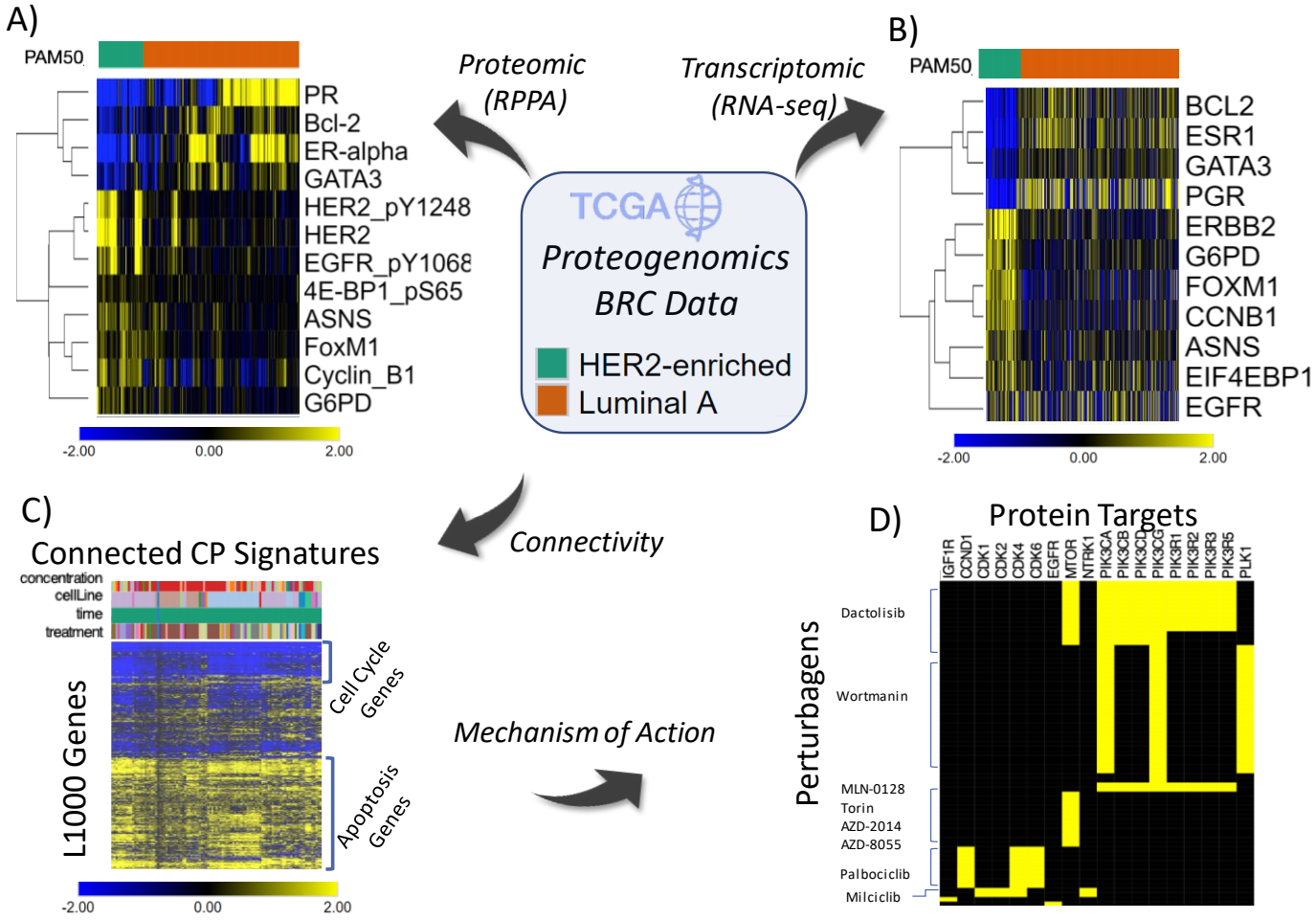
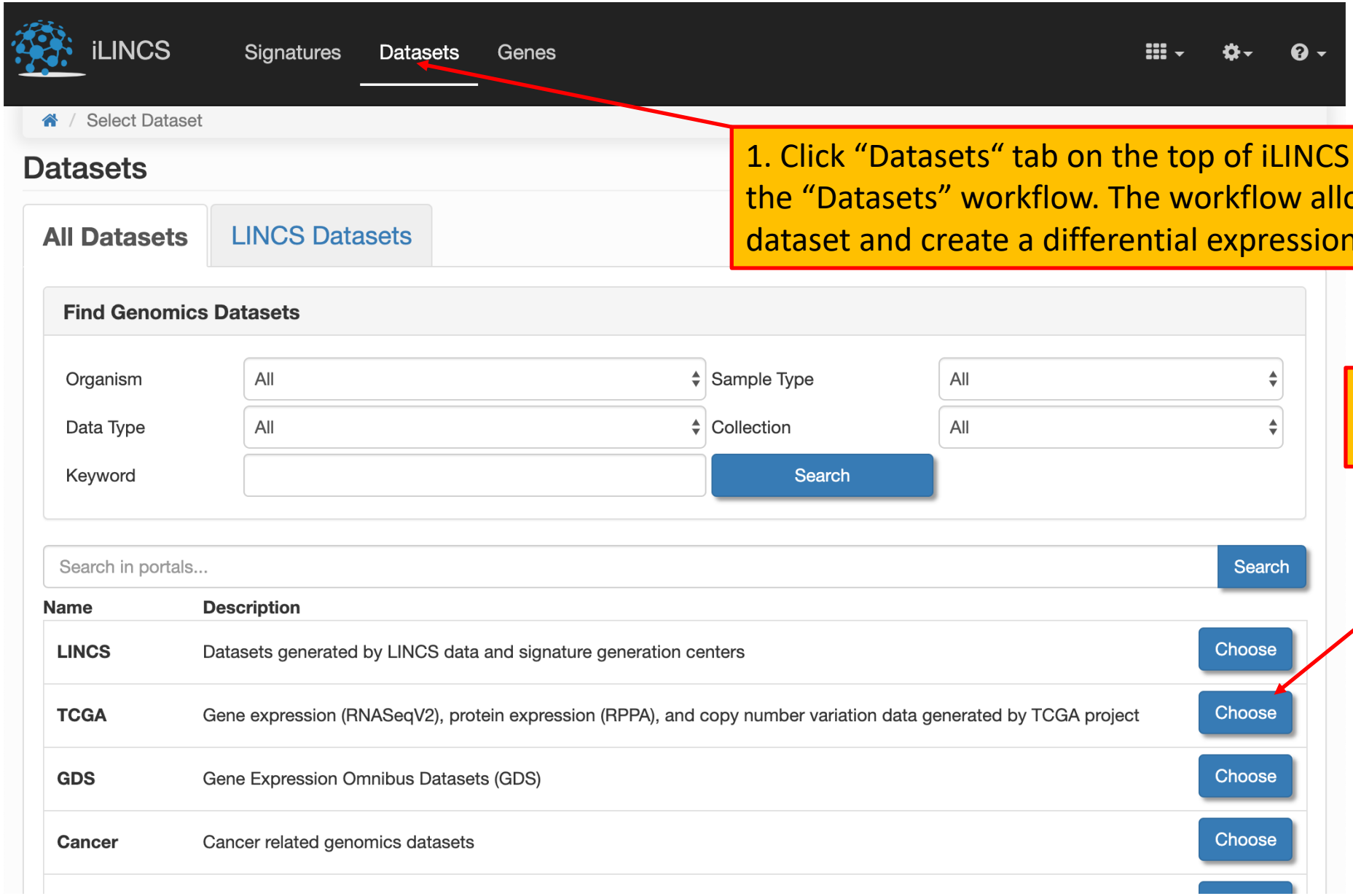
Use case 3: Proteo-genomics analysis of cancer driver events in breast cancer

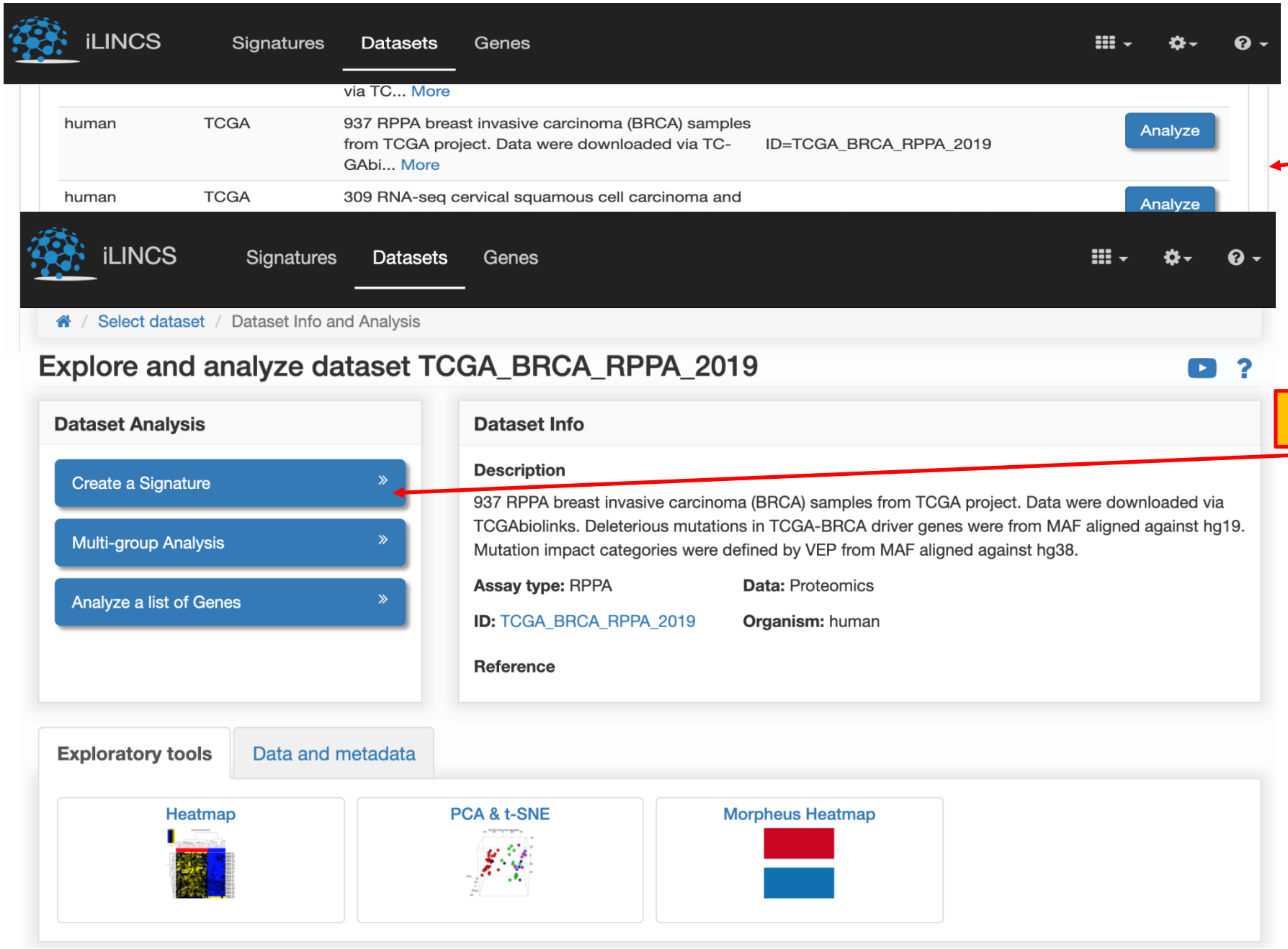
The use case describes the analysis of TCGA breast cancer RNA-seq and RPPA data using the iLINCS “Datasets” workflow to construct and analyze the differential gene and protein expression signatures contrasting Luminal A and Her2 enriched (Her2E) breast tumors. The analysis results are depicted in the 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).





