# GENIE BPC Analytic Data Guide

## Colorectal Cancer v1.1-consortium

# Contents

# Overview

The Biopharma Collaborative (BPC) Core Team has created analytic datasets that include data from the AACR Project GENIE Tier 1 registry, institutions' tumor registries, manual curation based on institutions' electronic health records (PRISSMM phenomic data model), and derived variables based on these data elements. This data guide serves as a resource for all data elements included in the analytic datasets. These datasets containing demographic and phenomic information can be combined with the genomic data that the institutions have submitted to GENIE.

*Release Notes:* This document accompanies release **Colorectal Cancer v1.1-consortium**.

## BPC Projects

The BPC sponsors six projects that involve augmenting the genomic data in Project GENIE to include PRISSMM phenomic data and cancer-related outcomes. Each project encompasses a single cancer site. While all projects are based on curating data under the PRISSMM phenomic data curation model, the variables available for each project may vary.

The six projects are: bladder cancer, breast cancer, colon/rectal cancer, non-small cell lung cancer, pancreas cancer, and prostate cancer. Note that a patient can only be selected for one project (e.g. if patient is selected for the CRC project, then the patient cannot be selected for the breast cancer project).

*Eligibility*

Patients are eligible for curation based on having a BPC Project-specific cancer. Cases meeting specified eligibility criteria are randomly selected from the AACR Project GENIE Cancer Registry for enhanced PRISSMM phenomic data curation.

The BPC Project-specific eligibility criteria for the CRC project are as follows:

- Eligible OncoTree Diagnoses:
    - Colorectal Adenocarcinoma (COADREAD)
    - Colon Adenocarcinoma (COAD)
    - Mucinous Adenocarcinoma of the Colon and Rectum (MACR)
    - Signet Ring Cell Adenocarcinoma of the Colon and Rectum (SRCCR)
    - Rectal Adenocarcinoma (READ)
- Stage I-IV at diagnosis
- Genomic sequencing report at a participating institution between January 1, 2015 and December 31, 2018
- Aged 18 or older at the time of genomic sequencing
- Minimum of two years of follow-up after sequencing

These selection criteria could affect the ability to generalize to the entirety of patients with CRC. Genomic sequencing is not always performed at diagnosis and therefore may lead to several forms of bias, including lead time bias and immortal time bias. Investigators who wish to test specific hypotheses should work with a statistician to perform analyses that account for these biases. Failure to do so may result in incorrect inferences.

*Data Privacy*

Compliance with data privacy required redaction of the name and duration of investigational drugs as well as date intervals that could lead to identification of a patient as >89 years of age at any time point.

## Format

The format of the data guide is as follows:
***Field name***
[Variable name]
Value (character/numeric/date/date-time)
Description of the field
Data Standard (where applicable)

Field names shaded in gray indicate that the variable is provided for completeness, but that another variable is preferred. For example, age at diagnosis is captured from the tumor registry and from curation, but the derived variable combining the two sources is the variable that should be used for analysis. In this case, the tumor registry and curated age at diagnosis are shown with gray shading.

Variables have been color coded to help users understand their provenance.
Variables shown in orange indicate variables obtained from the AACR Project GENIE Tier 1 data.
Variables shown in green indicate variables obtained directly from the tumor registry.
Variables shown in blue indicate curated variables as well as variables pertaining to curation and quality assurance (QA).
Variables shown in purple indicate derived variables. The following table provides further information about the types of variables.

| Dataset | File Name |
|---|---|
| Patient-Level Dataset | patient_level_dataset.csv |
| BPC Project Cancer Diagnosis Dataset | cancer_level_dataset_index.csv |
| Non-BPC Project Cancer Diagnosis Dataset | cancer_level_dataset_non_index.csv |
| Cancer-Directed Regimen Dataset | regimen_cancer_level_dataset.csv |
| PRISSMM Pathology | pathology_report_level_dataset.csv |
| PRISSMM Imaging Dataset | imaging_level_dataset.csv |
| PRISSMM Medical Oncologist Assessment Dataset | med_onc_note_level_dataset.csv |
| PRISSMM Tumor Marker Dataset | tm_level_dataset.csv |
| Cancer Panel Test | cancer_panel_test_level_dataset.csv |

## Abbreviations

**AACR** American Association for Cancer Research
**BPC** Biopharma Collaborative
**CRC** Colon/Rectal Cancer
**DFCI** Dana-Farber Cancer Institute, Boston, MA, USA
**EHR** Electronic Health Record
**GENIE** Genomics Evidence Neoplasia Information Exchange
**HIPAA** Health Insurance Portability and Accountability Act
**MSK** Memorial Sloan Kettering Cancer Center, New York, NY, USA
**NAACCR** North American Association of Central Cancer Registries
**NGS** Next Generation Sequencing
**QA** Quality Assurance
**SDV** Source Data Verification
**VICC** Vanderbilt-Ingram Cancer Center, Nashville, TN, USA

# Patient-Level Dataset

The patient-level dataset is structured as one record per patient.

## *BPC project cohort*

[cohort]

Value (Character)

- CRC

Description

- Indicates the BPC project cancer type

## *Record ID*

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets. Note that AACR Project GENIE refers to participating institutions as centers. The institutions include:
  - DFCI
  - MSK
  - VICC

## *Institution*

[institution]

Value (Character)

- DFCI = Dana Farber Cancer Institute
- MSK = Memorial Sloan Kettering Cancer Center
- VICC = Vanderbilt Ingram Cancer Center

Description

- Indicates the patient’s internal institution (i.e. the same institution as the [institution] variable) of cancer care.

## *Patient has Redacted Time Interval Data to Comply with Health Insurance Portability and Accountability Act (HIPAA)*

[redacted]

Value (Character)

- Yes
- No

Description

- Indicates whether any portion of the patient's data (across all datasets) was redacted to comply with HIPAA.
- Time intervals indicating age >89 years or that could be used to identify age >89 will not be available for patients who have [redacted] = "Yes"

## *Year of Birth*

[birth_year]

Value (Numeric)

- YYYY

Description

- Patient's year of birth

## *Ethnicity: Spanish/Hispanic Origin*

[naaccr_ethnicity_code]

Value (Character)

- Non-Spanish; non-Hispanic
- Mexican (includes Chicano)
- Puerto Rican
- Cuban
- South or Central American (except Brazil)
- Other specified Spanish/Hispanic origin (includes European, excludes Dominican Republic)
- Spanish NOS or Hispanic NOS or Latino NOS
- Spanish surname only
- Dominican Republic
- Unknown whether Spanish or not

Description

- Ethnicity of patient, independent of patient's race
- Institutions not collecting Spanish/Hispanic origin have set this field to "Unknown whether Spanish or not"

Data Standard: NAACCR #190

## *Race (Primary)*

[naaccr_race_code_primary]

Value (Character)

- White
- Black
- American Indian, Aleutian, or Eskimo
- Chinese
- Japanese
- Filipino
- Hawaiian
- Korean
- Vietnamese
- Laotian
- Hmong
- Kampuchean (Cambodian)
- Thai
- Asian Indian or Pakistani NOS
- Asian Indian
- Pakistani
- Micronesian NOS
- Chamorro/Chamoru
- Guamanian NOS
- Polynesian NOS
- Tahitian
- Samoan
- Tongan
- Melanesian NOS
- Fiji Islander
- New Guinean
- Other Asian
- Pacific Islander NOS
- Other
- Unknown

Description

- First race specified, independent of ethnicity
- For institutions collecting more than one race category, this race variable indicates the primary race for the patient.
- Institutions not collecting race set this field to “Unknown”

Data Standard: NAACCR #160

## *Race (Secondary)*

[naaccr_race_code_secondary]

Value (Character)

- White
- Black
- American Indian, Aleutian, or Eskimo
- Chinese
- Japanese
- Filipino
- Hawaiian
- Korean
- Vietnamese
- Laotian
- Hmong
- Kampuchean (Cambodian)
- Thai
- Asian Indian or Pakistani NOS
- Asian Indian
- Pakistani
- Micronesian NOS
- Chamorro/Chamoru
- Guamanian NOS
- Polynesian NOS
- Tahitian
- Samoan
- Tongan
- Melanesian NOS
- Fiji Islander
- New Guinean
- Other Asian
- Pacific Islander NOS
- No further race documented
- Other
- Unknown

Description

- Second race specified, independent of ethnicity
- Institutions not collecting secondary race set this field to “No further race documented”

Data Standard: NAACCR #161

## *Race (Tertiary)*

[naaccr_race_code_tertiary]

Value (Character)

- White
- Black
- American Indian, Aleutian, or Eskimo
- Chinese
- Japanese
- Filipino
- Hawaiian
- Korean
- Vietnamese
- Laotian
- Hmong
- Kampuchean (Cambodian)
- Thai
- Asian Indian or Pakistani NOS
- Asian Indian
- Pakistani
- Micronesian NOS
- Chamorro/Chamoru
- Guamanian NOS
- Polynesian NOS
- Tahitian
- Samoan
- Tongan
- Melanesian NOS
- Fiji Islander
- New Guinean
- Other Asian
- Pacific Islander NOS
- No further race documented
- Other
- Unknown

Description

- Third race specified, independent of ethnicity
- Institutions not collecting tertiary race set this field to “No further race documented”

Data Standard: NAACCR #162

## ***Sex***

[naaccr_sex_code]

Value (Character)

- Male
- Female
- Other intersex, disorders of sexual development/DSD
- Transsexual NOS
- Transsexual natal male
- Transsexual natal female

Description

- Patient’s sex at time of diagnosis of index cancer

Data Standard: NAACCR #220

## *Time (Days) from Date of Birth to Death*

[hybrid_death_int]

Value (Numeric)

Description

- Interval in days from date of birth to date of death, if applicable

## *Source of Death Information*

[hybrid_death_source]

Value (Character)

- Curation
- EHR
- NDI (National Death Index)
- Tumor Registry
- Other

Description

- Indicates source of death information
- Populated only if patient is known to be dead at the time of curation

## *Time (Days) from Date of Birth to Date of Last Oncology Visit to Internal Institution*

[last_oncvisit_int]

Value (Numeric)

Description

- Interval in days from date of birth to most recent date that the patient had an in-person or tele-visit visit with an oncology provider at the institution; these visits may include medical oncology, surgical oncology, radiation oncology, palliative care, social work, lab draws, imaging scans, emergency room, or hospital visits.
- Completeness of data is subject to availability of visit information at each participating institution. Consequently, the preferred variable for determining the date of last contact is [dob_lastalive_int].

## *Time (Days) from Date of Birth to Date of Last Contact*

[last_alive_int]

Value (Numeric)

Description

- Interval in days from date of birth to most recent date that there was confirmation that the patient was alive on a specific date, such as documentation of a phone call or email exchange with the patient or a family member.
- Completeness of data is subject to availability of visit information at each participating institution. Consequently, the preferred variable for determining the date of last contact is [dob_lastalive_int].

## *Time (Days) from Date of Birth to Date of Last Known Non-Oncology Visit to Internal Institution*

[last_anyvisit_int]

Value (Numeric)

Description

- Interval in days from date of birth to most recent date that the patient had an in-person or tele-visit visit to the internal health care network for non-oncology related care, including visits with primary care or cardiology.
- Completeness of data is subject to availability of visit information at each participating institution. Consequently, the preferred variable for determining the date of last contact is [dob_lastalive_int].

## *Time (Days) from Date of Birth to Enrollment in Hospice Care*

[enroll_hospice_int]

Value (Numeric)

Description

- Based on interval in days from date of birth to the day that the patient was enrolled in hospice care.
- Completeness of data is subject to availability of visit information at each participating institution. Consequently, the preferred variable for determining the date of last contact is [dob_lastalive_int].

### *Time (Days, Months, Years) from Date of Birth to Last Known Alive Date*

[dob_lastalive_int], [dob_lastalive_int_mos], [dob_lastalive_int_yrs]

Value (Numeric)

Description

Interval in days [dob_lastalive_int], months [dob_lastalive_int_mos], and years [dob_lastalive_int_yrs] from date of birth to last known alive date.
- This variable is recommended for censoring in most survival analyses.
- Based on the most recent date that the patient 1) received oncology care at the internal institution, 2) received any care at the internal institution, 3) documentation from any source that the patient is alive, and 4) the patient is enrolled in hospice.

### *Age (Years) at Death Date*

[age_death_yrs]

Value (Numeric)

Description

- Age at death, in years
- Populated only if patient is known to be dead at the time of curation

### *Age (Years) at Last Known Alive Date*

[age_last_fu_yrs]

Value (Numeric)

Description

- Derived data element representing the age of the patient at the time last known to be alive.
- Based on the most recent date that the patient 1) received oncology care at the internal institution, 2) received any care at the internal institution, 3) documentation from any source that the patient is alive, and 4) the patient is enrolled in hospice.

### *Number of Cancers, Any Type*

[n_cancers]

Value (Numeric)

Description

- Number of invasive and non-invasive/in situ cancer diagnoses ever experienced by the patient
- Based on the count of records in Cancer Diagnosis dataset for each record ID

## *Number of BPC Project Cancers (Index Cancers)*

[n_cancers_index]

Value (Numeric)

Description

- Number of BPC Project cancers that were identified for a patient
- The BPC Project cancer is defined as the cancer that met eligibility criteria, underwent genomic sequencing and was submitted to AACR Project GENIE.
- Each patient has at least one BPC Project cancer. Patients may have multiple BPC Project cancers, though this is rare.
- PRISSMM data elements are curated for BPC Project cancers.
- The terms “BPC Project cancer” and “index cancer” are used interchangeably.
- Further details regarding the definition of BPC Project and non-BPC Project cancers can be found in the Appendix: Additional Details Regarding BPC Project and Non-BPC Project Cancers

## *Number of Cancer-Directed Drug Regimens Curated*

[n_regimens_pt]

Value (Numeric)

Description

- The total number of cancer-directed drug regimens, including anti-neoplastic, immunotherapy, and hormone therapy that the patient has ever received for any cancer diagnosis
- Based on the count of records in the cancer-directed drugs dataset for each record ID
- This number includes cancer drug regimens given for non-BPC Project cancers.

## *Number of Imaging Reports*

[n_imaging_reports_pt]

Value (Numeric)

Description

- The total number of imaging scans reviewed starting in the month/year of the BPC Project cancer diagnosis
- Based on the count of records in the imaging dataset for each record ID
- Imaging scans include: CTs, MRIs, PET, PET/CTs, Bone Scans, and Nuclear Medicine scans
- This number includes scans done for non-BPC Project cancers
- For patients with a diagnosis of cancer in the breast, Mammograms are reviewed starting at the month/year of the BPC Project cancer diagnosis

## *Number of CT Scans*

[n_scans_ct_pt]

Value (Numeric)

Description

- The number of CT imaging scans reviewed, starting in the month/year of the BPC Project cancer diagnosis
- Based on the count of CT scans for each record ID
- This number includes CT scans done for non-BPC Project cancers

### *Number of MRIs*

[n_scans_mri_pt]

Value (Numeric)

Description

- The number of MRI imaging scans reviewed starting in the month/year of the BPC Project cancer diagnosis
- Based on the count of MRIs for each record ID
- This number includes MRIs done for non-BPC Project cancers

### *Number of PET or PET-CT Scans*

[n_scans_pet_ct_pt]

Value (Numeric)

Description

- The number of PET or PET-CT imaging scans reviewed starting in the month/year of the BPC Project cancer diagnosis
- Based on the count of PET and PET-CT scans for each record ID
- This number includes PET or PET-CT scans done for non-BPC Project cancers

### *Number of Mammograms (Breast Cancer Only)*

[n_scans_mammog_pt]

Value (Numeric)

Description

- The number of mammograms reviewed starting in the month/year of the BPC Project cancer diagnosis for patients with any (BPC Project or non-BPC Project) diagnosis of breast cancer
- Based on the count of mammograms for each record ID
- This number includes mammograms done for non-BPC Project cancers

### *Number of Bone Scans*

[n_scans_bone_pt]

Value (Numeric)

Description

- The number of bone scans reviewed starting in the month/year of the BPC Project cancer diagnosis
- Based on the count of bone scans for each record ID
- This number includes bone scans done for non-BPC Project cancers

## *Number of Other Scans*

[n_scans_other_pt]

Value (Numeric)

Description

- The number of other imaging scans, including other Nuclear Medicine scans, reviewed starting in the month/year of the BPC Project cancer diagnosis
- Based on the count of other scans for each record ID
- This number includes scans done for non-BPC Project cancers

## *Number of Pathology Reports*

[n_path_reports_pt]

Value (Numeric)

Description

- The number of pathology reports starting in the month/year of the BPC Project cancer diagnosis
- Based on the count of records in the pathology dataset for each record ID
- This number includes pathology reports for non-BPC Project cancers

## *Number of Medical Oncologist Assessments Curated*

[n_md_notes_pt]

Value (Numeric)

Description

- The number of medical oncologist assessments curated starting in the month/year of the BPC Project cancer diagnosis
- Based on the count of records in the medical oncologist assessment dataset for each record ID
- One medical oncologist assessment per month was curated. Curation instructions are provided in Appendix 3.

## *Number of Tumor Marker Results*

[n_tm_pt]

Value (Numeric)

Description

- The number of tumor marker results that were curated
- The Tumor Markers curated include: AFP, BhCG, CA125, CA15-3, CA19-9, CA2729, Calcitonin, CEA, Chromogranin A, LDH, PSA, NSE, Testosterone and Thyroglobulin
- Based on the count of records in the tumor marker dataset for each record ID

### *Number of Tumor Marker CA19-9 Results*

[n_tm_ca19_9_pt]

Value (Numeric)

Description

- The number of tumor marker CA19-9 results that were curated
- Based on the count of CA19-9 results in the tumor marker dataset for each record ID

### *Number of Tumor Marker CEA Results*

[n_tm_cea_pt]

Value (Numeric)

Description

- The number of tumor marker CEA results that were curated
- Based on the count of CEA results in the tumor marker dataset for each record ID

### *Number of Eligible Cancer Panel Tests Curated*

[n_cpt_pt]

Value (Numeric)

Description

- The number of cancer panel tests that met the eligibility criteria and were curated
- Not all sequenced specimens included in the AACR Project GENIE repository appear in this dataset due to eligibility requirements.
- Based on the count of records in the cancer panel test REDCap instrument

# Cancer-Diagnosis Dataset

Two cancer diagnosis datasets are provided.

1. The BPC Project cancer diagnosis dataset contains one record per BPC Project cancer diagnosis, per patient. A BPC Project cancer is the cancer that met the eligibility criteria for the project and was selected at random for PRISSMM phenomic data curation.
   This dataset can be linked to the following datasets:
   - Cancer-directed regimen dataset using variables [record_id] and [ca_seq]
   - Cancer panel test dataset using variables [record_id] and [ca_seq]
   - Patient-level, PRISSMM Pathology, PRISSMM Imaging, and PRISSMM Medical Oncologist Assessment datasets using [record_id].
2. The second cancer diagnosis dataset contains one record per non-BPC Project cancer diagnosis, per patient. This dataset includes two types of cancer diagnoses: 1) non-BPC Project invasive cancer and in situ/non-invasive cancer diagnoses, and 2) other tumors.
   This dataset can be linked to the following datasets:
   - Cancer-directed regimen dataset using variables [record_id] and [ca_seq]
   - Patient-level, PRISSMM Pathology, PRISSMM Imaging, and PRISSMM Medical Oncologist Assessment datasets using [record_id].
   - Cannot be linked to the cancer panel test dataset because non-BPC Project cancer diagnoses were not genomically sequenced (Appendix 1)

Further details regarding the definition of BPC Project and non-BPC Project cancers can be found in Appendix 1.

Field names shaded in gray indicate that an alternative variable is recommended for analysis. The recommended variable is noted in the description.

## *BPC Project Cohort*

[cohort]

Value (Character)

- CRC

Description

- Indicates the BPC Project cancer type

## *Record ID*

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets. Note that AACR Project GENIE refers to participating institutions as centers. The institutions include:
    - DFCI
    - MSK
    - VICC

### *Institution*

[institution]

Value (Character)

- DFCI
- MSK
- VICC

Description

- Indicates the patient’s internal institution of cancer care.

### *Cancer Sequence Identifier*

[ca_seq]

Value (Numeric)

- 0 = first and only cancer
- 1 = first of two or more primaries
- 2 = second of two or more primaries
- …10 = tenth of ten or more primaries

Description

- Sequentially numbers (first, second, third, etc.) cancer diagnoses based on the date of diagnosis
- This variable is missing if the diagnosis associated with the cancer-directed regimen is unknown.
- Primary key for the cancer diagnosis, cancer-directed regimen and cancer panel test datasets.

