cancer Center	Chicago, IL
Peng	Wang	The University of Chicago Comprehensive Cancer Center	Chicago, IL
Moung	Christine	University of California-San Francisco Helen Diller Family Comprehensive Cancer Center	San Francisco, CA
Eric	Talevich	University of California-San Francisco Helen Diller Family Comprehensive Cancer Center	San Francisco, CA
Laura	Van't Veer	University of California-San Francisco Helen Diller Family Comprehensive Cancer Center	San Francisco, CA
Alejandro	Sweet-Cordero	University of California-San Francisco Helen Diller Family Comprehensive Cancer Center	San Francisco, CA
Michelle	Turski	University of California-San Francisco Helen Diller Family Comprehensive Cancer Center	San Francisco, CA
Amanda	Wren	University of California-San Francisco Helen Diller Family Comprehensive Cancer Center	San Francisco, CA

(Continued on following page)

(Continued)

First Name and Middle Initial(s)	Last Name	Institution	Location (city, state/providence, country)
Susana	Aguilar Izquierdo	Vall d' Hebron Institute of Oncology	Barcelona, Spain
Rodrigo	Dienstmann	Vall d' Hebron Institute of Oncology	Barcelona, Spain
Francesco	Mancuso	Vall d' Hebron Institute of Oncology	Barcelona, Spain
Cristina	Viaplana Donato	Vall d' Hebron Institute of Oncology	Barcelona, Spain
Ingrid	Anderson	Vanderbilt-Ingram Cancer Center	Nashville, TN
Sandip	Chaugai	Vanderbilt-Ingram Cancer Center	Nashville, TN
Joseph	Coco	Vanderbilt-Ingram Cancer Center	Nashville, TN
Daniel	Fabbri	Vanderbilt-Ingram Cancer Center	Nashville, TN
Yuanchu	James Yang	Vanderbilt-Ingram Cancer Center	Nashville, TN
Leigh	Jones	Vanderbilt-Ingram Cancer Center	Nashville, TN
Michele	LeNoue Newton	Vanderbilt-Ingram Cancer Center	Nashville, TN
Christine	Lovly	Vanderbilt-Ingram Cancer Center	Nashville, TN
Christine	Micheel	Vanderbilt-Ingram Cancer Center	Nashville, TN
Sanjay	Mishra	Vanderbilt-Ingram Cancer Center	Nashville, TN
Ben	Park	Vanderbilt-Ingram Cancer Center	Nashville, TN
Thomas	Stricker	Vanderbilt-Ingram Cancer Center	Nashville, TN
Lucy	Wang	Vanderbilt-Ingram Cancer Center	Nashville, TN
Jeremy	Warner	Vanderbilt-Ingram Cancer Center	Nashville, TN
Li	Wen	Vanderbilt-Ingram Cancer Center	Nashville, TN
Yuanchu James	Yang	Vanderbilt-Ingram Cancer Center	Nashville, TN
Chen	Ye	Vanderbilt-Ingram Cancer Center	Nashville, TN
Meijian	Guan	Wake Forest University Health Sciences (Wake Forest Baptist Medical Center)	Winston-Salem, NC
Guangxu	Jin	Wake Forest University Health Sciences (Wake Forest Baptist Medical Center)	Winston-Salem, NC
Liang	Liu	Wake Forest University Health Sciences (Wake Forest Baptist Medical Center)	Winston-Salem, NC
Umit	Topaloglu	Wake Forest University Health Sciences (Wake Forest Baptist Medical Center)	Winston-Salem, NC
Wei	Zhang	Wake Forest University Health Sciences (Wake Forest Baptist Medical Center)	Winston-Salem, NC
Kaya	Bilguvar	Yale University Cancer Center	New Haven, CT
Michael	D'Eletto	Yale University Cancer Center	New Haven, CT
Daniel	Dykas	Yale University Cancer Center	New Haven, CT
James	Knight	Yale University Cancer Center	New Haven, CT
Shrikant	Mane	Yale University Cancer Center	New Haven, CT
E. Zeynep	Ormay	Yale University Cancer Center	New Haven, CT
Walther	Zenta	Yale University Cancer Center	New Haven, CT