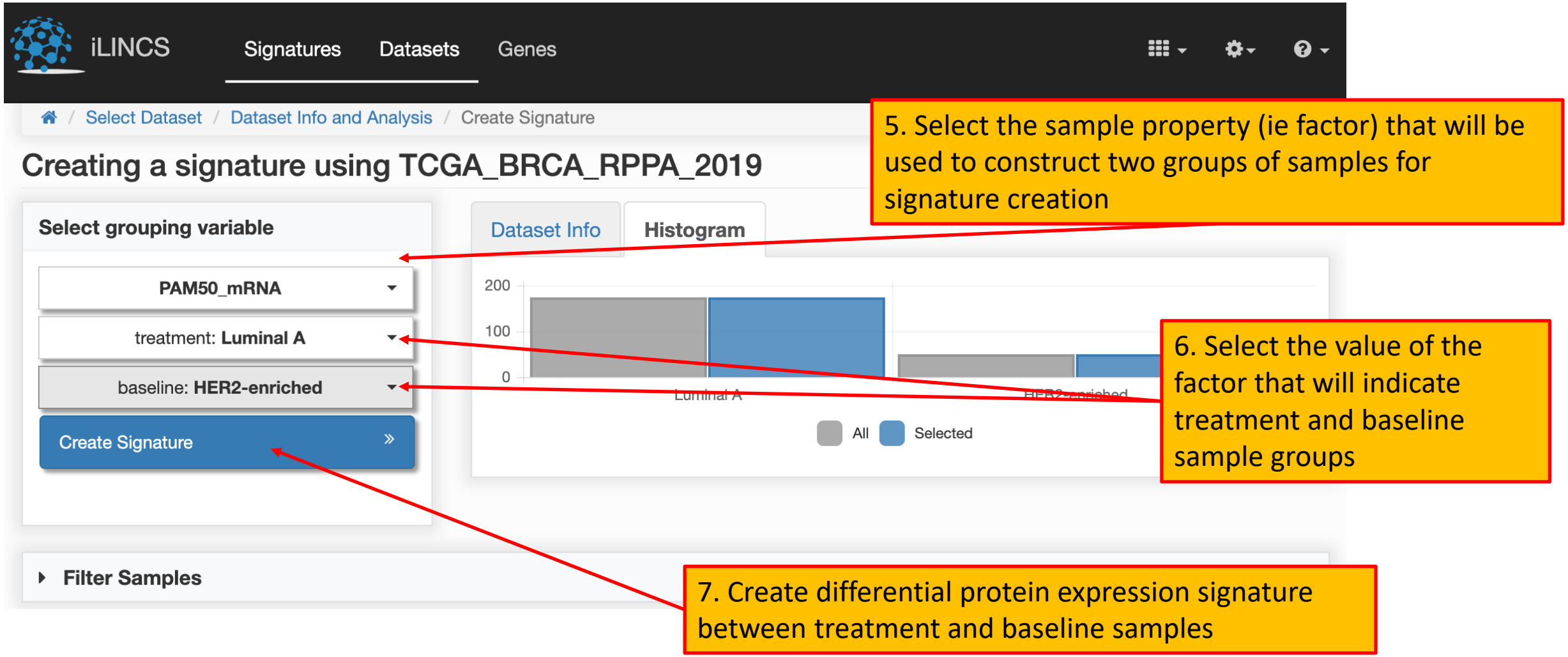
1. Click “Datasets“ tab on the top of iLINCS homepage to start with the “Datasets” workflow. The workflow allows you to select a dataset and create a differential expression gene/protein signature.

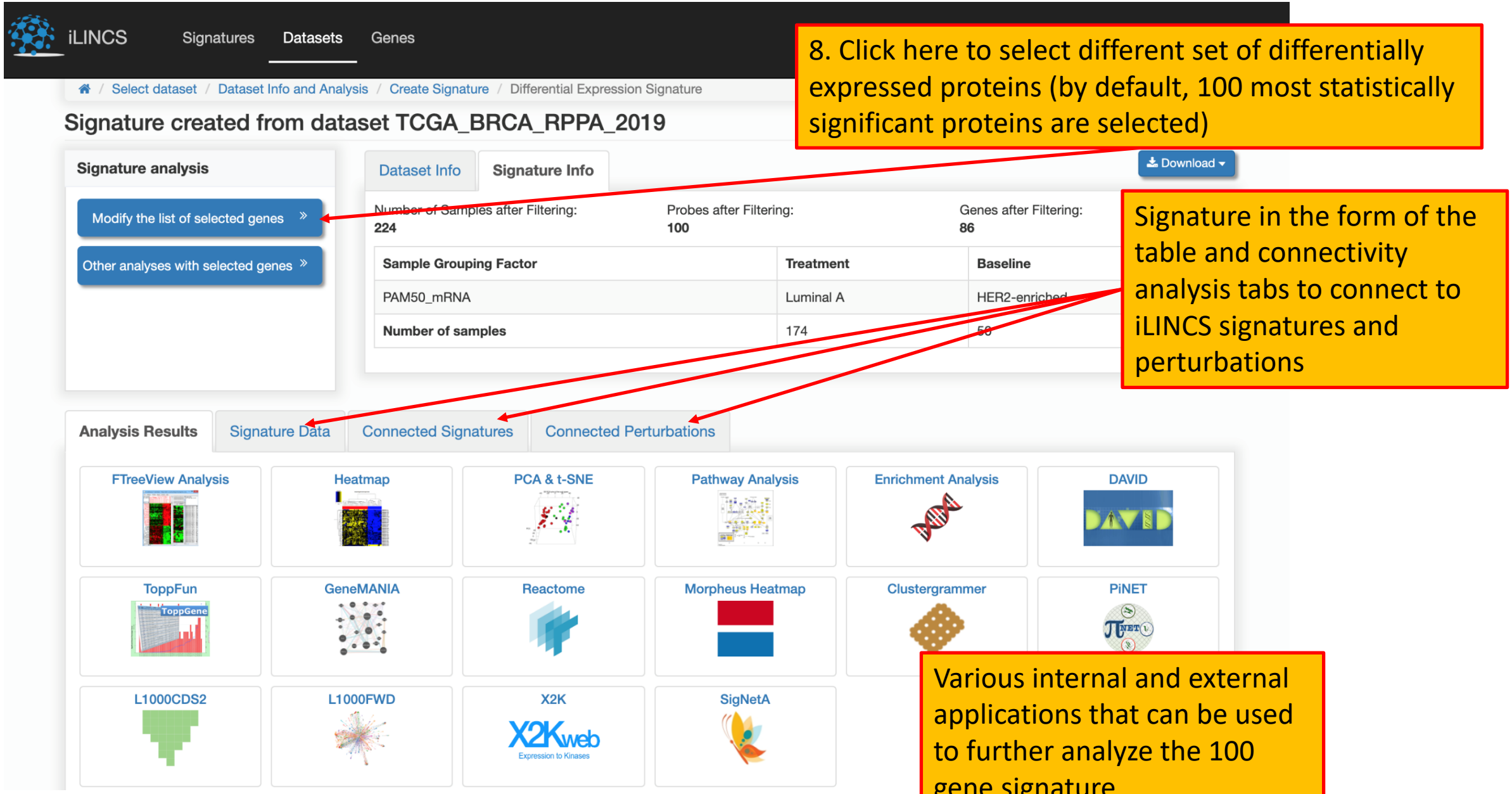
2. To browse TCGA datasets click “Choose” button.



