Analyzing Clinical and Genomic Oncological Data with {genieBPC} and {gnomeR}

R\Medicine Demo
June 7, 2023

Sammi Brown
Jessica Lavery
Karissa Whiting

Agenda


Projects GENIE & GENIE BPC


Clinico-Genomic Data Processing Pipeline


Case study

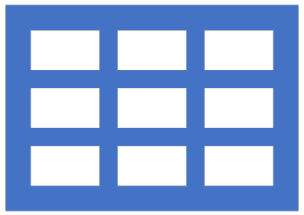

Clinical data processing with {genieBPC}


Genomic data processing with {gnomeR}


Conclusion

Projects GENIE & GENIE BPC

The goal of **Project GENIE Biopharma Collaborative (BPC)** is to augment the existing registry genomic data from AACR Project GENIE with enhanced clinical (phenomic) data to support clinical-genomics analyses.



Phenomic data are curated using the PRISSMM curation model to capture detailed information on cancer diagnosis, drug regimens, disease status from radiology reports, pathology reports and medical oncologist assessments, structured in several datasets with over 700 feature variables.

Analyses using linked clinico-genomic databases – including GENIE BPC – will help to drive advancements in precision oncology in identifying the genomic alterations and drug therapies that optimize clinical outcomes.

Genomic data included in GENIE

Researchers receive genomic data in different formats and types

The AACR Project GENIE data repository is comprised of one type of genomic data called **tumor DNA sequencing assays**

- Collected from tumor samples via biopsy/resection
- Compare DNA sequence in cancer cells with that in normal cells

Sequencing assays can be broad or targeted

- Broad regions: whole genome/whole exome sequencing
- Targeted regions: gene panels
  - GENIE data consists of data from targeted gene panels from high-throughput (huge amounts of data) sequencing assays, also referred to as next-generation sequencing (NGS)

GENIE BPC Data

- Data are publicly released by cancer cohort: non-small cell lung (NSCLC), colorectal (CRC), breast, pancreas, prostate, bladder
- New versions of data are released periodically to include additional patients and variables and to incorporate data corrections
- .csv and .txt data files are available for download from Sage Bionetworks' Synapse data sharing platform
- Downloading each file individually poses challenges for efficient and reproducible workflows

{genieBPC} & {gnomeR} R Packages

The {genieBPC} package is a pipeline to programmatically access the data corresponding to each release from Synapse to support reproducibility, and to create datasets linking clinical and genomic data for analysis.

Created and developed by

Samantha Brown

Michael Curry

Hannah Fuchs

Jessica Lavery

Axel Martin

Dan Sjoberg

Karissa Whiting

The {gnomeR} package provides a consistent framework for genetic data wrangling, processing, visualization and analysis.