### *BPC Project (Index) Cancer Indicator*

[redcap_ca_index]

Value (Character)

- Yes
- No

Description

- The BPC Project cancer is defined as the cancer that met eligibility criteria, underwent genomic sequencing and was submitted to AACR Project GENIE.
    - Each patient has at least one BPC Project cancer. Patients may have multiple BPC Project cancers, though this is rare.
    - PRISSMM data elements are curated for BPC Project cancers.
    - The terms "BPC Project cancer" and "index cancer" are used interchangeably.
    - Further details regarding the definition of BPC Project and non-BPC Project cancers can be found in the Appendix: Additional Details Regarding BPC Project and Non-BPC Project Cancers.
- Populated only if diagnosis is eligible for curation (Appendix 1)

## *Time (Days) from Date of Birth to Cancer Diagnosis*

[ca_cadx_int]

Value (Numeric)

Description

- Interval in days between date of birth and curated date of cancer diagnosis

## *Source of Cancer Diagnosis Date*

[ca_dx_how]

Value (Character)

- Pathology Report
- Imaging Report
- Physical Exam
- Other

Description

- Source of the curated diagnosis date
- If [ca_dx_how] = "Pathology Report", the cancer diagnosis date is based upon a review of the pathology report indicating the first histologic confirmation of cancer.
- Populated only if diagnosis is eligible for curation (Appendix 1)

## *Time (Days) from Date of Birth to Next Curated Cancer Diagnosis*

[dob_next_ca_days]

Value (Numeric)

Description

- Time from date of birth to the next curated cancer diagnosis
- Populated only if patient has a subsequent cancer diagnosis

## *Time (days) to First BPC Project Cancer Diagnosis*

[tt_first_index_ca]

Value (Numeric)

Description

- Time to first BPC project cancer diagnosis from other cancer diagnosis
- Time is negative if the BPC project cancer occurred prior to the comparative cancer; time is positive if BPC Project cancer occurred after the comparative cancer
- Populated only if cancer diagnosis is not the first BPC Project cancer. This field is blank for the first BPC Project cancer.

## *Tumor Registry Time (Days) from Date of Birth to First Contact at Institution*

[naaccr_first_contact_int]

Value (Numeric)

Description

- Interval in days from date of birth to first contact at institution based on tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #580

## *Curated Age (Years) at Diagnosis*

[ca_age]

Value (Numeric)

Description

- Curated patient age at diagnosis
- Populated only if diagnosis is eligible for curation (Appendix 1)
- Derived variable [age_dx] incorporates both tumor registry and curated age.

## *Derived Age (Years) at Diagnosis*

[age_dx]

Value (Numeric)

Description

- Patient age at diagnosis
- Derived from curated age, if available. If unavailable, tumor registry age was used.

## *Tumor Registry Primary Cancer Site*

[naaccr_site_cd]

Value (Character)

- ICD-O-3 topography code

Description

- Primary cancer site from tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.
- Derived variable [ca_d_site] incorporates both tumor registry and curated cancer site.

Data Standard: NAACCR #521

## *Curated Primary Cancer Site*

[ca_site]

Value (Character)

- ICD-O-3 topography code

Description

- Curated primary cancer site
- Populated only if cancer diagnosis is curated and cancer type is present ([ca_type]). This field is blank for cancer diagnoses obtained from the tumor registry.
- Derived variable [ca_d_site] incorporates both tumor registry and curated cancer site.

## *Derived Primary Cancer Site*

[ca_d_site]

Value (Character)

- ICD-O-3 topography code

Description

- Derived from curated primary cancer site [ca_site], if available. If unavailable, tumor registry primary cancer site was used [naaccr_site_cd].

## *SEER Recode of Primary Cancer Site*

[site_recode]

Value (Character)

- Bones and Joints
- Brain and Other Nervous System
- Breast
- Colon and Rectum
- Digestive System
- Endocrine System
- Eye and Orbit
- Female Genital System
- Leukemia
- Liver and Bile Duct
- Lymphoma
- Male Genital System
- Oral Cavity and Pharynx
- Respiratory System
- Skin excluding Basal and Squamous
- Soft Tissue including Heart
- Urinary System
- Miscellaneous

Description

- Cancer site grouped by ICD-O-3 topography codes according to the SEER cancer registry

Data Standard: SEER Site Recode ICD-O-3/WHO (2008) Definition

## ***Cancer Type***

[ca_type]

Value (Character)

- Adrenocorical Carcinoma
- Anal Cancer
- Appendix Cancer
- Bile Duct Cancer
- Bladder Cancer
- Brain Cancer
- Breast Cancer
- NET or Carcinoid
- Cervical Cancer
- Colon Cancer
- Colon/Rectum Cancer
- Esophagus Cancer
- Ewing Sarcoma
- Fallopian Tube Cancer
- Gallbladder Cancer
- Germ Cell Tumor
- GIST
- Head and Neck Cancer

- Mesothelioma
- Ill Defined/Cancer of Unknown Primary
- Liver Cancer
- Lung Cancer, NOS
- Melanoma
- Merkel Cell
- Neuroblastoma
- Non-Small Cell Lung Cancer
- Osteosarcoma
- Ovarian Cancer
- Pancreatic Cancer
- Parathyroid Cancer
- Penis Cancer
- Peritoneum Cancer
- Placenta Cancer
- Prostate Cancer
- Rectum and Rectosigmoid Cancer
- Renal Kidney Cancer
- Renal Pelvis Cancer
- Retinoblastoma
- Rhabdomyosarcoma
- Scrotum Cancer
- Small Cell Lung Cancer
- Small Intestine Cancer
- Stomach Cancer
- Testis Cancer
- Thymus Cancer
- Thyroid Cancer
- Uterus Cancer
- Vagina Cancer
- Vulva Cancer
- Wilms Tumor
- Other

Description

- Cancer type was characterized based on information that includes ICD-O-3 topography and morphology codes.
- Mappable to AJCC Collaborative Stage v2.05, AJCC staging v7 and v8, SEER, and NCI.
- Populated only if diagnosis is not a hematopoietic or lymphoid neoplasm or pre-malignancy ([ca_heme_malig] = “No”)

## *Brain cancer type*

[ca_type_brain]

Value (Character)

- Diffuse astrocytic and oligodendroglial tumors
- Other astrocytic tumors
- Astrocytic tumors
- Oligodendroglial tumors
- Mixed gliomas
- Other gliomas
- Ependymal Tumors
- Medulloblastomas
- Pineal Parenchymal Tumors
- Choriod plexus tumors
- Neuronal and mixed neuronal-gilial tumors
- Embryonal tumors
- Meningeal Tumors
- Craniopharyngioma (Grade I)
- Mesenchymal non meningothelial tumors
- Tumors of cranial and paraspinal nerves
- Tumors of the pineal region
- Tumors of the sellar region
- Germ cell tumors
- Other

Description

- Populated only if:
    - Cancer type is recorded as brain cancer ([ca_type] = “Brain Cancer”)
    - Non-BPC Project cancer

### *Cancer type other*

[ca_type_oth]

Value (Character)

- Free-text

Description

- Populated only if:
    - Cancer type is recorded as “other” ([ca_type] = “Other”)
    - Non-BPC Project cancer

### *Hematopoietic or lymphoid neoplasm or pre-malignancy*

[ca_heme_malig]

Value (Character)

- Yes
- No

Description

- Indicates whether cancer diagnosis is a hematologic malignancy
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Non-BPC Project cancer

## *Type of hematopoietic or lymphoid neoplasm*

[ca_heme_type]

Value (Character)

- Leukemia
- Lymphoma
- Langerhans Cell Histiocytosis
- MDS Myelodysplastic Syndrome
- MGUS Monoclonal gammopathy of undetermined significance
- Multiple Myeloma
- Plasmacytoma
- Other hematopoietic or lymphoid neoplasm
- Unspecified

Description

- Populated only if:
  - Hematopoietic or lymphoid neoplasm or pre-malignancy ([ca_heme_malig] = “Yes”).
  - Non-BPC Project cancer

Data Standard: National Cancer Institute Cancer Types: Hematologic/Blood

## *Tumor Registry Histology*

[naaccr_histology_cd]

Value (Numeric)

- ICD-O-3 morphology code

Description

- Histology code from tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #552

## *Histology Category*

[ca_hist_adeno_squamous]

Value (Character)

- Adenocarcinoma
- Squamous cell
- Sarcoma
- Small cell carcinoma
- Other histologies/mixed tumor

Description

- Broad histology group based on [naaccr_histology_cd]
- Populated for BPC Project cancers only.

## *Tumor Registry ICD-O-3 Behavior Code*

[naaccr_behavior_cd]

Value (Numeric)

- ICD-O-3 morphology code

Description

- Behavior code from tumor registry (ICD-O-3)
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #523

## *Tumor Registry Laterality Code*

[naaccr_laterality_cd]

Value (Character)

- 0 = Not a paired site
- 1 = Right: origin of primary
- 2 = Left: origin of primary
- 3 = Only one side involved, right or left origin unspecified
- 4 = Bilateral involvement at time of diagnosis, lateral origin unknown for a single primary; or both ovaries involved simultaneously, single histology; bilateral retinoblastomas; bilateral Wilms' tumors
- 5 = Paired site: midline tumor 9 = Paired site, but no information concerning laterality

Description

- Laterality code from the tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #410

## Tumor Registry Grade

[naaccr_grade]

Value (Numeric)

- 1 = Grade I
- 2 = Grade II
- 3 = Grade III
- 4 = Grade IV
- 5 = T-cell
- 6 = B-cell
- 7 = Null cell
- 8 = NK (natural killer) cell
- 9 = Grade/differentiation unknown, not stated or not applicable

Description

- The grade or degree of differentiation of the tumor at diagnosis from the tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.
- Derived variable [ca_grade] incorporates both tumor registry and curated grade.

Data Standard: NAACCR #440

## Curated Grade or Differentiation of the Tumor

[ca_grade_diff]

Value (Character)

- I
- II
- III
- IV
- Grade differentiation unknown or not stated or not applicable

Description

- The curated grade or degree of differentiation of the tumor
- Populated only if cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
- Derived variable [ca_grade] incorporates both tumor registry and curated grade.

Data Standard: NAACCR #440

## Derived Grade or Differentiation of Tumor

[ca_grade]

Value (Character)

- I
- II
- III
- IV

Description

- Derived from curated grade [ca_grade_diff], if available. If unavailable, tumor registry grade was used [naaccr_grade].

### *Tumor Registry Best Group Stage*

[best_ajcc_stage_cd]

Value (Character)

- Free-text

Description

- Best stage group calculated by each institution’s tumor registry software (METRIQ Tumor Registry Algorithm)
- Populated only if:
  - Cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.
  - BPC Project cancer diagnosis
- Derived variable [stage_dx] incorporates both tumor registry and curated stage.

### *Curated Stage IV at Diagnosis*

[ca_stage_iv]

Value (Character)

- No
- Yes
- Not Applicable
- Unknown

Description

- Indicates whether cancer was diagnosed as stage IV
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry, unless the tumor registry Best Group Stage is not available [best_ajcc_stage_cd] ≠ 88, 99)
  - Diagnosis is not a hematopoietic or lymphoid neoplasm or pre-malignancy ([ca_heme_malig] = “No”).
- Derived variable [stage_dx] incorporates both tumor registry and curated stage.

## *Curated Group Stage at Diagnosis*

[ca_stage]

Value (Character)

- 0
- 0A
- 0is
- I
- IA
- IB
- IC
- II
- IIA
- IIB
- IIC
- III
- IIIA
- IIIB
- IIIC
- IV
- IVA
- IVB
- IVC
- Not Applicable
- Unknown

Description

- Curated stage group at diagnosis
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Cancer diagnosis was not diagnosed at stage IV ([ca_stage_iv] = “No”).
- Derived variable [stage_dx] incorporates both tumor registry and curated stage.

## *Derived Stage at Diagnosis*

[stage_dx]

Value (Character)

- Stage I
- Stage II
- Stage III
- Stage I-III NOS
- Stage IV

Description

- Cancer stage at diagnosis
- Derived from a combination of tumor registry and curated stage variables as below:
  - Tumor registry best group stage [best_ajcc_stage_cd], if available.
  - If tumor registry best group stage was unavailable, curated group stage was used [ca_stage_iv], [ca_stage] to determine whether the patient was stage IV or stage I-III. If the patient was stage I-III or if any of the above variables were missing:
    - If patient received neoadjuvant chemotherapy or radiation therapy before pathological stage diagnosis, then clinical staging was used, if available. If unavailable, then pathologic staging was used.
    - If patient did not receive neoadjuvant chemotherapy or radiation therapy before pathological stage diagnosis, then pathologic staging was used, if available. If unavailable, then clinical staging was used.
- Populated for BPC Project cancers only.

## *Derived Stage IV at Diagnosis*

[stage_dx_iv]

Value (Character)

- Stage 0
- Stage I-III
- Stage IV

Description

- Grouped cancer stage at diagnosis
- Derived from tumor registry best group stage [best_ajcc_stage_cd], if available. If unavailable, curated group stage was used [ca_stage_iv].

## *Tumor Registry TNM Pathologic Stage*

[naaccr_tnm_path_desc]

Value (Character)

- Free-text

Description

- TNM pathology from tumor registry
  - T describes the size of the tumor and any spread of cancer into nearby tissue
  - N describes spread of cancer to nearby lymph nodes
  - M describes distant metastasis
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #920

## Tumor Registry Pathologic T Stage

[naaccr_path_t_cd]

Value (Character)

- Free-text

Description

- Pathologic T stage from tumor registry
- T describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #880

## Curated Pathologic T Stage

[ca_path_t_stage]

Value (Character)

- TX
- T0
- T1
- T2
- T3
- T3
- Not Applicable
- Unknown

Description

- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Cancer was not diagnosed at stage IV ([ca_stage_iv] = “No”).

## Curated Pathologic T1 Stage Detail

[ca_path_t1_det]

Value (Character)

- T1mic
- T1a
- T1a2
- T1b
- T1b1
- T1b2
- T1c
- T1d
- Not Applicable
- Unknown

Description

- Curated pathologic T1 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Pathologic T stage is T1 ([ca_path_t_stage] = “T1”).

Data Standard: NAACCR #880

## Curated Pathologic T2 Stage Detail

[ca_path_t2_det]

Value (Character)

- T2a
- T2a1
- T2a2
- T2b
- T2c
- T2d
- Not Applicable
- Unknown

Description

- Curated pathologic T2 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Pathologic T stage is T2 ([ca_path_t_stage] = “T2”).

Data Standard: NAACCR #880

## Curated Pathologic T3 Stage Detail

[ca_path_t3_det]

Value (Character)

- T3a
- T3b
- T3c
- T3d
- Not Applicable
- Unknown

Description

- Curated pathologic T3 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Pathologic T stage is T3 ([ca_path_t_stage] = “T3”).

Data Standard: NAACCR #880

## *Curated Pathologic T4 Stage Detail*

[ca_path_t4_det]

Value (Character)

- T4a
- T4b
- T4c
- T4e
- Not Applicable
- Unknown

Description

- Curated pathologic T4 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Pathologic T stage is T4 ([ca_path_t_stage] = “T4”).

Data Standard: NAACCR #880

## *Tumor Registry Pathologic N Stage*

[naaccr_path_n_cd]

Value (Character)

- Free-text

Description

- Pathologic N stage from tumor registry
- N describes spread of cancer to nearby lymph nodes
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #890

## ***Curated Pathologic N Stage***

[ca_path_n_stage]

Value (Character)

- NX
- N0
- N1
- N2
- N3
- N4
- Not Applicable
- Unknown

Description

- Curated pathologic N stage
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Cancers was not diagnosed at stage IV ([ca_stage_iv] = “No”).

## ***Curated Pathologic N1 Stage Detail***

[ca_path_n1_det]

Value (Character)

- N1mi
- N1a
- N1b
- N1c
- Not Applicable
- Unknown

Description

- Curated pathologic N1 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Pathologic N stage is N1 ([ca_path_n_stage] = “N1”).

Data Standard: NAACCR #890

### *Curated Pathologic N2 Stage Detail*

[ca_path_n2_det]

Value (Character)

- N2a
- N2b
- N2c
- Not Applicable
- Unknown

Description

- Curated pathologic N2 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Pathologic N stage is N2 ([ca_path_n_stage] = “N2”).

Data Standard: NAACCR #890

### *Curated Pathologic N3 Stage Detail*

[ca_path_n3_det]

Value (Character)

- N3a
- N3b
- N3c
- Not Applicable
- Unknown

Description

- Curated pathologic N3 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Pathologic N stage is N3 ([ca_path_n_stage] = “N3”).

Data Standard: NAACCR #890

### Tumor Registry Pathologic M Stage

[naaccr_path_m_cd]

Value (Character)

- Free-text

Description

- Pathologic M stage from tumor registry
- M describes metastasis
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #900

### Tumor Registry Pathologic Group Stage

[naaccr_path_stage_cd]

Value (Character)

- Free-text

Description

- Pathologic group stage from tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #910

### Curated Pathologic Group Stage

[ca_path_group_stage]

Value (Character)

- 0
- 0A
- 0is
- I
- IA
- IA1
- IA2
- IB
- IB1
- IB2

- IC
- IS
- II
- IIA
- IIA1
- IIA2
- IIB
- IIC
- III
- IIIA
- IIIB
- IIIC
- IIIC1
- IIIC2
- IV
- IVA
- IVA1
- IVA2
- IVB
- IVC
- Occult
- Not Applicable
- Unknown

Description

- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Cancer was not diagnosed at stage IV ([ca_stage_iv] = "No").

### *Tumor Registry Clinical T Stage*

[naaccr_clin_t_cd]

Value (Character)

- Free-text

Description

- Clinical T stage from tumor registry
- T describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #940

## *Curated Clinical T Stage*

[ca_clin_t_stage]

Value (Character)

- TX
- T0
- T1
- T2
- T3
- T4
- Not Applicable
- Unknown

Description

- Curated clinical T stage
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - If cancer was not diagnosed at stage IV ([ca_stage_iv] = “No”).

## *Curated Clinical T1 Stage Detail*

[ca_clin_t1_det]

Value (Character)

- T1mic
- T1a
- T1a2
- T1b
- T1b1
- T1b2
- T1c
- T1d
- Not Applicable
- Unknown

Description

- Curated clinical T1 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Curated clinical T stage is T1 ([ca_clin_t_stage] = “T1”).

Data Standard: NAACCR #940

## *Curated Clinical T2 Stage Detail*

[ca_clin_t2_det]

Value (Character)

- T2a
- T2a1
- T2a2
- T2b
- T2c
- T2d
- Not Applicable
- Unknown

Description

- Curated clinical T2 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Curated clinical T stage is T2 ([ca_clin_t_stage] = “T2”).

Data Standard: NAACCR #940

## *Curated Clinical T3 Stage Detail*

[ca_clin_t3_det]

Value (Character)

- T3a
- T3b
- T3c
- T3d
- Not Applicable
- Unknown

Description

- Curated clinical T3 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Curated clinical T stage is T3 ([ca_clin_t_stage] = “T3”).

Data Standard: NAACCR #940

## *Curated Clinical T4 Stage Detail*

[ca_clin_t4_det]

Value (Character)

- T4a
- T4b
- T4c
- T4e
- Not Applicable
- Unknown

Description

- Curated clinical T4 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Curated clinical T stage is T4 ([ca_clin_t_stage] = “T4”).

Data Standard: NAACCR #940

## *Tumor registry clinical N*

[naaccr_clin_n_cd]

Value (Character)

- Free-text

Description

## *Curated Clinical N Stage*

[ca_clin_n_stage]

Value (Character)

- NX
- N0
- N1
- N2
- N3
- N4
- Not Applicable
- Unknown

Description

- Curated clinical N stage
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Cancer was not diagnosed at stage IV ([ca_stage_iv] = “No”).

### ***Curated Clinical N1 Stage Detail***

[ca_clin_n1_det]

Value (Character)

- N1a
- N1b
- N1c
- Not Applicable
- Unknown

Description

- Curated clinical N1 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Curated clinical N stage is N1 ([ca_clin_n_stage] = “N1”).

Data Standard: NAACCR #950

### ***Curated Clinical N2 Stage Detail***

[ca_clin_n2_det]

Value (Character)

- N2a
- N2b
- N2c
- Not Applicable
- Unknown

Description

- Curated clinical N2 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Curated clinical N stage is N2 ([ca_clin_n_stage] = “N2”).

Data Standard: NAACCR #950

### *Curated Clinical N3 Stage Detail*

[ca_clinical_n3_det]

Value (Character)

- N3a
- N3b
- N3c
- Not Applicable
- Unknown

Description

- Curated clinical N3 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Curated clinical N stage is N3 ([ca_clin_n_stage] = “N3”).

