

An Introduction to the ISB-CGC Web App

brought to you by

The ISB Cancer Genomics Cloud



Main Landing Page

- Handy place to access documentation, code, and send feedback
- You may only log in using a Google Managed identity by clicking the signin button

<https://isb-cgc.appspot.com>

Cancer Genomics Cloud

The ISB Cancer Genomics Cloud (ISB-CGC) is democratizing access to **TCGA** data and coupling it with unprecedented computational power to allow researchers to explore and analyze this vast data-space.

[Documentation](#)

[GitHub](#)

[Feedback](#)

1. Log into the system:

- The Dashboard provides an overview of the different workbooks and cohorts you create.
- Workbooks contain worksheets, where you can create analyses.
- Gene and Variable favorites is where you can define lists of interest to yourself.
- On top of this, there is a Menu button next to your username that you can use to easily jump from page to page.

Your Dashboard

Saved Workbooks (0)

Workbooks store the Analyses you create -- and their related data.

[+ Create A New Workbook](#)

Saved Cohorts (0)

You don't have any saved Cohorts.

[Create Cohort](#)

Gene Favorites (0)

You don't have any saved Gene Favorites.

[Create Gene Favorites](#)

Variable Favorites (0)

You don't have any saved Variable Favorites.

[Create Variable Favorites](#)

2. Click "Create Cohort"

Your Dashboard

Saved Workbooks (0)

Workbooks store the Analyses you create -- and their related data.

[+ Create A New Workbook](#)

Saved Cohorts (0)

You don't have any saved Cohorts.

[Create Cohort](#)

Gene Favorites (0)

You don't have any saved Gene Favorites.

[Create Gene Favorites](#)

Variable Favorites (0)

You don't have any saved Variable Favorites.

[Create Variable Favorites](#)

3. Cohort Creation

- On the left side, there are panels of features that you can use to define your cohort.
- The Details panel will show you how many samples and participants you currently have selected in your cohort. So initially we start with all of TCGA.
- The Clinical Features panel displays a visual breakdown of a few features of the current cohort you've specified.
- The Public Data Availability panel shows what kind of data is available for your current cohort.
- Note that there are two public data projects: TCGA and CCLE.

Create Cohort

Save As New Cohort

| DONOR | DATA TYPE |
|--|-----------|
| ▼ PUBLIC PROJECTS | |
| <input checked="" type="checkbox"/> TCGA (23688) | |
| <input type="checkbox"/> CCLE (1750) | |
| ▶ PUBLIC STUDIES | |
| ▶ VITAL STATUS | |
| ▶ GENDER | |
| ▶ AGE AT DIAGNOSIS | |
| ▶ SAMPLE TYPE | |
| ▶ TUMOR TISSUE SITE | |
| ▶ HISTOLOGICAL TYPE | |
| ▶ PRIOR DIAGNOSIS | |
| ▶ PATHOLOGIC STAGE | |
| ▶ TUMOR STATUS | |
| ▶ NEW TUMOR EVENT AFTER INITIAL TREATMENT | |
| ▶ HISTOLOGICAL GRADE | |
| ▶ RESIDUAL TUMOR | |
| ▶ TOBACCO SMOKING HISTORY | |
| ▶ ICD-10 | |
| ▶ ICD-O-3 SITE | |