Created and developed by

Arshi Arora

Michael Curry

Hannah Fuchs

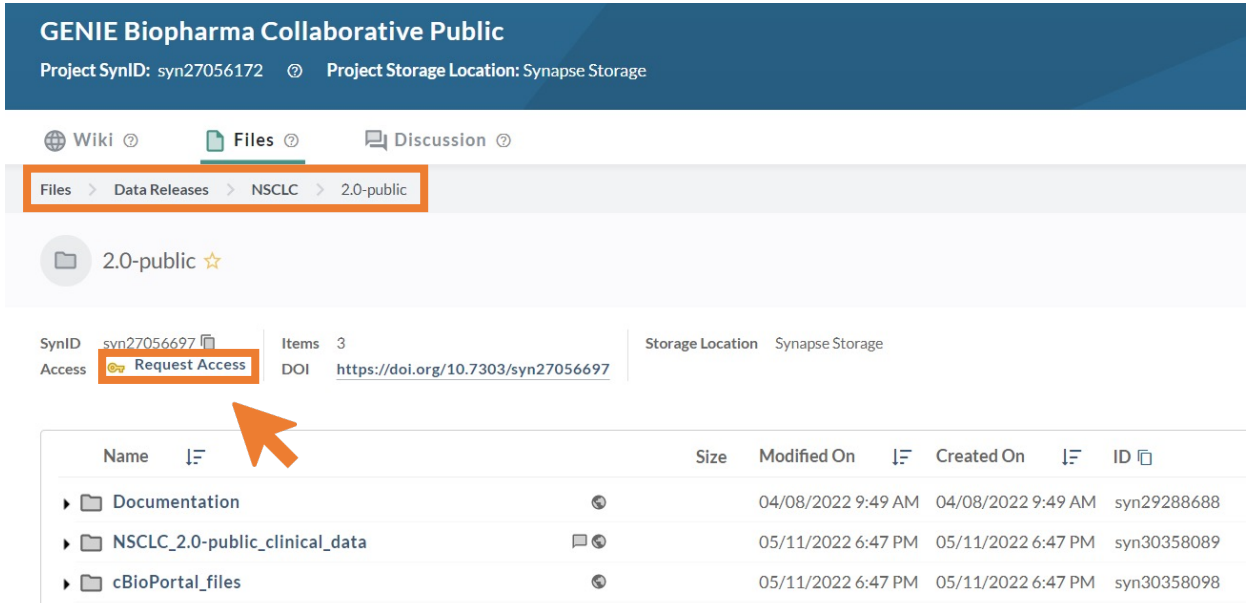
Axel Martin

Karissa Whiting

Register for Synapse Account

Instructions:

1. Register for a 'Synapse' account. Be sure to create a username and password. **Do NOT connect via your Google account.**
   a) https://www.synapse.org/#
2. Accept the **Synapse account terms of use.**
3. Navigate to GENIE Biopharma Collaborative Public page
   a) https://www.synapse.org/#!Synapse:syn27056172/wiki/616601
4. In the Files folder, navigate to Data Releases → NSCLC → 2.0-public
5. Select *Request Access*, review the **terms of data use** and click *Accept*



Installation Instructions

Installing {genieBPC}:

```
install.packages("GENIE-BPC/genieBPC")
```

Installing {gnomeR}:

```
# install.packages("devtools")
devtools::install_github("MSKCC-Epi-Bio/gnomeR")
```

- These instructions are also included in the Demo.R script on our GitHub repository: https://github.com/GENIE-BPC/intro_to_genieBPC_and_gnomeR
- Further R package details are available on the {genieBPC} GitHub repo & website and the {gnomeR} GitHub repo & website
- Note: Both R packages require R version >=3.6

Clinico-Genomic Data Processing Pipeline

Data stored on SYNAPSE

Data import

`pull_data_synapse()`

Imports GENIE BPC data from Synapse into the R environment

`synapse_version()`

Indicates the versions of the data that are available to be specified in `pull_data_synapse()`

Data processing

`create_analytic_cohort()`

Selects an analytic cohort based on cancer diagnosis information and/or cancer-directed drug regimen information

`select_unique_ngs()`

Selects a unique next generation sequencing (NGS) test corresponding to the selected diagnoses

Data visualization

`drug_regimen_sunburst()`

Creates a sunburst figure of drug regimen information corresponding to the selected diagnoses in the order that the regimens were administered

Genomic Processing

{gnomeR}

create_gene_binary()

Processes data on mutation, CNA and fusions into analytic format

tbl_genomic()

Summarizes gene alterations across clinical variables of interest

Case Study

Create a cohort of patients who were diagnosed with Stage IV adenocarcinoma non-small cell lung cancer (NSCLC) and received Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab as their first cancer-directed drug regimen after diagnosis.