Data Standard: NAACCR #950

### *Tumor Registry Clinical M Stage*

[naaccr_clin_m_cd]

Value (Character)

- Free-text

Description

- Clinical M stage from tumor registry
- M describes metastasis
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #960

### *Tumor Registry Clinical Group Stage*

[naaccr_clin_stage_cd]

Value (Character)

- Free-text

Description

- Clinical group stage from tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #970

### *Tumor Registry General Summary Stage*

[naaccr_seer_sum_stage]

Value (Character)

- Free-text

Description

- General summary stage from tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #959

### *Neoadjuvant Chemotherapy or Radiation Therapy Before Pathologic Stage Diagnosis*

[ca_tx_pre_path_stage]

Value (Character)

- Yes
- No
- Not Applicable Unknown

Description

- Indicates whether patient received neoadjuvant chemotherapy or radiation therapy before pathologic stage diagnosis
- Populated only if cancer was not diagnosed at stage IV ([ca_stage_iv] = “No”)

### *Curated General Summary Stage*

[ca_gen_sum_stage_2]

Value (Character)

- Unstaged
- In situ
- Localized
- Regional direct extension only
- Regional lymph nodes only
- Regional direct extension and regional lymph nodes
- Regional NOS
- Distant
- Not applicable

Description

- Summary stage includes all information through completion of surgery in the first course of treatment or within 4 months of diagnosis in the absence of disease progression, whichever is longer.
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Cancer type is Neuroblastoma, Brain Cancer, Ewing Sarcoma, Retinoblastoma, Rhabdomyosarcoma, or Wilms Tumor ([ca_type] = “Neuroblastoma”, “Brain Cancer”, “Ewing Sarcoma”, “Retinoblastoma”, “Rhabdomyosarcoma”, or “Wilms Tumor”).

### *Tumor Registry TNM Edition Number*

[naaccr_tnm_edition_num]

Value (Character)

- Free-text

Description

- TNM edition number from tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #1060

### *Sites of Distant Metastasis at the Time of Cancer Diagnosis (Stage IV Patients)*

[ca_dmets_yn]

Value (Character)

- Yes
- No – patient is stage IV with no distant metastases
- NA

Description

- Indicates whether stage IV patient had distant metastases at time of cancer diagnosis
- Populated only if cancer diagnosed at stage IV ([best_ajcc_stage_cd] = “4”, “4A” “4B” or [ca_stage_iv] = “Yes”)

### *Sites of Distant Metastases at Cancer Diagnosis*

[ca_first_dmets1]-[ca_first_dmets10]

Value (Character)

- ICD-O-3 topography code

Description

- Site of distant metastases
- Up to 10 sites of distant metastasis are recorded
- Populated only if distant metastases are present at time of diagnosis ([ca_dmets_yn] = "Yes")

### Distant Metastasis at Diagnosis: Adrenal

[dmets_dx_adrenal]

Value (Character)

- Yes
- No

Description

- Indicates whether a distant adrenal metastasis was identified at diagnosis
- Identified by ICD-O-3 topography codes C74.0, C74.1, C74.9 in sites of distant metastases at cancer diagnosis ([ca_first_dmets1]-[ca_first_dmets10])
- Populated only if:
    - Distant metastases are present at time of diagnosis ([ca_dmets_yn] = "Yes")
    - BPC Project cancer

### Distant Metastasis at Diagnosis: Bone

[dmets_dx_bone]

Value (Character)

- Yes
- No

Description

- Indicates whether a distant bone metastasis was identified at diagnosis
- Identified by ICD-O-3 topography codes C40.0, C40.1, C40.2, C40.3, C40.8, C40.9, C41.0, C41.1, C41.2, C41.3, C41.4, C41.8, C41.9 in sites of distant metastases at cancer diagnosis ([ca_first_dmets1]-[ca_first_dmets10])
- Populated only if:
    - Distant metastases are present at time of diagnosis ([ca_dmets_yn] = "Yes")
    - BPC Project cancer

### Distant Metastasis at Diagnosis: Brain

[dmets_dx_brain]

Value (Character)

- Yes
- No

Description

- Indicates whether a distant brain metastasis was identified at diagnosis
- Identified by ICD-O-3 topography codes C71.0, C71.1, C71.2, C71.3, C71.4, C71.5, C71.6, C71.7, C71.8, C71.9 in sites of distant metastases at cancer diagnosis ([ca_first_dmets1]-[ca_first_dmets10])
- Populated only if:
    - Distant metastases are present at time of diagnosis ([ca_dmets_yn] = “Yes”)
    - BPC Project cancer

### *Distant Metastasis at Diagnosis: Liver*

[dmets_dx_liver]

Value (Character)

- Yes
- No

Description

- Indicates whether a distant liver metastasis was identified at diagnosis
- Identified by ICD-O-3 topography codes C22.0, C22.1 in sites of distant metastases at cancer diagnosis ([ca_first_dmets1]-[ca_first_dmets10])
- Populated only if:
    - Distant metastases are present at time of diagnosis ([ca_dmets_yn] = “Yes”)
    - BPC Project cancer

### *Distant Metastasis at Diagnosis: Lung*

[dmets_dx_lung]

Value (Character)

- Yes
- No

Description

- Indicates whether a distant lung metastasis was identified at diagnosis
- Identified by ICD-O-3 topography codes C34.0, C34.1, C34.2, C34.3, C34.8, C34.9 in sites of distant metastases at cancer diagnosis ([ca_first_dmets1]-[ca_first_dmets10])
- Populated only if:
    - Distant metastases are present at time of diagnosis ([ca_dmets_yn] = “Yes”)
    - BPC Project cancer

### *Distant Metastasis at Diagnosis: Lymph Nodes*

[dmets_dx_lymph]

Value (Character)

- Yes
- No

Description

- Indicates whether a distant lymph node metastasis was identified at diagnosis
- Identified by ICD-O-3 topography codes C77.0, C77.1, C77.2, C77.3, C77.4, C77.5, C77.8, C77.9 in sites of distant metastases at cancer diagnosis ([ca_first_dmets1]-[ca_first_dmets10])
- Populated only if:
    - Distant metastases are present at time of diagnosis ([ca_dmets_yn] = "Yes")
    - BPC Project cancer

### *Distant Metastasis at Diagnosis: Pleura*

[dmets_dx_pleura]

Value (Character)

- Yes
- No

Description

- Indicates whether a distant pleura metastasis was identified at diagnosis
- Identified by ICD-O-3 topography codes C38.4, F30 in sites of distant metastases at cancer diagnosis ([ca_first_dmets1]-[ca_first_dmets10])
- Populated only if:
    - Distant metastases are present at time of diagnosis ([ca_dmets_yn] = "Yes")
    - BPC Project cancer

### *Distant Metastasis at Diagnosis: Subcutaneous Tissue*

[dmets_dx_subc_tissue]

Value (Character)

- Yes
- No

Description

- Indicates whether a distant subcutaneous tissue metastasis was identified at diagnosis
- Identified by ICD-O-3 topography codes C49.0, C49.2, C49.3, C49.4 in sites of distant metastases at cancer diagnosis ([ca_first_dmets1]-[ca_first_dmets10])
- Populated only if:
    - Distant metastases are present at time of diagnosis ([ca_dmets_yn] = "Yes")
    - BPC Project cancer

### *Distant Metastasis at Diagnosis: Other*

[dmets_dx_other]

Value (Character)

- Yes
- No

Description

- Indicates whether another distant metastasis was identified at diagnosis
- Identified by any other ICD-O-3 topography code for distant metastasis not meeting above criteria in sites of distant metastases at cancer diagnosis (variables with prefix [dmets_dx])
- Populated only if:
    - Distant metastases are present at time of diagnosis ([ca_dmets_yn] = "Yes")
    - BPC project cancer

## *Colorectal Cancer: Tumor Deposits*

[ca_crc_td]

Value (Character)

- 0 = surgical resection of the primary site is performed, the pathology report is available for review, and tumor deposits are not mentioned
- 1-99
- 100 or more
- Tumor Deposits identified, number unknown
- Not applicable: Information not collected for this case
- Not documented in medical record

Description

- Count of colorectal cancer tumor deposits
- Populated for colon/rectum cancer diagnoses only; for other diagnoses, this variable is blank.

Data Standard: NAACCR #3934

## *Colorectal Cancer: Circumferential Radial Margin*

[ca_crc_crm]

Value (Character)

- 0-99
- Margins cannot be assessed
- Described as at least 1 mm
- Described as at least 2 mm
- Described as at least 3 mm
- Described as greater than 3 mm
- No resection of primary site
- 100 mm or greater
- Margins clear distance from tumor not stated
- Not applicable
- Not documented in medical record

Description

- Colorectal cancer circumferential radial margin
- Populated for colon/rectum cancer diagnoses only; for other diagnoses, this variable is blank.

Data Standard: NAACCR #2930

### *Colorectal Cancer: Perineural Invasion*

[ca_crc_peri_inv]

Value (Character)

- None
- Perineural invasion present
- Not applicable
- No histologic examination
- Unknown not documented

Description

- Colorectal cancer perineural invasion
- Populated for colon/rectum cancer diagnoses only; for other diagnoses, this variable is blank.

Data Standard: NAACCR #2862

### *Number of Regimens Associated with the Cancer Diagnosis*

[ca_n_regimens]

Value (Numeric)

Description

- Count of cancer-directed regimens that were associated with the cancer diagnosis

### *Time (Days, Months, Years) from Cancer Diagnosis to Last Known Alive Date*

[dx_lastalive_int], [dx_last_alive_int_mos], [dx_last_alive_int_yrs]

Value (Numeric)

Description

- Interval in days [dx_lastalive_int]; months [dx_last_alive_int_mos]; or years [dx_last_alive_int_yrs] from cancer diagnosis to last known alive date

### *Time (Days, Months, Years) from Cancer Diagnosis to Death*

[dx_death_int], [dx_death_int_mos], [dx_death_int_yrs]

Value (Numeric)

Description

- Interval in days [dx_death_int]; months [dx_death_int_mos]; or years [dx_death_int_yrs] from cancer diagnosis to death

### *Overall Survival from Diagnosis (Days, Months, Years)*

[tt_os_dx_days], [tt_os_dx_mos], [tt_os_dx_yrs]

Value (Numeric)

Description

- Time from diagnosis to death or the date last known alive
- Interval in days [tt_os_dx_days]; months [tt_os_dx_mos]; or years [tt_os_dx_yrs] from cancer diagnosis to last known alive date or death

### *Overall Survival from Diagnosis: Status Indicator*

[os_status_dx]

Value (Numeric)

- 1 = Dead
- 0 = Censored

Description

- An event is defined by death
- Patients were censored if not known to be dead

### *Progression Free Survival-Imaging (PFS-I) from Diagnosis: Status Indicator*

[pfs_i_dx_status]

Value (Numeric)

- 1 = Progression according to radiologist assessment; death
- 0 = Censored

Description

- An event is defined by:
  - Radiologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
  - Death
- Patients were censored at the time of next cancer diagnosis or end of follow up
  - In the rare instance that there is a second cancer diagnosis on the same date, the second cancer diagnosis is not used to censor PFS-I
- Populated only for BPC Project cancers with stage IV disease at diagnosis.
  - For patients with multiple stage IV BPC Project cancers diagnosed on the same date, PFS-I is calculated for all stage IV BPC Project cancers. However, only one record per patient should be included in time-to-event analyses of PFS-I

### ***Progression Free Survival-Imaging (PFS-I) from diagnosis***

[tt_pfs_i_dx_days], [tt_pfs_i_dx_mos], [tt_pfs_i_dx_yrs]

Value (Numeric)

Description

- Time from diagnosis to radiologist impression of progressing/worsening/enlarging cancer status, subsequent cancer diagnosis, last known alive date, or death
- Interval in days [tt_pfs_i_dx_days]; months [tt_pfs_i_dx_mos]; or years [tt_pfs_i_dx_yrs] from diagnosis to:
  - Radiologist assessment of progression
  - Death
  - Subsequent cancer diagnosis or last known alive date, if patient is censored according to PFS-I status indicator [pfs_i_dx_status]
- Populated only for BPC Project cancers with stage IV disease at diagnosis.
  - For patients with multiple stage IV BPC Project cancers diagnosed on the same date, PFS-I is calculated for all stage IV BPC Project cancers. However, only one record per patient should be included in time-to-event analyses of PFS-I.

### ***Progression Free Survival-Imaging (PFS-I) from Diagnosis: Status Indicator (Including Radiologist Impression of Mixed Cancer Status)***

[pfs_i_dx_mixed_status]

Value (Numeric)

- 1 = Progression according to radiologist assessment; death
- 0 = Censored

Description

- An event is defined by:
  - Radiologist assessment of the change in the patient's cancer status as either mixed (i.e. some cancer sites progressing and some cancer sites improving) or progressing/worsening/enlarging
  - Death
- Patients were censored at the time of next cancer diagnosis or end of follow up
  - In the rare instance that there is a second cancer diagnosis on the same date, the second cancer diagnosis is not used to censor PFS-I
- Populated only for BPC Project cancers with stage IV disease at diagnosis.
  - For patients with multiple stage IV BPC Project cancers diagnosed on the same date, PFS-I is calculated for all stage IV BPC Project cancers. However, only one record per patient should be included in time-to-event analyses of PFS-I.

## *Progression Free Survival-Imaging (PFS-I) from Diagnosis (Including Radiologist Impression of Mixed Cancer Status)*

[tt_pfs_i_dx_mixed_days], [tt_pfs_i_dx_mixed_mos], [tt_pfs_i_dx_mixed_yrs]

Value (Numeric)

Description

- Time from diagnosis to radiologist impression of progressing/worsening/enlarging/mixed cancer status, subsequent cancer diagnosis, last known alive date, or death
- Interval in days [tt_pfs_i_dx_mixed_days]; months [tt_pfs_i_dx_mixed_mos]; or years [tt_pfs_i_dx_mixed_yrs] from diagnosis to:
  - Radiologist assessment of mixed or progressing disease
  - Death
  - Subsequent cancer diagnosis or last known alive date, if patient is censored according to PSF-I status indicator [pfs_i_dx_mixed_status]
- Populated only for BPC Project cancers with stage IV disease at diagnosis.
  - For patients with multiple stage IV BPC Project cancers diagnosed on the same date, PFS-I is calculated for all stage IV BPC Project cancers. However, only one record per patient should be included in time-to-event analyses of PFS-I.

## *Progression Free Survival - Medical Oncologist Assessment (PFS-M) from Diagnosis: Status Indicator*

[pfs_m_dx_status]

Value (Numeric)

- 1 = Progression according to medical oncologist assessment; death
- 0 = Censored

Description

- An event is defined by:
  - Medical oncologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
  - Death
- Patients were censored at the time of next cancer diagnosis or end of follow up
  - In the rare instance that there is a second cancer diagnosis on the same date, the second cancer diagnosis is not used to censor PFS-M
- Populated only for BPC Project cancers with stage IV disease at diagnosis.
  - For patients with multiple stage IV BPC Project cancers diagnosed on the same date, PFS-M is calculated for all stage IV BPC Project cancers. However, only one record per patient should be included in time-to-event analyses of PFS-M.

## *Progression Free Survival - Medical Oncologist Assessment (PFS-M) from Diagnosis*

[tt_pfs_m_dx_days], [tt_pfs_m_dx_mos], [tt_pfs_m_dx_yrs]

Value (Numeric)

Description

- Time from diagnosis to medical oncologist assessment of progressing/worsening/enlarging cancer status, subsequent cancer diagnosis, last known alive date or death
- Interval in days [tt_pfs_m_dx_days]; months [tt_pfs_m_dx_mos]; or years [tt_pfs_m_dx_yrs] from diagnosis to:
    - Medical oncologist assessment of progressing disease
    - Death
    - Subsequent cancer diagnosis or last known alive date, if patient is censored according to PSF-M status indicator [pfs_m_dx _status]
- Populated only for BPC Project cancers with stage IV disease at diagnosis.
    - For patients with multiple stage IV BPC Project cancers diagnosed on the same date, PFS-M is calculated for all stage IV BPC Project cancers. However, only one record per patient should be included in time-to-event analyses of PFS-M.

## *Progression Free Survival - Medical Oncologist Assessment (PFS-M) Status Indicator (Including Medical Oncologist Assessment of Mixed Cancer Status)*

[pfs_m_dx_mixed_status]

Value (Numeric)

- 1 = Progression according to medical oncologist assessment; death
- 0 = Censored

Description

- An event is defined by:
    - Medical oncologist assessment of the change in the patient's cancer status is either mixed (with some cancer sites progressing and some cancer sites improving) or progressing/worsening/enlarging
    - Death
- Patients were censored at the time of next cancer diagnosis or end of follow up
    - In the rare instance that there is a second cancer diagnosis on the same date, the second cancer diagnosis is not used to censor PFS-M
- Populated only for BPC Project cancers with stage IV disease at diagnosis.
    - For patients with multiple stage IV BPC Project cancers diagnosed on the same date, PFS-M is calculated for all stage IV BPC Project cancers. However, only one record per patient should be included in time-to-event analyses of PFS-M.

## *Progression Free Survival - Medical Oncologist Assessment (PFS-M) from Diagnosis (Including Medical Oncologist Assessment of Mixed Cancer Status)*

[tt_pfs_m_dx_mixed_days], [tt_pfs_m_dx_mixed_mos], [tt_pfs_m_dx_mixed_yrs]

Value (Numeric)

Description

- Time from diagnosis to medical oncologist assessment of progressing/worsening/enlarging/mixed cancer status, subsequent cancer diagnosis, last known alive date or death
- Interval in days [tt_pfs_m_dx_mixed_days]; months [tt_pfs_m_dx_mixed_mos]; or years [tt_pfs_m_dx_mixed_yrs] from diagnosis to:
    - Medical oncologist assessment of mixed or progressing disease
    - Death
    - Subsequent cancer diagnosis or last known alive date, if patient is censored according to PSF-M status indicator [pfs_m_dx_mixed_status]
- Populated only for BPC Project cancers with stage IV disease at diagnosis.
    - For patients with multiple stage IV BPC Project cancers diagnosed on the same date, PFS-M is calculated for all stage IV BPC Project cancers. However, only one record per patient should be included in time-to-event analyses of PFS-M.

## Cancer-Directed Regimen Dataset

The cancer-directed regimen dataset is structured as one record per regimen-associated cancer diagnosis, per patient. For example, if a regimen is associated with a single cancer diagnosis, there will be one corresponding record in this dataset. If a regimen is associated with two cancer diagnoses, then there will be two corresponding records in this dataset: one for the first associated cancer diagnosis and another for the second associated cancer diagnosis, etc.

Cancer-directed regimens were curated for all cancer diagnoses, including both BPC Project and non-BPC Project cancers. A regimen can consist of one drug or up to five drugs given together. Cancer-directed drugs include anti-neoplastic drugs, immunotherapies, targeted therapies, and hormone therapies. A break in treatment of ≥8 weeks was used to indicate the end of a regimen; even if all drugs in the regimen were re-initiated 8+ weeks later, this was considered a new regimen.

If the cancer-directed drug was part of an investigational drug trial, the drug name(s) [drugs_drug_1-drugs_drug_5] will be set to “Investigational Drug” and the end date interval will match the start date interval. Identification of investigational drugs varies by institution depending on contractual obligations. For some institutions, all drugs that are part of an investigational trial are required to be masked, even if standard of care. For other institutions, only the investigational drug(s) are required to be masked.

This dataset can be linked to the following datasets:

- BPC Project and non-BPC Project cancer diagnosis datasets using the variables [record_id] and [ca_seq]
- Cancer panel test dataset using the variables [record_id] and [ca_seq]
- Patient-level, PRISSMM Pathology, PRISSMM Imaging, and PRISSMM Medical Oncologist Assessment datasets using [record_id]. Field names shaded in gray indicate that an alternative variable is recommended for analysis. The recommended variable is noted in the description.

### *BPC Project Cohort*

[cohort]

Value (Character)

- CRC

Description

- Indicates the BPC Project cancer type

### *Record ID*

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets. Note that AACR Project GENIE refers to participating institutions as centers. The institutions include:
  - DFCI
  - MSK
  - VICC

### *Institution*

[institution]

Value (Character)

- DFCI
- MSK
- VICC

Description

- Indicates the patient’s internal institution of cancer care.

### *Regimen Number*

[regimen_number]

Value (Numeric)

- 1 = first regimen
- 2 = second regimen
- …n = nth regimen

Description

- Order for cancer-directed regimens based on the start date of the first cancer-directed drug
- The cancer-directed regimen dataset is structured as one row per regimen-associated cancer diagnosis, per patient. For example, if a regimen is associated with a single cancer diagnosis, there will be one corresponding record in this dataset. If a regimen is associated with two cancer diagnoses, then there will be two corresponding records in this dataset: one for the first associated cancer diagnosis and another for the second associated cancer diagnosis.
  - This implies that multiple records will have the same regimen number in this dataset when they are associated with a different cancer diagnosis.

### *Cancer Sequence Associated with Regimen*

[ca_seq]

Value (Numeric)

- 0 = first and only cancer
- 1 = first of two or more primaries
- 2 = second of two or more primaries
- …10 = tenth of ten or more primaries

Description

- The associated cancer sequence number for this regimen.
- When regimens are associated with multiple cancer diagnoses, each associated cancer diagnosis will be a separate row in the dataset.
- Populated only if it is known which diagnosis is associated with this cancer-directed regimen.
- Primary key for the cancer diagnosis, cancer-directed regimen and cancer panel test datasets.