Selected Filters

Clear All

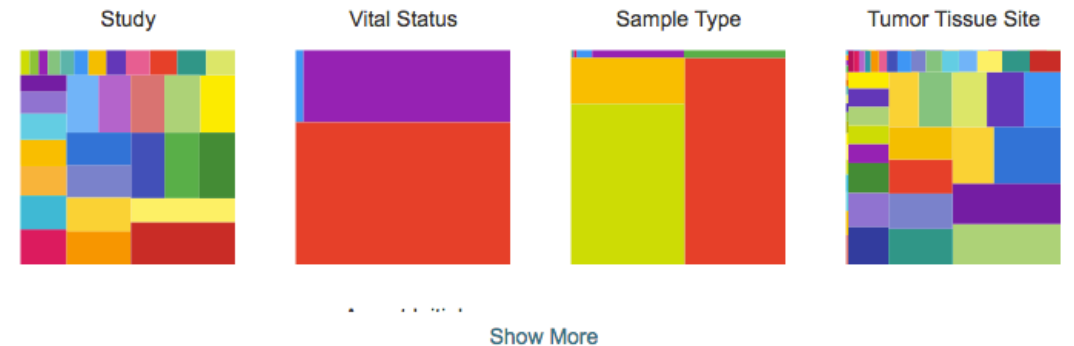
Project: TCGA ✕

Details

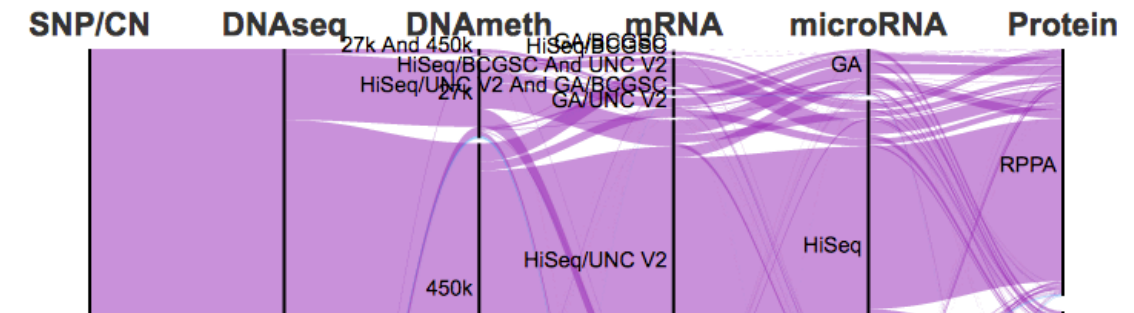
Total Number of Samples: 23688

Total Number of Participants: 11311

Clinical Features



Public Data Availability



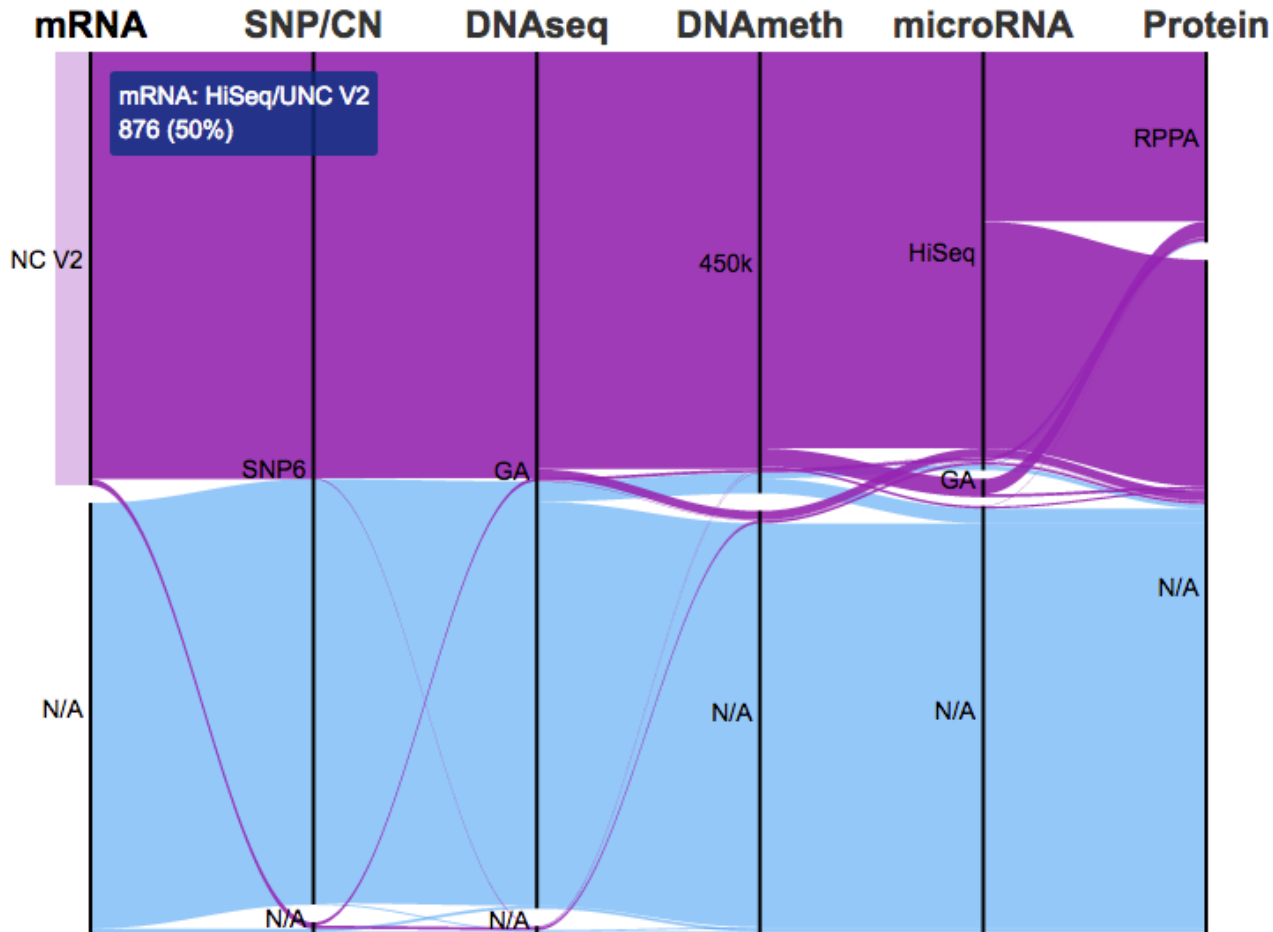
4. Create TCGA Head and Neck (HNSC), and Cervical (CESC) Cohort