Follow along using the Demo.R script on our GitHub repository: https://github.com/GENIE-BPC/intro_to_genieBPC_and_gnomeR

Clinico-Genomic Data Processing Pipeline

Data stored on SYNAPSE

Data import

`pull_data_synapse()`

Imports GENIE BPC data from Synapse into the R environment

`synapse_version()`

Indicates the versions of the data that are available to be specified in `pull_data_synapse()`

Data processing

`create_analytic_cohort()`

Selects an analytic cohort based on cancer diagnosis information and/or cancer-directed drug regimen information

`select_unique_ngs()`

Selects a unique next generation sequencing (NGS) test corresponding to the selected diagnoses

Data visualization

`drug_regimen_sunburst()`

Creates a sunburst figure of drug regimen information corresponding to the selected diagnoses in the order that the regimens were administered

Genomic Processing

{gnomeR}

create_gene_binary()

Processes data on mutation, CNA and fusions into analytic format

tbl_genomic()

Summarizes gene alterations across clinical variables of interest

Set Synapse Credentials

To pull data from Synapse, users must create a Synapse account and store their Synapse credentials in the R environment. The **`set_synapse_credentials()`** function will store credentials during each R session:

```
set_synapse_credentials(username = 'your_username',
                        password = 'your_password')
```

`synapse_version()`

- Helper function that returns a table of GENIE BPC data releases that are currently available
- `synapse_version()` has one input: most_recent = TRUE/FALSE
  - Calling `genieBPC::synapse_version(most_recent = TRUE)` will return a table with each cancer cohort and its latest data release version
  - Calling `genieBPC::synapse_version(most_recent = FALSE)` will return a table with all cancer cohorts and data releases available

`synapse_version()`

- Helper function that returns a table of GENIE BPC data releases that are currently available
- `synapse_version()` has one input: most_recent = TRUE/FALSE
  - Calling `genieBPC::synapse_version(most_recent = TRUE)` will return a table with each cancer cohort and its latest data release version
  - Calling `genieBPC::synapse_version(most_recent = FALSE)` will return a table with all cancer cohorts and data releases available

`synapse_version(most_recent = TRUE)`

cohort	version	release_date	all_versions
BLADDER	v1.1-consortium	November 2022	Most Recent Versions
BrCa	v1.2-consortium	October 2022	Most Recent Versions
CRC	v1.2-consortium	August 2021	Most Recent Versions
CRC	v2.0-public	October 2022	Most Recent Versions
NSCLC	v2.1-consortium	August 2021	Most Recent Versions
NSCLC	v2.0-public	May 2022	Most Recent Versions
PANC	v1.2-consortium	January 2023	Most Recent Versions
Prostate	v1.2-consortium	January 2023	Most Recent Versions

`pull_data_synapse()`

- Pull GENIE BPC clinical and genomic data directly from Synapse into R
- Can specify cancer type (`cohort`) and version of data (`version`)
  - Version of the data is updated periodically on Synapse with re-releases (new variables available, additional QA, etc.)
- Returns a nested list of data frames for each cancer site for the accompanying version

Argument	Description	Acceptable Values
`cohort`	• GENIE BPC Project cancer • Currently, NSCLC and CRC are the only two publicly available datasets	• NSCLC • CRC • BrCa • PANC • Prostate • BLADDER
`version`	Version of the data (e.g v1.1-consortium, v2.0-public)	• Values can be found in `synapse_version()`

Demo: Run **`pull_data_synapse()`** for case study

Demo: Run **`pull_data_synapse()`** for case study

```
library(genieBPC)
```

Demo: Run **`pull_data_synapse()`** for case study

```
library(genieBPC)
set_synapse_credentials()
```

Demo: Run **`pull_data_synapse()`** for case study

```
library(genieBPC)
set_synapse_credentials()
nsclc_synapse_data <- pull_data_synapse(cohort = "NSCLC", version = "v2.0-public")
```

Demo: Run **`pull_data_synapse()`** for case study

```
library(genieBPC)

set_synapse_credentials()

nsclc_synapse_data <- pull_data_synapse(cohort = "NSCLC", version = "v2.0-public")
```

Calling **`nsclc_synapse_data$NSCLC_v2.0`** returns a list of datasets in `nsclc_synapse_data`:

- pt_char
- ca_dx_index
- ca_dx_non_index
- ca_drugs
- prissmm_pathology
- prissmm_imaging
- prissmm_md
- cpt
- mutations_extended
- cna
- fusions