## *BPC Project (Index) Cancer Indicator*

[redcap_ca_index]

Value (Character)

- Yes
- No

Description

- The BPC Project cancer is defined as the cancer that met eligibility criteria, underwent genomic sequencing and was submitted to AACR Project GENIE.
  - Each patient has at least one BPC Project cancer. Patients may have multiple BPC Project cancers, though this is rare.
  - PRISSMM data elements are curated for BPC Project cancers.
  - The terms “BPC Project cancer” and “index cancer” are used interchangeably.
  - Further details regarding the definition of BPC Project and non-BPC Project cancers can be found in the Appendix: Additional Details Regarding BPC Project and Non-BPC Project Cancers.
- Populated only if diagnosis is eligible for curation (Appendix 1)

## *Number of Cancer-Directed Drugs in a Regimen*

[drugs_num]

Value (Numeric)

- 1-5

Description

- Number of cancer-directed drugs in a regimen; up to 5 recorded

## *Treatment Route for Non-Standard Drug Administrations*

[drugs_admin]

Value (Character)

- Intraperitoneal (direct into the abdominal cavity)
- Intrapleural (direct into the pleural space around the lung)
- Intralesional (direct into a lesion)
- Intrathecal (by spinal catheter)
- Intravesical (bladder installation)
- Other

Description

- The route of administration for non-standard drug administrations
- Populated only if the drugs were not administered orally, intravenously or by infusion

## *Institution Which Administered/Ordered Regimen*

[drugs_inst]

Value (Character)

- At the internal/native institution only
- Split across both the internal and external institution
- At external institution only

Description

- Location where regimen was administered/ordered
- Indicates whether regimen was administered/ordered at the internal (i.e. the same institution as the [institution] variable) or external institution.
  - Details regarding external institutions are not curated.

## *Institution Which First Ordered the Regimen*

[drugs_firstinst]

Value (Character)

- Internal institution
- External institution

Description

- Location where regimen was first ordered
- Indicates whether regimen was administered/ordered at the internal (i.e. the same institution as the [institution] variable) or external institution.
  - Details regarding external institutions are not curated.
- Populated only if the regimen was administered/ordered across internal and external institutions ([drugs_inst] = “Split across internal and external institution”)

## *Regimen Was Part of a Clinical Trial*

[drugs_ct_yn]

Value (Character)

- Yes
- No

Description

- Indicator for whether cancer-directed drug was part of a clinical trial
- If the cancer-directed drug was part of an investigational drug trial, the drug name(s) [drugs_drug_1-drugs_drug_5] will be set to "Investigational drug" and the end date interval will match the start date interval.
  - Identification of investigational drugs varies by institution depending on contractual obligations. For some institutions, all drugs that are part of an investigational trial are required to be masked, even if standard of care. For other institutions, only the investigational drug(s) are required to be masked.

### *Regimen Discontinuation Status*

[drugs_dc_ynu]

Value (Character)

- Yes
- No
- Unknown

Description

- Indicates whether the drug regimen was discontinued. If not discontinued, the patient was still receiving cancer-directed regimen at time of curation, or it is unknown whether the regimen had ended.
- The response to this variable specifies whether there will be an end date [drugs_enddt_int_drug_1 – drugs_enddt_int_drug_5] or a last day of administration [drugs_lastdt_int_drug_1 – drugs-lastdt_int_drug_5] because the drug therapy is ongoing at curation.

### *Names of Drugs in Regimen*

[regimen_drugs]

Value (Character)

Description

- Names of cancer-directed drugs received together
- Concatenation of variables [drugs_drug_1]-[drugs_drug_5]

### *Name of Cancer-Directed Drug in Regimen, Drugs 1-5*

[drugs_drug_1] – [drugs_drug_5]

Value (Character)

- Name of cancer-directed drug 1 through 5 in each regimen (Appendix 2)

Description

- The cancer drug label contains the generic/ingredient name with the synonyms in parentheses (e.g. Nivolumab (BMS936558, MDX1106, NIVO, ONO4538, Opdivo))
- If the cancer-directed drug was part of an investigational drug trial, the drug name(s) [drugs_drug_1-drugs_drug_5] will be set to "Investigational Drug" and the end date interval will match the start date interval.
    - Identification of investigational drugs varies by institution depending on contractual obligations. For some institutions, all drugs that are part of an investigational trial are required to be masked, even if standard of care. For other institutions, only the investigational drug(s) are required to be masked.

Data Standard: National Cancer Institute Thesaurus: Antineoplastic Agents

## *Name of Other Cancer-Directed Drug in Regimen, Drugs 1-5*

[drugs_drug_oth_1] – [drugs_drug_oth_5]

Value (Character)

- Free-text

Description

- Name of cancer-directed drug 1 through 5 in each regimen
- Populated only if the drug name is not available on the National Cancer Institute Thesaurus: List of Antineoplastic Drugs ([drugs_drug_1] – [drugs_drug_5] = "Other NOS", "Other antineoplastic", "Other hormone" or "Clinical Trial Drug not specified")

## *Time (Days) from Date of Birth to Start of Cancer-Directed Drug in Regimen, Drugs 1-5*

[drugs_startdt_int_1] – [drugs_startdt_int_5]

Value (Numeric)

Description

- Interval in days from date of birth to start of cancer-directed drug 1 – drug 5

## *Time (Days) from Date of Birth to End of Cancer-Directed Drug in Regimen, Drugs 1-5*

[drugs_enddt_int_1] – [drugs_enddt_int_5]

Value (Numeric)

Description

- Interval in days from date of birth to end of cancer-directed drug 1 – drug 5
- For investigational drugs, the end date interval is set to the start date interval.

### *Time (Days) from Date of Birth to Last Known Date of Administration/Order of Cancer-Directed Drugs*

[drugs_lastdt_int_1] – [drugs-lastdt_int_5]

Value (Numeric)

Description

- Interval in days from date of birth to last known administration of cancer-directed drug 1 – drug 5
- If the drug is an intravenously administered drug, the last known date recorded is the last known date of administration of the drug.
- If the drug is an orally administered drug, the last known date recorded is the last known date that the drug was prescribed.
- Populated only if drug discontinuation status is unknown ([drugs_dc_ynu] = "Unknown, no documentation found")

### *Time (Days, Months) from Cancer Diagnosis to Start of Cancer-Directed Drug in Regimen, Drugs 1-5*

[dx_drug_start_int_1] – [dx_drug_start_int_5], [dx_drug_start_int_mos_1] – [dx_drug_start_int_mos_5]

Value (Numeric)

Description

- Interval in days [dx_drug_start_int_1] – [dx_drug_start_int_5] or months [dx_drug_start_int_mos_1] – [dx_drug_start_int_mos_5] from cancer diagnosis to start of cancer-directed drug 1 – drug 5

### *Duration (Days) of Cancer-Directed Drug in Regimen, Drugs 1-5*

[drug_start_end_int_1] – [drug_start_end_int_5]

Value (Numeric)

Description

- Interval in days from start of drug ([drugs_startdt_int_1] – [drugs_startdt_int_5]) to end of drug [drugs_enddt_int_1] – [drugs_enddt_int_5]
- Drugs starting and ending on the same day have a duration of 1 day
- Populated only if:
  - Regimen is known to be discontinued ([drugs_dc_ynu] = "Yes")
  - Drug is not an investigational drug

### *Duration (Days) from Start of Cancer-Directed Drug to End Date or Last Known Administration Date of Drug, Drugs 1-5*

[drug_start_end_or_lastadm_int_1] - [drug_start_end_or_lastadm_int_5]

Value (Numeric)

Description

- Interval in days from start of drug to
  - End date, if regimen was discontinued
  - Last known administration date, if regimen was not known to be discontinued at time of curation
- Populated only if drug is not an investigational drug

### *Time from Associated Cancer Diagnosis to Start of Cancer-Directed Drug Regimen*

[dx_reg_start_int], [dx_reg_start_int_mos], [dx_reg_start_int_yrs]

Value (Numeric)

Description

- Interval in days [dx_reg_start_int]; months [dx_reg_start_int_mos]; or years [dx_reg_start_int_yrs] from associated cancer diagnosis to the start of the first drug in cancer-directed drug regimen

### *Time (Days, Months, Years) from Start of Cancer-Directed Drug Regimen to End of First Drug Discontinued in Regimen*

[reg_start_end_any_int], [reg_start_end_any_int_mos], [reg_start_end_any_int_yrs]

Value (Numeric)

Description

- Interval in days [reg_start_end_any_int]; months [reg_start_end_any_int_mos]; or years [reg_start_end_any_int_yrs] from associated cancer diagnosis to the end of first drug discontinued in cancer-directed drug regimen
- Populated only if:
  - Regimen is known to be discontinued ([drugs_dc_ynu] = “Yes”)
  - Regimen does not contain an investigational drug

### *Time (Days, Months, Years) from Start of Cancer-Directed Drug Regimen to End Date or Last Known Administration Date of First Drug Discontinued in Regimen*

[reg_start_end_or_lastadm_any_int], [reg_start_end_or_lastadm_any_mos], [reg_start_end_or_lastadm_any_yrs]

Value (Numeric)

Description

- Interval in days [reg_start_end_or_lastadm_any_int]; months [reg_start_end_or_lastadm_any_mos]; or years [reg_start_end_or_lastadm_any_yrs] from start of cancer-directed drug regimen to:
    - End date of first drug discontinued in cancer-directed regimen, if regimen was known to be discontinued at time of curation
    - Earliest last known administration date across all drugs in cancer-directed regimen, if regimen was not known to be discontinued at time of curation
- Populated only if regimen does not contain an investigational drug

## *Time (Days, Months, Years) from Start of Cancer-Directed Drug Regimen to End of All Drugs in Regimen*

[reg_start_end_all_int], [reg_start_end_all_int_mos], [reg_start_end_all_int_yrs]

Value (Numeric)

Description

- Interval in days [reg_start_end_all_int]; months [reg_start_end_all_int_mos]; or years [reg_start_end_all_int_yrs] from the start of the cancer-directed drug regimen to the end of all drugs in cancer-directed drug regimen
- Populated only if:
    - Regimen is known to be discontinued ([drugs_dc_ynu] = "Yes")
    - Regimen does not contain an investigational drug

## *Time (Days, Months, Years) from Start of Cancer-Directed Drug Regimen to End Date or Last Known Administration Date Across All Drugs in Regimen*

[reg_start_end_or_lastadm_all_int], [reg_start_end_or_lastadm_all_mos], [reg_start_end_or_lastadm_all_yrs]

Value (Numeric)

Description

- Interval in days [reg_start_end_or_lastadm_all_int]; months [reg_start_end_or_lastadm_all_mos]; or years [reg_start_end_or_lastadm_all_yrs] from the start of cancer-directed drug regimen to:
    - End date of all drugs in cancer-directed drug regimen, if regimen was known to be discontinued at time of curation
    - Last known administration date across all drugs administered as part of cancer-directed regimen, if regimen was not known to be discontinued at time of curation
- Populated only if regimen does not contain an investigational drug

## *Overall Survival from Start of Cancer-Directed Drug (Days, Months, Years)*

[tt_os_d1_days] – [tt_os_d5_days], [tt_os_d1_mos] – [tt_os_d5_mos], [tt_os_d1_yrs] – [tt_os_d5_yrs]

Value (Numeric)

Description

- Time from start of cancer-directed drug to death/last known alive date
- Interval in days [tt_os_d1_days- tt_os_d5_days]; months [tt_os_d1_mos- tt_os_d5_mos]; or years [tt_os_d1_yrs- tt_os_d5_yrs] from start of cancer-directed drug [drugs_startdt_int_drug_1-drugs_startdt_int_drug_5] to death [hybrid_death_int] or last known alive date [dob_lastalive_int]

### *Overall Survival from Start of Cancer-Directed Drug Regimen: Status Indicator*

[os_status_g]

Value (Numeric)

- 1 = Dead
- 0 = Censored

Description

- An event is defined by death
- Patients were censored if not known to be dead

### *Overall Survival from Start of Cancer-Directed Drug: Status Indicator*

[os_status_d]

Value (Numeric)

- 1 = Dead
- 0 = Censored

Description

- An event is defined by death
- Patients were censored if not known to be dead

### *Overall Survival from Start of Cancer-Directed Drug Regimen (Days, Months, Years)*

[tt_os_g_days, tt_os_g_mos, tt_os_g_yrs]

Value (Numeric)

Description

- Time from start of cancer-directed drug regimen to death/last known alive date
- Interval in days [tt_os_g_days]; months [tt_os_g_mos]; or years [tt_os_g_yrs] from the start of the cancer-directed drug regimen [dob_reg_start_int] to death [hybrid_death_int] or last known alive date [dob_lastalive_int]

### ***Progression Free Survival – Imaging (PFS-I) from Start of Cancer-Directed Drug Regimen: Status Indicator***

[pfs_i_status_g]

Value (Numeric)

- 1 = Progression or death
- 0 = Censored

Description

- An event is defined by:
  - Radiologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
  - Death
- Patients were censored at the time of next cancer diagnosis or end of follow up
- Populated only for regimens associated with BPC Project cancers with stage IV disease at diagnosis. For patients with multiple BPC Project cancers, PFS is only defined if the first cancer was stage IV at diagnosis.

### ***Progression Free Survival – Imaging (PFS-I) from Start of Cancer-Directed Drug Regimen***

[tt_pfs_i_g_days, tt_pfs_i_g_mos, tt_pfs_i_g_yrs]

Value (Numeric)

Description

- Time from start of cancer-directed drug regimen to progression, subsequent cancer diagnosis, last known alive date, or death
- Interval in days [tt_pfs_i_g_days]; months [tt_pfs_i_g_mos]; or years [tt_pfs_i_g_yrs] from cancer-directed drug regimen to:
  - Radiologist assessment of progression
  - Death
  - Subsequent cancer diagnosis or last known alive date, if patient is censored according to PFS-I status indicator [pfs_i_status_g]
- Populated only for regimens associated with BPC Project cancers with stage IV disease at diagnosis. For patients with multiple BPC Project cancers, PFS is only defined if the first cancer was stage IV at diagnosis.

### ***Progression Free Survival – Medical Oncologist Assessment (PFS-M) from Start of Cancer-Directed Drug Regimen: Status Indicator***

[pfs_m_status_g]

Value (Numeric)

- 1 = Progression according to medical oncologist assessment; death
- 0 = Censored

Description

- An event is defined by:
  - Medical oncologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
  - Death
- Patients were censored at the time of next cancer diagnosis or end of follow up
- Populated only for regimens associated with BPC Project cancers with stage IV disease at diagnosis. For patients with multiple BPC Project cancers, PFS is only defined if the first cancer was stage IV at diagnosis

## *Progression Free Survival – Medical Oncologist Assessment (PFS-M) from Start of Cancer-Directed Regimen*

[tt_pfs_m_g_days, tt_pfs_m_g_mos, tt_pfs_m_g_yrs]

Value (Numeric)

Description

- Time from start of cancer-directed drug regimen to progression, subsequent cancer diagnosis, last known alive date, or death
- Interval in days [tt_pfs_g_days]; months [tt_pfs_g_mos]; or years [tt_pfs_g_yrs] from start of cancer-directed drug regimen [dob_reg_start_int] to:
  - Medical oncologist assessment of mixed or progressing disease
  - Death
  - Subsequent cancer diagnosis or last known alive date, if patient is censored according to PSF-M status indicator [pfs_m_status_g]
- Populated only for regimens associated with BPC Project cancers with stage IV disease at diagnosis. For patients with multiple BPC Project cancers, PFS is only defined if the first cancer was stage IV at diagnosis.

# PRISSMM Pathology Dataset

The pathology dataset is structured as one record per pathology report, per patient. All pathology reports beginning with the month and year of the first BPC Project cancer diagnosis are curated. All subsequent pathology reports were recorded (including pathology reports corresponding to non-BPC Project cancer and subsequent BPC Project cancer diagnoses; Appendix 1). Additionally, all non-BPC Project CRC pathology reports are curated.

The PRISSMM Pathology dataset can be linked to the following datasets:

- Cancer panel test dataset using [record_id], [ca_seq], [path_proc_number] and [path_report_number]
- Patient-level, BPC Project and non-BPC Project cancer diagnosis, cancer-directed regimen, PRISSMM Imaging, and PRISSMM Medical Oncologist Assessment datasets using [record_id].

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The preferred variable is noted in the description.

## *BPC Project Cohort*

[cohort]

Value (Character)

- CRC

Description

- Indicates the BPC Project cancer type

## *Record ID*

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets. Note that AACR Project GENIE refers to participating institutions as centers. The institutions include:
  - DFCI
  - MSK
  - VICC

## *Institution*

[institution]

Value (Character)

- DFCI
- MSK
- VICC

Description

- Indicates the patient's internal institution of cancer care.

### *Pathology Procedure Number*

[path_proc_number]

Value (Numeric)

Description

- Order for pathology procedures based on the date of the procedure [path_proc_int]
- Pathology procedures occurring on the same date have the same procedure number

### *Pathology Report Number*

[path_rep_number]

Value (Numeric)

Description

- Order for pathology reports from the same pathology procedure [path_proc_number]
- For example, a pathology procedure with three associated reports will have pathology report numbers 1-3.

### *Institution Where Procedure Was Performed*

[path_proc_inst]

Value (Character)

- Internal Institution
- External Institution

Description

- Indicates whether the pathology procedure was performed at the internal (i.e. the same institution as the [institution] variable) or external institution.
    - Details regarding external institutions are not curated.

### *Institution Where Pathology was Reviewed*

[path_rep_inst]

Value (Character)

- Internal Institution
- External Institution

Description

- Indicates whether the pathology procedure was reviewed at the internal (i.e. the same institution as the [institution] variable) or external institution.
    - Details regarding external institutions are not curated.

## *Time (Days) from Date of Birth to Pathology Procedure Date*

[path_proc_int]

Value (Numeric)

Description

- Interval in days from date of birth to pathology procedure date

## *Time (Days, Months, Years) from First BPC Project Cancer to Pathology Procedure Date*

[dx_path_proc_days], [dx_path_proc_mos], [dx_path_proc_yrs]

Value (Numeric)

Description

- Interval in days [dx_path_proc_days]; months [dx_path_proc_mos]; or years [dx_path_proc_yrs] from first BPC Project cancer diagnosis to pathology procedure date

## *Pathology Type*

[path_proc_type]

Value (Character)

- Cytology
- Surgical pathology
- Other

Description

- Type of pathology

## *Number of Specimens Included in the Pathology Report*

[path_num_spec]

Value (Numeric)

- 1-45

Description

- The number of distinct specimens from one procedure that are included in the report
- Up to 45 specimens can be curated from each pathology report

## *Anatomic Site for Each Specimen 1-45*

[path_site1] – [path_site45]

Value (Character)

- ICD-O-3 topography code

Description

- The anatomic site of the specimen is often different than the type of invasive cancer. For example, a specimen from any body part may still be lung cancer.

## *In Situ Cancer Found in at Least One Specimen in the Pathology Report*

[path_insitu_any]

Value (Character)

- Yes
- No

Description

- Indicates whether in situ cancer was found in any of the 45 specimens on the pathology report
- Based on [path_insitu1] – [path_insitu45]

## *Number of Specimens with In Situ Cancer in Pathology Report*

[n_specimen_insitu]

Value (Numeric)

- 0-45

Description

- Number of specimens with in situ cancer on the pathology report
- Based on [path_insitu1] – [path_insitu45]

## *In Situ Cancer Identified in Specimen 1-45*

[path_insitu1] – [path_insitu45]

Value (Character)

- Yes
- No

Description

- Indicates whether any in situ cancer (ICD-O-3 Behavior Code = 2) is present in the specimen

### *Invasive Cancer Found in at Least One Specimen in the Pathology Report*

[path_ca_inv_any]

Value (Character)

- Yes
- No

Description

- Indicates whether invasive cancer was found in any of the 45 specimens on the pathology report
- Based on [path_ca1] – [path_ca45]

### *Number of Specimens with Invasive Cancer in Pathology Report*

[n_specimen_inv]

Value (Numeric)

- 0-45

Description

- Number of specimens with invasive cancer in pathology report
- Based on [path_ca1] – [path_ca45]

### *Invasive Cancer Identified in Specimen 1-45*

[path_ca1] – [path_ca45]

Value (Character)

- Yes
- No

Description

- Indicates whether any invasive cancer (ICD-O-3 Behavior Code = 3) is present in the specimen

### *Invasive Cancer Type for Each Specimen 1-45*

[path_ca_type1] – [path_ca_type45]

Value (Character)

- Adrenocorical Carcinoma
- Anal Cancer
- Appendix Cancer
- Bile Duct Cancer
- Bladder Cancer
- Brain Cancer
- Breast Cancer
- NET or Carcinoid
- Cervical Cancer
- Colon Cancer
- Colon/Rectum Cancer
- Esophagus Cancer
- Ewing Sarcoma
- Fallopian Tube Cancer
- Gallbladder Cancer
- Germ Cell Tumor
- GIST
- Head and Neck Cancer
- Mesothelioma
- Ill Defined/Cancer of Unknown Primary
- Liver Cancer
- Lung Cancer, NOS
- Melanoma
- Merkel Cell
- Neuroblastoma
- Non Small Cell Lung Cancer
- Osteosarcoma
- Ovarian Cancer
- Pancreatic Cancer
- Parathyroid Cancer
- Penis Cancer
- Peritoneum Cancer
- Placenta Cancer
- Prostate Cancer
- Rectum and Rectosigmoid Cancer
- Renal Kidney Cancer
- Renal Pelvis Cancer
- Retinoblastoma
- Rhabdomyosarcoma
- Scrotum Cancer
- Small Cell Lung Cancer
- Small Intestine Cancer
- Stomach Cancer
- Testis Cancer
- Thymus Cancer
- Thyroid Cancer

- Uterus Cancer
- Vagina Cancer
- Vulva Cancer
- Wilms Tumor
- Other
- Not stated
- Unknown

Description

- Indicates cancer type identified in the specimen with invasive cancer
- Populated only if invasive cancer is present ([path_ca1] –[path_ca25] = “Yes”)

### *Cancer Histology Type for Specimen 1-45*

[path_ca_hist1] – [path_ca_hist45]

Value (Character)

- ICD-O-3 morphology code

Description

- Cancer histology type associated with specimen with invasive cancer
- Populated only if invasive cancer is present ([path_ca1] –[path_ca45] = “Yes”)

## Biomarkers

Up to three tests for PD-L1 can be associated with a pathology report. The biomarker information corresponding to each of the three tests is consistent across variables, i.e. information corresponding to the first PD-L1 test is stored in variables [pdl1_yn], [pdl1_prepaint], [pdl1_test], etc. and the information corresponding to the second PD-L1 test is stored in variables [pdl1_yn_2], [pdl1_prepaint_2], [pdl1_test_2], etc. Summary variables indicating any testing and any positive result across all three instances are also provided (i.e. [pdl1_testing] and [pdl1_positive_any]).