| DONOR | DATA TYPE | DONOR | DATA TYPE | DONOR | DATA TYPE |
|--|-----------|---|-----------|---|-----------|
| <input checked="" type="checkbox"/> PUBLIC PROJECTS | | <input checked="" type="checkbox"/> PUBLIC PROJECTS | | <input checked="" type="checkbox"/> PUBLIC PROJECTS | |
| <input checked="" type="checkbox"/> TCGA (23688) | | <input checked="" type="checkbox"/> TCGA (23688) | | <input checked="" type="checkbox"/> TCGA (23688) | |
| <input type="checkbox"/> CCLE (1750) | | <input type="checkbox"/> CCLE (1750) | | <input type="checkbox"/> CCLE (1750) | |
| <input type="checkbox"/> PUBLIC STUDIES | | <input type="checkbox"/> PUBLIC STUDIES | | <input type="checkbox"/> PUBLIC STUDIES | |
| <input type="checkbox"/> VITAL STATUS | | <input type="checkbox"/> BRCA (2403) | | <input type="checkbox"/> BRCA (2403) | |
| <input type="checkbox"/> GENDER | | <input type="checkbox"/> LUSC (1425) | | <input type="checkbox"/> LUSC (1425) | |
| <input type="checkbox"/> AGE AT DIAGNOSIS | | <input type="checkbox"/> LUAD (1301) | | <input type="checkbox"/> LUAD (1301) | |
| <input type="checkbox"/> SAMPLE TYPE | | <input type="checkbox"/> OV (1300) | | <input type="checkbox"/> OV (1300) | |
| <input type="checkbox"/> TUMOR TISSUE SITE | | <input type="checkbox"/> HNSC (1186) | | <input type="checkbox"/> HNSC (1186) | |
| <input type="checkbox"/> HISTOLOGICAL TYPE | | <input type="checkbox"/> GBMLGG (1174) | | <input type="checkbox"/> GBMLGG (1174) | |
| <input type="checkbox"/> PRIOR DIAGNOSIS | | <input type="checkbox"/> LGG (1174) | | <input type="checkbox"/> LGG (1174) | |
| <input type="checkbox"/> PATHOLOGIC STAGE | | <input type="checkbox"/> KIRC (1144) | | <input type="checkbox"/> KIRC (1144) | |
| <input type="checkbox"/> TUMOR STATUS | | <input type="checkbox"/> UCEC (1134) | | <input type="checkbox"/> UCEC (1134) | |
| <input type="checkbox"/> NEW TUMOR EVENT AFTER INITIAL TREATMENT | | <input type="checkbox"/> COAD (1097) | | <input type="checkbox"/> COAD (1097) | |
| <input type="checkbox"/> HISTOLOGICAL GRADE | | <input type="checkbox"/> PRAD (1076) | | <input type="checkbox"/> PRAD (1076) | |
| <input type="checkbox"/> RESIDUAL TUMOR | | <input type="checkbox"/> THCA (1075) | | <input type="checkbox"/> THCA (1075) | |
| <input type="checkbox"/> TOBACCO SMOKING HISTORY | | <input type="checkbox"/> SKCM (1041) | | <input type="checkbox"/> SKCM (1041) | |
| <input type="checkbox"/> ICD-10 | | <input type="checkbox"/> STAD (1018) | | <input type="checkbox"/> STAD (1018) | |
| <input type="checkbox"/> ICD-O-3 SITE | | <input type="checkbox"/> BLCA (895) | | <input type="checkbox"/> BLCA (895) | |
| <input type="checkbox"/> ICD-O-3 HISTOLOGY | | <input type="checkbox"/> LIHC (859) | | <input type="checkbox"/> LIHC (859) | |
| | | <input type="checkbox"/> CESC (670) | | <input type="checkbox"/> CESC (670) | |
| | | <input type="checkbox"/> KIPAN (614) | | <input type="checkbox"/> KIPAN (614) | |
| | | <input type="checkbox"/> SARC (602) | | <input type="checkbox"/> SARC (602) | |
| | | <input type="checkbox"/> PAAD (458) | | <input type="checkbox"/> PAAD (458) | |
| | | <input type="checkbox"/> ESCA (424) | | <input type="checkbox"/> ESCA (424) | |
| | | <input type="checkbox"/> PCPG (366) | | <input type="checkbox"/> PCPG (366) | |
| | | <input type="checkbox"/> READ (349) | | <input type="checkbox"/> READ (349) | |
| | | <input type="checkbox"/> TGCT (306) | | <input type="checkbox"/> TGCT (306) | |
| | | <input type="checkbox"/> THYM (250) | | <input type="checkbox"/> THYM (250) | |
| | | <input type="checkbox"/> DLBC (227) | | <input type="checkbox"/> DLBC (227) | |
| | | <input type="checkbox"/> KICH (226) | | <input type="checkbox"/> KICH (226) | |

- For the purposes of this analysis, we will create a cohort comprised of all TCGA Head and Neck and Cervical samples.
- To do this we select those from the Public Studies.
- It is important to note that if we had not selected the TCGA Project, our cohort could include samples that are also from the CCLE Project.

