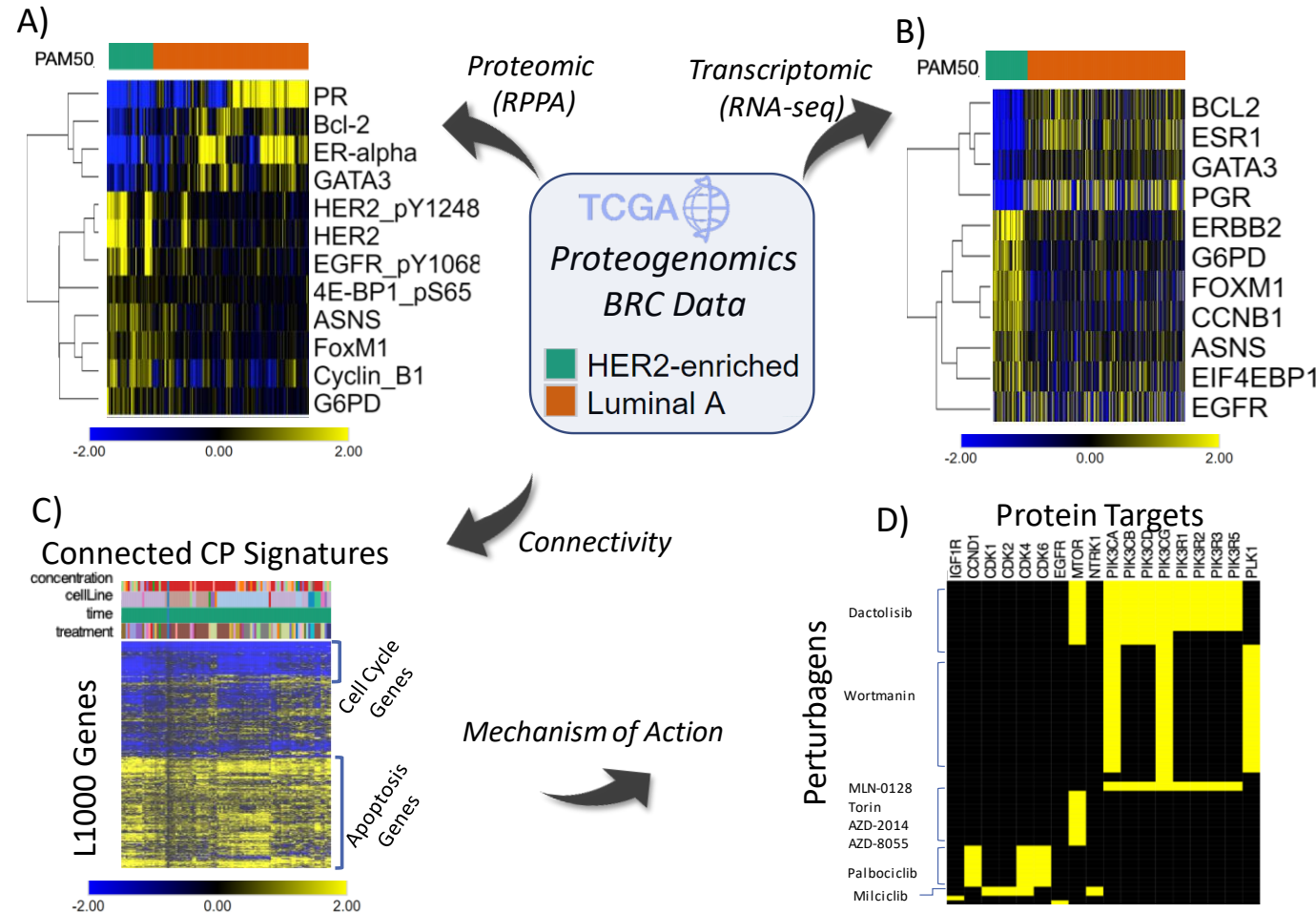


Use case 3: Proteo-genomics analysis of cancer driver events in breast cancer

The use case describes the analysis of TCGA breast cancer RNA-seq and RPPA data using the iLINCS “Datasets” workflow to construct and analyze the differential gene and protein expression signatures contrasting Luminal A and Her2 enriched (Her2E) breast tumors. The analysis results are depicted in the Figure. A) Most differentially expressed proteins in the proteomics signatures constructed by comparing RPPA profiles of Her2E and Luminal A BRC samples; B) Gene expression profile of the genes corresponding to proteins in A) based on RNA-seq data; C) Top 100 CP signatures most connected with the transcriptional signature constructed by comparing RNA-seq profiles of Her2E and Luminal A samples; D) Selected chemical perturbagens and their targets for CP signatures in C).





1. Click "Datasets" tab on the top of iLINCS homepage to start with the "Datasets" workflow. The workflow allows you to select a dataset and create a differential expression gene/protein signature.

Datasets

All Datasets

LINCS Datasets

Find Genomics Datasets

Organism

All

Sample Type

All

Data Type

All

Collection

All

Keyword

Search

Search in portals...