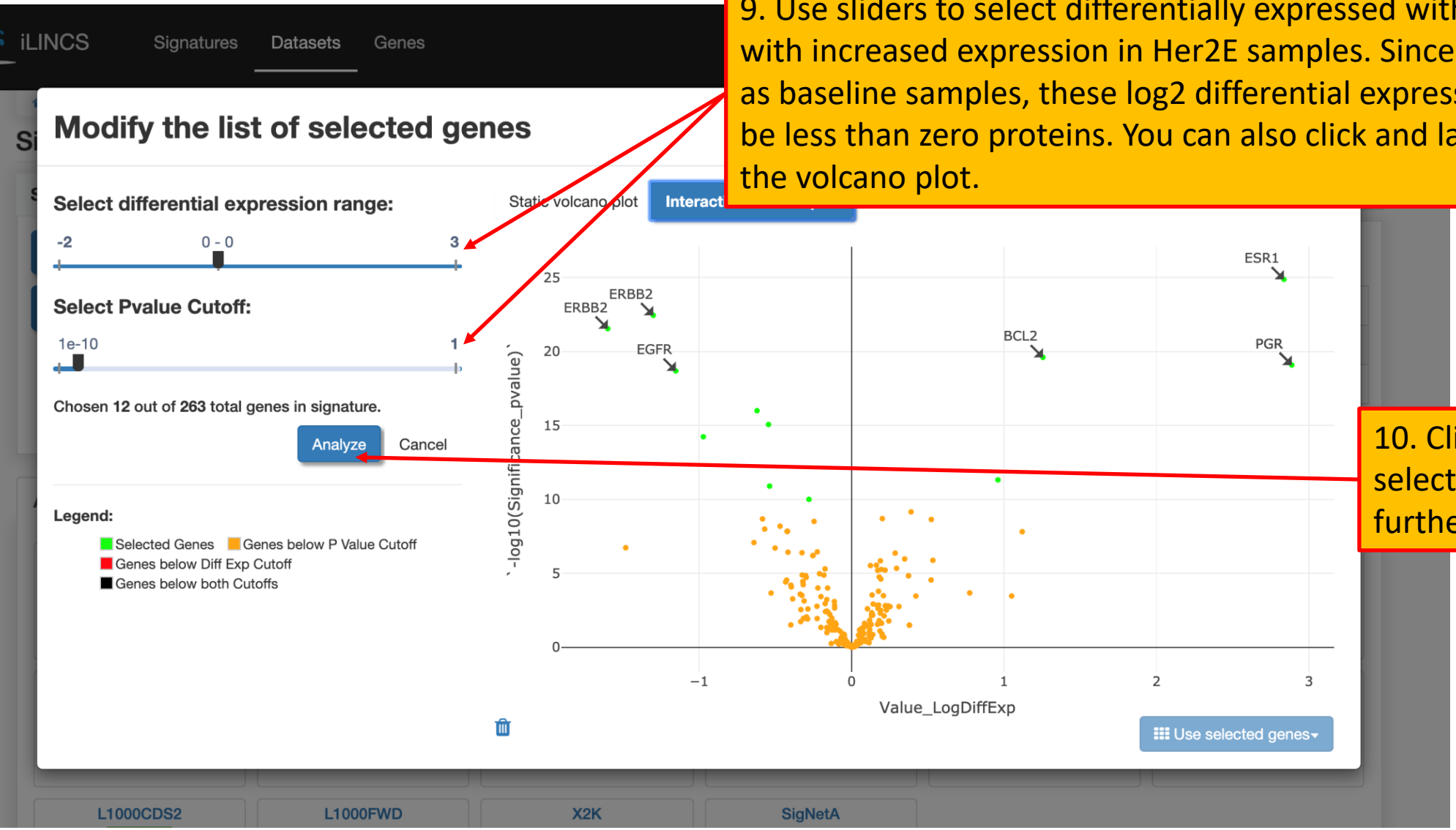
3. Select the BRC RPPA dataset for the analysis.

4. To create signature.

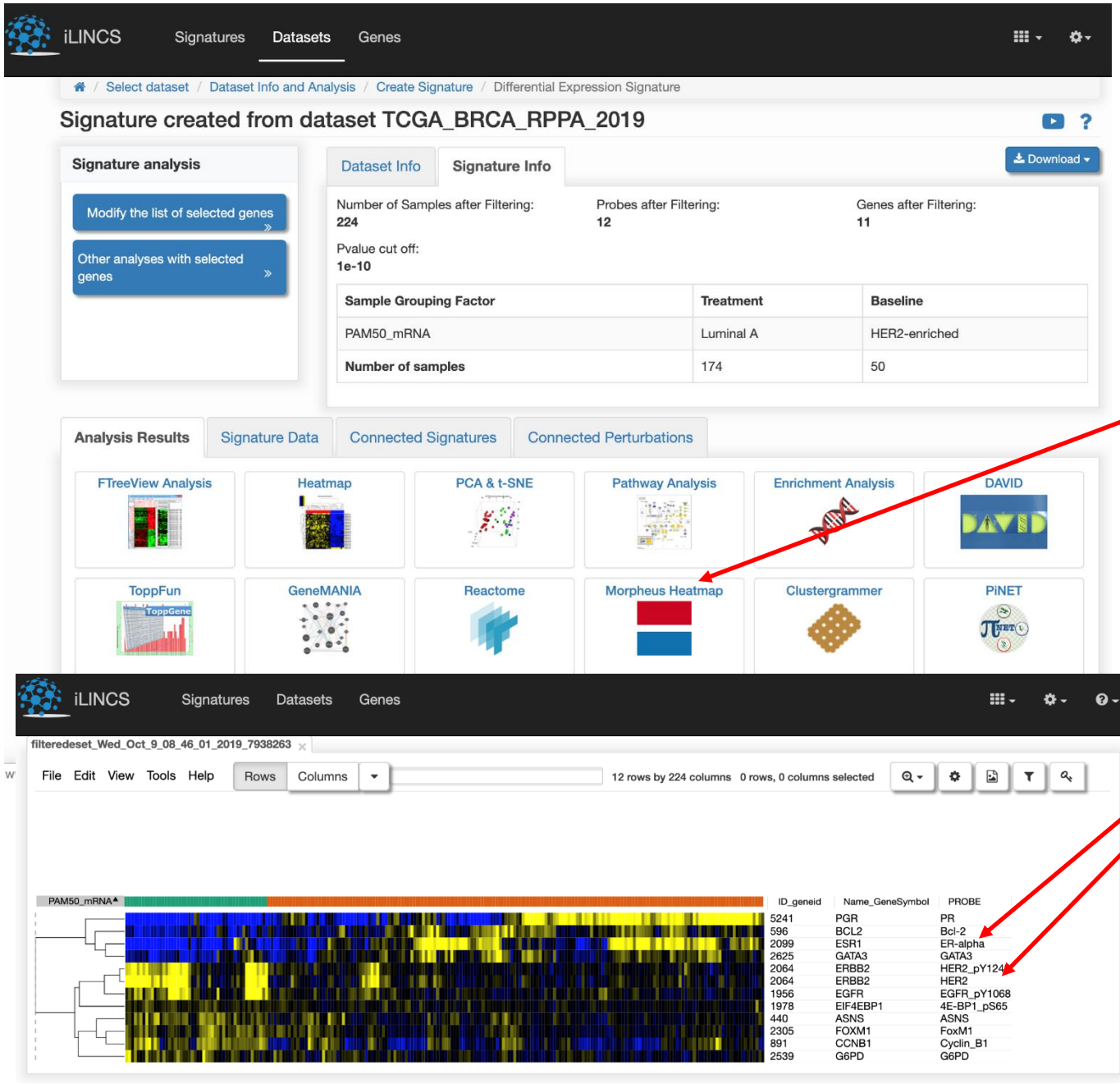

iLINCS
Signatures
Datasets
Genes
Select Dataset / Dataset Info and Analysis / Create Signature
Creating a signature using TCGA_BRCA_RPPA_2019
Select grouping variable
PAM50_mRNA
treatment: Luminal A
baseline: HER2-enriched
Create Signature
Dataset Info
Histogram
200
100
0
Luminal A
All
Selected
Filter Samples
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
7. Create differential protein expression signature between treatment and baseline samples


iLINCS
Signatures
Datasets
Genes
Select dataset / Dataset Info and Analysis / Create Signature / Differential Expression Signature
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
58
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
L1000CDS2
L1000FWD
X2K
X2Kweb
Expression to Kinases
SigNetA
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
Various internal and external applications that can be used to further analyze the 100 gene signature

9. Use sliders to select differentially expressed with p-value<10e-10 that with increased expression in Her2E samples. Since Her2E samples were used as baseline samples, these log2 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



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

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

iLINCS Signatures Datasets Genes

Select dataset / Dataset Info and Analysis / Create Signature / Differential Expression Signature