Clinico-Genomic Data Processing Pipeline

Data stored on SYNAPSE

Data import

`pull_data_synapse()`

Imports GENIE BPC data from Synapse into the R environment

`synapse_version()`

Indicates the versions of the data that are available to be specified in `pull_data_synapse()`

Data processing

`create_analytic_cohort()`

Selects an analytic cohort based on cancer diagnosis information and/or cancer-directed drug regimen information

`select_unique_ngs()`

Selects a unique next generation sequencing (NGS) test corresponding to the selected diagnoses

Data visualization

`drug_regimen_sunburst()`

Creates a sunburst figure of drug regimen information corresponding to the selected diagnoses in the order that the regimens were administered

Genomic Processing

{gnomeR}

create_gene_binary()

Processes data on mutation, CNA and fusions into analytic format

tbl_genomic()

Summarizes gene alterations across clinical variables of interest

`create_analytic_cohort()`

Create a cohort from the GENIE BPC data

Cancer diagnosis information such as cancer cohort, treating institution, histology, and stage at diagnosis

Cancer-directed regimen information including regimen name and regimen order.

This function returns all clinical and genomic data for the selected patients

`create_analytic_cohort()`

Argument	Description	Acceptable Values
`data_synapse`	List returned from `pull_data_synapse()`	• Name of object in global environment that was returned from `pull_data_synapse()`

`create_analytic_cohort()`

Argument	Description	Acceptable Values
`data_synapse`	List returned from `pull_data_synapse()`	• Name of object in global environment that was returned from `pull_data_synapse()`
`index_ca_seq`	Index cancer sequence. Default is 1, indicating the patient's first index cancer. This is the cancer that met the eligibility criteria for the project and was selected at random for PRISSMM phenomic data curation.	• Numeric (1+)

`create_analytic_cohort()`

Argument	Description	Acceptable Values
`data_synapse`	List returned from `pull_data_synapse()`	• Name of object in global environment that was returned from `pull_data_synapse()`
`index_ca_seq`	Index cancer sequence. Default is 1, indicating the patient's first index cancer. This is the cancer that met the eligibility criteria for the project and was selected at random for PRISSMM phenomic data curation.	• Numeric (1+)
`institution`	GENIE BPC participating institution. Default selection is all institutions. *Note that not all institutions curated data for all cancer sites.*	• DFCI • MSK • UHN • VICC

create_analytic_cohort()

Argument	Description	Acceptable Values
stage_dx	Stage at diagnosis. Default selection is all stages.	• Stage I • Stage II • Stage III • Stage I-III NOS • Stage IV

create_analytic_cohort()

Argument	Description	Acceptable Values
stage_dx	Stage at diagnosis. Default selection is all stages.	• Stage I • Stage II • Stage III • Stage I-III NOS • Stage IV
histology	Cancer histology. Default selection is all histologies.	• Adenocarcinoma • Squamous cell • Sarcoma • Small cell carcinoma • Other histologies/mixed tumor

create_analytic_cohort()

Argument	Description	Acceptable Values
regimen_drugs	Vector with names of drugs in cancer-directed regimen, separated by a comma. For example, to specify a regimen consisting of Carboplatin and Pemetrexed Disodium, specify regimen_drugs = "Carboplatin, Pemetrexed Disodium".	Acceptable values are found in the drug_names_by_cohort dataset provided with this package.
regimen_type	Indicates whether the regimen(s) specified in regimen_drugs indicates the exact regimen to return, or if regimens containing the drugs listed in regimen_drugs should be returned.	• Exact • Containing

Example: `regimen_drugs` and `regimen_type`

`regimen_drugs`	`regimen_type`	Example regimens returned
Carboplatin	Exact	• Carboplatin
Carboplatin	Containing	• Carboplatin • Carboplatin, Cisplatin • Carboplatin, Paclitaxel • Carboplatin, Pemetrexed Disodium • etc.

create_analytic_cohort()