### *Any PD-L1 Testing Reported on Pathology Report*

[pdl1_testing]

Value (Character)

- Yes
- No

Description

- Indicates whether any PD-L1 testing was reported on a given pathology report
- Combines variables [pdl1_yn], [pdl1_yn_2], [pdl1_yn_3]

## *Any Positive PD-L1 Result Reported on Pathology Report*

[pdl1_positive_any]

Value (Character)

- Yes
- No

Description

- Indicates whether any PD-L1 testing on a given pathology report returned a positive result
- Based on any of the following criteria being met:
    - A low, high, or positive summary score
    - A percentage or percentage range tumor cells greater than 0
    - A percentage or percentage range of infiltrating immune cells greater than 0
- Populated only if PD-L1 testing is indicated on pathology report (i.e. [pdl1_testing] = “Yes”)

## *PD-L1 Testing Reported*

[pdl1_yn], [pdl1_yn_2], [pdl1_yn_3]

Value (Character)

- Yes
- No

Description

- Indicates whether PD-L1 testing was reported on a given pathology report
- Up to three PD-L1 tests on a single pathology report are curated
- Populated only if any in situ or invasive cancer identified in specimens 1-25 ([path_insitu1]-[path_insitu25] = “Yes” or [path_ca1]-[path_ca25] = “Yes”)
- Derived variable [pdl1_testing] incorporates all three PD-L1 tests.

## *Time (Days) from Date of Birth to PD-L1 Report Date*

[pdl1_prepaint], [pdl1_prepaint_2], [pdl1_prepaint_3]

Value (Numeric)

Description

- Interval in days from date of birth to pathology report PD-L1 test result
- Populated only if PD-L1 testing reported ([pdl1_yn], [pdl1_yn_2], [pdl1_yn_3] = “Yes”)

## *PD-L1 Antibody Test Type*

[pdl1_test], [pdl1_test_2], [pdl1_test_3]

Value (Character)

- 22C3
- 28-2
- E1L3N
- SP142
- Other
- Unknown

Description

- PD-L1 test identified in pathology report
- Up to three tests can be recorded per pathology report
- Populated only if PD-L1 testing reported ([pdl1_yn], [pdl1_yn_2], [pdl1_yn_3] = “Yes”)

### *PD-L1 Results Presented as a Percentage or Percentage Range of Tumor Cells*

[pdl1_type___1], [pdl1_type_2___1], [pdl1_type_3___1]

Value (Character)

- Percentage or Percentage Range of Tumor Cells

Description

- Indicates that PD-L1 test results are represented as percentage or percentage range of tumor cells
- Populated only if PD-L1 results represented as the percentage or percentage range of tumor cells

### *PD-L1 Results Presented as a Percentage or Percentage Range of Infiltrating Immune Cells*

[pdl1_type___2], [pdl1_type_2___2], [pdl1_type_3___2]

Value (Character)

- Percentage or Percentage Range of Infiltrating Immune Cells

Description

- Indicates that PD-L1 test results represented as the percentage or percentage range of infiltrating immune cells
- Populated only if PD-L1 results represented as the percentage or percentage range of infiltrating immune cells

### *PD-L1 Results Presented as a Numeric Combined Positive Score*

[pdl1_type___3], [pdl1_type_2___3], [pdl1_type_3___3]

Value (Character)

- Numeric (Combined Positive Score)

Description

- Indicates that PD-L1 test results represented as a numeric combined positive score
- Populated only if PD-L1 results represented as a numeric combined positive score

## *PD-L1 Results Presented as Summary Assessment*

[pdl1_type___4], [pdl1_type_2___4], [pdl1_type_3___4]

Value (Character)

- Summary Assessment

Description

- Indicates that PD-L1 test results represented as a summary assessment
- Populated only if PD-L1 results represented as a summary assessment

## *Percentage of Tumor Cells Positive for PD-L1*

[pdl1_perc], [pdl1_perc_2], [pdl1_perc_3]

Value (Numeric)

- 0-100

Description

- The percentage of tumor cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of tumor cells ([pdl1_type___1], [pdl1_type_2___1], [pdl1_type_3___1] = “Percentage or Percentage Range of Tumor Cells”)

## *Lower Range (%) of Tumor Cells Positive for PD-L1*

[pdl1_tclrange], [pdl1_tclrange_2], [pdl1_tclrange_3]

Value (Character)

- <1
- 0-100
- Other
- Not Applicable

Description

- Lower range (%) of tumor cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of tumor cells ([pdl1_type___1], [pdl1_type_2___1], [pdl1_type_3___1] = “Percentage or Percentage Range of Tumor Cells”)

## *Upper Range (%) of Tumor Cells Positive for PD-L1*

[pdl1_tcurange], [pdl1_tcurange_2], [pdl1_tcurange_3]

Value (Character)

- <1
- 0-100
- Other
- Not Applicable

Description

- Upper range (%) of tumor cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of tumor cells ([pdl1_type___1], [pdl1_type_2___1], [pdl1_type_3___1] = “Percentage or Percentage Range of Tumor Cells”)

## *Percentage of Infiltrating Immune Cells Positive for PD-L1*

[pdl1_icperc], [pdl1_icperc_2], [pdl1_icperc_3]

Value (Numeric)

- 0-100

Description

- The percentage of infiltrating immune cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of infiltrating immune cells ([pdl1_type___2], [pdl1_type_2___2], [pdl1_type_3___2] = “Percentage or Percentage Range of Infiltrating Immune Cells”)

## *Lower Range (%) of Infiltrating Immune Cells Positive for PD-L1*

[pdl1_iclrange], [pdl1_iclrange_2], [pdl1_iclrange_3]

Value (Character)

- <1
- 0-100
- Other
- Not Applicable

Description

- Lower range (%) of infiltrating immune cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of infiltrating immune cells ([pdl1_type___2], [pdl1_type_2___2], [pdl1_type_3___2] = “Percentage or Percentage Range of Infiltrating Immune Cells”)

## *Upper Range (%) of Infiltrating Immune Cells Positive for PD-L1*

[pdl1_icurange], [pdl1_icurange_2], [pdl1_icurange_3]

Value (Character)

- <1
- 0-100
- Other
- Not Applicable

Description

- Upper range (%) of infiltrating immune cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of infiltrating immune cells ([pdl1_type___2], [pdl1_type_2___2], [pdl1_type_3___2] = “Percentage or Percentage Range of Infiltrating Immune Cells”)

## *Numeric Combined Positive Score (CPS) for PD-L1*

[pdl1_num], [pdl1_num_2], [pdl1_num_3]

Value (Numeric)

Description

- The Combined Positive Score (CPS) value
- The minimum CPS is 0
- Populated only if PD-L1 testing reported as a numeric combined positive score ([pdl1_type___3], [pdl1_type_2___3], [pdl1_type_3___3] = “Numeric (Combined Positive Score)”)

## *Lower Range of the Combined Positive Score (CPS) for PD-L1*

[pdl1_lcpsrange], [pdl1_lcpsrange_2], [pdl1_lcpsrange_3]

Value (Character)

- <1
- 0-11
- Other
- Not Applicable

Description

- The lower range of the Combined Positive Score (CPS)
- Populated only if PD-L1 testing reported as a numeric combined positive score ([pdl1_type___3], [pdl1_type_2___3], [pdl1_type_3___3] = “Numeric (Combined Positive Score)”)

## *Upper Range of the Combined Positive Score (CPS) for PD-L1*

[pdl1_ucpsrange], [pdl1_ucpsrange_2], [pdl1_ucpsrange_3]

Value (Character)

- <1
- 0-11
- Other
- Not Applicable

Description

- The upper range of the Combined Positive Score (CPS)
- Populated only if PD-L1 testing reported as a numeric combined positive score ([pdl1_type___3], [pdl1_type_2___3], [pdl1_type_3___3] = “Numeric (Combined Positive Score)”)

## *Summary Assessment of PD-L1*

[pdl1_sum], [pdl1_ sum _2], [pdl1_ sum _3]

Value (Character)

- High
- Low
- Positive
- Negative
- Indeterminate/Not stated

Description

- Overall summary assessment value of PD-L1 as stated in pathology report; not specific to immune cells or tumor cells
- Populated only if PD-L1 testing reported as a summary score ([pdl1_type___4], [pdl1_type_2___1], [pdl1_type_3___4] = “Summary Assessment”)

## *Any Microsatellite Instability (MSI) Testing Reported on Pathology Report*

[msi_testing]

Value (Character)

- Yes
- No

Description

- Indicates whether any microsatellite instability (MSI) testing was reported on pathology report
- Combines variables [msi_yn], [msi_yn_2], [msi_yn_3]

## *Microsatellite Instability (MSI) Testing Using Polymerase Chain Reaction (PCR) Reported*

[msi_yn], [msi_yn_2], [msi_yn_3]

Value (Character)

- Yes
- No

Description

- Indicates whether microsatellite instability (MSI) testing using PCR was reported for a specimen containing cancer

## *Time (Days) from Date of Birth to Microsatellite Instability (MSI) Testing Report Date*

[msi_prepaint], [msi_prepaint_2], [msi_prepaint_3]

Value (Numeric)

Description

- Interval in days from date of birth to report date for pathology report Microsatellite Instability (MSI) test result
- Populated only if MSI testing reported ([msi_yn], [msi_yn_2], [msi_yn_3] = “Yes”)

## *Microsatellite Instability (MSI) Result*

[msi_result], [msi_result_2], [msi_result_3]

Value (Character)

- MSS: Stable
- MSI-H: High
- MSI-L: Low or instability in <30% of the microsatellite markers
- Indeterminate/Not stated

Description

- Microsatellite Instability (MSI) result from pathology report
- Populated only if MSI testing reported ([msi_testing] = “Yes”)

## *Any MSI-H test result*

[msi_high_any]

Value (Character)

- Yes
- No

Description

- Indicates whether any Microsatellite Instability (MSI) result is high ([msi_result], [msi_result_2], [msi_result_3] = “MSI-H: High)

## *Microsatellite Instability (MSI) – H: High Details*

[msi_high], [msi_high_2], [msi_high_3]

Value (Character)

- >=30% of the markers exhibit instability
- 2 or more of the 5 markers exhibit instability
- Other

Description

- Details of result if Microsatellite Instability (MSI) testing is MSI-H: High
- Populated only if MSI result is high ([msi_result], [msi_result_2], [msi_result_3] = "MSI-H: High")

## *Microsatellite Instability (MSI) – L: Low Details*

[msi_low], [msi_low_2], [msi_low_3]

Value (Character)

- 1-29% of the markers exhibit instability
- 1 of 5 markers exhibit instability
- Other

Description

- Details of result if Microsatellite Instability (MSI) testing is MSI-L: Low
- Populated only if MSI result is low ([msi_result], [msi_result_2], [msi_result_3] = "MSI-L: Low")

## *Any Mismatch Repair (MMR) Testing Reported on Pathology Report*

[mmr_testing]

Value (Character)

- Yes
- No

Description

- Indicates whether any mismatch repair (MMR) testing was reported on pathology report
- Combines variables [mmr_yn], [mmr_yn_2], [mmr_yn_3]

## *Mismatch Repair (MMR) or Expression/Presence of MH1, or MSH2 or MSH6 or PMS2 Testing*

[mmr_yn], [mmr_yn_2], [mmr_yn_3]

Value (Character)

- Yes
- No

Description

- Indicates whether testing was done for Mismatch Repair or expression/presence of MLH1, MSH2 or MSH6 or PMS2 for a specimen containing cancer

## *Mismatch Repair (MMR): Time (Days) from Date of Birth to MMR Testing Report Date*

[mmr_prepaint], [mmr_prepaint_2], [mmr_prepaint_3]

Value (Numeric)

Description

- Interval in days from date of birth to report date of pathology report with Mismatch Repair (MMR) testing
- Populated only if MMR testing reported ([mmr_yn], [mmr_yn_2], [mmr_yn_3] = “Yes”)

## *Mismatch Repair (MMR): MLH1 Expression*

[mmr_mlh1], [mmr_mlh1_2], [mmr_mlh1_3]

Value (Character)

- Intact expression/present
- Loss of expression/absent
- Cannot be determined/Not mentioned

Description

- Specifies MLH1 expression from Mismatch Repair (MMR) testing
- Populated only if MMR testing reported ([mmr_testing] = “Yes”)

## *Mismatch Repair (MMR): MSH2 Expression*

[mmr_msh2], [mmr_msh2_2], [mmr_msh2_3]

Value (Character)

- Intact expression/present
- Loss of expression/absent
- Cannot be determined/Not mentioned

Description

- Specifies MSH2 expression from Mismatch Repair (MMR) testing
- Populated only if MMR testing reported ([mmr_testing] = “Yes”)

## *Mismatch Repair (MMR): MSH6 Expression*

[mmr_msh6], [mmr_msh6_2], [mmr_msh6_3]

Value (Character)

- Intact expression/present
- Loss of expression/absent
- Cannot be determined/Not mentioned

Description

- Specifies MSH6 expression from Mismatch Repair (MMR) testing
- Populated only if MMR testing reported ([mmr_testing] = “Yes”)

## *Mismatch Repair (MMR): PMS2 Expression*

[mmr_pms2], [mmr_pms2_2], [mmr_pms2_3]

Value (Character)

- Intact expression/present
- Loss of expression/absent
- Cannot be determined/Not mentioned

Description

- Specifies PMS2 expression from Mismatch Repair (MMR) testing
- Populated only if MMR testing reported ([mmr_testing] = “Yes”)

## *Mismatch Repair (MMR): Overall Interpretation*

[mmr_result], [mmr_result_2], [mmr_result_3]

Value (Character)

- No loss of nuclear expression of MMR proteins: No evidence of deficient mismatch repair (low probability of MSI-H): proficient
- Loss of nuclear expression of one or more MMR proteins; deficient mismatch repair
- Indeterminate/Not stated

Description

- Overall Mismatch Repair (MMR) interpretation
- Populated only if MMR testing reported ([mmr_testing] = “Yes”)

## *Mismatch Repair Proficient (MMR-P): Report Details*

[mmrp_det], [mmrp_det_2], [mmrp_det _3]

Value (Character)

- No loss of nuclear expression of MLH1, MSH2, MSH6 or PMS2 proteins
- Described as MMR intact, preserved, normal

Description

- Details of MMR status that were recorded in pathology report if interpretation of MMR testing was MMR-P: Proficient
- Populated only if MMR result is proficient ([mmr_result], [mmr_result_2], [mmr_result_3] = “No loss of nuclear expression of MMR proteins: No evidence of deficient mismatch repair (low probability of MSI-H): proficient”)

### *MMR Deficient (MMR-D): Report Details*

[mmrd_det], [mmrd_det _2], [mmrd_det _3]

Value (Character)

- Loss of nuclear expression (or absence of) one or more of MLH6, MSH2, MSH6 or MPS2 proteins
- Described as abnormal

Description

- Details of MMR status that were recorded in pathology report if interpretation of MMR testing was MMR Deficient (MMR-D)
- Populated only if MMR result is deficient ([mmr_result], [mmr_result_2], [mmr_result_3] = “Loss of nuclear expression of one or more MMR proteins; deficient mismatch repair”)

# PRISSMM Imaging Dataset

The PRISSMM Imaging dataset is structured as one record per imaging report, per patient. Imaging reports were curated beginning with the month and year of the first BPC Project cancer diagnosis. All subsequent imaging reports were recorded (including imaging reports corresponding to non-BPC Project cancer and subsequent BPC Project cancer diagnoses; Appendix 1).

The PRISSMM Imaging dataset can be linked to all datasets using the variable [record_id].

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The preferred variable is noted in the description.

## *BPC Project Cohort*

[cohort]

Value (Character)

- CRC

Description

- Indicates the BPC Project cancer type

## *Record ID*

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets. Note that AACR Project GENIE refers to participating institutions as centers. The institutions include:
  - DFCI
  - MSK
  - VICC

## *Institution*

[institution]

Value (Character)

- DFCI
- MSK
- VICC

Description

- Indicates the patient’s internal institution of cancer care.

## *Imaging Report Number*

[scan_number]

Value (Numeric)

Description

- Unique identifier for imaging reports based on the date of the scan [image_scan_int]
- Different scans occurring on the same date have distinct scan numbers (e.g. a CT and MRI occurring on the same date may be numbered as scan 1 and scan 2, though the ordering is arbitrary)

## *Time (Days) from Date of Birth to Imaging Date*

[image_scan_int]

Value (Numeric)

Description

- Interval in days from date of birth to scan date

## *Time (Days, Months, Years) from First BPC Project Cancer to Imaging Date*

[dx_scan_days], [dx_scan_mos], [dx_scan_yrs]

Value (Numeric)

Description

- Interval in days [dx_scan_days]; months [dx_scan_mos]; or years [dx_scan_yrs] from first BPC Project cancer diagnosis to scan date

## *Time (Days) from Date of Birth to Reference Imaging Date*

[image_ref_scan_int]

Value (Numeric)

Description

- Interval days from date of birth to reference imaging date
- Populated only if there is evidence of cancer on the imaging report ([image_ca] = “Yes”)

## *Time (Days, Months, Years) from First BPC Project Cancer to Reference Imaging Date*

[dx_ref_scan_days], [dx_ref_scan_mos], [dx_ref_scan_yrs]

Value (Numeric)

Description

- Interval in days [dx_ref_scan_days]; months [dx_ref_scan_mos]; or years [dx_ref_scan_yrs] from first BPC Project cancer diagnosis to imaging date

## *Institution Where Scan was Performed*

[image_inst_perf]

Value (Character)

- Internal institution
- External institution

Description

- Indicates whether scan was performed at the internal (i.e. the same institution as the [institution] variable) or external institution.
    - Details regarding external institutions are not curated.