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

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

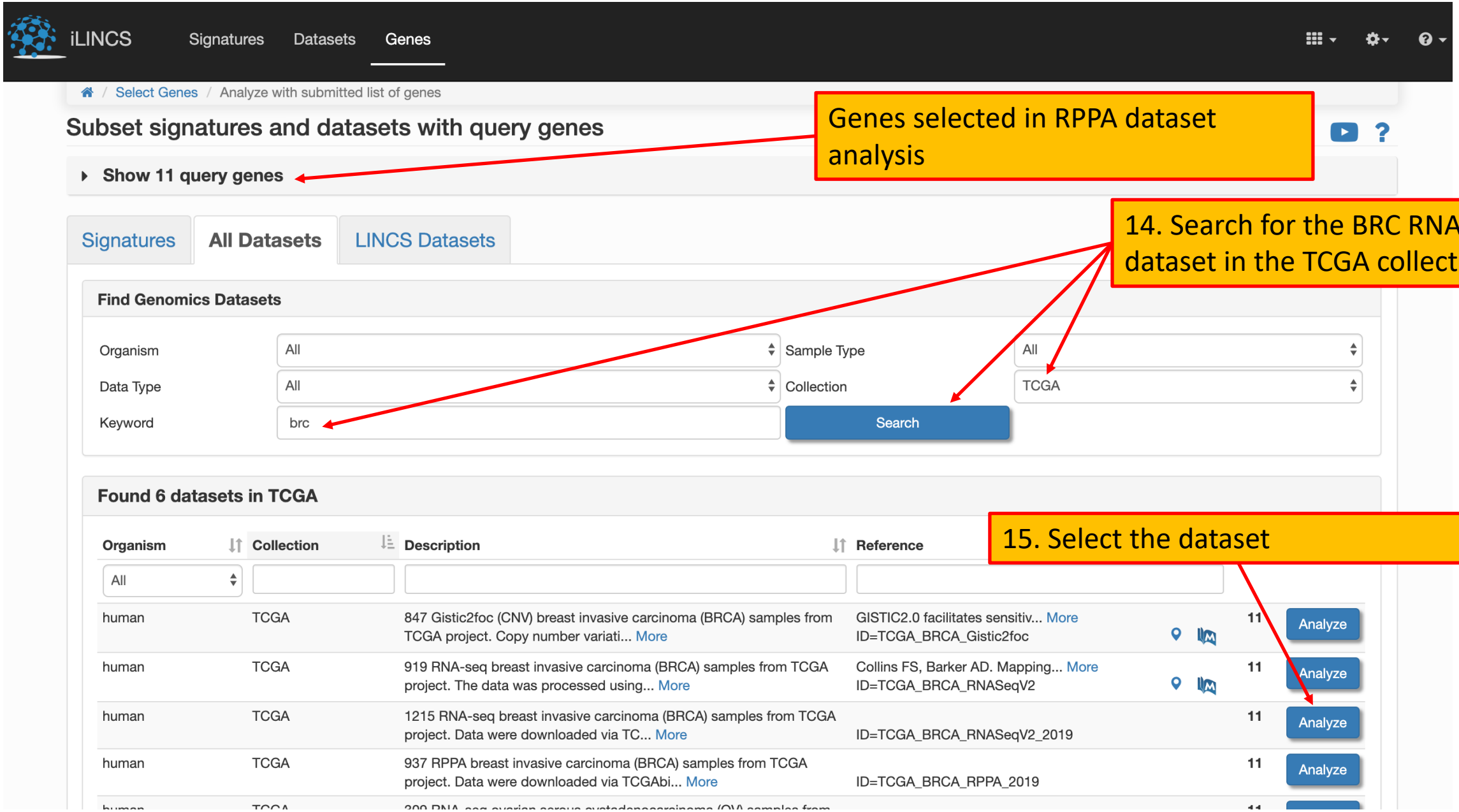
5 | 10 | 20

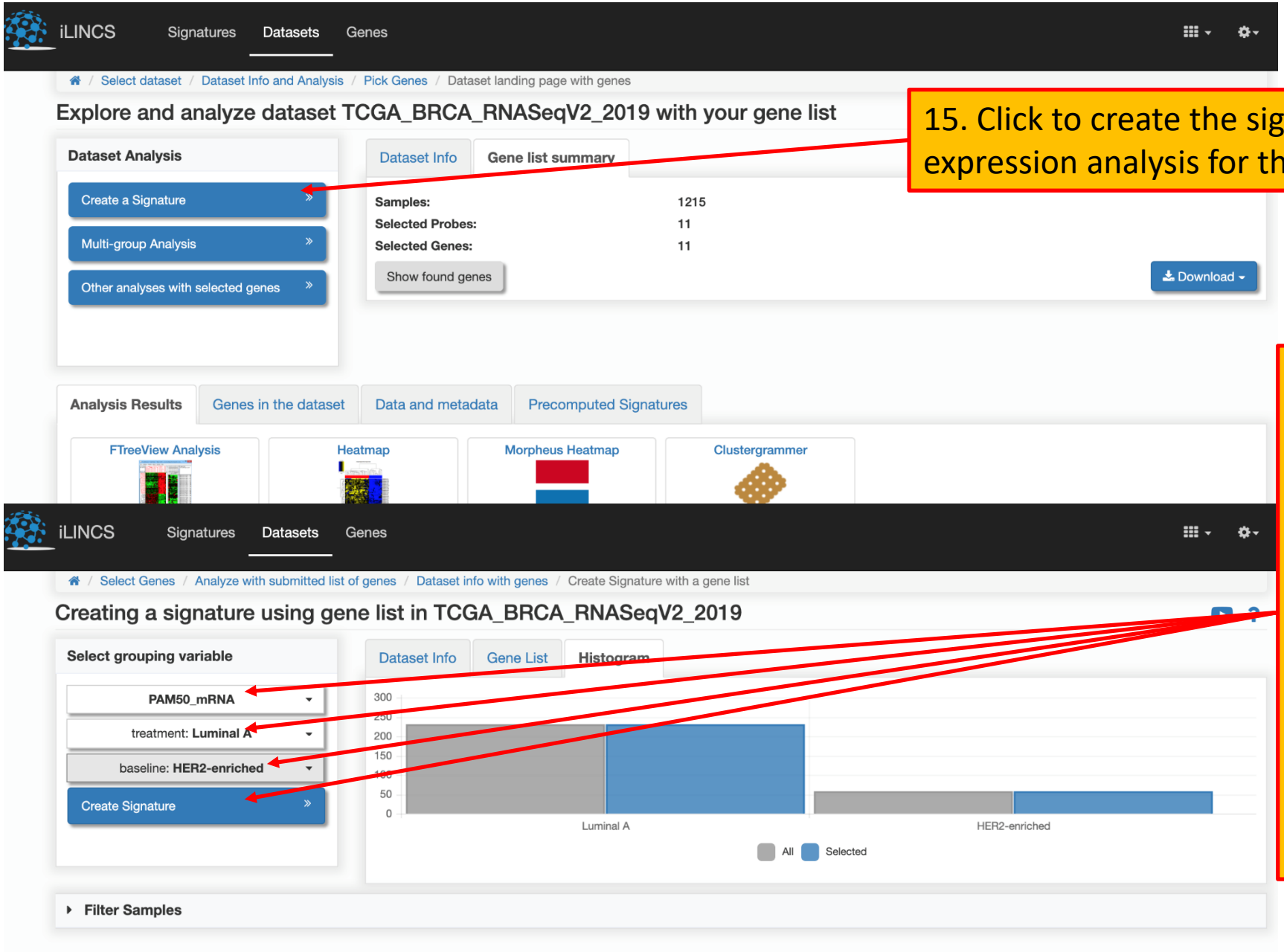
First « 1 2 3 » Last

Page 1

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


iLINCS
Signatures
Datasets
Genes
Select Genes / Analyze with submitted list of genes
Subset signatures and datasets with query genes
Genes selected in RPPA dataset analysis
Show 11 query genes
14. Search for the BRC RNA-seq dataset in the TCGA collection
Signatures
All Datasets
LINCS Datasets
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
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

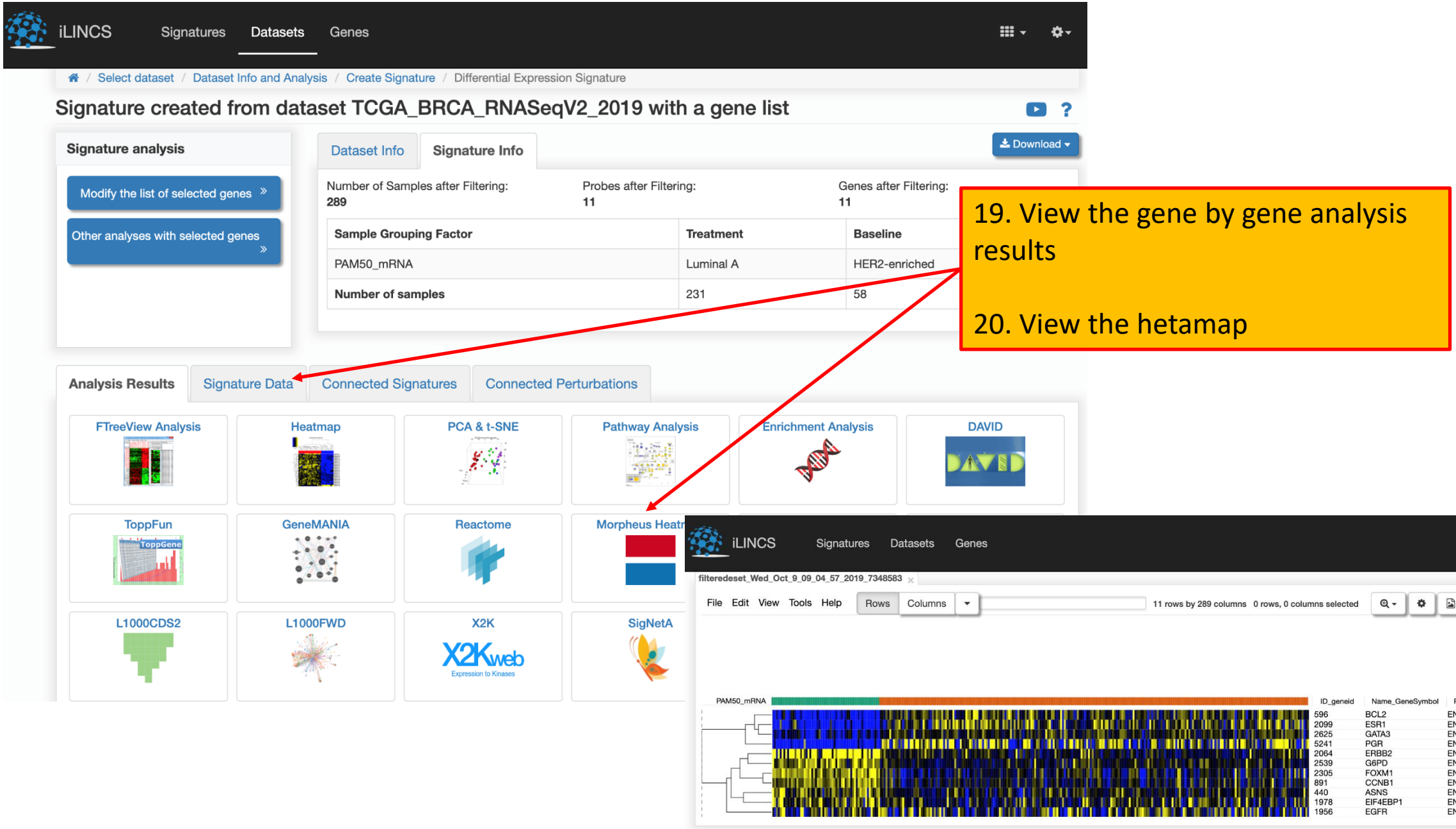


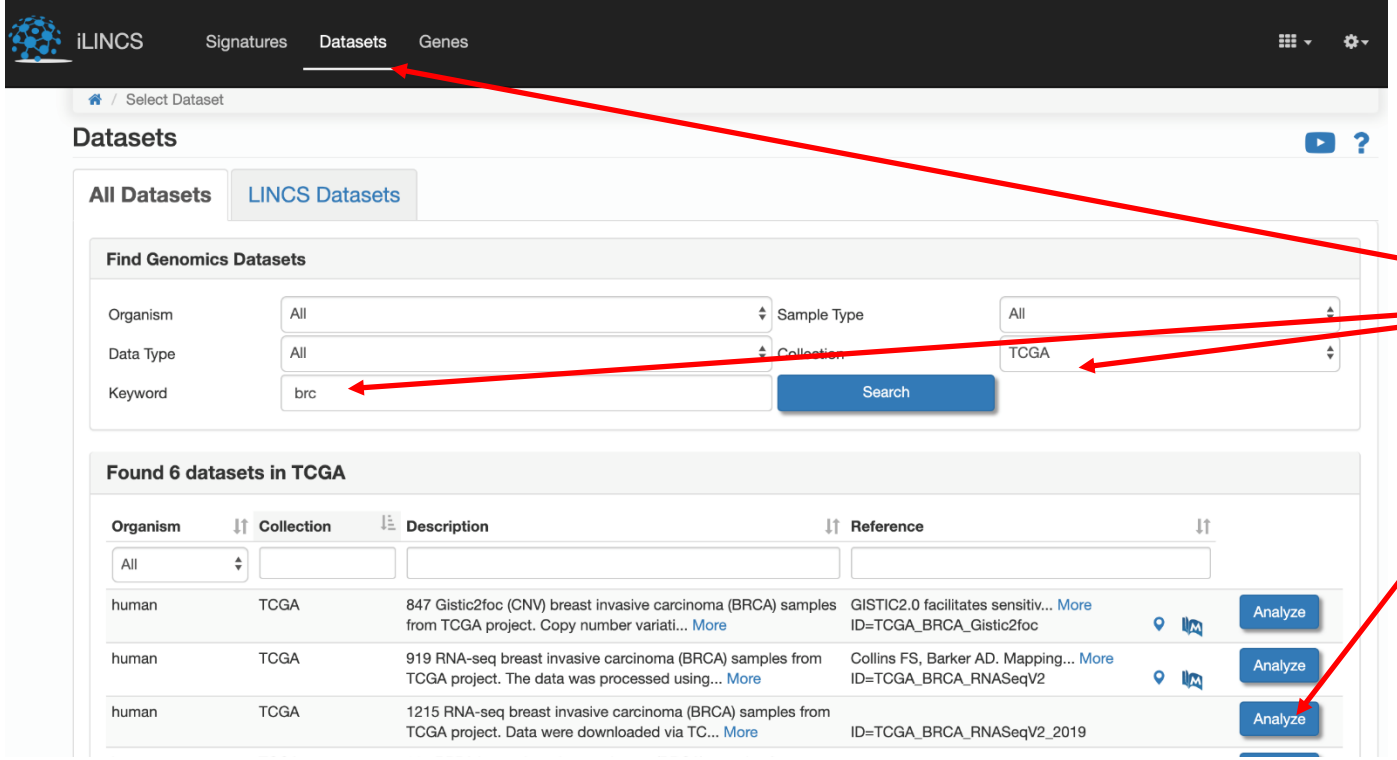
15. Click to create the signature (ie perform differential expression analysis for the selected 11 genes)