Argument	Description	Acceptable Values
regimen_order	Order of cancer-directed regimen. If multiple drugs are specified, regimen_order indicates the regimen order for all drugs; different values of regimen_order cannot be specified for different drug regimens.	• Numeric (1+)
regimen_order_type	Specifies whether the 'regimen_order' parameter refers to the order of receipt of the drug regimen within the cancer diagnosis (across all other drug regimens; "within cancer") or the order of receipt of the drug regimen within the times that that drug regimen was administered (e.g. the first time carboplatin pemetrexed was received, out of all times that the patient received carboplatin pemetrexed; "within regimen").	• Within cancer • Within regimen

create_analytic_cohort()

Argument	Description	Acceptable Values
return_summary	Specifies whether summary tables are returned using {gtsummary}. Default is FALSE.	• TRUE • FALSE

Demo: `create_analytic_cohort()` for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

```
nsclc_cohort <- create_analytic_cohort(
```

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
```

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
  stage_dx = c("Stage IV"),
```

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
  stage_dx = c("Stage IV"),
  histology = "Adenocarcinoma",
```

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
  stage_dx = c("Stage IV"),
  histology = "Adenocarcinoma",
  regimen_drugs = c("Carboplatin, Pemetrexed Disodium",
```

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
  stage_dx = c("Stage IV"),
  histology = "Adenocarcinoma",
  regimen_drugs = c("Carboplatin, Pemetrexed Disodium",
                    "Cisplatin, Pemetrexed Disodium",
```

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
  stage_dx = c("Stage IV"),
  histology = "Adenocarcinoma",
  regimen_drugs = c("Carboplatin, Pemetrexed Disodium",
                    "Cisplatin, Pemetrexed Disodium",
                    "Bevacizumab, Carboplatin, Pemetrexed Disodium",
```

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
  stage_dx = c("Stage IV"),
  histology = "Adenocarcinoma",
  regimen_drugs = c("Carboplatin, Pemetrexed Disodium",
                    "Cisplatin, Pemetrexed Disodium",
                    "Bevacizumab, Carboplatin, Pemetrexed Disodium",
                    "Bevacizumab, Cisplatin, Pemetrexed Disodium"),
```

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
  stage_dx = c("Stage IV"),
  histology = "Adenocarcinoma",
  regimen_drugs = c("Carboplatin, Pemetrexed Disodium",
                    "Cisplatin, Pemetrexed Disodium",
                    "Bevacizumab, Carboplatin, Pemetrexed Disodium",
                    "Bevacizumab, Cisplatin, Pemetrexed Disodium"),
  regimen_type = "Exact",
```

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
  stage_dx = c("Stage IV"),
  histology = "Adenocarcinoma",
  regimen_drugs = c("Carboplatin, Pemetrexed Disodium",
                    "Cisplatin, Pemetrexed Disodium",
                    "Bevacizumab, Carboplatin, Pemetrexed Disodium",
                    "Bevacizumab, Cisplatin, Pemetrexed Disodium"),
  regimen_type = "Exact",
  regimen_order = 1,
```

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
  stage_dx = c("Stage IV"),
  histology = "Adenocarcinoma",
  regimen_drugs = c("Carboplatin, Pemetrexed Disodium",
                    "Cisplatin, Pemetrexed Disodium",
                    "Bevacizumab, Carboplatin, Pemetrexed Disodium",
                    "Bevacizumab, Cisplatin, Pemetrexed Disodium"),
  regimen_type = "Exact",
  regimen_order = 1,
  regimen_order_type = "within cancer",
```