## *Institution Where Image was Interpreted*

[image_inst_inter]

Value (Character)

- Internal institution
- External institution

Description

- Indicates whether image was interpreted at the internal (i.e. the same institution as the [institution] variable) or external institution.
    - Details regarding external institutions are not curated.
- Populated only if imaging was performed at an external institution ([image_inst_perf] = “External institution”)

## *Imaging Scan Type*

[image_scan_type]

Value (Character)

- CT
- MRI
- PET or PET-CT
- Bone Scan
- Other Nuclear Medicine Scan
- Mammogram – Use for Breast Cancer Only
- Other CA-specific scan

Description

- Type of imaging scan

### If Other Scan Type, Type of Scan

[oth_scan]

Value (Character)

- Free-text

Description

- Specifies scan type if imaging scan type is other
- Populated only if scan type is listed as other ([image_scan_type] = “Other CA-specific scan”)

### Imaging Site: Brain/Head

[image_scansite___1]

Value (Character)

- Brain/Head

Description

- Indicates a scan of the brain/head
- Populated only if scan is of brain/head

### Imaging Site: Spine

[image_scansite___2]

Value (Character)

- Spine

Description

- Indicates a scan of the spine
- Populated only if scan is of the spine

### *Imaging Site: Neck*

[image_scansite___3]

Value (Character)

- Neck

Description

- Indicates a scan of the neck
- Populated only if scan is of the neck

### *Imaging Site: Chest*

[image_scansite___4]

Value (Character)

- Chest

Description

- Indicates a scan of the chest
- Populated only if scan is of the chest

### *Imaging Site: Abdomen*

[image_scansite___5]

Value (Character)

- Abdomen

Description

- Indicates a scan of the abdomen
- Populated only if scan is of the abdomen

### *Imaging Site: Pelvis*

[image_scansite___6]

Value (Character)

- Pelvis

Description

- Indicates a scan of the pelvis
- Populated only if scan is of the pelvis

### ***Imaging Site: Extremity***

[image_scansite___7]

Value (Character)

- Extremity

Description

- Indicates a scan of an extremity
- Populated only if scan is of an extremity

### ***Imaging Site: Full body***

[image_scansite___8]

Value (Character)

- Full body

Description

- Indicates a full body scan
- Populated only if a full body scan

### ***Imaging Sites***

[scan_sites]

Value (Character)

Description

- List of sites scanned on this imaging report
- Concatenation of sites scanned [image_scansite___1] – [image_scansite___8]

### ***Radiologist Assessment of any Evidence of Cancer on this Imaging Report***

[image_ca]

Value (Character)

- Yes, the Impression states or implies there is evidence of cancer
- No, the Impression states or implies there is no evidence of cancer
- The Impression is uncertain, indeterminate, or equivocal
- The Impression does not mention cancer

Description

- Indicates whether the radiologist assessment indicates any evidence of cancer

## *Radiologist Assessment of Change in Cancer Status*

[image_overall]

Value (Character)

- Improving/Responding
- Stable/No change
- Mixed
- Progressing/Worsening/Enlarging
- Not stated/Indeterminate

Description

- Radiologist's assessment of the change in the patient's cancer status
- Populated only if there is evidence of cancer on the imaging report ([image_ca] = "Yes")

## *Location of Cancer Based on Imaging Report*

[image_casite1]-[image_casite15]

Value (Character)

- ICD-O-3 topography code

Description

- Location of cancer on imaging report based on ICD-O-3 topography code
- Populated only if there is evidence of cancer on the imaging report ([image_ca] = "Yes")

# PRISSMM Medical Oncologist Assessment Dataset

The PRISSMM Medical Oncologist Assessment dataset is structured as one record per curated medical oncologist assessment, per patient. Medical oncologist assessments were curated beginning with the month and year of the first BPC Project cancer diagnosis (Appendix 1). One medical oncologist assessment per month was curated; curation instructions regarding the selection of the assessment to curate are provided in Appendix 3.

The PRISSMM Medical Oncologist Assessment dataset can be linked to all datasets using the variable [record_id].

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The preferred variable is noted in the description.

## *BPC Project Cohort*

[cohort]

Value (Character)

- CRC

Description

- Indicates the BPC Project cancer type

## *Record ID*

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets. Note that AACR Project GENIE refers to participating institutions as centers. The institutions include:
    - DFCI
    - MSK
    - VICC

## *Institution*

[institution]

Value (Character)

- DFCI
- MSK
- VICC

Description

- Indicates the patient’s internal institution of cancer care.

## *Medical Oncologist Visit Identifier*

[md_visit_number]

Value (Numeric)

Description

- Unique identifier for curated medical oncologist assessments based on the visit date [md_onc_visit_int]
- Only one medical oncologist assessment per month was curated (Appendix 3)

## *Time (Days) from Date of Birth to Medical Oncologist Visit*

[md_onc_visit_int]

Value (Numeric)

Description

- Interval in days from date of birth to medical oncologist visit
- Based on the date the visit occurred, not the date the assessment was signed or uploaded.

## *Time (Days, Months, Years) from First BPC Project Cancer to Medical Oncologist Visit*

[dx_md_visit_days], [dx_md_visit_mos], [dx_md_visit_yrs]

Value (Numeric)

Description

- Interval in days [dx_md_visit_days]; months [dx_md_visit_mos]; or years [dx_md_visit_yrs] from first BPC Project cancer diagnosis to medical oncologist visit.
    - Based on the date the visit occurred, not the date the assessment was signed or uploaded.

## *Cancer Diagnosis Assessed by Medical Oncologist*

[md_type_ca_cur]

Value (Character)

- Adrenocorical Carcinoma
- Anal Cancer
- Appendix Cancer
- Bile Duct Cancer
- Bladder Cancer
- Brain Cancer
- Breast Cancer
- NET or Carcinoid
- Cervical Cancer
- Colon Cancer
- Colon/Rectum Cancer
- Esophagus Cancer
- Ewing Sarcoma
- Fallopian Tube Cancer
- Gallbladder Cancer
- Germ Cell Tumor
- GIST
- Head and Neck Cancer
- Mesothelioma
- Ill Defined/Cancer of Unknown Primary
- Liver Cancer
- Lung Cancer, NOS
- Melanoma
- Merkel Cell
- Neuroblastoma
- Non Small Cell Lung Cancer
- Osteosarcoma
- Ovarian Cancer
- Pancreatic Cancer
- Parathyroid Cancer
- Penis Cancer
- Peritoneum Cancer
- Placenta Cancer
- Prostate Cancer
- Rectum and Rectosigmoid Cancer
- Renal Kidney Cancer
- Renal Pelvis Cancer
- Retinoblastoma
- Rhabdomyosarcoma
- Scrotum Cancer
- Small Cell Lung Cancer
- Small Intestine Cancer
- Stomach Cancer
- Testis Cancer
- Thymus Cancer
- Thyroid Cancer

- Uterus Cancer
- Vagina Cancer
- Vulva Cancer
- Wilms Tumor
- Other

Description

- Indicates the cancer diagnosis associated with the given medical oncologist assessment
- This variable can be used to identify the medical oncologist assessments that are associated with the BPC Project cancer

## ***Medical Oncologist Assessment of Evidence of Cancer***

[md_ca]

Value (Character)

- Yes, the Impression/Plan states or implies there is evidence of cancer
- No, the Impression/Plan states or implies there is no evidence of cancer
- Impression/Plan is uncertain, indeterminate, or equivocal
- Impression/Plan does not mention cancer

Description

- Medical oncologist’s assessment of whether there is evidence of cancer

## ***Medical Oncologist Assessment of Change in Cancer Status***

[md_ca_status]

Value (Character)

- Improving/Responding
- Stable/No change
- Mixed
- Progressing/Worsening/Enlarging
- Not stated/Indeterminate

Description

- Medical oncologist’s assessment of the change in the patient’s cancer status

# PRISSMM Tumor Marker Dataset

The PRISSMM Tumor Marker dataset is structured as one record per curated tumor marker result, per patient. Serum-based tumor markers that are related to the diagnosis/prognosis of cancer were curated. The tumor markers for patients in the BPC Project colorectal cancer cohort include CEA and CA19-9.

Note: variables pertaining to PD-L1, MSI and MMR are recorded in the pathology dataset.

The PRISSMM Tumor Marker dataset can be linked to all datasets using the variable [record_id].

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The preferred variable is noted in the description.

## *BPC Project Cohort*

[cohort]

Value (Character)

- CRC

Description

- Indicates the BPC Project cancer type

## *Record ID*

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets. Note that AACR Project GENIE refers to participating institutions as centers. The institutions include:
    - DFCI
    - MSK
    - VICC

## *Institution*

[institution]

Value (Character)

- DFCI
- MSK
- VICC

Description

- Indicates the patient’s internal institution of cancer care.

### *Tumor Marker Sequence Number*

[tm_number]

Value (Numeric)

Description

- Unique identifier for tumor marker assessment based on the specimen collection date [tm_spec_collect_int]

### *Time (Days) from Date of Birth to Date of Tumor Marker Specimen Collection*

[tm_spec_collect_int]

Value (Numeric)

Description

- Interval in days from date of birth to tumor marker specimen collection

### *Time (Days, Months, Years) from First BPC Project Cancer to Tumor Marker Collection*

[dx_tm_days], [dx_tm_mos], [dx_tm_yrs]

Value (Numeric)

Description

- Interval in days [dx_tm_days]; months [dx_tm_mos]; or years [dx_tm_yrs] from first BPC Project cancer diagnosis to tumor marker specimen collection for each tumor marker that is recorded

### *Tumor Marker Lab Test*

[tm_type]

Value (Character)

- CA19-9
- CEA

Description

- The type of tumor marker test
- There may be multiple tumor markers obtained on the same date

### *Tumor Marker Result*

[tm_num_result]

Value (Numeric)

Description

- Numeric value of tumor marker result

### *Tumor Marker Result Unit*

[tm_result_units]

Value (Character)

- mIU/mL
- ng/mL
- U/mL

Description

- The units corresponding to the tumor marker results [tm_num_results]

### *Tumor Marker Lower Limit of Normal*

[tm_normal_range_lower]

Value (Numeric)

Description

- Lower limit of normal for tumor marker results

### *Tumor Marker Upper Limit of Normal*

[tm_normal_range_upper]

Value (Numeric)

Description

- Upper limit of normal for tumor marker results

# Cancer Panel Test (Next Generation Sequencing) Dataset

The cancer panel test refers to the multi-gene panels that have been performed through next generation sequencing (NGS) assays. The Cancer Panel Test dataset is structured as one record per cancer panel test and its associated cancer diagnosis, per patient. For example, if a cancer panel test was associated with one cancer diagnosis, there will be one corresponding record in this dataset. If a cancer panel test was associated with two cancer diagnoses, then there will be two corresponding records in this dataset: one for the first associated cancer diagnosis and another for the second associated cancer diagnosis, etc. The terms “cancer panel test (CPT)” and “next generation sequencing (NGS)” are used interchangeably.

The Cancer Panel Test dataset can be linked to the following datasets:

- BPC Project Cancer Diagnosis dataset using the variables [record_id] and [ca_seq]
- Cancer-Directed Drug Regimen dataset using the variables [record_id] and [ca_seq]
- PRISSMM Pathology dataset using [record_id], [ca_seq], [path_proc_number] and [path_report_number]
- Patient-Level, PRISSMM Imaging, and PRISSMM Medical Oncologist Assessment datasets using [record_id]
    - Cannot be linked to the Non-BPC Project Cancer Diagnosis dataset because non-BPC Project cancer diagnoses were not genomically sequenced (Appendix 1)

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The preferred variable is noted in the description.

### *BPC Project Cohort*

[cohort]

Value (Character)

- CRC

Description

- Indicates the BPC Project cancer type

### *Record ID*

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets. Note that AACR Project GENIE refers to participating institutions as centers. The institutions include:
    - DFCI
    - MSK
    - VICC

## *Institution*

[institution]

Value (Character)

- DFCI
- MSK
- VICC

Description

- Indicates the patient’s internal institution of cancer care.

## *Cancer Panel Test (Next Generation Sequencing) Number*

[cpt_number]

Value (Numeric)

- 1 = First curated next generation sequencing (NGS) test for this patient
- 2 = Second curated NGS test for this patient
- … n = nth NGS test for this patient

Description

- Order for the curated next generation sequencing (NGS) test based on the report date

## *Time (Days) from Date of Birth to Sequencing Report*

[cpt_report_int]

Value (Numeric)

Description

- Interval in days from date of birth to date of NGS report
- Not populated for all institutions. Consequently, the preferred variable for determining the cancer panel test report date is [dob_cpt_report_days].

### *Cancer Diagnosis Number Associated with Next Generation Sequencing Test*

[ca_seq]

Value (Numeric)

- 0 = first and only cancer
- 1 = first of two or more primaries
- 2 = second of two or more primaries
- …10 = tenth of ten or more primaries

Description

- The associated cancer sequence number for this NGS
- This variable is missing if the diagnosis associated with the NGS test is unknown.
- Note that NGS tests may be associated with multiple cancer diagnoses. If that is the case, each associated cancer diagnosis will be a separate row in the dataset.
- Primary key for the cancer diagnosis, cancer-directed regimen and cancer panel test datasets.

### *Pathology Procedure Number of Next Generation Sequencing Specimen*

[path_proc_number]

Value (Numeric)

Description

- Pathology procedure in which the specimen is described
- Primary key for PRISSMM pathology and cancer panel test datasets.

### *Pathology Report Number of Next Generation Sequencing Specimen*

[path_rep_number]

Value (Numeric)

Description

- Pathology report in which the specimen is described
- Primary key for PRISSMM pathology and cancer panel test datasets.

### *GENIE Sample ID*

[cpt_genie_sample_id]

Value (Character)

Description

- GENIE sample ID corresponding to specimen
- Corresponds to variable [sample_id] in AACR Tier 1 data

### *Next Generation Sequencing Specimen OncoTree Diagnosis Code*

[cpt_oncotree_code]

Value (Character)

Description

- The primary cancer diagnosis code based on the OncoTree ontology
- Corresponds to variable [oncotree_code] in AACR Tier 1 data

Data Standard: OncoTree Ontology

### *Specimen Sample Type*

[cpt_sample_type]

Value (Character)

- 1 = Primary tumor
- 2 = Lymph node metastasis
- 3 = Distant organ metastasis
- 4 = Metastasis site unspecified
- 5 = Local recurrence
- 6 = Not otherwise specified
- 7 = Not applicable or hematologic malignancy
- Local recurrence
- Metastasis site unspecified
- Not applicable or hematologic malignancy
- Primary tumor

Description

- Sample type associated with specimen on which NGS was performed
- Variable not coded consistently across institutions
- Corresponds to variable [sample_type] in AACR Tier 1 data

### *Sequencing Assay ID*

[cpt_seq_assay_id]

Value (Character)

Description

- The institutional assay identifier for the NGS genomic testing platform.
- Components are separated by hyphens, with the first component corresponding to the institution's abbreviation.
- All specimens tested by the same platform should have the same identifier.
- Corresponds to variable [seq_assay_id] in AACR Tier 1 data

## *Time (Days) from Date of Birth to Cancer Panel Test Order Date*

[cpt_order_int]

Value (Numeric)

Description

- Interval in days from date of birth to cancer order date

## *Year of Next Generation Sequencing*

[cpt_seq_date]

Value (Character)

Description

- Year of NGS

## *Derived Time (Days) from Date of Birth to Sequencing Report*

[dob_cpt_report_days]

Value (Numeric)

Description

- Interval in days from date of birth to date of sequencing report

## *Next Generation Sequencing Test Report Returned On or After Date of Death*

[cpt_report_post_death]

Value (Numeric)

- 1 = Yes, cancer panel test was returned on or after patient's date of death
- 0 = No, cancer panel test was not returned on or after patient's date of death

Description

- Indicates whether cancer panel test report was returned on or after patient's date of death

## *Time (Days, Months, Years) from BPC Project Cancer to Sequencing Report*

[dx_cpt_rep_days], [dx_cpt_rep_mos], [dx_cpt_rep_yrs]

Value (Numeric)

Description

- Interval in days [dx_cpt_rep_days]; months [dx_cpt_rep_mos]; or years [dx_cpt_rep_yrs] from BPC Project cancer diagnosis to sequencing report

## *Time (Days, Months, Years) from Diagnosis to Pathology Procedure Corresponding to the Next Generation Sequencing*

[dx_path_proc_cpt_days], [dx_path_proc_cpt_mos], [dx_path_proc_cpt_yrs]

Value (Numeric)

Description

- Interval in days [dx_path_proc_cpt_days]; months [dx_path_proc_cpt_mos]; or years [dx_path_proc_cpt_yrs] from BPC Project cancer diagnosis to pathology procedure corresponding to cancer panel test

## *Time (Days, Months, Years) from Pathology Procedure to the Next Generation Sequencing Report Date*

[path_proc_cpt_rep_days], [path_proc_cpt_rep_mos], [path_proc_cpt_rep_yrs]

Value (Numeric)

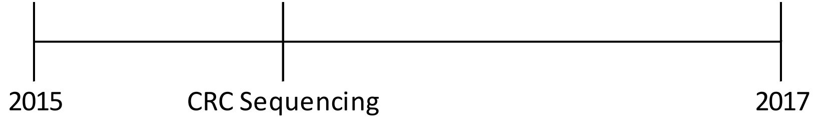
Description

- Interval in days [path_proc_cpt_rep_days]; months [path_proc_cpt_rep_mos]; or years [path_proc_cpt_rep_yrs] from pathology procedure corresponding to cancer panel test to cancer panel test report date

# Appendix 1. BPC Project and Non-BPC Project Cancers

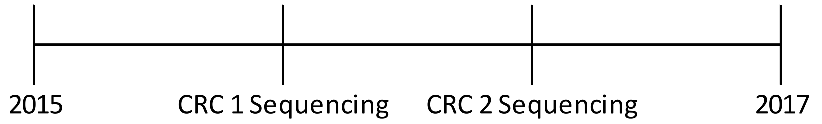
**Definition of BPC Project Cancer:** A BPC Project cancer is the cancer that met the eligibility criteria for the project (i.e. genomic sequencing reported) and was selected at random for PRISSMM phenomic data curation. The terms “BPC Project cancer” and “index cancer” are used interchangeably.

Scenario 1: Single BPC Project Cancer, No Second Primary CRC



Some patients may have more than one BPC Project cancer because they have multiple sequenced cancers that met the eligibility criteria. For example, a patient with a diagnosis of CRC that was sequenced in 2015 and a second primary of CRC that was sequenced in 2017 will have each diagnosis classified as a BPC Project cancer in the CRC BPC Project.

Scenario 2: Multiple BPC Project Cancers

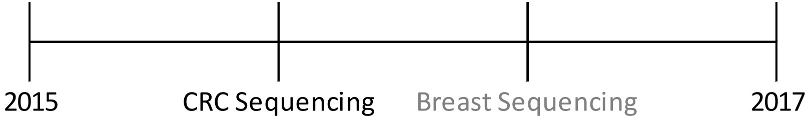


**Definition of Non-BPC Project Cancer:** A non-BPC Project cancer diagnosis can be a diagnosis of the same or different cancer type as the BPC Project cancer that occurs prior to, simultaneous with, or after the BPC Project cancer. Non-BPC Project cancers curated included: 1) non-BPC Project invasive cancer and in situ/non-invasive cancer diagnoses, and 2) other benign tumors. In a very small number of instances, some benign tumors (behavior codes 0 and 1), such as a benign brain tumor or a hemangioma, are obtained from the institution’s tumor registry. These tumors were ineligible for curation, in which case only the tumor registry data is available.

The non-BPC Project cancers do not have associated genomic sequencing.

A non-BPC Project cancer diagnosis could be a different cancer type altogether, such as a breast cancer diagnosis in a patient included in the CRC BPC Project. The CRC cancer is the cancer that made the patient eligible for the BPC Project, but some information regarding the breast cancer diagnosis was curated.

### Scenario 3: Single BPC Project Cancer, Single Non-BPC Project Cancer



In Scenario 3, the breast cancer that was sequenced is classified as a non-BPC Project cancer for the CRC BPC Project. Note that a patient can only be selected for one BPC project (e.g. if patient is selected for the CRC project, then the patient cannot be selected for the breast cancer project).