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

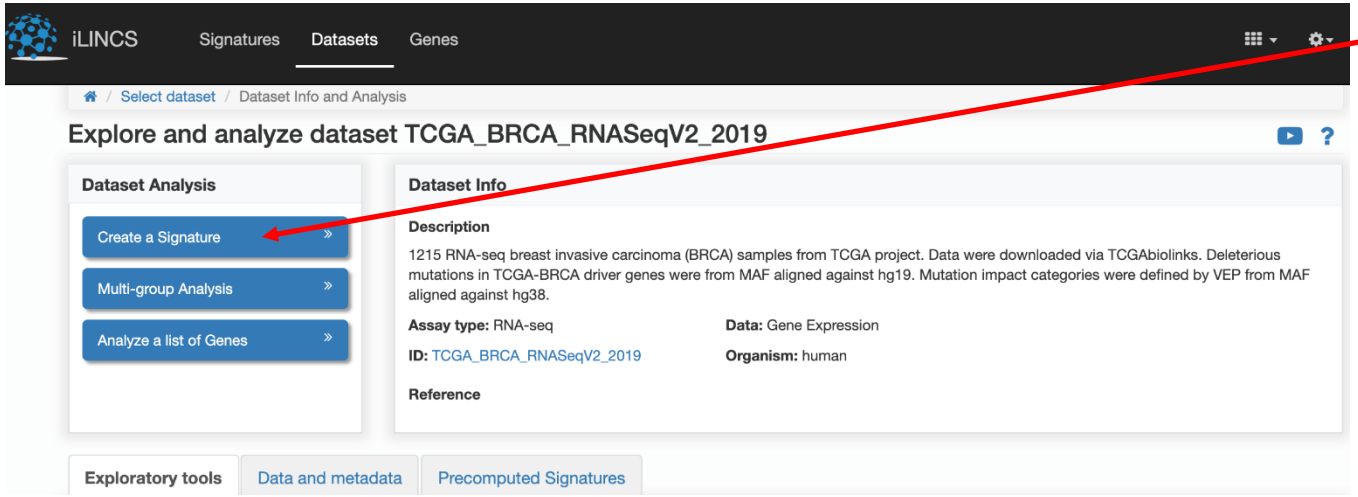

iLINCS
Signatures
Datasets
Genes
Select dataset / Dataset Info and Analysis / Create Signature / Differential Expression Signature
Signature created from dataset TCGA_BRCA_RNASeqV2_2019 with a gene list
Signature analysis
Modify the list of selected genes
Other analyses with selected genes
Dataset Info
Signature Info
Download
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
L1000CDS2
L1000FWD
X2K
X2Kweb
Expression to Kinases
SigNetA
iLINCS
Signatures
Datasets
Genes
filteredeset_Wed_Oct_9_09_04_57_2019_7348583
File Edit View Tools Help
Rows
Columns
11 rows by 289 columns 0 rows, 0 columns selected
PAM50_mRNA
ID_geneid
Name_GeneSymbol
PROBE
596 BCL2 ENSG00000171791
2099 ESR1 ENSG00000091831
2625 GATA3 ENSG00000107485
5241 PGR ENSG00000082175
2064 ERBB2 ENSG00000141736
2539 G6PD ENSG00000160211
2305 FOXM1 ENSG00000111206
891 CCNB1 ENSG00000134057
440 ASNS ENSG00000070669
1978 EIF4EBP1 ENSG00000187840
1956 EGFR ENSG00000146648



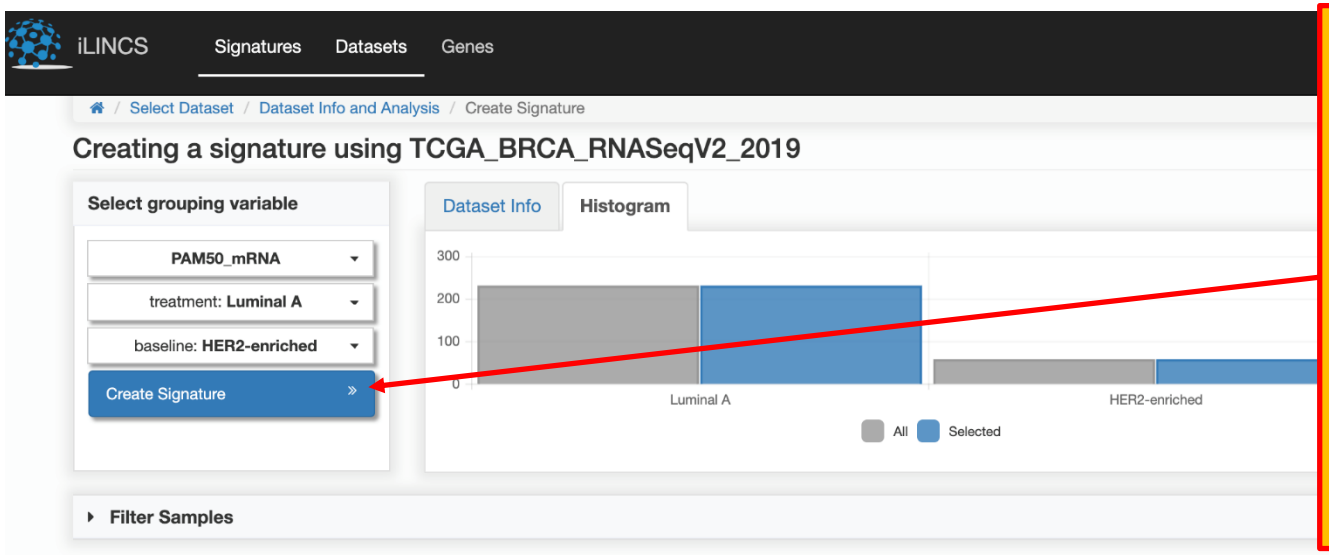
21. Start with the Datasets workflow

22. Search Datasets for BRC in the TCGA collection

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



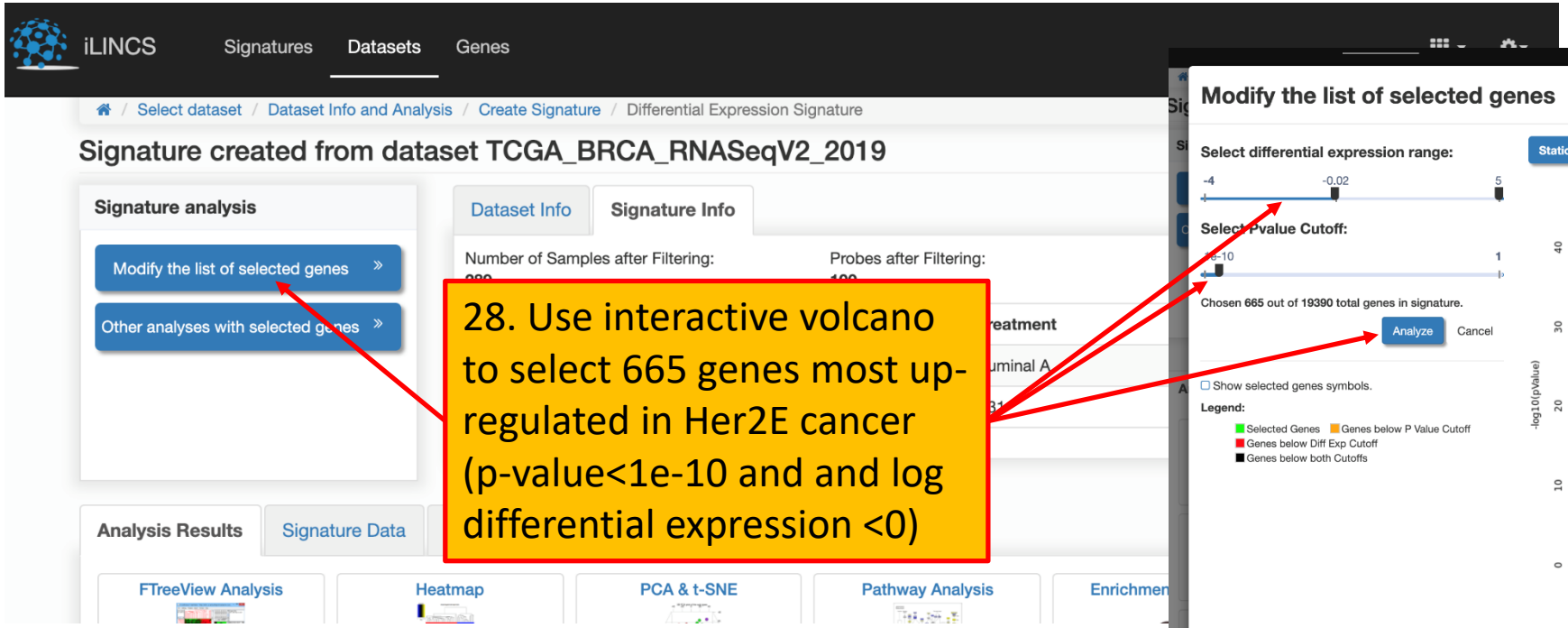
24. Initiate the full signature creation