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
  stage_dx = c("Stage IV"),
  histology = "Adenocarcinoma",
  regimen_drugs = c("Carboplatin, Pemetrexed Disodium",
                    "Cisplatin, Pemetrexed Disodium",
                    "Bevacizumab, Carboplatin, Pemetrexed Disodium",
                    "Bevacizumab, Cisplatin, Pemetrexed Disodium"),
  regimen_type = "Exact",
  regimen_order = 1,
  regimen_order_type = "within cancer",
  return_summary = TRUE
```

Demo: **`create_analytic_cohort()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_cohort <- create_analytic_cohort(
  data_synapse = nsclc_synapse_data$NSCLC_v2.0,
  stage_dx = c("Stage IV"),
  histology = "Adenocarcinoma",
  regimen_drugs = c("Carboplatin, Pemetrexed Disodium",
                    "Cisplatin, Pemetrexed Disodium",
                    "Bevacizumab, Carboplatin, Pemetrexed Disodium",
                    "Bevacizumab, Cisplatin, Pemetrexed Disodium"),
  regimen_type = "Exact",
  regimen_order = 1,
  regimen_order_type = "within cancer",
  return_summary = TRUE
)
```

nsclc_cohort $tbl_overall_ summary

Characteristic	N = 241 patients[1]
Number of diagnoses per patient in cohort_ca_dx data frame	
1	241 (100%)
Number of regimens per patient in cohort_ca_drugs data frame	
1	241 (100%)
Number of CPTs per patient in cohort_ngs data frame	
1	222 (92%)
2	18 (7.5%)
4	1 (0.4%)

[1] n (%)

nsclc_cohort
$tbl_cohort

Characteristic	N = 241 Diagnoses[1]
Cohort (cohort)	
NSCLC	241 (100%)
Institution (institution)	
DFCI	92 (38%)
MSK	118 (49%)
VICC	31 (13%)
Stage at diagnosis (stage_dx)	
Stage IV	241 (100%)
Histology (ca_hist_adeno_squamous)	
Adenocarcinoma	241 (100%)

[1] n (%)

nsclc_cohort $tbl_drugs

Characteristic	N = 241 Regimens[1]
Cohort (cohort)	
NSCLC	241 (100%)
Institution (institution)	
DFCI	92 (38%)
MSK	118 (49%)
VICC	31 (13%)
Drugs in regimen (regimen_drugs)	
Bevacizumab, Carboplatin, Pemetrexed Disodium	52 (22%)
Bevacizumab, Cisplatin, Pemetrexed Disodium	27 (11%)
Carboplatin, Pemetrexed Disodium	124 (51%)
Cisplatin, Pemetrexed Disodium	38 (16%)

[1] n (%)

nsclc_cohort $tbl_ngs

Characteristic	N = 262 Cancer Panel Tests[1]
Cohort (cohort)	
NSCLC	262 (100%)
Institution (institution)	
DFCI	99 (38%)
MSK	126 (48%)
VICC	37 (14%)
OncoTree code (cpt_oncotree_code)	
LCLC	1 (0.4%)
LUAD	253 (97%)
LUAS	1 (0.4%)
LUSC	1 (0.4%)
NSCLC	4 (1.5%)
NSCLCPD	2 (0.8%)
Sequence assay ID (cpt_seq_assay_id)	
DFCI-ONCOPANEL-1	1 (0.4%)
DFCI-ONCOPANEL-2	57 (22%)
DFCI-ONCOPANEL-3	41 (16%)
MSK-IMPACT341	3 (1.1%)
MSK-IMPACT410	61 (23%)
MSK-IMPACT468	62 (24%)
VICC-01-SOLIDTUMOR	26 (9.9%)
VICC-01-T5A	1 (0.4%)
VICC-01-T7	10 (3.8%)

[1] n (%)

Clinico-Genomic Data Processing Pipeline

Data stored on SYNAPSE

Data import

`pull_data_synapse()`

Imports GENIE BPC data from Synapse into the R environment

`synapse_version()`

Indicates the versions of the data that are available to be specified in `pull_data_synapse()`

Data processing

`create_analytic_cohort()`

Selects an analytic cohort based on cancer diagnosis information and/or cancer-directed drug regimen information

`select_unique_ngs()`

Selects a unique next generation sequencing (NGS) test corresponding to the selected diagnoses

Data visualization

`drug_regimen_sunburst()`

Creates a sunburst figure of drug regimen information corresponding to the selected diagnoses in the order that the regimens were administered

Genomic Processing

{gnomeR}

create_gene_binary()