Search

Name

Description

LINCS

Datasets generated by LINCS data and signature generation centers

Choose

TCGA

Gene expression (RNASeqV2), protein expression (RPPA), and copy number variation data generated by TCGA project

Choose

GDS

Gene Expression Omnibus Datasets (GDS)

Choose

Cancer

Cancer related genomics datasets

Choose

2. To browse TCGA datasets click "Choose" button.

human	TCGA	937 RPPA breast invasive carcinoma (BRCA) samples from TCGA project. Data were downloaded via TC-GAbi... More	ID=TCGA_BRCA_RPPA_2019	Analyze
human	TCGA	309 RNA-seq cervical squamous cell carcinoma and		Analyze

3. Select the BRC RPPA dataset for the analysis.



Select dataset / Dataset Info and Analysis

Explore and analyze dataset TCGA_BRCA_RPPA_2019



Dataset Analysis

- Create a Signature >>
- Multi-group Analysis >>
- Analyze a list of Genes >>

Dataset Info

Description
937 RPPA breast invasive carcinoma (BRCA) samples from TCGA project. Data were downloaded via TCGAAbiolinks. Deleterious mutations in TCGA-BRCA driver genes were from MAF aligned against hg19. Mutation impact categories were defined by VEP from MAF aligned against hg38.

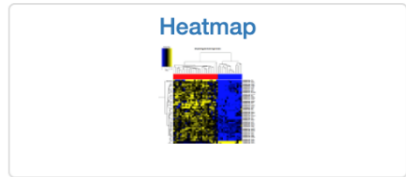
Assay type: RPPA **Data:** Proteomics
ID: TCGA_BRCA_RPPA_2019 **Organism:** human

Reference

4. To create signature.



Exploratory tools Data and metadata





Creating a signature using TCGA_BRCA_RPPA_2019

Select grouping variable

PAM50_mRNA

treatment: Luminal A

baseline: HER2-enriched

Create Signature

Dataset Info

Histogram



5. Select the sample property (ie factor) that will be used to construct two groups of samples for signature creation

6. Select the value of the factor that will indicate treatment and baseline sample groups

Filter Samples

7. Create differential protein expression signature between treatment and baseline samples

Signature created from dataset TCGA_BRCA_RPPA_2019

Signature analysis

Modify the list of selected genes >>

Other analyses with selected genes >>

Dataset Info

Signature Info

Download

Number of Samples after Filtering:	224	Probes after Filtering:	100	Genes after Filtering:	86
Sample Grouping Factor		Treatment		Baseline	
PAM50_mRNA		Luminal A		HER2-enriched	
Number of samples		174		50	

8. Click here to select different set of differentially expressed proteins (by default, 100 most statistically significant proteins are selected)

Signature in the form of the table and connectivity analysis tabs to connect to iLINCS signatures and perturbations

Analysis Results

Signature Data

Connected Signatures

Connected Perturbations

- FTreeView Analysis
- Heatmap
- PCA & t-SNE
- Pathway Analysis
- Enrichment Analysis
- DAVID
- TopFun
- GeneMANIA
- Reactome
- Morpheus Heatmap
- Clustergrammer
- PINET
- L1000CDS2
- L1000FWD
- X2K
- X2Kweb
- SigNetA

Various internal and external applications that can be used to further analyze the 100 gene signature

Modify the list of selected genes

Select differential expression range:



Select Pvalue Cutoff:



Chosen 12 out of 263 total genes in signature.

Analyze

Cancel

Legend:

- Selected Genes
- Genes below P Value Cutoff
- Genes below Diff Exp Cutoff
- Genes below both Cutoffs



9. Use sliders to select differentially expressed with $p\text{-value} < 10e-10$ that with increased expression in Her2E samples. Since Her2E samples were used as baseline samples, these \log_2 differential expression of these proteins will be less than zero proteins. You can also click and label individual proteins in the volcano plot.

10. Click analyze to use the selected 12 proteins in further analyses

L1000CDS2

L1000FWD

X2K

SigNetA

Signature created from dataset TCGA_BRCA_RPPA_2019

Signature analysis

Modify the list of selected genes >>

Other analyses with selected genes >>

Dataset Info **Signature Info** Download

Number of Samples after Filtering: 224
 Probes after Filtering: 12
 Genes after Filtering: 11