5. Let's look at data availability for this cohort

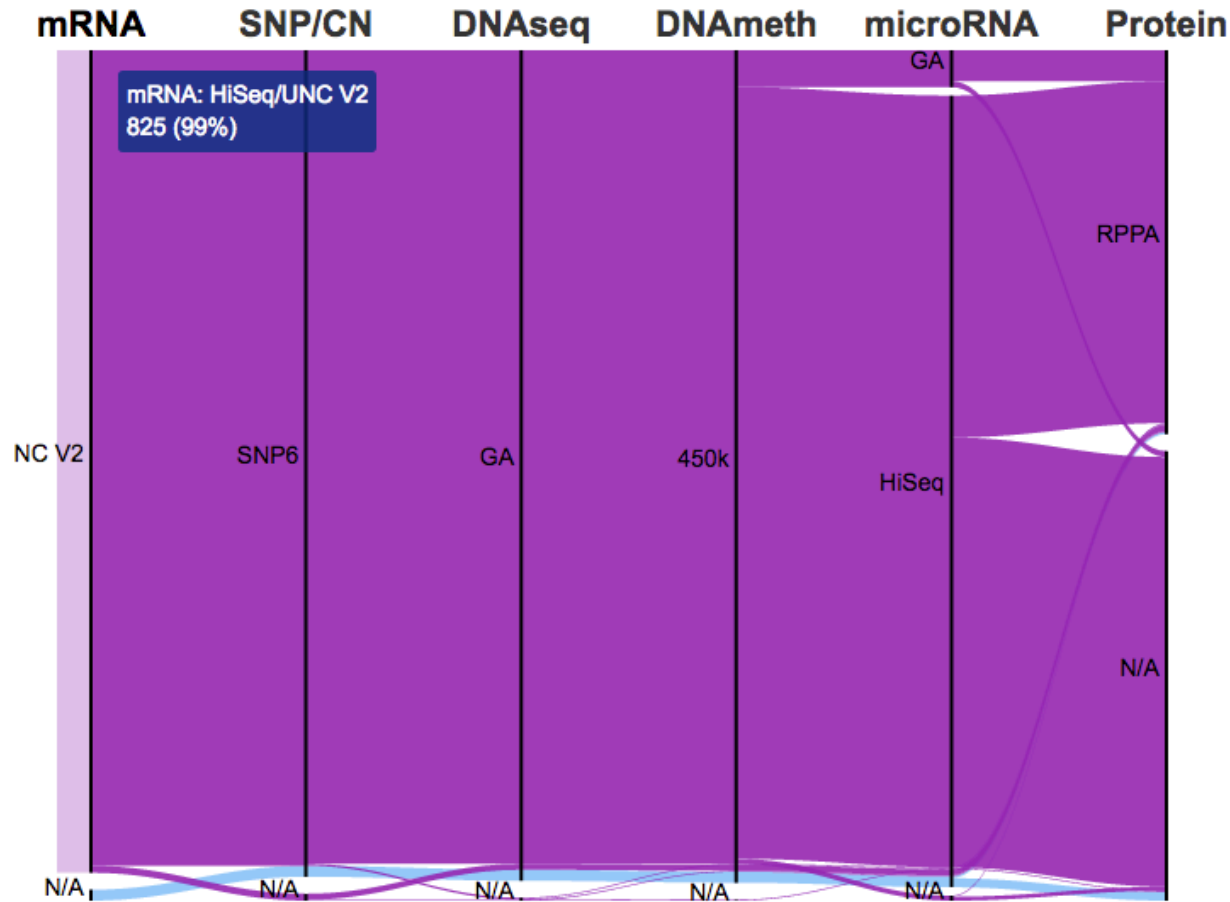
Public Data Availability



- This is called a parallel sets graph. It shows the distribution of data for the samples selected.
- 50% of our participants have HiSeq/UNC V2 gene expression data available
- Of those 876 samples, we can see that a large portion of them have SNP6 data, and a small sliver do not.
- Of the samples that have both HiSeq/UNC V2 and SNP6 data, another large portion also have DNaseq: GA data.
- The data availability graph can be re-ordered based on what you're most interested in. Here, we use gene expression data as our main focus.

6. Select the Sample Type 'Primary tumor Tissue'

Public Data Availability



- Notice that now 99% of our samples have HiSeq/UNC V2 gene expression data.
- After selecting only Primary tumor Tissue, we can see that most of our samples have gene expression data.