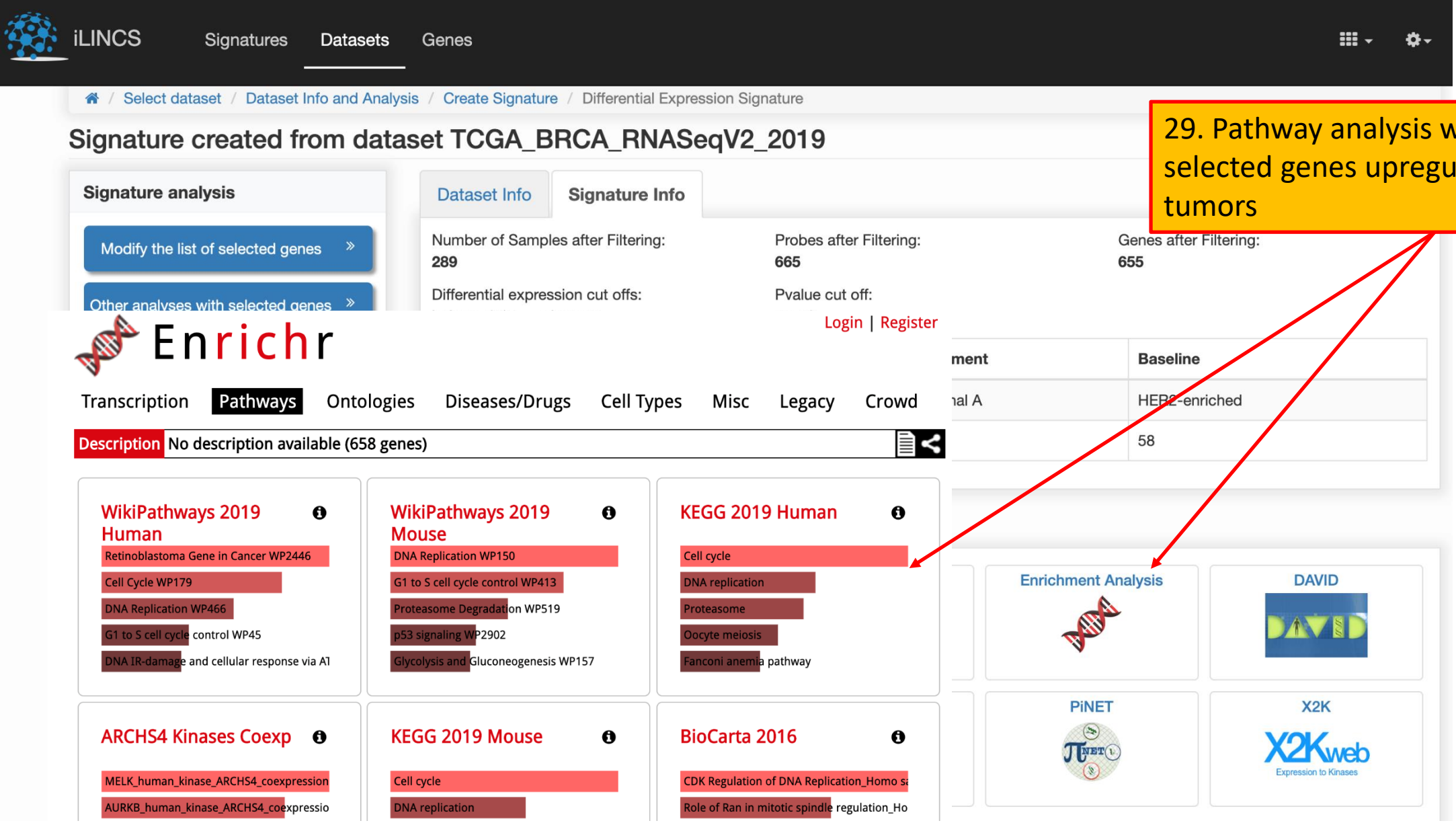
25. Select the sample property (ie factor) that will be used to construct two groups of samples for signature creation

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

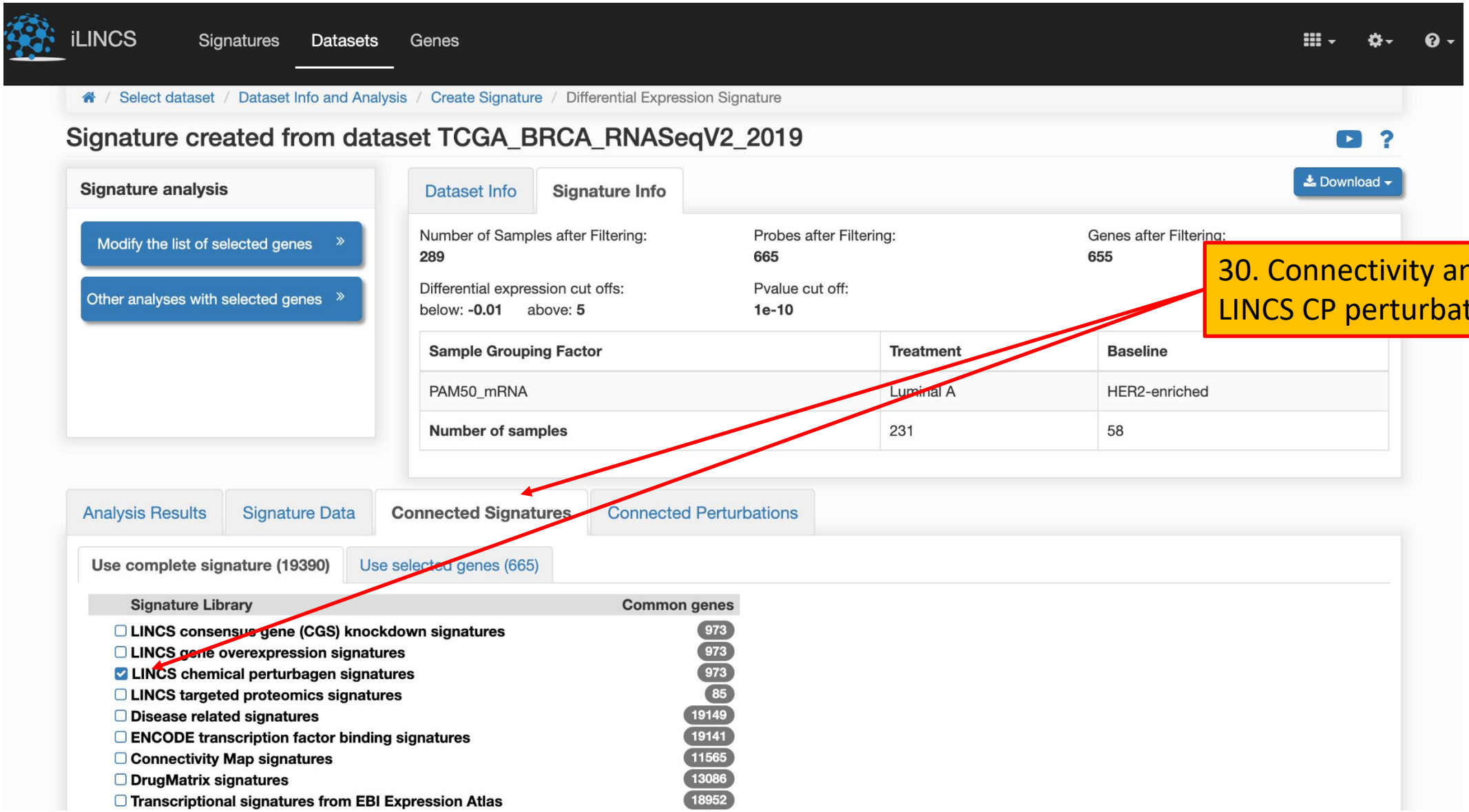
27. Perform differential gene expression analysis between treatment and baseline samples

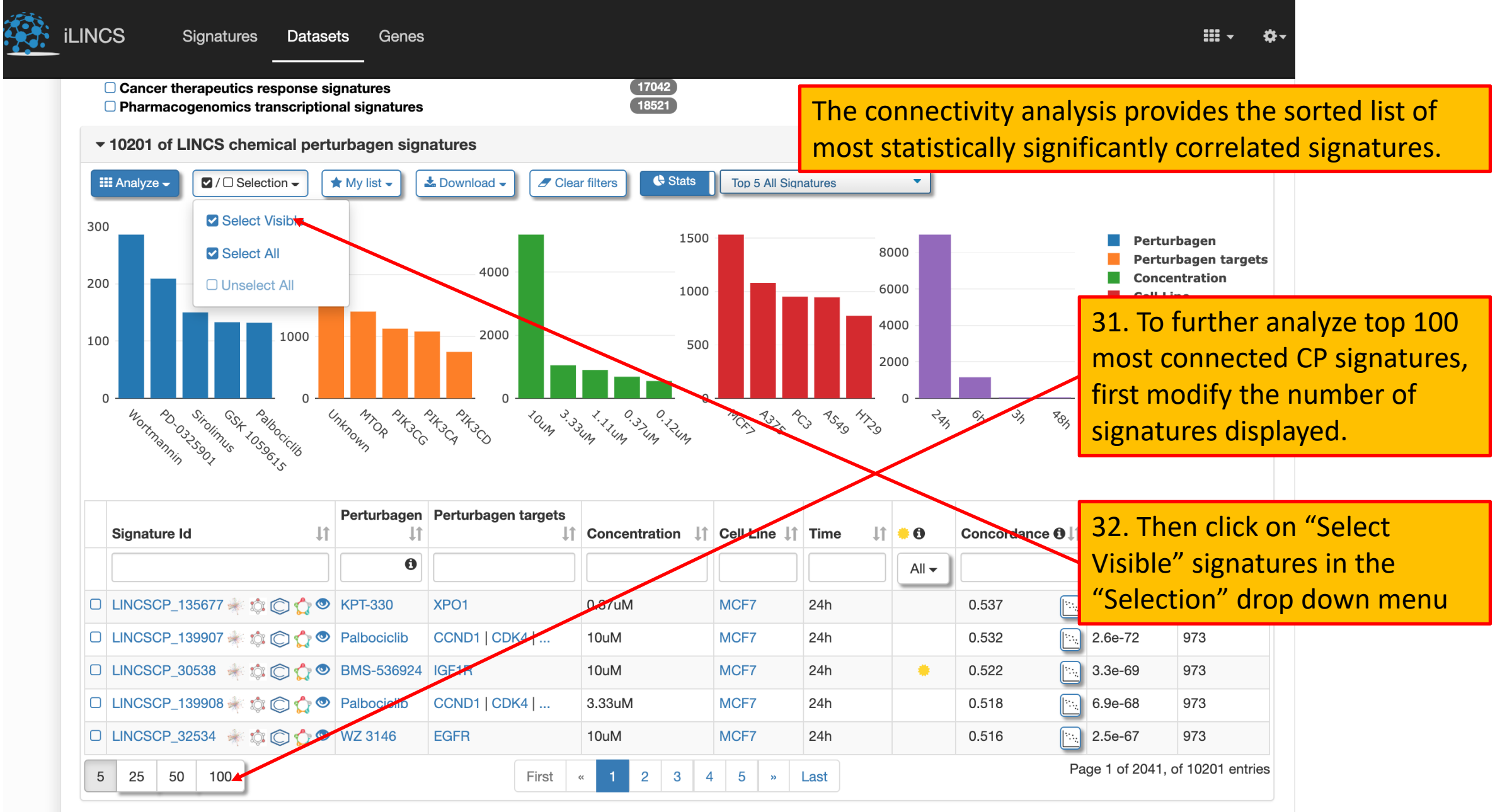


28. Use interactive volcano to select 665 genes most up-regulated in Her2E cancer (p-value<1e-10 and and log differential expression <0)