Pvalue cut off: 1e-10

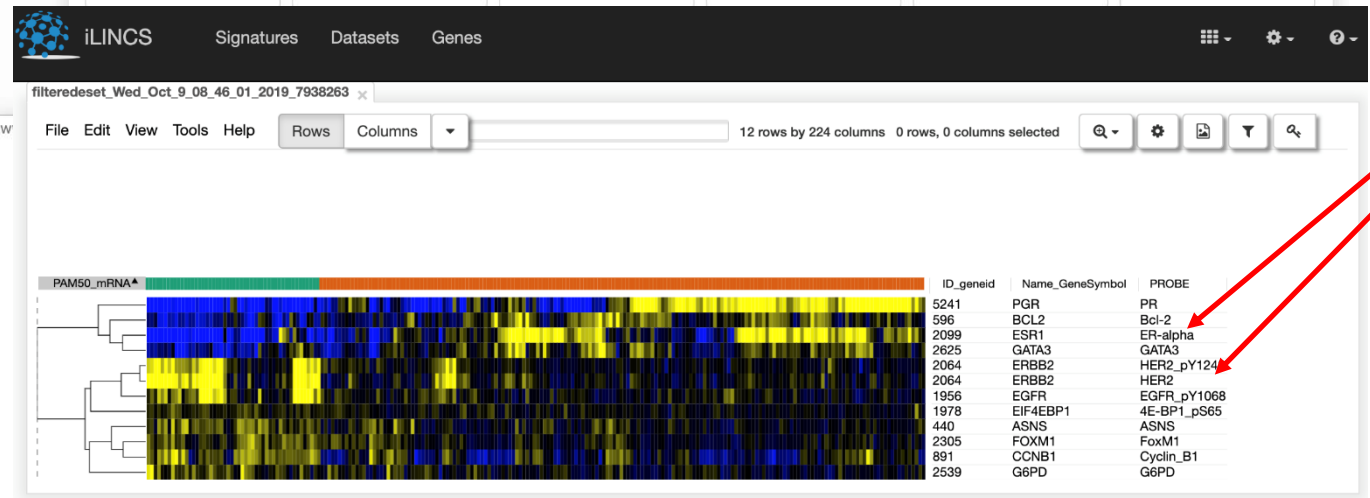
Sample Grouping Factor	Treatment	Baseline
PAM50_mRNA	Luminal A	HER2-enriched
Number of samples	174	50

Analysis Results Signature Data Connected Signatures Connected Perturbations

FTreeView Analysis Heatmap PCA & t-SNE Pathway Analysis Enrichment Analysis DAVID

ToppFun GeneMANIA Reactome **Morpheus Heatmap** Clustergrammer PiNET

11. Click to view the heatmap of the 12 proteins in the analysis of the corresponding RNA-seq dataset



Most differentially expressed proteins are known drivers of Her2E and Luminal A BRC subtypes

Use Case 3: Proteo-genomics analysis of cancer driver events in breast cancer

Signature created from dataset TCGA_BRCA_RPPA_2019

Signature analysis

- Modify the list of selected genes >>
- Other analyses with selected genes >>

Dataset Info Signature Info

Number of Samples after Filtering: **224** Probes after Filtering: **12** Genes after Filtering: **11**

P value cut off: **1e-10**

Sample Grouping Factor	Treatment	Baseline
PAM50_mRNA	Luminal A	HER2-enriched
Number of samples	174	50

13. Click to proceed with the analysis of the corresponding RNA-seq dataset using selected 12 genes

12. Interactive box plot for distribution of sample expression levels

Analysis Results

Signature Data

Connected Signatures

Connected Perturbations

Show complete signature (263)

Show selected genes (12)

Probe ID	Gene ID	Symbol	Differential Expression	P Value
ER-alpha	2099	ESR1	2.836	1.3e-25
HER2_pY1248	2064	ERBB2	-1.300	3.6e-23
HER2	2064	ERBB2	-1.599	2.9e-22
Bcl-2	596	BCL2	1.255	2.5e-20
PR	5241	PGR	2.887	8.2e-20

5 10 20

First << 1 2 3 >> Last

Page 1





Subset signatures and datasets with query genes

Show 11 query genes

Genes selected in RPPA dataset analysis

Signatures

All Datasets

LINCS Datasets

14. Search for the BRC RNA-seq dataset in the TCGA collection

Find Genomics Datasets

Organism: All | Sample Type: All

Data Type: All | Collection: TCGA

Keyword: brc

Search

Found 6 datasets in TCGA

15. Select the dataset

Organism	Collection	Description	Reference		Analyze
human	TCGA	847 Gistic2foc (CNV) breast invasive carcinoma (BRCA) samples from TCGA project. Copy number variati... More	GISTIC2.0 facilitates sensitiv... More ID=TCGA_BRCA_Gistic2foc	11	Analyze
human	TCGA	919 RNA-seq breast invasive carcinoma (BRCA) samples from TCGA project. The data was processed using... More	Collins FS, Barker AD. Mapping... More ID=TCGA_BRCA_RNASeqV2	11	Analyze
human	TCGA	1215 RNA-seq breast invasive carcinoma (BRCA) samples from TCGA project. Data were downloaded via TC... More	ID=TCGA_BRCA_RNASeqV2_2019	11	Analyze
human	TCGA	937 RPPA breast invasive carcinoma (BRCA) samples from TCGA project. Data were downloaded via TCGAbi... More	ID=TCGA_BRCA_RPPA_2019	11	Analyze

iLINC Signatures Datasets Genes

Select dataset / Dataset Info and Analysis / Pick Genes / Dataset landing page with genes

Explore and analyze dataset TCGA_BRCA_RNASeqV2_2019 with your gene list

Dataset Analysis

- Create a Signature
- Multi-group Analysis
- Other analyses with selected genes

Dataset Info **Gene list summary**

Samples:	1215
Selected Probes:	11
Selected Genes:	11

Show found genes Download

Analysis Results Genes in the dataset Data and metadata Precomputed Signatures

FTreeView Analysis Heatmap Morpheus Heatmap Clustergrammer

iLINC Signatures Datasets Genes

Select Genes / Analyze with submitted list of genes / Dataset info with genes / Create Signature with a gene list