7. The resulting cohort

Your Dashboard > Cohorts >

Create Cohort

Save As New Cohort

| DONOR | DATA TYPE |
|---------------------------------|-----------|
| ▶ PUBLIC PROJECTS | |
| ▶ PUBLIC STUDIES | |
| ▶ VITAL STATUS | |
| ▶ GENDER | |
| ▶ AGE AT DIAGNOSIS | |
| ▶ SAMPLE TYPE | |
| ▶ TUMOR TISSUE SITE | |
| ▶ HISTOLOGICAL TYPE | |
| ▶ PRIOR DIAGNOSIS | |
| ▶ PATHOLOGIC STAGE | |
| ▶ TUMOR STATUS | |
| ▶ NEW TUMOR EVENT AFTER INITIAL | |

Selected Filters

Clear All

Project: TCGA ✕

Study: HNSC ✕

Study: CESC ✕

SampleTypeCode: 01 ✕

Details

Total Number of Samples: 836

Total Number of Participants: 836

Clinical Features



Show More

8. Save the cohort and provide it a name: TCGA Head and Neck, and Cervical

Your Dashboard > Cohorts >

Create Cohort

Save As New Cohort

| DONOR | DATA TYPE |
|---------------------------------|-----------|
| ▶ PUBLIC PROJECTS | |
| ▶ PUBLIC STUDIES | |
| ▶ VITAL STATUS | |
| ▶ GENDER | |
| ▶ AGE AT DIAGNOSIS | |
| ▶ SAMPLE TYPE | |
| ▶ TUMOR TISSUE SITE | |
| ▶ HISTOLOGICAL TYPE | |
| ▶ PRIOR DIAGNOSIS | |
| ▶ PATHOLOGIC STAGE | |
| ▶ TUMOR STATUS | |
| ▶ NEW TUMOR EVENT AFTER INITIAL | |

Selected Filters

Clear All

Project: TCGA ✕ Study: HNSC ✕ Study: CESC ✕ SampleTypeCode: 01 ✕

Details

Total Number of Samples: 836

Total Number of Participants: 836

Clinical Features



Show More

9. Cohort Listing Page

- This is where you can see all of the cohorts you've created and that have been shared with you.
- Notice that you also have access to Public Cohorts. These are cohorts that we've created for you. So far it's just one, but we plan on adding more.
- Another way that we could have created our cohort is by taking the union of two previously created cohorts. In this example, you can see that there is already a TCGA HNSC and TCGA CESC cohort. I could select those and click the Set Operations button, We currently support Unions, Intersects, and Set Complements.
- To start an analysis, we're going to select our cohort and click the New Workbook button. We're going to use this cohort and explore differential gene expression conditional on HPV Status.

Your Dashboard >

Cohorts

[+ Create New Cohort](#)

SAVED COHORTS PUBLIC COHORTS

[New Workbook](#) [Delete](#) [Set Operations](#) [Share](#)

| <input type="checkbox"/> | Cohort Name | # Samples | # Patients | Ownership | Shared With | Last Modified | ^ |
|--------------------------|----------------------------------|-----------|------------|-----------|-------------|-----------------------|---|
| <input type="checkbox"/> | TCGA Head and Neck, and Cervical | 836 | 836 | You | (0) | 05/18/2016 9:32 a.m. | |
| <input type="checkbox"/> | TCGA HNSC CESC | 1744 | 836 | You | (0) | 05/16/2016 12:13 p.m. | |
| <input type="checkbox"/> | TCGA HNSC | 1123 | 528 | You | (0) | 05/16/2016 12:12 p.m. | |
| <input type="checkbox"/> | TCGA CESC | 621 | 308 | You | (0) | 05/12/2016 12:02 p.m. | |
| <input type="checkbox"/> | DNA Meth | 213 | 214 | You | (0) | 04/12/2016 2:30 p.m. | |
| <input type="checkbox"/> | GBM | 1174 | 605 | You | (0) | 03/14/2016 11:10 a.m. | |

10. New Workbook

- When you create a new workbook, it is automatically populated with one worksheet.
- A worksheet is comprised of different data sources that you will use in your analysis. You can see that the Cohort we selected is already available.
- Let's first edit some details of our workbook by giving it a more meaningful name and then a short description.

Your Dashboard > Saved Workbooks >

Untitled Workbook

This is a workbook created with cohorts added to the first worksheet. Click Edit Details to change your workbook title and description.

Edit Details

Duplicate

Delete

Share

Shared With (0)

Worksheet 1



Comments (0)

Source Data

Genes



Variables



Cohorts



✕ TCGA Head and Neck, and Cervical

Analysis Type

-- select an analysis --

⚙ Edit Analysis Settings

📌 To Complete this Analysis:

- You must select an Analysis Type (above)
- You must select Genes or Variables (or, optionally, both)
- You must select a Cohorts

Resubmit Plot

11. Add Variables to your worksheet

Your Dashboard > Saved Workbooks >

Untitled Workbook

This is a workbook created with cohorts added to the first worksheet. Click Edit Details to change your workbook title and description.

[Edit Details](#) [Duplicate](#) [Delete](#) | [Share](#) Shared With (0)

Worksheet 1 +

Comments (0)

Source Data

- Genes +
- Variables** +
- Cohorts +
- ✕ TCGA Head and Neck, and Cervical

Analysis Type

-- select an analysis --

⚙ Edit Analysis Settings

ⓘ To Complete this Analysis:

- You must select an Analysis Type (above)
- You must select Genes or Variables (or, optionally, both)
- You must select a Cohorts

[Resubmit Plot](#)

12. Creating a new Variable List

- If you don't already have variable lists created, you will be taken here. If you do, then you will be taken to the your list of previously created variable lists. To get to this page, click the Apply New Variable List button.
- The idea behind this concept is for you to be able to create a list of variables you might use in your analysis and save it all together. It will also allow you to reuse that list in other analyses.
- Here, you can select variables that are **not** gene specific, so mainly clinical and miRNA.