# Appendix 2. Cancer-Directed Drugs

- Mercaptopurine(AltiMercaptopurine,Azathiopurine,BW57323H,Flocofil,Ismipur,Leukerin,Leupurin,Mern,NCIC04886,PuriNethol,Purimethol,Purinethiol,Purinethol,U4748,WR2785)
- Acetylcysteine(Acetadote,Airbron,Broncholysin,Brunac,Fabrol,Fluatox,Fluimucetin,Fluimucil,Fluprowit,MucoSanigen,Mucocedyl,Mucolator,Mucolyticum,Mucomyst,Mucosolvin,Mucret,NAC,NeoFluimucil,Parvolex,Respaire,Tixair)
- Recombinant Interferon Alfa(Alferon,Alpha Interferon,Leukocyte Interferon,Lymphoblast Interferon,Lymphoblastoid Interferon)
- Aminoglutethimide(Cytadren,Aminoblastin,Elipten,Orimenten,Rodazol)
- Asparaginase(ASP1,Asparaginase II,Asparaginase E.Coli,Colaspase,Elspar,Kidrolase,LASP,LAsnase,Laspar,Leucogen,Leunase,MK965,Paronal,Re82TAD15,Serasa,Spectrila)
- Azacitidine(5 AZC,5AC,Mylosar,U18496,Vidaza)
- BCG Vaccine(BCG TICE,Bacille CalmetteGuerin Live,ImmuCyst,Imovax BCG,Monovax,Oncotice,Pacis,Pastimmun)
- Bleomycin(BLEO,BLM)
- Busulfan(BUS,Busulfanum,Busulfex,CB2041,GT41,Glyzophrol,Joacamine ester,Mielucin,Misulban, Misulfan,Mitosan,Myeleukon,Myelosan,Mylecytan,Myleran,Sulfabutin,WR19508)
- Captopril,Capoten
- Carmustine(BCNU,Becenum,BiCNU, Carmubris,Carmustin,Carmustinum,FDA 0345,Nitrourean,Nitrumon,K27702,SRI1720,WR139021)
- Chlorambucil,Chlorambucilum,Chloraminophen,Chloraminophene,Chlorbutine,Chlorobutin,ChlorobutineEcloril,Leukeran,Leukersan,Leukoran,Linfolizin,Linfolysin
- Cimetidine(Tagamet)
- Cisplatin(Cisplatina,Cisplatinum,Cisplatyl,Citoplatino,Citosin,Cysplatyna,DDP,Lederplatin,Metaplatin,Neoplatin,Peyrone's Chloride,Peyrone's Salt,Placis,Plastistil,Platamine,Platiblastin,PlatiblastinS, Platinex,Platinol,Platinoxan,Platinum Diamminodichloride,Platiran,Platistin,Platosin)
- Clomiphene Citrate(Serophene)
- Cyclophosphamide(CTX,CYCLOcell,Carloxan,Clafen,Claphene,Cyclophosphamide,Cyclophosphan,Cyclophosphane,Cyclostine,Cytophosphan,Cytophosphane,Cytoxan,Endoxan,Fosfaseron,Genoxal,Genuxal,Ledoxina,Mitoxan,Neosar,Revimmune,Syklofosfamid,WR138719)
- Cytarabine(ARAcell,Alexan,AraC,Arabine,Aracytidine,Aracytin,Aracytine,CHX3311,Cytarabinum,Cytarbel,Cytosar,CytosarU,Cytosine Arabinoside,Erpalfa,Starasid,Tarabine PFS,U 19920,U19920,Udicil,WR28453)
- Dacarbazine(Asercit,Biocarbazine,DIC,DTIC,DTICDome, Dacarbazina,Dacarbazine DTIC,Dacatic,Dakarbazin,Deticene,Detimedac,Fauldetic,Imidazole Carboxamide,WR139007
- Dactinomycin(Actinomycin A IV,Actinomycin C1,Actinomycin D,Actinomycin I1,Actinomycin IV,Actinomycin X 1,Cosmegen,DACT,Lyovac Cosmegen)
- Epirubicin Hydrochloride(Ellence,IMI28,Pharmorubicin PFS)
- Estramustine(Leo 275,RO 218837)
- Amifostine(APAETP,Cytofos,Ethiofos,Ethyol,Gammaphos,WR2721,YM08310)
- Etoposide(EPEG,Lastet,Toposar,VP 16213,VP16,VP16213,Vepesid)
- Floxuridine(FDUR,FUDR,WR138720)
- Fluorouracil(5FU,5Fluorouracil,5Fluracil,AccuSite,Adrucil,Carac,Fluouracil,Flurablastin,Fluracedyl,Fluracil,Fluril,Fluroblastin,Ribofluor,Ro29757)
- Flutamide(Apimid,Cebatrol,Chimax,Cytomid,Drogenil,Euflex,Eulexin,Eulexine,FLUT,Flucinom,Flucinome,Flugerel,Fluken,Flulem,FlutaGry,Flutabene,Flutacan,Flutamex,Flutamin,Flutan,Flutaplex,Fugerel,Grisetin,Niftolide,Oncosal,Profamid,Prostacur,Prostadirex,Prostica,Prostogenat,SCH 13521,Tafenil,Tecnoflut,Testotard)
- Altretamine(ENT50852,HMM,HXM,Hemel,Hexalen,Hexaloids,Hexastat)
- Hydroxyurea(Droxia,Hydrea,Hydroxycarbamide,Litalir, OncoCarbideOxeron,SQ1089,Syrea,WR83799)
- Ifosfamide(Cyfos,Holoxan,Holoxane,IFO,IFOCell,IFX,Ifex,Ifolem,Ifomida,Ifomide,Ifosfamidum,Ifoxan,MJF9325, Mitoxana,Naxamide,Seromida,Tronoxal,Z4942)
- Isotretinoin(Absorica,Accure,Accutane,Amnesteem,Cistane,Claravis,Isotrex,Isotrexin,Myorisan,Neovitamin A,Oratane,Ro43780,Roaccutan,Roaccutane,Roacutan,Sotret,ZENATANE)
- Leucovorin Calcium(Calfolex,Calinat,Cehafolin,Citofolin,Citrec,Citrovorum Factor,Cromatonbic Folinico,Dalisol,Disintox,Divical,Ecofol,Emovis,FOLIcell, Flynoken A,Folaren,Folaxin,Foliben,Folidan,Folidar,Folinac,Folinate Calcium,Folinoral,Folinvit,Foliplus,Folix,Imo,Lederfolat, Lederfolin,Leucosar,Rescufolin,Rescuvolin,Tonofolin,Wellcovorin)
- Lomustine(Belustin,CCNU,Cecenu,CeeNU,Citostal,Gleostine,Lomeblastin,Lomustinum,Lucostin,Lucostine,Prava,RB1509,WR139017)
- Mechlorethamine Hydrochloride(Caryolysine,Chlorethamine HCl,Chlorethazine Hydrochloride,Chloromethine HCl,Cloramin,Erasol,HN 2 Hydrochloride,Mustargen,Mustargen HCl,Mustargen Hydrochloride,Mustine Hydrochloride,NLost,OncoCloramin,WR147650)
- Melphalan(L Sarcolysine,Melphalanum,Phenylalanine Mustard,Sarcoclorin,Sarkolysin,WR19813)
- Methotrexate(Abitrexate,AlphaMethopterin,Amethopterin,Brimexate,CL14377,Emtexate,Farmitrexat,Fauldexato,Folex,Lantarel,Ledertrexate,Lumexon,MTX,Maxtrex,Medsatrexate,Metex,Methoblastin, Methylaminopterin,Methotrexatum,Metotrexato,Metrotex,Mexate,Novatrex,Texate,Tremetex,Trexeron,Trixilem,WR19039)
- Plicamycin(A2371,Aureolic acid,MTH,Mithracin,Mithracine,Mithramycin,PA144)
- Mitotane(CB313,Chloditan,Chlodithane,DDD,Mytotan, WR1304)5

- Mitoxantrone Hydrochloride(CL232315,DHAD,DHAQ,DHAQ,Mitroxone,Neotalem,Novantrone,Onkotrone,Pralifan)
- Pentostatin(CI825,CoVidarabine,Covidarabine,DCF,Deoxycoformycin,Nipent,PD81565)
- Procarbazine Hydrochloride(MIH hydrochloride,Matulane,NCIC01810,Natulan,Natulanar,PCB,PCZ,Ro46467 1)
- Streptozocin(STZ,U 9889,U9889,Zanosar)
- Sulindac(Aflodac,Algocetil,ApoSulin,Arthrocine,Artribid,Citireuma,Clinoril,Clisundac,Imbaral,MK231,NovoSundac,Reumofil,Reumyl,Sulinol,Sulreuma)
- Tamoxifen Citrate(ApoTamox,Clonoxifen,Dignotamoxi,Ebefen,Emblon,Estroxyn,Fentamox,GenTamoxifen,Genox,ICI 46,474,ICI46474,Jenoxifen,Kessar,Ledertam,Lesporene,Nolgen,Noltam,Nolvadex,NolvadexD,Nourytam,NovoTamoxifen,Novofen,Noxitem,Oestrifen,Oncotam,PMSTamoxifen,Soltamox,TAM,Tamax,Tamaxin,Tamifen,Tamizam,Tamofen,Tamoxasta,Tamoxifeni Citras,Zemide)
- Teniposide(EPT,PTG,Thenylidene Lignan,VM 26,VM26,Vehem,Vumon)
- Thalidomide(Contergan,Distaval,Kevadon,Neurosedyn,Pantosediv,Pantosediv,Sedalis,Sedoval K17,Softenon,Softenon,Synovir,Talimol,Thalomid
- Thiotepa(Oncotiotepa,STEPA,TESPA,TIO TEF,TSPA)
- Thioguanine(2Amino 6MP,BW 5071,Lanvis,Tabloid,Tioguanin,WR1141,U3B,X27)
- Tretinoin(ATRA,Aberel,Airol,Airol,Aknoten,Avita,Cordes Vas,Dermairol,EpiAberel,Eudyna,Renova,RetinA,RetinA MICRO,RetisolA,Ro 5488,StievaA,StievaA Forte,Trans Retinoic Acid)
- Trifluridine(F3TdR,Triflorothymidine)
- Coenzyme Q10(CoQ10,CoQ10,Coenzyme Q10,UBIDECARENONE,Ubiquinone 10,coenzyme Q10,vitamin Q10)
- Verapamil
- Vinblastine Sulfate,29060LE,Exal,VINCALEUKOBLASTINE,Velban,Velbe,Velsar
- Gemcitabine Hydrochloride(FF 10832,FF10832,FF10832,Gemzar,LY188011,LY188011,dFdCyd)
- SarCNU(sarcosinamide nitrosourea)
- 2Methoxyestradiol(2MeE2,2MeOE2,2Methoxy Estradiol,2ME2,Panzem)
- Decitabine(Dacogen,Deoxyazacytidine,Dezocitidine)
- Acitretin(Etretin,Neotigason,Ro 101670,Soriatane
- Arsenic Trioxide(ATO,Trisenox)
- Bryostatin 1(B705008K112)
- Clarithromycin(Abbott56268,Biaxin)
- Etoposide Phosphate(Etopophos)
- Exemestane(Aromasin,FCE24304)
- Fenretinide(HPR4)
- Fludarabine Phosphate(Beneflur,Fludara,Oforta,SH T 586)
- Omacetaxine Mepesuccinate(CGX635,Ceflatonin,HHT,Synribo)
- Mafosfamide
- Megestrol Acetate(Maygace,Megace,Megestat,Megestil,Niagestin,Ovaban,Pallace,SC10363)
- Nilutamide(Anandron,Nilandron,RU23908)
- Oxaliplatin(1OHP,Ai Heng,DACPLAT,Dacotin,ELOXATIN,Eloxatine,JM83)
- Pegaspargase(Oncaspar,PEGLA)
- Temozolomide(M & B 39831,Methazolastone,RP46161,SCH52365,Temcad,Temodal,Temodar,Temomedac)
- Toremifene(Farestone)
- Triptorelin(AY25650,CL118532,Decapeptyl,Detryptoreline)
- Carboplatin(Blastocarb,CBDCA,Carboplat,Carboplatin Hexal,Carboplatino,Carboplatinum,Carbosin,Carbosol,Carbotec,Displata,Ercar,JM8,Nealorin,Novoplatinum,Paraplat,Paraplatin,Paraplatine,Platinwas,Ribocarbo)
- AGM1470,TNP470
- O6Benzylguanine
- Leuprolide Acetate(A43818,Carcinil,DepoEligard,Enanton,Enantone,EnantoneGyn,Ginecrin,LEUP,Leuplin,Lucrin Depot,Lupron,Lupron Depot,Procren,Procrin,Prostap,TAP144,Trenantone,UnoEnantone,Viadur)
- Doxorubicin Hydrochloride,ADM,Adriacin,Adriamycin,Adriamycin Hydrochloride,Adriamycin RDF,Adriamycine,Adriblastina,Adriblastine,Adrimedac,Chloridrato de Doxorrubicina,DOX,DOXOCELL,Doxorubin,FI 106,FI106,Rubex
- Cladribine,CdA(Leustat,Leustatin,Leustatine,RWJ26251)
- Valrubicin,AD32,AD 32,AD32,Valstar,Valtaxin)
- Mivobulin Isethionate(CI980)
- Fulvestrant(Faslodex,ICI182780,ZD9238)
- Recombinant Interleukin12(Cytotoxic Lymphocyte Maturation Factor,IL12,Interleukin 12,NMIL12,Ro247472)
- Irinotecan Hydrochloride(CPT11,Campto,Camptosar,U101440E)
- Vinorelbine Tartrate(Biovelbin,Eunades,KW2307,NVB,Navelbine,Vinorelbine Ditartrate)
- Paclitaxel(Anzatax,Asotax,Bristaxol,Praxel,Taxol)
- Goserlin Acetate(Zoladex,ZDX)
- Imiquimod(Aldara,R837,S26308,Zyclara)
- Recombinant CD40Ligand,CD154 antigen,CD40L,TBAM,rhu CD40L
- Denileukin Diftitox(DAB389IL2,LY335348,Ontak)
- Rubitecan(Camptogen,Nitrocamptothecin,Orathecin,RFS 2000
- Aldesleukin(Proleukin,Recombinant Human IL2,rserHuIL2)
- Docetaxel (Taxotere,RP56976)
- Letrozole(CGS 20267,Femara)
- Pemetrexed Disodium(Alimta,LY231514)
- Tezacitabine(FMdC,MDL 101,731)
- Romidepsin(Depsipeptide,FK228,FR901228,Istodax)
- Edrecolomab(MOAB171A,Panorex)
- Pegylated Liposomal Doxorubicin Hydrochloride(ATI0918,Caelyx,DOXSL,DOXIL,Doxorubicin HCl Liposomal,Evacet,LipoDox,LipoDox,Lipodox 50,Liposomal Adriamycin,Pegylated Liposomal Doxorubicin Hydrochloride,SLiposomal Doxorubicin,Stealth Liposomal Doxorubicin,TLC D99)
- Dinutuximab(Ch14.18,Unituxin)

- Alvocidib Hydrochloride(Flavopiridol Hydrochloride,HL275,L868275)
- Alitretinoin(ALRT1057,LGD1057,Panretin,Panretyn,Panrexin)
- Eflornithine Hydrochloride,MDL 71782,Ornidyl,RMI71782)
- Daunorubicin Hydrochloride(CERUBIDINE,Ondena,RP13057,Rubidomycin Hydrochloride,Rubilem)
- Idarubicin Hydrochloride(IMI30,Idamycin,SC33428,Zavedos)
- Bicalutamide(Casodex,Cosudex,ICI176334)
- Anastrozole(Arimidex,ICID1033,ZD1033)
- DX521
- KRN5500
- Elinafide(LU79553)
- Becatecarin(BMS181176,BMY2755714,DEAE-Rebeccamycin,NSC655649,Rebeccamycin Analogue,XL119)
- Bexarotene(LGD1069,Targretin)
- Trastuzumab(ABP980,ALT02,Antip185HER2,HER2 Monoclonal Antibody,Herceptin,Herceptin PF05280014,Herzuma, Ogivri, Ontruzant, PF05280014, RO0452317,Trastuzumab ABP 980,Trastuzumab ALT02,Trastuzumab HLX02, Trastuzumab PF05280014, Trastuzumab dkst, cerb2 Monoclonal Antibody, rhuMAb HER2,trastuzumab EG12014, Trastuzumab-pkrb, Trazimera)
- Alemtuzumab(Campath,LDP03,Lemtrada)
- Imatinib Mesylate(CGP57148B, Gleevec, STI571)
- Plitidepsin(APLD, Aplidin,Aplidine,DDB,Dehydrodemnin B)
- Trabectedin(ET743,Ecteinascidin,Yondelis)
- Rituximab(ABP798,BI 695500,C2B8 Monoclonal Antibody,CTP10,IDEC102,IDECC2B8,MabThera,PF05280586,RTXM83,Rituxan,Rituximab ABP 798,Rituximab BI 695500,Rituximab CTP10,Rituximab GB241,Rituximab IBI301,Rituximab PF05280586,Rituximab RTXM83,Rituximab SAIT101,Truxima,rituximab abbs)
- Tipifarnib(R115777,Zarnestra)
- Nelarabine(506U78,Arranon,Compound 506U78,GW506U78)
- Irofulven,HMAF,Hydroxymethylacylfulvene,MGI 114,MGI114
- Cetuximab,Cetuximab CDP1,Cetuximab CMAB009,Cetuximab KL 140,Chimeric MoAb C225,Erbitux,IMCC225
- Vincristine Sulfate,Kyocristine,Leurocristine Sulfate,Leurocristine sulfate,Oncovin,Vincasar,Vincosid,Vincrex)
- Interferon Alfacon1(Advaferon,CIFN,Consensus Interferon,IFN Alfacon1,Infergen,rmetHuIFNCon1)
- ISIS3521(Affinitac, Affinitak, CGP 64128A, LY900003)
- Capecitabine(Ro091978 000,Xeloda)
- Paclitaxel Poliglumex(CT2103, CT2103, PGTXL, Xyotax)
- Vorinostat(SAHA,Suberanilohydroxamic Acid, Zolinza)
- Gemtuzumab Ozogamicin(CDP771,CMA676,Mylotarg)
- Yttrium Y90 Ibritumomab Tiuxetan(IDECY2B8,Y 90 Zevalin)
- Mitomycin(Ametycine,MITO,MITOC,MITOMYCIN C, MitoMedac, Mitocin, MitocinC,Mitolem,MitomycinX,Mitomycine C,Mitosol,Mitozytrex,Mutamycin,Mutamycine,NCIC04706)
- Lonafarnib(SCH66336,Sarasar,lonafarnib)
- Semaxanib(Semoxind, Sugen 541)
- Tegafurgimeraciloteracil Potassium(BMS247616,S1,STS1,Teysuno)
- Cilengitide(EMD121974)
- L778123
- Temsirolimus(CCI779, Torisel)
- Peginterferon Alfa2b(PEG IFN Alfa2b,PEGIntron,CH54031,Sylatron)
- Bortezomib(LDP341,MLN341,PS341,Velcade)
- Gefitinib(Iressa,ZD1839)
- BMS214662
- Panitumumab(ABXEGF, Vectibix)
- Ribozyme RPI.4610(Angiozyme,AntiFlt1 Ribozyme,RPI.4610,RPI.4610,RPI4610,Ribozyme RPI.4610)
- Vatalanib(CGP79787,PTK787,ZK232934)
- Midostaurin(CGP41251, PKC412,Rydapt)
- Lurtotecan Liposome(NX211,OSI211)
- Canertinib Dihydrochloride(CI1033,PD0183805002B,PD183805)
- Epratuzumab(LymphoCide,hLL2)
- NM3
- Recombinant Interferon Alfa2a(Alpha 2 Interferon,IFN alpha2A,Laroferon,A,rHuIFNa 2a)
- Sipuleucel T(APC8015,PA2024 PAP GMCSFLoaded Dendritic Cell Vaccine, Provenge,SipT)
- Brentuximab(CAC10,SGN30)
- Matuzumab(EMD72000)
- Abarelix(PPI149,Plenaxis,R3827)
- Bevacizumab(AntiVEGF rhuMAb, BEVZ92,Bevacizumab-awwb,Bevacizumab BI 695502,Bevacizumab CBT 124,Bevacizumab FKB238,Bevacizumab HD204,Bevacizumab HLX04,Bevacizumab MIL60,Bevacizumab QL 1101,HD204,Avastin,Mvasi)
- Sorafenib Tosylate(BAY439006 Tosylate,BAY549085,Nexavar)
- Liposomal Daunorubicin Citrate(DaunoXome)
- Interleukin12 Gene(IL12 gene,Pralatrexate,FOLOTYN,PDX)
- Testolactone(Fludestrin,SQ9538,Teslac)
- Etanercept(Enbrel,TNFR:Fc)
- gp100 Antigen(glycoprotein 100)
- Tretinoin Liposome(AR623,All transretinoic acid liposomal,Atragen,Tretinoin Liposomal,TretinoinLF)
- Recombinant Vaccinia PSA Vaccine(rVPSA,rVPSA Vaccine)
- Iodine I 131 Tositumomab(Bexxar)
- Aminocamptothecin Colloidal Dispersion
- Carmustine Implant(BCNU Wafer,Carmustine Copolymer,Carmustine Wafer,Gliadel,Gliadel Wafer)
- Tositumomab(AntiCD20 Antibody,MoAb AntiB1)
- Canfosfamide Hydrochloride(TLK286,Telcyta)
- Methyl 5 Aminolevulinate Hydrochloride Cream(Metvix,Metvixia)
- Ipilimumab(BMS734016,MDX010,MDXCTLA4,Yervoy)
- Edotecarin(J107088)
- Lenalidomide(CC5013,CDC501,Revlimid)
- Recombinant FowlpoxProstate Specific Antigen Vaccine(rFPSA)
- Ziv Aflibercept(AFLIBERCEPT,AVE0005,Eylea,VEGF Trap,Zaltrap)