Creating a signature using gene list in TCGA_BRCA_RNASeqV2_2019

Select grouping variable

- PAM50_mRNA
- treatment: Luminal A
- baseline: HER2-enriched

Create Signature

Dataset Info **Gene List** **Histogram**

300
200
150
100
50
0

Luminal A HER2-enriched

All Selected

Filter Samples

15. Click to create the signature (ie perform differential expression analysis for the selected 11 genes)

16. Select the sample property (ie factor) that will be used to construct two groups of samples for signature creation

17. Select the value of the factor that will indicate treatment and baseline sample groups

18. Perform differential gene expression analysis between treatment and baseline samples

Signature created from dataset TCGA_BRCA_RNASeqV2_2019 with a gene list

Signature analysis

- Modify the list of selected genes >
- Other analyses with selected genes >

Dataset Info

Signature Info

Number of Samples after Filtering: 289	Probes after Filtering: 11	Genes after Filtering: 11
Sample Grouping Factor	Treatment	Baseline
PAM50_mRNA	Luminal A	HER2-enriched
Number of samples	231	58


19. View the gene by gene analysis results

20. View the hetamap


Analysis Results

- Signature Data
- Connected Signatures
- Connected Perturbations


FTreeView Analysis




Heatmap




PCA & t-SNE




Pathway Analysis



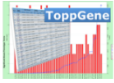
Enrichment Analysis




DAVID




ToppFun




GeneMANIA




Reactome




Morpheus Heatmap




L1000CDS2




L1000FWD

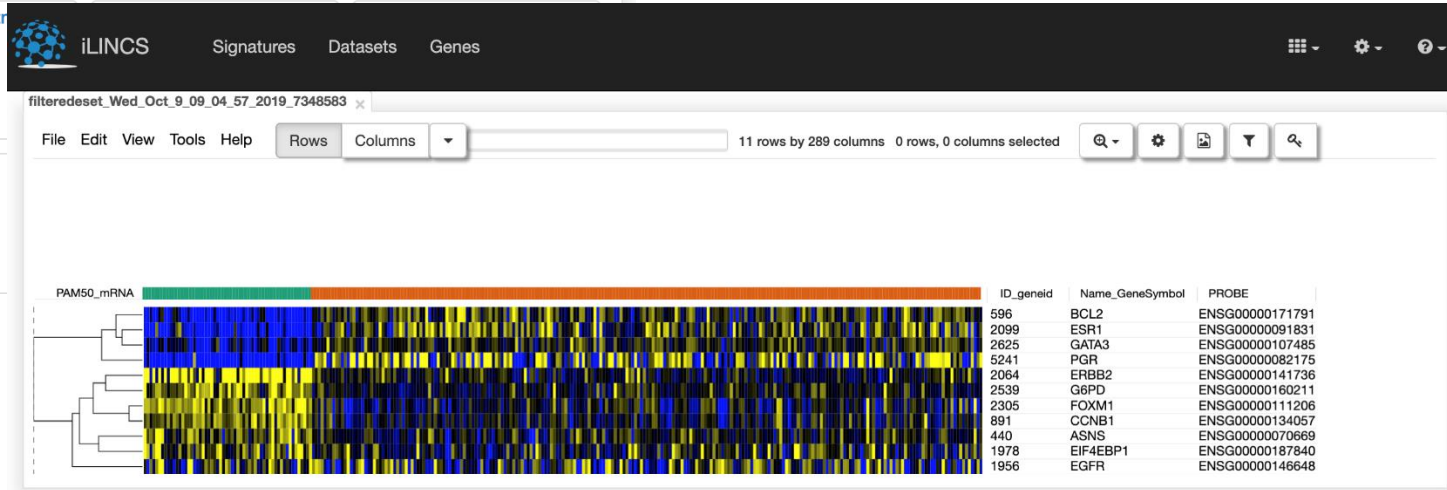


X2K



SigNetA





Use Case 3: Proteo-genomics analysis of cancer driver events in breast cancer

The screenshot shows the iLINC'S 'Datasets' workflow. At the top, the navigation bar includes 'Signatures', 'Datasets', and 'Genes'. The 'Datasets' tab is active. Below the navigation, there are tabs for 'All Datasets' and 'LINCS Datasets'. A search section titled 'Find Genomics Datasets' contains filters for 'Organism' (set to 'All'), 'Sample Type' (set to 'All'), 'Data Type' (set to 'All'), and 'Collection' (set to 'TCGA'). A 'Keyword' field contains 'brc'. A blue 'Search' button is located to the right of the keyword field. Below the search filters, a section titled 'Found 6 datasets in TCGA' displays a table of results. The table has columns for 'Organism', 'Collection', 'Description', and 'Reference'. Three rows are visible, each with an 'Analyze' button to its right. Red arrows point from the yellow callout boxes to the 'Datasets' tab, the 'Collection' dropdown, the 'Keyword' field, the 'Search' button, and the 'Analyze' button for the third dataset.