Data Source | Variables

Apply To Worksheet

Back To Workbook

Name of Favorite (Required)

My new favorite variable

COMMON

FAVORITES (8)

CLINICAL

MIRNA

VITAL STATUS

GENDER

AGE AT DIAGNOSIS

TUMOR TISSUE SITE

HISTOLOGICAL TYPE

PRIOR DIAGNOSIS

TUMOR STATUS

NEW TUMOR EVENT AFTER INITIAL TREATMENT

HISTOLOGICAL GRADE

RESIDUAL TUMOR

TOBACCO SMOKING HISTORY

ICD-10

ICD-O-3 SITE

ICD-O-3 HISTOLOGY

Selected Variables

Clear All

Select your favorite variables from the left panel.

13. Provide a name and select the following variables from the Common tab.

- We provide a name for our variable list: HPV Variables
- And select the following variables on the common tab:
 - Vital Status
 - Gender
 - Age at Diagnosis
 - Tumor Tissue Site
 - Histological Type
 - Prior Diagnosis
 - Tumor Status
 - Tobacco Smoking History
- You'll notice that they will appear in the Selected Variables panel.

The screenshot shows a web interface for selecting variables. At the top, there is a text input field labeled "Name of Favorite (Required)" containing the text "HPV Variables". Below this is a list of variable categories: "COMMON", "FAVORITES (8)", "CLINICAL", and "MIRNA". The "COMMON" tab is active. A list of variables is shown with checkboxes: VITAL STATUS, GENDER, AGE AT DIAGNOSIS, TUMOR TISSUE SITE, HISTOLOGICAL TYPE, PRIOR DIAGNOSIS, TUMOR STATUS, NEW TUMOR EVENT AFTER INITIAL TREATMENT, HISTOLOGICAL GRADE, RESIDUAL TUMOR, TOBACCO SMOKING HISTORY, ICD-10, ICD-O-3 SITE, and ICD-O-3 HISTOLOGY. The first seven variables are checked. To the right, a "Selected Variables" panel contains a "Clear All" button and a list of the selected variables: Vital Status, Gender, Age at Diagnosis, Tumor Tissue Site, Histological Type, Prior Diagnosis, Tumor Status, and Tobacco Smoking History.

14. Select HPV Calls, HPV Status, and Study from the Clinical tab

Name of Favorite (Required)

HPV Variables

COMMON FAVORITES (8) **CLINICAL** MIRNA

Feature Search

Study

Selected Variables Clear All

Select your favorite variables from the left panel.

Vital Status ✕ Gender ✕ Age at Diagnosis ✕ Tumor Tissue Site ✕

Histological Type ✕ Prior Diagnosis ✕ Tumor Status ✕

Tobacco Smoking History ✕ Hpv Calls ✕ Hpv Status ✕ Study ✕

This is an autocomplete box, so try typing in 'hpv' to get the HPV specific variables

- We also want some less common clinical variables, so we move on to the Clinical tab.
- Here we can start typing in the variable we're interested in. In our case it's 'hpv'
- To get the Study variable, try using just part of the work like 'tud'
- We hit save and are brought back to the worksheet.

15. Save the list by clicking the "Apply to Worksheet" button

16. Add Genes to your worksheet

Your Dashboard > Saved Workbooks >

HPV Workbook

Edit Details

Duplicate

Delete

Share

Shared With (0)

Worksheet 1



Comments (0)

Source Data

Genes +

Variables +

✗ Vital Status

✗ Gender

✗ Age at Diagnosis

✗ Tumor Tissue Site

✗ Histological Type

✗ Prior Diagnosis

Analysis Type

-- select an analysis --

⚙ Edit Analysis Settings

📌 To Complete this Analysis:

- You must select an Analysis Type (above)
- You must select **Genes** or **Variables** (or, optionally, both)
- You must select a Cohorts

Resubmit Plot

17. Create a gene list for your HPV analysis

Your Dashboard > Saved Workbooks > HPV Workbook > Saved Gene Favorites >

Create Gene List

Name of Favorite (Required)
HPV Genes

Selected Genes (required) Clear All | Upload Genes List

PVT1 x RAD51L1 x TMPRSS3 x ERBB2 x FN1 x SERPINB11 x Enter your favorite gene's name

Apply To Worksheet Cancel View Gene Identifiers

This is an autocomplete box, so try typing in 'RAD51'

- Similarly to variables, if you have gene lists created, you will be taken to a listing of your gene lists.
- If you're unsure of what your gene might be called, you can use the View Gene Identifiers to help.
- We are going to use this list of genes:
 - PVT1, RAD51L1, TMPRSS3, ERBB2, FN1, SERPINB11
- We provide a name, and click the Apply To Worksheet button.