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


iLINCS
Signatures
Datasets
Genes
Select dataset / Dataset Info and Analysis / Create Signature / Differential Expression Signature
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
Analysis Results
Signature Data
Connected Signatures
Connected Perturbations
Use complete signature (19390)
Use selected genes (665)
Signature Library | Common genes
LINCS consensus gene (CGS) knockdown signatures 973
LINCS gene overexpression signatures 973
LINCS chemical perturbagen signatures 973
LINCS targeted proteomics signatures 85
Disease related signatures 19149
ENCODE transcription factor binding signatures 19141
Connectivity Map signatures 11565
DrugMatrix signatures 13086
Transcriptional signatures from EBI Expression Atlas 18952
30. Connectivity analysis with LINCS CP perturbation signatures



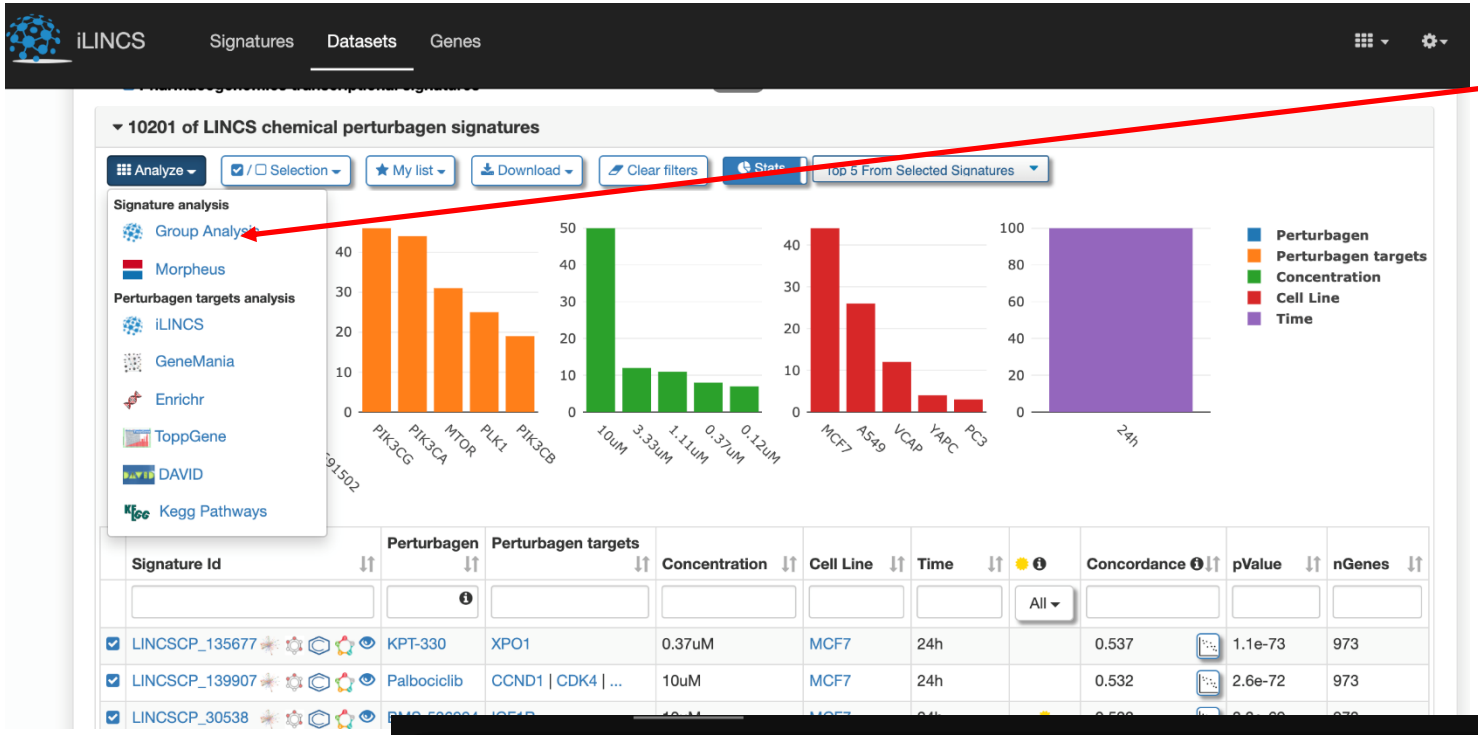
The connectivity analysis provides the sorted list of most statistically significantly correlated signatures.

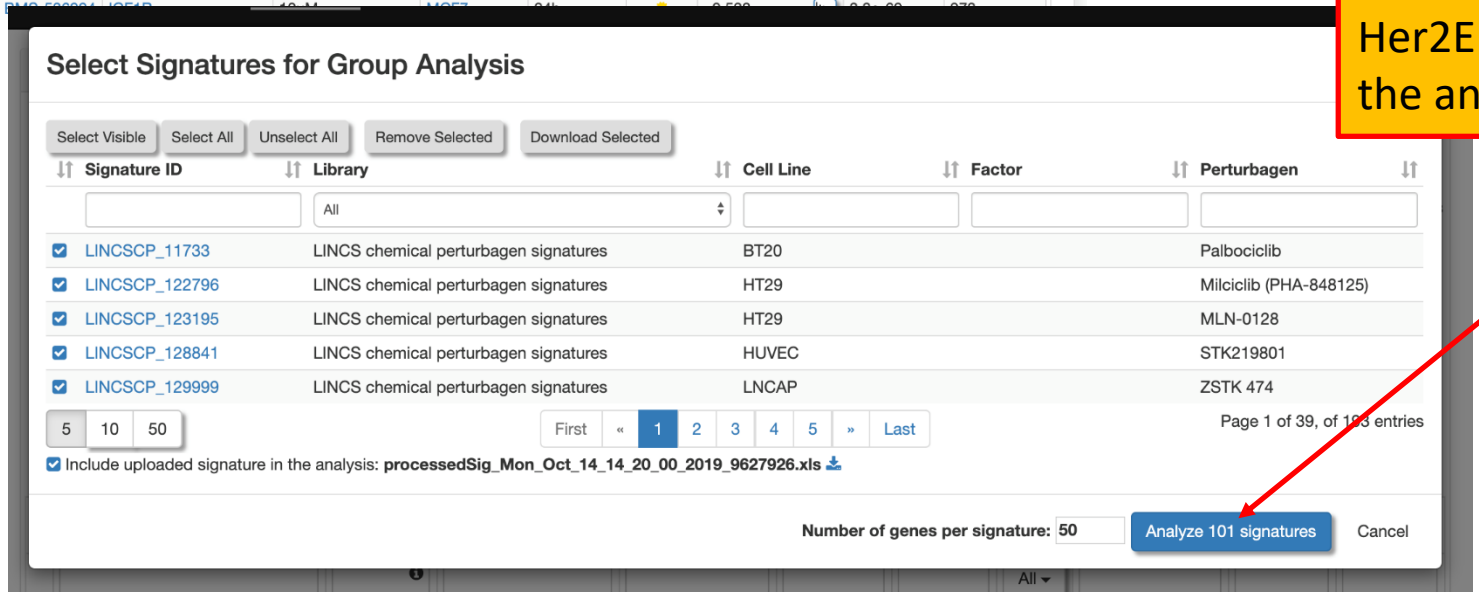
31. To further analyze top 100 most connected CP signatures, first modify the number of signatures displayed.