Organism	Collection	Description	Reference
human	TCGA	847 Gistic2foc (CNV) breast invasive carcinoma (BRCA) samples from TCGA project. Copy number variati... More	GISTIC2.0 facilitates sensitiv... More ID=TCGA_BRCA_Gistic2foc
human	TCGA	919 RNA-seq breast invasive carcinoma (BRCA) samples from TCGA project. The data was processed using... More	Collins FS, Barker AD. Mapping... More ID=TCGA_BRCA_RNASeqV2
human	TCGA	1215 RNA-seq breast invasive carcinoma (BRCA) samples from TCGA project. Data were downloaded via TC... More	ID=TCGA_BRCA_RNASeqV2_2019

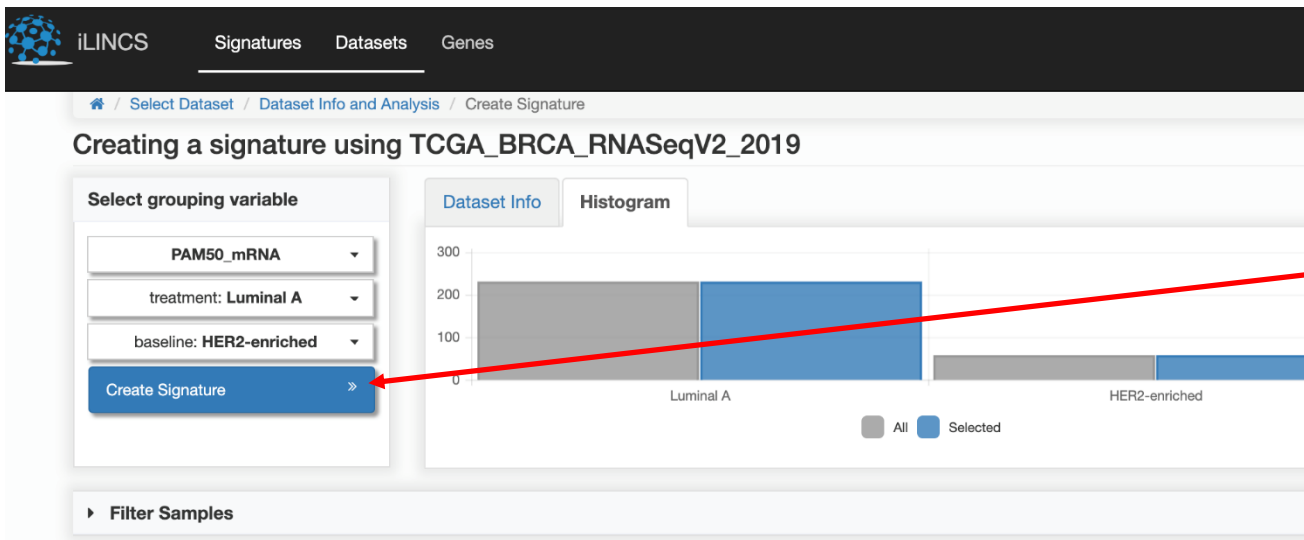
21. Start with the Datasets workflow

22. Search Datasets for BRC in the TCGA collection

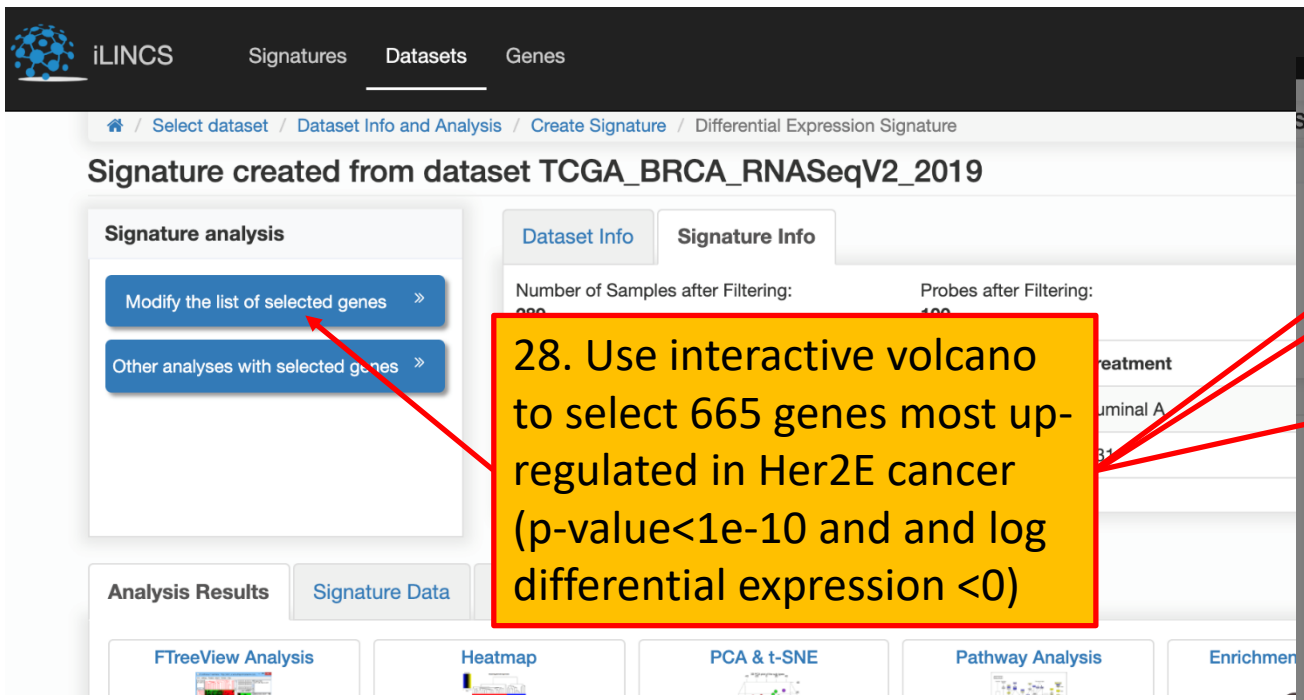
23. Select the BRC RNA-seq dataset for the analysis