- Nabpaclitaxel(ABI 007,ABI007,Abraxane,Albumin bound Paclitaxel,Nab paclitaxel,Nanoparticle Albumin bound Paclitaxel)
- Erlotinib Hydrochloride(CP358,774,OSI774,Tarceva)
- Vincristine Sulfate Liposome(Marqibo)
- Recombinant Vaccinia(MUC1 Vaccine,rVMUC1 Vaccine)
- Vandetanib(AZD6474,Caprelsa,ZD6474,ZD6474,Zactima)
- Topotecan Hydrochloride(Hycamptamine, Hycamtin,SKF S104864A)
- ONYX015(CI1042)
- Indisulam(E7070)
- Clofarabine(Clofarex,Clolar)
- Eribulin Mesylate(B1939 Mesylate,E7389,ER086526,Halaven)
- Sunitinib Malate(SU011248, Sutent)
- T900607
- Agatolimod Sodium(CpG 7909,PF3512676,ProMune)
- Lapachone(beta Lapachone)
- MART1 Antigen(Antigen LB39AA,Antigen SK29AA,MART1,MLANA,Melan A,MelanA,MelanA Protein)
- Atorvastatin Calcium(CI981,Lipitor)
- Tazarotene(AGN190168,Avage,Tazorac)
- Valproic Acid(Depakene,Stavzor,Valproate)
- 2,6 Diaminopurine(DAP)
- Peginterferon Alfa2a(Pegasys)
- ABT510
- Ixabepilone(mRNA2416)
- Tanespimycin(KOS953)
- PPI2458
- Ispinesib(CK0238273,SB715992)
- Pertuzumab(2C4, Omnitarg, Perjeta, RO4368451, rhuMAb2C4)
- OSI7904L(GS7904L)
- Dasatinib(BMS354825, Sprycel)
- Lorvotuzumab Mertansine(BB10901,IMGN901,huN901DM1)
- Anakinra(Kinaret,rIL1ra,rIL1RN)
- Axitinib(AG013736,Inlyta)
- Recombinant Human Endostatin(Endostar,rhEndostatin)
- Cinacalcet Hydrochloride(Mimpara,Sensipar)
- Motesanib Diphosphate(AMG 706)
- Cediranib Maleate(AZD2171,AZD2171 Maleate,Recentin)
- Bardoxolone(CDDO,RTA401)
- Degarelix(FE200486,Firmagon)
- Everolimus(Afinitor,Certican,RAD001,Votubia,Zortress)
- Retaspimycin Hydrochloride(IPI50)
- Lestaurtinib(CEP701,KT5555,SPM924)
- Tandutinib(CT53518,MLN518)
- Lucatumumab(CHIR1212,HCD122)
- Paclitaxel Loaded Polymeric Micelle(Cynviloq TM,Genexol PM,IG001)
- TPI287
- Volociximab(M200)
- Lenvatinib Mesylate(E7080,Lenvima)
- Belinostat(Beleodaq,PXD101)
- gp100:209217 210M Peptide Vaccine
- Ridaforolimus(AP23573,Deforolimus,MK8669)
- Tremelimumab(ticilimumab, CP675, CP-675206)
- Bavituximab(Tarvacin)
- XL820
- BI2536
- Neratinib(HKI272,PB272,Nerlynx)
- MK0752
- Palbociclib(Ibrance,PD0332991)
- AG024322
- Atiprimod(SK&F106615,azaspirane)
- Figitumumab(CP751871)
- GMK562 Cell Vaccine
- Carfilzomib(Kyprolis,PR171)
- Brivanib Alaninate(BMS582664)
- Dacomitinib(PF00299804, PF0029980403, Vizimpro)
- Veliparib(ABT888)
- Pelareorep(POBB0209,Reolysin,Reovirus Serotype 3,Wildtype Reovirus)
- Pazopanib Hydrochloride(GW786034B,Votrient)
- Bosutinib(SKI606, Bosulif)
- Sagopilone(DE03757,Epothilone ZK219477,SHY03757A,ZKEPO,ZKEpothilone)
- 1018ISS(CPG1018)
- Oportuzumab Monatox(Proxinium,VB4845,Vicinium)
- BRaf VEGFR2 Inhibitor RAF265(CHIR265,RAF265)
- Talimogene Laherparepvec(Imlygic,OncoVEX GMCSF,TVEC)
- Tozasertib Lactate(L001281814,MK0457,VX680)
- Recombinant Thyrotropin Alfa(TSHalpha,Thyrogen)
- Mapatumumab(HGSETR1,HGSETR1,TRM1 mAb)
- Bendamustine Hydrochloride(Bendeka,Cytostasan Hydrochloride,Levact,Ribomustin,SyB L0501,Treanda)
- Chloroquine
- Pegvisomant(Somavert,Trovert)
- Dacetuzumab(SGN40,huS2C6)
- Dovitinib Lactate(CHIR258,TKI258)
- Tivantinib(ARQ197)
- Plinabulin(NPI2358)
- Catumaxomab(Removab)
- Urelumab(BMS663513)
- Tesevatinib(EXEL7647,KD019,KD019,XL647)
- Barasertib(AZD1152,AZD2811)
- Mogamulizumab(KW-0761, KM8761,Poteligeo)
- CDK Inhibitor SNS032(BMS387032)
- Vintafolide(EC145)
- Iniparib(BSI201,SAR240550)
- Blinatumomab(Blincyto,MEDI538,MT103)
- CRLX101(IT101)
- Tertomotide(GV1001,PrimoVax)
- PM00104(Zalypsis)
- Obatoclax Mesylate(GX15070MS)
- Seliciclib(CYC202,R-roscovitine)
- Pegdinetanib(Angiocept,BMS844203,CT322)
- Tasisulam(LY573636)
- Fresolimumab(GC1008)
- Asparaginase Erwinia chrysanthemi(Crisantaspasum,Cristantaspase,Erwinase,Erwinaze)
- ZYC300
- Marizomib,(ML858,NPI0052,Salinosporamide A)
- Crenolanib(CP868596)
- Efatutazone Dihydrochloride(CS7017,Inolitazone Dihydrochloride)
- E2F1 Pathway Activator ARQ171
- Vemurafenib(PLX4032,RG7204,RO5185426,Zelboraf)

- Tagraxofusp-erzs(SL-401, Tagraxofusp, Elzonris, DT388IL3 fusion protein)
- Navitoclax(A855071.0,ABT263)
- Auranofin(Ridaura)
- Gimatecan(LBQ707, ST1481)
- Oblimersen Sodium(Augmerosen,G3139,Genasense)
- Lapatinib Ditosylate(Tykerb)
- Cabazitaxel(Jevtana, RPR116258A,Taxoid XRP6258)
- Brentuximab Vedotin(ADC SGN35, Adcetris, SGN35)
- Panobinostat(Faridak,Farydak,LBH589)
- Ofatumumab(GSK1841157, Arzerra)
- AZD7762
- Elotuzumab(BMS901608,Empliciti,HuLuc63,PDL063)
- Rilotumumab(AMG102)
- Liposome encapsulated Daunorubicin Cytarabine(CPX351,Liposomal AraC Daunorubicin,Vyxeos)
- Nivolumab(BMS936558,MDX1106,NIVO,ONO4538,Opdivo)
- Moxetumomab Pasudotox(HA22, CAT8015, GCR8015, Lumoxiti)
- Abiraterone Acetate(CB7630,Yonsa,Zytiga)
- Abexinostat(CRA024781,PCI24781)
- Aldoxorubicin(DOXOEMCH,DoxorubicinEMCH,INNO206)
- Cobimetinib(Cotellic,GDC0973,XL518)
- Quizartinib(AC010220,AC220)
- Rindopepimut(CDX110,PF04948568)
- PRLX93936
- Obinutuzumab(Afutuzumab,GA101, R7159,RO5072759, Gazyva)
- Ramucirumab(Cyramza,IMC1121B,LY3009806)
- OSI930
- GCS100
- Pidilizumab(CT011,MDV9300)
- Conatumumab(AMG 655,XG1048)
- Anti PSCA Monoclonal Antibody AGS1C4D4(MK4721)
- AVE9633
- Ganitumab(AMG479)
- Inotuzumab Ozogamicin(CMC544,WAY207294, Besponsa)
- Pioglitazone
- Voxtalisib(SAR245409,XL765)
- Pilaralisib(SAR245408,XL147)
- Alisertib(MLN8237)
- Olaparib(AZD2281,KU0059436,Lynparza)
- Apatorsen(ISIS306053,OGX427)
- GS9219
- Enzalutamide(ASP9785, MDV3100, Xtandi)
- Linifanib(ABT869)
- hTERT I540 R572Y D988Y Multipeptide Vaccine
- Pomalidomide(Actimid,CC4047,Imnovid,Pomalyst)
- Bazedoxifene(TSE424,WAY140424)
- AntiCD19 Monoclonal Antibody MDX1342
- Daratumumab(Darzalex,HuMaxCD38,JNJ54767414)
- Apatinib(YN968D1)
- Zoptarelin Doxorubicin(AEZS108,AN152,ZEN008)
- Vismodegib(Erivedge,GDC0449)
- Crizotinib(PF2341066,Xalkori)
- BGT226
- Lanreotide Acetate(Somatuline Depot)
- Corticorelin Acetate(Xerecept,hCRF)
- Tasidotin(ILX651)
- Histrelin Acetate(Supprelin, Vantas)
- Ganetespib(STA9090)
- Glesatinib(MG90265, MG90265X)
- Tucatinib(ARRY380, Irbinitinib, ONT380)
- Pevonedistat(MLN4924)
- Trametinib(GSK1120212, JTP74057, Mekinist)
- ENMD 2076
- Ponatinib Hydrochloride(Iclusig)
- Calaspargase Pegol(EZN2285,Oncaspar IV, SCPEG E Coli L Asparaginase)
- Regorafenib(BAY734506,Stivarga)
- PGG BetaGlucan(Imprime PGG)
- IPH2101
- PF04217903
- Idelalisib(Zydelig,CAL101, GS1101)
- Pacritinib(SB1518)
- BT062
- Patritumab(AMG888, U31287)
- Dinaciclib(MK7965,SCH727965)
- Ammonia N13
- Olaratumab(IMC3G3,Lartruvo)
- Cixutumumab(AntiIGF1R Recombinant MOAB IMCA12)
- Elesclomol Sodium(STA4783)
- Talmapimod(SCIO469)
- BHQ880
- Codrituzumab(GC33)
- Foretinib(GSK1363089,XL880)
- Niraparib(Zejula, MK4827)
- Patidegib(FIN5,IP9 Free Base,IPI926,IPI926 Free Base,Saridegib)
- Talotrexin Ammonium(PT523)
- Dacinostat(NVPLAQ824)
- Ibrutinib(CRA032765,Imbruvica,PCI32765)
- Alpha1 Proteinase Inhibitor Human(A1AT,A1PI,AAT,Alpha 1 Antitrypsin, Aralast, ProlastinC)
- Cold Contaminant-free Iobenguane I 131(Azedra, Ultratrace MIBG)
- GDC0941 Bismesylate
- Lurbinectedin(PM01183)
- PF03084014
- Sonidegib(Erismodegib,LDE225,Odomzo)
- Dabrafenib(Tafinlar, GSK2118436A, GSK2118436)
- AntiKSP AntiVEGF siRNAs ALNVSP02
- BMS863233
- Trastuzumab Emtansine(ADO TRASTUZUMAB EMTANSINE, Kadcyla, PRO132365, RO5304020, TDM1 ,Trastuzumab DM1,Trastuzumab MCCDM1)
- Ixazomib Citrate(MLN9708,Ninlaro)
- TAK901
- MEDI573
- Tovetumab(MEDI575)
- BCG Solution(Bacillus Calmette Guerin Solution,TICE BCG Solution)
- Tocilizumab(Actemra,MRA,R1569)
- Imetelstat Sodium
- Glasdegib(PF04449913, Daurismo)
- Binimetinib(ARRY162,ARRY438162,MEK162,Mektovi)
- Uprosertib(GSK2141795)
- Tivozanib(AV951)
- Onalespib(AT13387)
- Smac Mimetic GDC0152
- Necitumumab(IMC11F8,Portrazza)

- Infigratinib(BGJ398)
- Pexidartinib(PLX3397)
- CXCR4 Antagonist BL8040(BKT140)
- Momelotinib(CYT387,GS0387)
- Vistusertib(AZD2014)
- Sapanisertib(INK128,MLN0128,TAK228)
- Capmatinib(INC280,INCB28060,INCB028060)
- Buparlisib(BKM120)
- Autologous Melanoma Lysate Pulsed Dendritic Cell Vaccine
- XL019
- MK2206
- Smac Mimetic LCL161
- Mitoguazone Dihydrochloride(MGBG 2HCl)
- Urokinase Derived Peptide A6
- Trebananib(AMG386)
- AS1411
- MDM2 Antagonist RO5045337(R7112)
- Adavosertib(AZD1775,MK1775)
- Tabalumab(LY2127399)
- Gedatolisib(PKI587, PF05212384)
- Apalutamide(ARN509,Erleada,JNJ56021927)
- Epacadostat(INCDB024360)
- CP547632
- Alpelisib(BYL719, Piqray)
- AntiCD30 CD16A Monoclonal Antibody AFM13
- Lutetium Lu 177 Dotatate(Lutathera)
- Vosaroxin(AG7352,SPC595)
- AMG337
- Nilotinib Hydrochloride Monohydrate(AMN107, Tasigna)
- Ribociclib(LEE011, Kisqali)
- MEDI3617
- Anti PSMA Monoclonal Antibody MMAE Conjugate
- PI3K Inhibitor ZSTK474
- Merestinib(LY2801653)
- LY2875358
- Talazoparib(BMN673, Talzenna)
- Ulocuplumab(BMS936564,MDX1338)
- Teprotumumab(R1507,RO4858696,RV001)
- Vedolizumab(Entyvio,LDP02,MLN0002,MLN02)
- ARRY382
- Ricolinostat(ACY1215)
- LFA102
- Afatinib Dimaleate(BIBW 2992MA2,Gilotrif)
- PF04136309
- Abemaciclib(LY2835219,Verzenio)
- PWT33597 Mesylate
- AGS22M6E
- Ruxolitinib Phosphate(INCB18424 Phosphate,Jakafi)
- Cabozantinib Smalate(Cabometyx,Cometriq,XL184)
- NS018
- Encorafenib(Braftovi,LGX818)
- Brigatinib(AP26113,Alunbrig)
- AS703988 MSC2015103B
- Varlilumab(CDX1127)
- Duvelisib(Copiktra,INK1197,IPI145)
- Paclitaxel Trevatide(ANG1005,GRN1005)
- Rociletinib(CO1686)
- IMGN529
- Pegylated Recom Lasparaginase Erwinia chrysanthemi(Asparec)
- MLN0264
- Molibresib(GSK525762)
- Alectinib(AF802,Alecensa,CH5424802,RG7853,RO5424802)
- Olmutinib(BI1482694,HM61713)
- Selinexor(KPT-330,Xpovio)
- Trifluridine and Tipiracil Hydrochloride(Lonsurf,TAS102)
- Ensartinib(X396)
- Tisagenlecleucel(CART-19,CTL019,Kymriah)
- Thioureidobutyronitrile(Kevetrin)
- Venetoclax(ABT0199,ABT199,GDC0199,RG7601,Venclexta)
- Durvalumab(MEDI4736, Imfinzi)
- Erdafitinib(JNJ-42756493, Balversa)
- Lirilumab(BMS986015,IPH2102)
- PVX410
- Polatuzumab Vedotin(RG7596,DCDS4501A,FCU 2711,Polivy)
- Umbralisib(RP5264,TGR12020)
- Ulixertinib(BVD523,VRT752271)
- Amcasertib(BBI503)
- Darolutamide(BAY1841788,ODM201,Nubeqa)
- Brilanestrant(ARN810,GDC0810,RO7056118)
- Atezolizumab(MPDL3280A,RG7446,RO5541267,Tecentriq)
- Pembrolizumab(Keytruda,Lambrolizumab,MK3475,SCH 900475)
- GDC0994
- Tazemetostat(Tazverik,E7438,EPZ6438)
- Vadastuximab Talirine(SGNCD33A)
- Enasidenib(AG221,CC90007,Idhifa)
- AKT 1 2 Inhibitor BAY1125976(AY1125976)
- CPI0610
- Relatlimab(BMS986016)
- BET Inhibitor RO6870810(RG6146,TEN010)
- TAK659
- Acalabrutinib(ACP196,Calquence)
- Lorlatinib(PF06463922, Lorbrena)
- UAE Inhibitor TAK243(AOB87172,MLN7243)
- Akt ERK Inhibitor ONC201(TIC10)
- AntiBCMA Conjugate GSK2857916(J6M0mcMMAF)
- Ivosidenib(AG120, Tibsovo)
- Asciminib(ABL001)
- Enfortumab Vedotin(Padcev, ASG22CE)
- TLR7 8 9 Antagonist IMO8400
- Entrectinib(RXDX101,Rozlytrek)
- Nazartinib(EGF816)
- Naquotinib(ASP8273)
- Ceritinib(LDK378, Zykadia)
- Larotrectinib(LOXO 101, ARRY470)
- Osimertinib(AZD9291, Mereletinib, Tagrisso)
- Utomilumab(PF5082566, PF2566)
- Gilteritinib(ASP2215, Xospata)
- DLYE5953A
- Pegylated Liposomal Nanoparticle Docetaxel Prodrug MNK010(MP35491)
- Avelumab(MSB0010718C, Bavencio)
- Taselisib(GDC0032,RO5537381)
- Oral Azacitidine(CC486)
- Altiratinib(DCC2701)
- MGD007
- STM434
- REGN1979

- RO6958688
- Mavelertinib(PF06747775)
- Axicabtagene Ciloleucel(KTEC19, Yescarta)
- CC90002
- VLX1570
- Cemiplimab(REGN2810)
- Spartalizumab(PDR001)
- Citarinostat(ACY241,CC96241,HDACIN2)
- PLX9486
- AntiFGFR3 Monoclonal Antibody B701
- Avapritinib(Ayvakit,BLU285)
- ASTX660
- SC003
- ABBV085
- Bintrafusp Alfa(MSB0011359C,M7824)
- Pegzilarginase(AEB1102,CoArgIPEG)
- Topical Fluorouracil(ActinoHermal, Arumel, Carac, Cytosafe, Efudex, Efurix, Fiverocil, Fluoroplex, Flurox, Timazin, Tolak)
- KTN0158
- IMGN779
- MIW815(ADUS100)
- GS5829
- INCAGN01876
- Apilimod Dimesylate Capsule(LAM 002A)
- AP32788
- TSR042
- AntiGITR Monoclonal Antibody GWN323
- Defactinib Hydrochloride(PF04554878, VS6063)
- LY3300054
- Trastuzumab Deruxtecan(DS8201a,Enhertu,WHO10516)
- H3B8800
- ATR Kinase Inhibitor VX803
- Aldesleukin Prodrug NKTR214
- Cereblon Modulator CC90009
- Rucaparib Camsylate(C0338, Rubraca, Rucaparib Phosphate)
- LY3214996
- PEN221
- Albumin binding Cisplatin Prodrug BTP114
- XMT1522
- BMS986179
- Vecabrutinib(BIIB062,BSK4841,FP182,SNS062)
- LOXO195
- Rituximab and Hyaluronidase Human(Rituxan Hycela)
- Copanlisib Hydrochloride(BAY806946,Aliqopa)
- Icotinib(BPI2009)
- Zanubrutinib(Brukinsa,BGD3111)
- INCB001158(CB1158)
- PF06863135
- SAR439459
- AZD1390
- Anti-TROP2/DXd Antibody-drug Conjugate DS-1062a
- Allogeneic GMCSF secreting Lethally Irradiated Pancreatic Tumor Cell Vaccine
- ASTX029
- DHES0815A
- Other NOS
- Other antineoplastic
- Other hormone
- Clinical Trial Drug not specified

# Appendix 3. Curation Instructions for Medical Oncologist Assessments

- Find the date of diagnosis of the cancer of interest.
- Some patients have more than one cancer diagnosis. **Do** review and curate only notes for the cancer of interest.
- **Do** review and curate one clinical assessment per month, beginning at time of diagnosis.
  - **Do** choose first Medical Oncology note of the month that is authored by an MD actively following the patient for the cancer of interest.
  - If there is no note by a medical oncologist (MD) that month. **Do** use the first note by a nurse practitioner or physician assistant (NP/PA) from a medical oncology practice.
    - **Rarely** look at more than one note per month. If the patient has imaging scans after the first note in the month, look on or up to 7 days following the imaging scans.
  - **Do** give priority to internal visit notes. Only curate notes from an outside institution if there is no DFCI/Partners oncology note in that month. Prioritize as follows:
    1. Internal MD
    2. Internal NP/PA
    3. External MD
    4. External NP/PA
  - If a fellow has a note and an attending physician adds an addendum, review any information in the Summary/impression/Evaluation/Plan, including the addendum.
  - **Do not** use notes from Radiation Oncology. Surgery/Surgical Oncology, inpatient care, primary care, or other specialists not related to cancer (e.g. dermatology, cardiology).
    - Patients with early stage cancers are sometimes only followed by a surgical oncologist. This may mean that they are many months without notes that qualify for curation. That is ok!
  - **Rarely** curate notes by an oncologist from a different specialty than the cancer of interest. Occasionally a patient will transfer care to a different type of oncologist due to the particulars of their disease. For example. a patient with brain metastases from their primary cancer may be followed by Neuro Oncology. **Do** curate these notes if notes from the primary oncologist are not available. **Do not** curate notes from a different oncology specialist that are pertaining to a different cancer diagnosis.
- **Do** review only the Impression/Plan section at the bottom of the note as well as the reason for visit.
  - **Rarely** a medical oncology provider will summarize the cancer status directly above the Impression/Plan section, which can be reviewed for curation.
- **Do not** review any of the other sections. including the physical exam, interval history, lab results. etc.
  - The Impression/Plan section may have a different name in medical oncologist progress notes; other section headers could include Assessment, Summary, Conclusion, Problem List Items Addressed this Visit.
  - When there are no section headings in a provider's note. **Do** review everything located beneath the physical exam.