18. When complete, click 'Apply to Worksheet' to save and return to your workbook

19. Creating a Violin Plot comparing HPV Status VS PVT1 Gene Expression

Worksheet 1 +

Source Data

Genes

- ERBB2
- FN1
- PVT1
- SERPINB11
- RAD51L1
- TMPRSS3

Variables

- Vital Status
- Gender
- Age at Diagnosis
- Tumor Tissue Site
- Histological Type
- Prior Diagnosis
- Tobacco Smoking History
- Tumor Status
- Residual Tumor
- Hpv Calls
- Hpv Status
- Study

Cohorts

- TCGA Head and Neck, and Cervical

Analysis Type

Violin Plot

Edit Analysis Settings

To Complete this Analysis:

- You must select an Analysis Type (above)
- You must select Genes or Variables (or, optionally, both)
- You must select a Cohorts

Resubmit Plot

Plot Settings

Comments (0)

X Axis Variable

-- select a variable--

Y Axis Variable

-- select a variable--

Color By Feature

-- select a variable--

Cohorts

TCGA Head and Neck, and Cervical

Update Plot

- We provide several different types of analyses (For more information please see our online documentation):
 - Barchart – 1 Categorical variable
 - Histogram – 1 Numerical variable
 - Scatterplot – 2 Numerical variables
 - Violin Plot – 1 Categorical and 1 Numerical variable
 - Cubby Hole Plot – 2 Categorical variables
 - SeqPeek – 1 Gene
- We want to plot HPV Status VS Gene Expressions for PVT1. Since that is a categorical feature versus a numerical feature, we choose a violin plot.

19. Creating a Violin Plot comparing HPV Status VS PVT1 Gene Expression

Worksheet 1 +

Comments (0)

Source Data

Genes +

- ERBB2
- FN1
- PVT1
- SERPINB11
- RAD51L1
- TMPRSS3

Variables +

- Vital Status
- Gender
- Age at Diagnosis
- Tumor Tissue Site
- Histological Type
- Prior Diagnosis
- Tobacco Smoking History
- Tumor Status
- Residual Tumor
- Hpv Calls
- Hpv Status
- Study

Cohorts +

- TCGA Head and Neck, and Cervical

Analysis Type

Violin Plot

Edit Analysis Settings

To Complete this Analysis:

- You must select an Analysis Type (above)
- You must select Genes or Variables (or, optionally, both)
- You must select a Cohorts

Resubmit Plot

Plot Settings x

X Axis Variable

Hpv Status

Y Axis Variable

-- select a variable--

Color By Feature

Please select an option

Cohorts

TCGA Head and Neck, and Cervical

Update Plot

Select 'PVT1'

19. Creating a Violin Plot comparing HPV Status VS RAD51L1 Gene Expression

Worksheet 1 +

Comments (0)

Source Data

Genes +

- ERBB2
- FN1
- PVT1
- SERPINB11
- RAD51L1
- TMPRSS3

Variables +

- Vital Status
- Gender
- Age at Diagnosis
- Tumor Tissue Site
- Histological Type
- Prior Diagnosis
- Tobacco Smoking History
- Tumor Status
- Residual Tumor
- Hpv Calls
- Hpv Status
- Study

Cohorts +

- TCGA Head and Neck, and Cervical

Analysis Type

Violin Plot

Edit Analysis Settings

To Complete this Analysis:

- You must select an Analysis Type (above)
- You must select Genes or Variables (or, optionally, both)
- You must select a Cohorts

Resubmit Plot

Plot Settings x

X Axis Variable

Hpv Status

Y Axis Variable

PVT1

Gene Expression

Platform Filter

Illumina HiSeq

Center Filter

UNC

Select Feature

PVT1 mRNA (Illumina HiSeq, UNC RSEM)

Color By Feature

Study

Cohorts

- TCGA Head and Neck, and Cervical

- Gene Expression
- Platform: Illumina HiSeq
- Center: UNC
- Feature: PVT1 mRNA (Illumina HiSeq, UNC RSEM)

- Without specifying the platform and filter, we could end up with a lot of potential variables to plot, but can only pick one at a time.

19. Creating a Violin Plot comparing HPV Status VS RAD51L1 Gene Expression

Worksheet 1

Comments (0)

Source Data

Genes

- ERBB2
- FN1
- PVT1
- SERPINB11
- RAD51L1
- TMPRSS3

Variables

- Vital Status
- Gender
- Age at Diagnosis
- Tumor Tissue Site
- Histological Type
- Prior Diagnosis
- Tobacco Smoking History
- Tumor Status
- Residual Tumor
- Hpv Calls
- Hpv Status
- Study

Cohorts

- TCGA Head and Neck, and Cervical

Analysis Type

Violin Plot

Edit Analysis Settings

To Complete this Analysis:

- You must select an Analysis Type (above)
- You must select Genes or Variables (or, optionally, both)
- You must select a Cohorts