The screenshot shows the 'Dataset Info and Analysis' page for the dataset 'TCGA_BRCA_RNASeqV2_2019'. The navigation bar includes 'Signatures', 'Datasets', and 'Genes'. The 'Datasets' tab is active. The page title is 'Explore and analyze dataset TCGA_BRCA_RNASeqV2_2019'. On the left, there is a 'Dataset Analysis' section with three buttons: 'Create a Signature', 'Multi-group Analysis', and 'Analyze a list of Genes'. The 'Create a Signature' button is highlighted with a red arrow. On the right, there is a 'Dataset Info' section with a 'Description' field containing text about the dataset, and 'Assay type: RNA-seq' and 'Organism: human' fields. Below the 'Dataset Info' section, there are tabs for 'Exploratory tools', 'Data and metadata', and 'Precomputed Signatures'. A red arrow points from the yellow callout box to the 'Create a Signature' button.

24. Initiate the full signature creation



- 25. Select the sample property (ie factor) that will be used to construct two groups of samples for signature creation
- 26. Select the value of the factor that will indicate treatment and baseline sample groups
- 27. Perform differential gene expression analysis between treatment and baseline samples



- 28. Use interactive volcano to select 665 genes most up-regulated in Her2E cancer (p-value < 1e-10 and log differential expression < 0)



Use Case 3: Proteo-genomics analysis of cancer driver events in breast cancer



Signature created from dataset TCGA_BRCA_RNASeqV2_2019

Signature analysis

Modify the list of selected genes >>

Other analyses with selected genes >>

Dataset Info

Signature Info

Number of Samples after Filtering:
289

Probes after Filtering:
665

Genes after Filtering:
655

Differential expression cut offs:

Pvalue cut off:

[Login](#) | [Register](#)



Transcription **Pathways** Ontologies Diseases/Drugs Cell Types Misc Legacy Crowd

Description No description available (658 genes)

Signature	Baseline
Her2-enriched	HEP2-enriched
	58

WikiPathways 2019 Human

- Retinoblastoma Gene in Cancer WP2446
- Cell Cycle WP179
- DNA Replication WP466
- G1 to S cell cycle control WP45
- DNA IR-damage and cellular response via A1

WikiPathways 2019 Mouse

- DNA Replication WP150
- G1 to S cell cycle control WP413
- Proteasome Degradation WP519
- p53 signaling WP2902
- Glycolysis and Gluconeogenesis WP157

KEGG 2019 Human

- Cell cycle
- DNA replication
- Proteasome
- Oocyte meiosis
- Fanconi anemia pathway

ARCHS4 Kinases Coexp

- MELK_human_kinase_ARCHS4_coexpression
- AURKB_human_kinase_ARCHS4_coexpression

KEGG 2019 Mouse

- Cell cycle
- DNA replication

BioCarta 2016

- CDK Regulation of DNA Replication_Homo sapiens
- Role of Ran in mitotic spindle regulation_Homo sapiens

Enrichment Analysis



DAVID



PiNET



X2K



29. Pathway analysis with Enrichr of selected genes upregulated in Her2E tumors

Signature created from dataset TCGA_BRCA_RNASeqV2_2019

Signature analysis

Modify the list of selected genes >>

Other analyses with selected genes >>

Dataset Info **Signature Info** Download

Number of Samples after Filtering: **289** Probes after Filtering: **665** Genes after Filtering: **655**

Differential expression cut offs: below: **-0.01** above: **5** Pvalue cut off: **1e-10**

Sample Grouping Factor	Treatment	Baseline
PAM50_mRNA	Luminal A	HER2-enriched
Number of samples	231	58

30. Connectivity analysis with LINCS CP perturbation signatures

Analysis Results **Signature Data** **Connected Signatures** **Connected Perturbations**

Use complete signature (19390) Use selected genes (665)

Signature Library	Common genes
<input type="checkbox"/> LINCS consensus gene (CGS) knockdown signatures	973
<input type="checkbox"/> LINCS gene overexpression signatures	973
<input checked="" type="checkbox"/> LINCS chemical perturbation signatures	973
<input type="checkbox"/> LINCS targeted proteomics signatures	85
<input type="checkbox"/> Disease related signatures	19149
<input type="checkbox"/> ENCODE transcription factor binding signatures	19141
<input type="checkbox"/> Connectivity Map signatures	11565
<input type="checkbox"/> DrugMatrix signatures	13086
<input type="checkbox"/> Transcriptional signatures from EBI Expression Atlas	18952