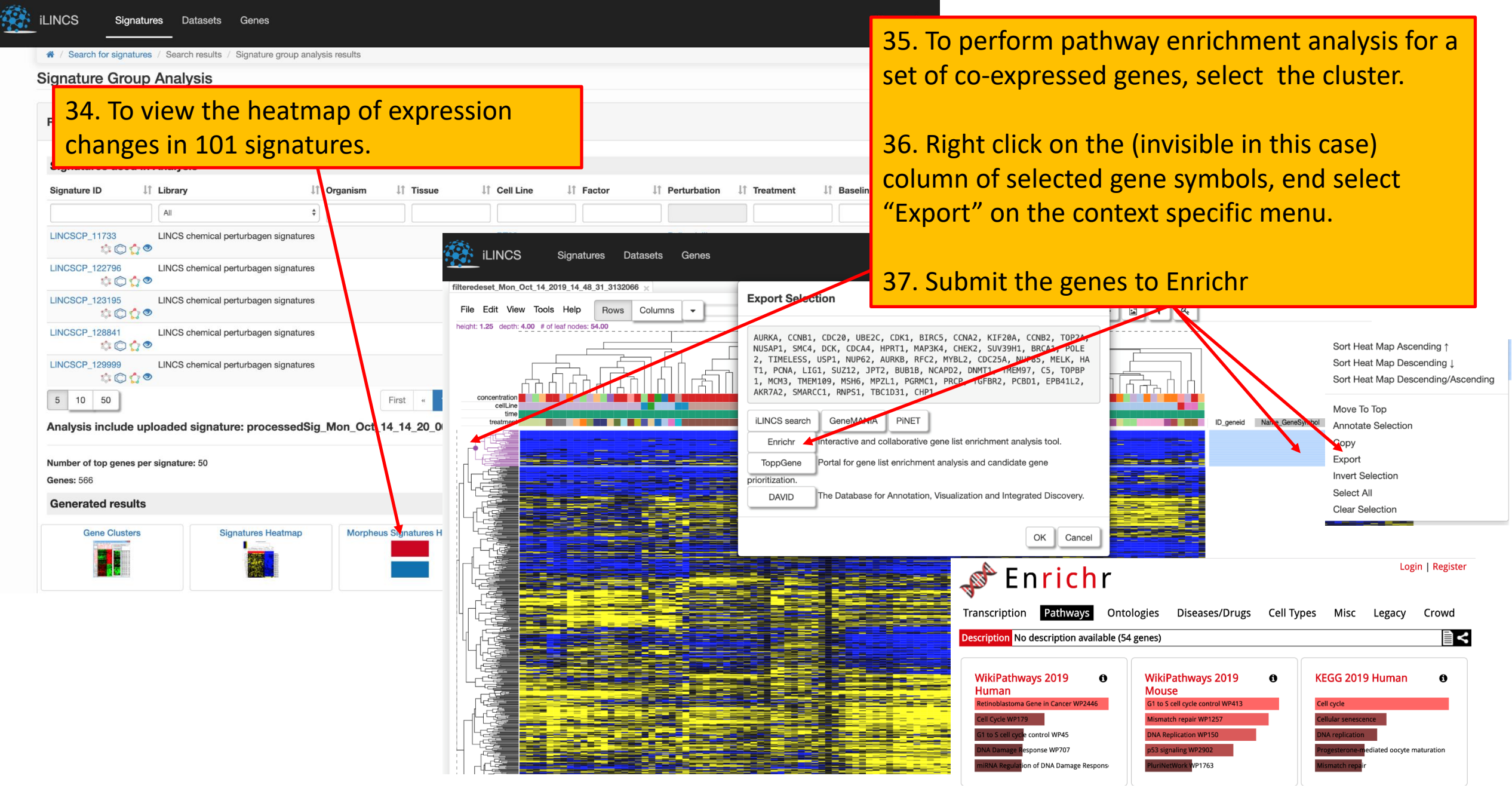
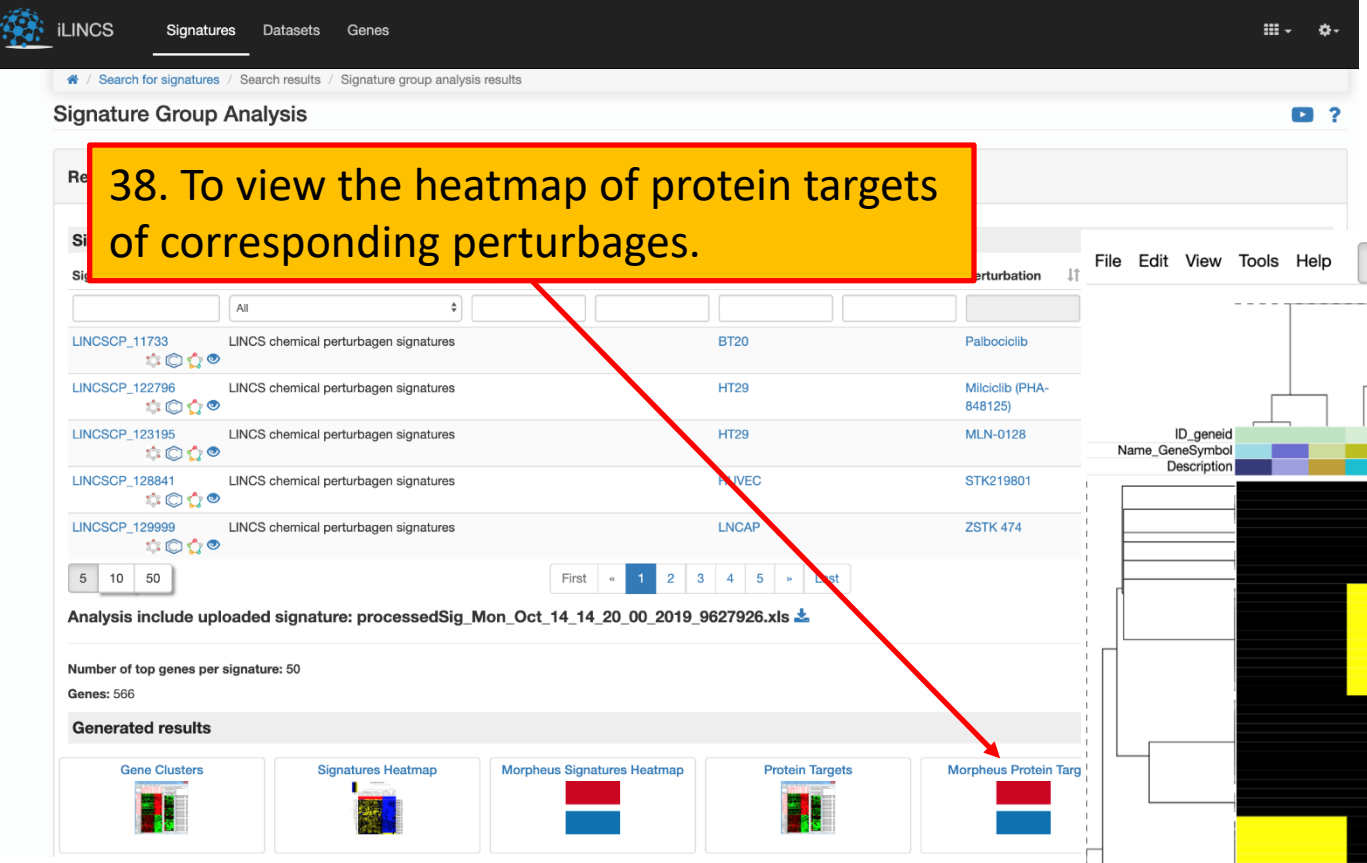
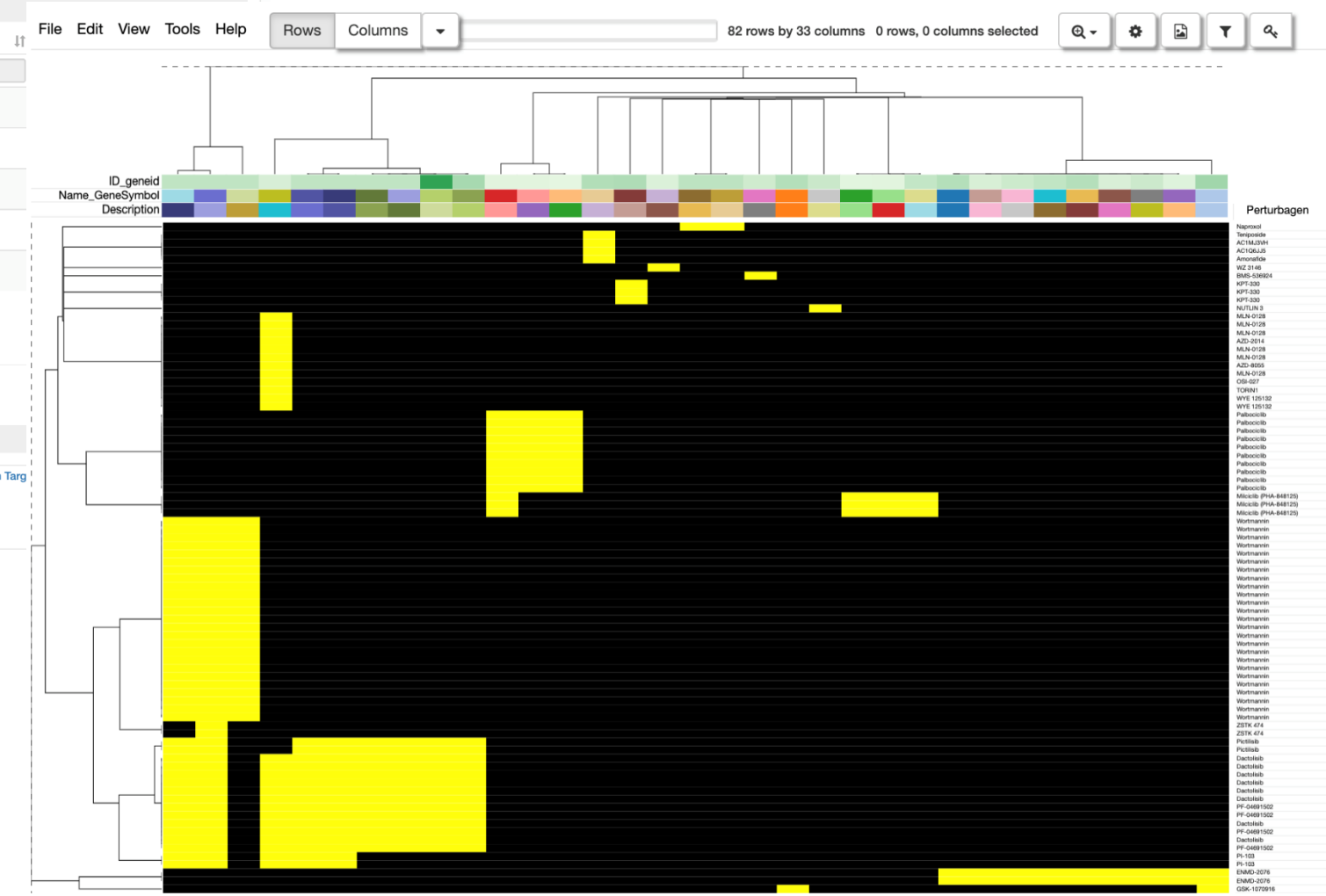
32. Then click on “Select Visible” signatures in the “Selection” drop down menu

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.






iLINCS
Signatures
Datasets
Genes
Search for signatures / Search results / Signature group analysis results
Signature Group Analysis
34. To view the heatmap of expression changes in 101 signatures.
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
Signature ID
Library
Organism
Tissue
Cell Line
Factor
Perturbation
Treatment
LINCSCP_11733
LINCS chemical perturbagen signatures
LINCSCP_122796
LINCS chemical perturbagen signatures
LINCSCP_123195
LINCS chemical perturbagen signatures
LINCSCP_128841
LINCS chemical perturbagen signatures
LINCSCP_129999
LINCS chemical perturbagen signatures
Analysis include uploaded signature: processedSig_Mon_Oct_14_14_20_0
Number of top genes per signature: 50
Genes: 566
Generated results
Gene Clusters
Signatures Heatmap
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, TGFBR2, PCBD1, EPB41L2, AKR7A2, SMARCC1, RNPS1, TBC1D31, CHP1
iLINCS search
GeneMANIA
PiNET
Enrichr
Interactive and collaborative gene list enrichment analysis tool.
ToppGene
Portal for gene list enrichment analysis and candidate gene prioritization.
DAVID
The Database for Annotation, Visualization and Integrated Discovery.
OK
Cancel
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
Login | Register
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
WikiPathways 2019 Mouse
G1 to S cell cycle control WP413
Mismatch repair WP1257
DNA Replication WP150
p53 signaling WP2902
KEGG 2019 Human
Cell cycle
Cellular senescence
DNA replication
Progesterone-mediated oocyte maturation
Mismatch repair

38. To view the heatmap of protein targets of corresponding perturbages.