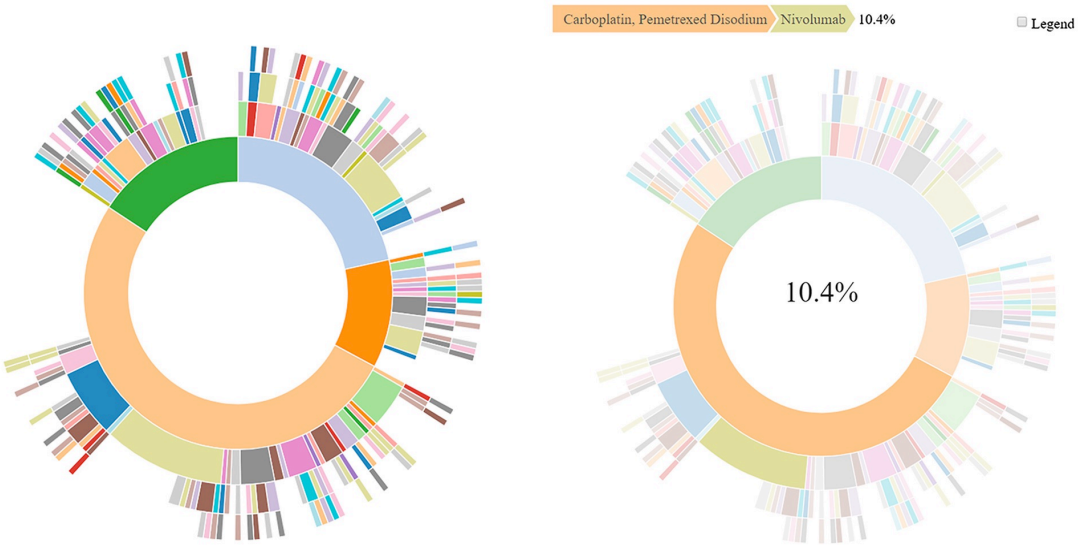
Processes data on mutation, CNA and fusions into analytic format

tbl_genomic()

Summarizes gene alterations across clinical variables of interest

drug_regimen_sunburst()

- Visualize the complete treatment course for selected cancer diagnoses
- Each ring corresponds to a regimen (i.e., innermost ring is first regimen, second innermost ring is second regimen, etc.)
- Interactive figure: Can hover to see regimen names and percent of patients receiving that regimen



drug_regimen_sunburst()