Cancer therapeutics response signatures

17042

Pharmacogenomics transcriptional signatures

18521

10201 of LINCS chemical perturbagen signatures

Analyze

Selection

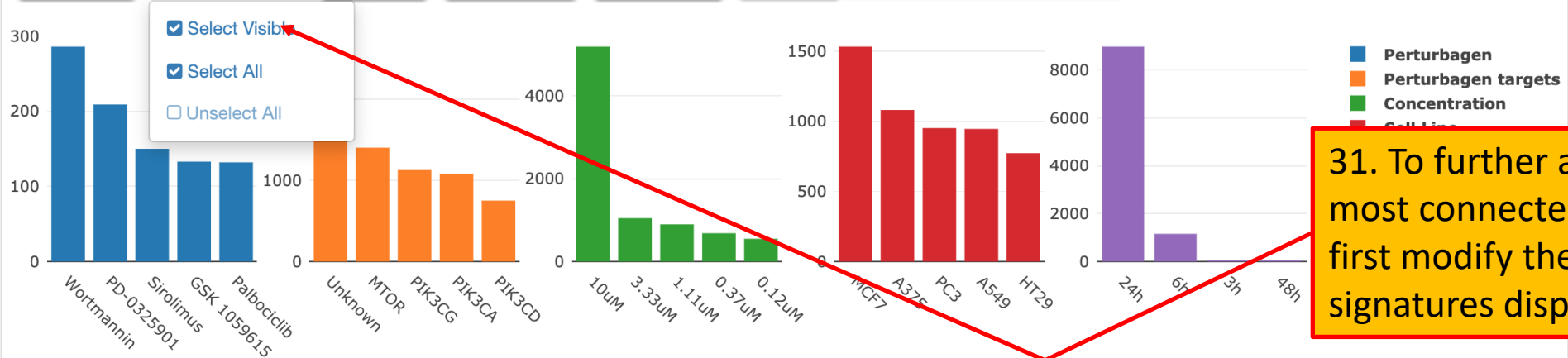
My list

Download

Clear filters

Stats

Top 5 All Signatures



The connectivity analysis provides the sorted list of most statistically significantly correlated signatures.

31. To further analyze top 100 most connected CP signatures, first modify the number of signatures displayed.

Signature Id	Perturbagen	Perturbagen targets	Concentration	Cell Line	Time	Concordance
<input type="checkbox"/> LINCSCP_135677	KPT-330	XPO1	0.87uM	MCF7	24h	0.537
<input type="checkbox"/> LINCSCP_139907	Palbociclib	CCND1 CDK4 ...	10uM	MCF7	24h	0.532
<input type="checkbox"/> LINCSCP_30538	BMS-536924	IGF1R	10uM	MCF7	24h	0.522
<input type="checkbox"/> LINCSCP_139908	Palbociclib	CCND1 CDK4 ...	3.33uM	MCF7	24h	0.518
<input type="checkbox"/> LINCSCP_32534	WZ 3146	EGFR	10uM	MCF7	24h	0.516

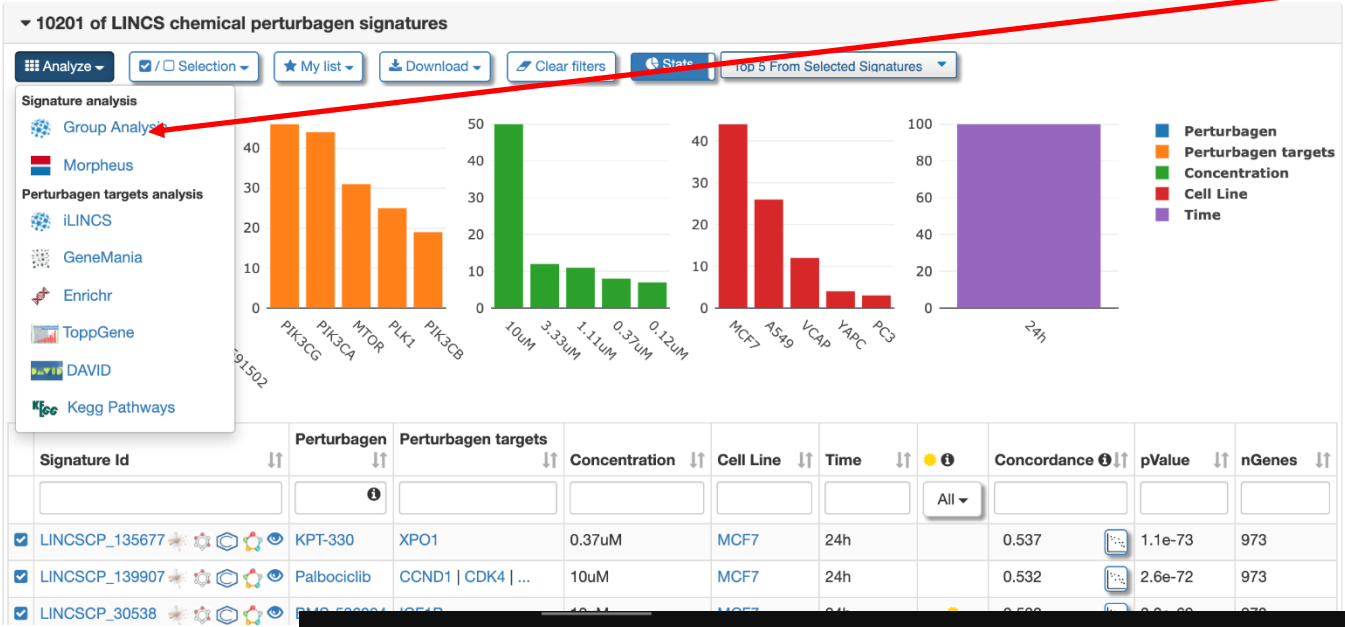
32. Then click on "Select Visible" signatures in the "Selection" drop down menu