Resubmit Plot

Plot Settings

HpV Status

Y Axis Variable

PVT1

Gene Expression

Platform Filter

Illumina HiSeq

Center Filter

UNC

Select Feature

PVT1 mRNA (Illumina HiSeq, UN

Color By Feature

Study

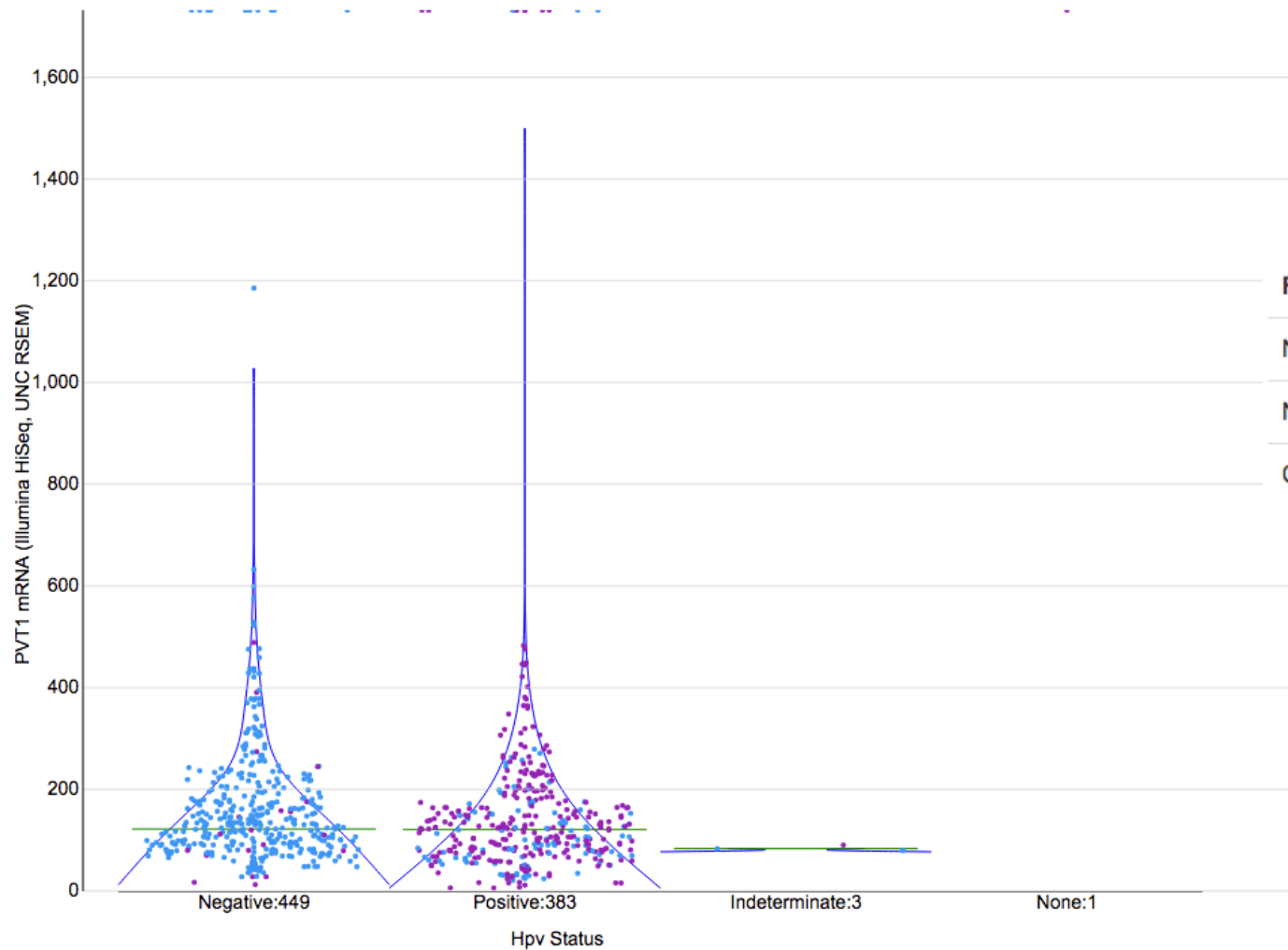
Cohorts

- TCGA Head and Neck, and Cervical

Update Plot

- Color By: Study
- Select Cohort
- Update Plot
- The violin plot will show each sample as a dot. By adding a color by, we are able to see an extra dimension of data.
- We also select the cohort we're interested in. If you had multiple cohorts in your data sources, you can select more than one.
- And we click the Update Plot button.

19. Creating a Violin Plot comparing HPV Status VS RAD51L1 Gene Expression



Legend



- This is the resulting violin plot.
- You can see that there are a lot more CESC samples that are HPV Positive

| Feature 1 | Feature 2 | logp | n |
|-------------------------------------|-------------------|---------|-----|
| N:GEXP:PVT1:mrna_unc_illumina_hiseq | C:CLIN:hpv_status | 0.752 | 815 |
| N:GEXP:PVT1:mrna_unc_illumina_hiseq | C:CLIN:Study | 0.378 | 815 |
| C:CLIN:hpv_status | C:CLIN:Study | 105.326 | 832 |

- Sample pairwise results for the features selected.