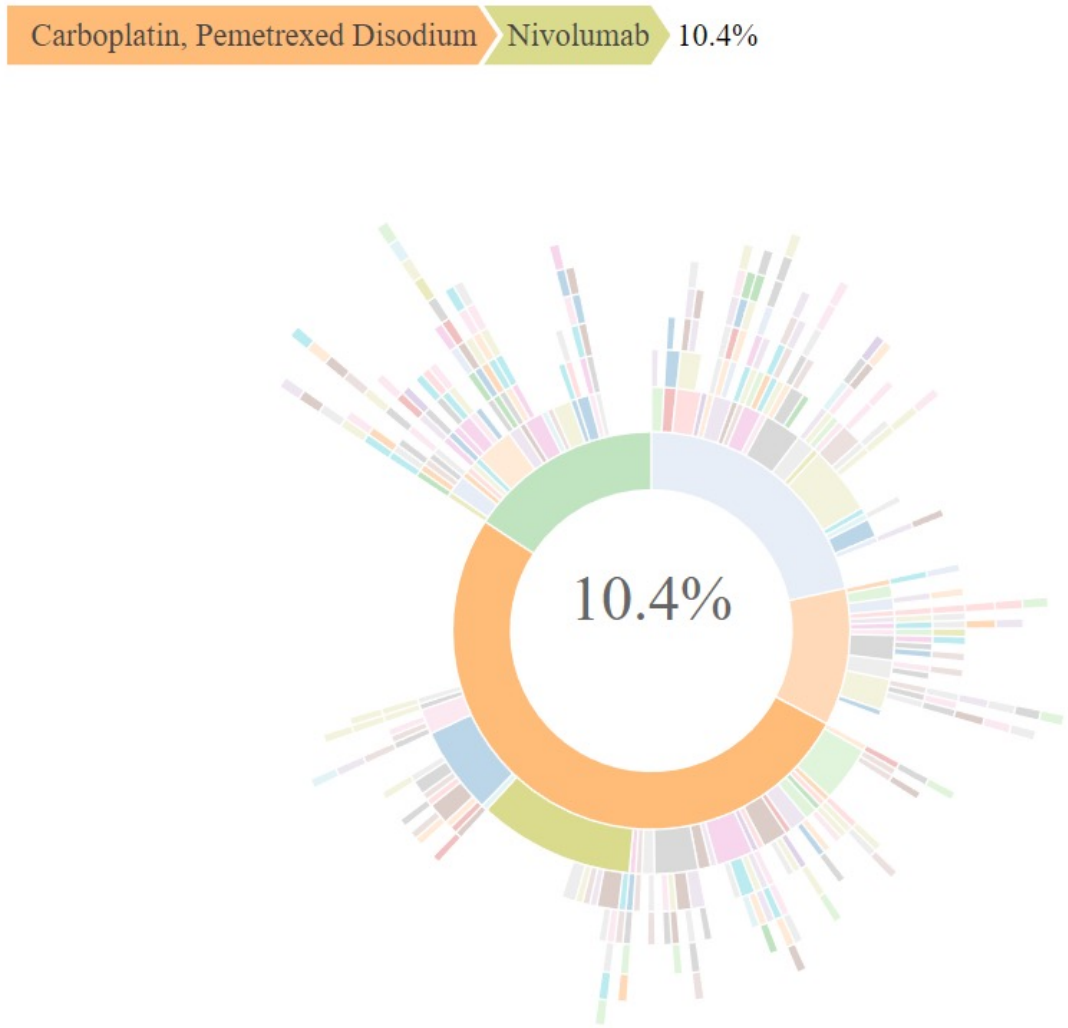
Argument	Description	Acceptable Values
data_synapse	List returned from pull_data_synapse()	• Name of object in global environment that was returned from pull_data_synapse()
data_cohort	The list returned from the create_analytic_cohort() function call	• Name of object in global environment that was returned from create_analytic_cohort()
max_n_regimens	The maximum number of regimens displayed in the sunburst plot	• Integer >0

Demo: **`drug_regimen_sunburst()`** for case study using NSCLC 2.0-public data

Case Study: Create a cohort of patients who were diagnosed with **Stage IV adenocarcinoma NSCLC** and received **Carboplatin and Pemetrexed +/- Bevacizumab or Cisplatin and Pemetrexed +/- Bevacizumab** as their **first** cancer-directed drug regimen after diagnosis

```
nsclc_sunburst <- drug_regimen_sunburst(
            data_synapse = nsclc_synapse_data$NSCLC_v2.0,
            data_cohort = nsclc_cohort)
```

nsclc_sunburst$
sunburst_plot



Genomic Data Processing With {gnomeR}

Clinico-Genomic Data Processing Pipeline

Data stored on SYNAPSE

Data import

`pull_data_synapse()`

Imports GENIE BPC data from Synapse into the R environment

`synapse_version()`

Indicates the versions of the data that are available to be specified in `pull_data_synapse()`

Data processing

`create_analytic_cohort()`

Selects an analytic cohort based on cancer diagnosis information and/or cancer-directed drug regimen information

`select_unique_ngs()`

Selects a unique next generation sequencing (NGS) test corresponding to the selected diagnoses

Data visualization

`drug_regimen_sunburst()`

Creates a sunburst figure of drug regimen information corresponding to the selected diagnoses in the order that the regimens were administered

Genomic Processing

{gnomeR}

create_gene_binary()

Processes data on mutation, CNA and fusions into analytic format

tbl_genomic()

Summarizes gene alterations across clinical variables of interest