5 25 50 100

First « 1 2 3 4 5 » Last

Page 1 of 2041, of 10201 entries

32. Select "Group Analysis" on the "Analysis" drop-down menu.



33. Confirm the selected signatures and the default number of most differentially expressed genes to be used from each signature. The number of signature is 101, because the original Luminal A vs Her2E signature is also included in the analysis.

Select Signatures for Group Analysis

Select Visible Select All Unselect All Remove Selected Download Selected

Signature ID	Library	Cell Line	Factor	Perturbagen
<input checked="" type="checkbox"/> LINCSCP_11733	LINCS chemical perturbagen signatures	BT20		Palbociclib
<input checked="" type="checkbox"/> LINCSCP_122796	LINCS chemical perturbagen signatures	HT29		Milciclib (PHA-848125)
<input checked="" type="checkbox"/> LINCSCP_123195	LINCS chemical perturbagen signatures	HT29		MLN-0128
<input checked="" type="checkbox"/> LINCSCP_128841	LINCS chemical perturbagen signatures	HUVEC		STK219801
<input checked="" type="checkbox"/> LINCSCP_129999	LINCS chemical perturbagen signatures	LNCAP		ZSTK 474

5 10 50 First « 1 2 3 4 5 » Last Page 1 of 39, of 193 entries

Include uploaded signature in the analysis: processedSig_Mon_Oct_14_14_20_00_2019_9627926.xls

Number of genes per signature: 50 Analyze 101 signatures Cancel



Signature Group Analysis

34. To view the heatmap of expression changes in 101 signatures.

Signature ID Library Organism Tissue Cell Line Factor Perturbation Treatment Baseline

LINCSCP_11733 LINCSCP_122796 LINCSCP_123195 LINCSCP_128841 LINCSCP_129999

Analysis include uploaded signature: processedSig_Mon_Oct_14_14_20_01

Number of top genes per signature: 50
Genes: 566

Generated results: Gene Clusters, Signatures Heatmap, Morpheus Signatures H

filteredeseet_Mon_Oct_14_2019_14_48_31_3132066

File Edit View Tools Help Rows Columns

height: 1.25 depth: 4.00 # of leaf nodes: 54.00

concentration cellLine time treatment

Export Selection

AURKA, CCNB1, CDC20, UBE2C, CDK1, BIRC5, CCNA2, KIF20A, CCNB2, TOP2A, NUSAP1, SMC4, DCK, CDCA4, HPRT1, MAP3K4, CHEK2, SUV39H1, BRCA1, POLE 2, TIMELESS, USP1, NUP62, AURKB, RFC2, MYBL2, CDC25A, NUP85, MELK, HA T1, PCNA, LIG1, SUZ12, JPT2, BUB1B, NCAPD2, DNMT1, TMEM97, C5, TOPBP 1, MCM3, TMEM109, MSH6, MPZL1, PGRMC1, PRCP, TGFB2, PCBD1, EPB41L2, AKR7A2, SMARCC1, RNPS1, TBC1D31, CHP1

iLINCS search GeneMANIA PINET

Enrichr Interactive and collaborative gene list enrichment tool.

ToppGene Portal for gene list enrichment analysis and candidate gene prioritization.

DAVID The Database for Annotation, Visualization and Integrated Discovery.

OK Cancel

35. To perform pathway enrichment analysis for a set of co-expressed genes, select the cluster.

36. Right click on the (invisible in this case) column of selected gene symbols, end select "Export" on the context specific menu.

37. Submit the genes to Enrichr

Sort Heat Map Ascending ↑
Sort Heat Map Descending ↓
Sort Heat Map Descending/Ascending

Move To Top
Annotate Selection
Copy
Export
Invert Selection
Select All
Clear Selection

Enrichr

Transcription Pathways Ontologies Diseases/Drugs Cell Types Misc Legacy Crowd

Description No description available (54 genes)

WikiPathways 2019 Human
Retinoblastoma Gene in Cancer WP2446
Cell Cycle WP179
G1 to S cell cycle control WP45
DNA Damage Response WP707
miRNA Regulation of DNA Damage Respons

WikiPathways 2019 Mouse
G1 to S cell cycle control WP413
Mismatch repair WP1257
DNA Replication WP150
p53 signaling WP2902
PluriNetWork WP1763

KEGG 2019 Human
Cell cycle
Cellular senescence
DNA replication
Progesterone-mediated oocyte maturation
Mismatch repair

