

HEMAP: Online resource for interactive exploration and e-staining of hematopoietic cancer data

08.01.2016

Table of Contents

HEMAP Overview	1
Home page	2
Explore:.....	2
e-staining the gene expression state with <i>IRX1</i>	4
e-staining Pathway ROSS_AML_WITH_PML_RARA_FUSION	5
e-staining acute precursor B cell vs. chronic lymphocytic leukemia clusters.....	6
Accessing pairwise analysis results.....	7
Search: <i>cancermap cluster CLL and GEXP type</i>	7
Search: <i>cancermap cluster aml and DrugSigDB type with GSK subfiltering</i>	9
Gene sets of pathway and drug signatures.....	10
TCGA AML Map	11
Additional data types for TCGA AML map: GNAB, CNVR, METH, MIRN and CLIN categories.....	12
e-staining <i>mir-125 across METH and MIRNA</i>	12
Lookup of copy number aberrations by gene name, vice versa.....	13
Annotation Search, e-staining, and box plotting.....	14
Main Category: <i>Leukemia AND Cytogenetics: crlf2 analysis to IRX3</i>	14
Sample groups from map cluster selection (mouse) and plotting combined with annotation search	16
Basic use case example: Annotation categories to gene expression e-staining	18
Advanced use case example: <i>In silico</i> drug screening using Hemap	24
Settings.....	29
Info.....	29

HEMAP Overview

Hemap (<http://compbio.uta.fi/hemap>) is the interactive online resource component of the project *Hemap: an Interactive Resource for Molecular Taxonomy and Comparative Analysis of Cancers with Hematopoietic Origin*. Please contact Merja Heinaniemi (merja.heinaniemi@uef.fi) and Matti Nykter (matti.nykter@uta.fi) for credentials. The HTML5 application is designed for interactive exploration of ~10,000 samples across cell types and hematological malignancies. For AML, additional ~100 samples of the TCGA dataset are included. The samples are curated with detailed annotations, including clinical and cytogenetic info. E-staining, dynamic coloring for interactive examination of sample clustering, is supported across annotation term searches, Pathway/Drug gene set scores and omics profiles. Alternatively, gene expression profiles can be

visualized with boxplots of custom sample groups and cluster selections. The main analysis feature of Hemap is the pairwise exploration of cancer cluster, drug, and genomic pathway associations with various feature types to characterize the sample groups. Hemap is optimized for HTML5 compatible browsers and has been extensively tested on recent versions of Chrome, Firefox and IE 11+.

Home page

The resource is organized into visualization, exploration (e-staining and analysis) and annotation tabs, followed by additional tools (boxplots), settings and information tabs. The map shown on the main page can be chosen using the links provided (upper left) and supports zoom/viewing of annotations and exporting data for summary statistic plots using mouse drags.

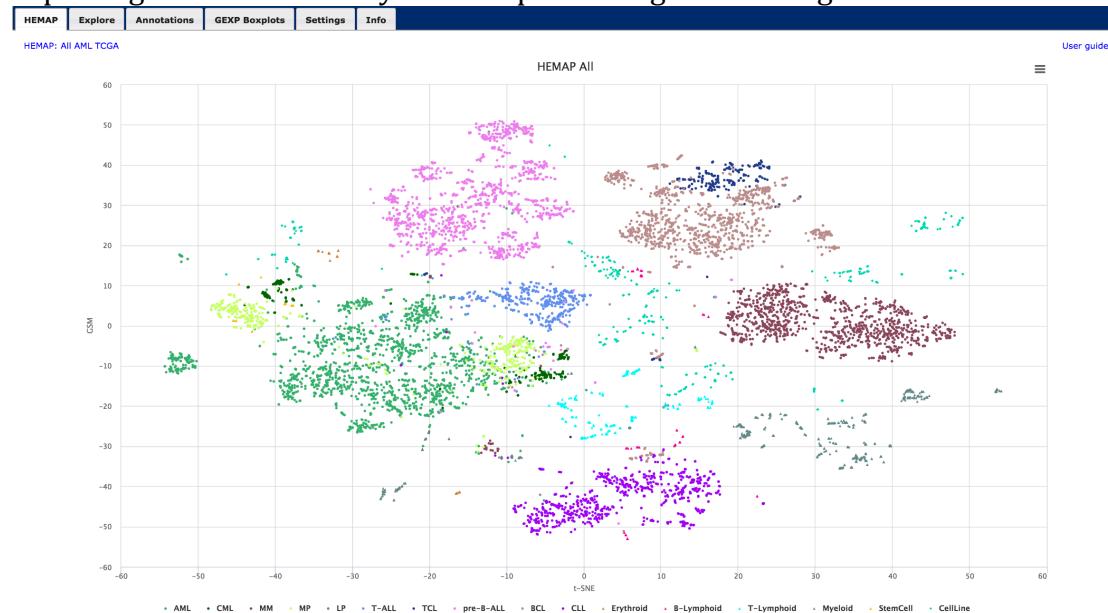


Figure 1. HEMAP home page shows samples (colored data points) on the 2D map obtained by t-SNE. The sample colors are from cell type annotations and disease classifications. The map supports zooming and accessing annotation views from selected samples.

Explore:



The Explore interface supports e-staining by gene, and pathway and drug marker signatures on maps [HEMAP ALL, AML and TCGA]. TCGA map allows staining on copy number, methylation and other omic types. In addition, samples annotated by class/clusters can also be selected. Different analysis results for clusters and cluster comparisons can be accessed using inputs provided for searching pairwise/pairwise results. These sortable records can be selected for e-stain for

visual inspection directly from the result table. The Settings tab allows customizing map and pairwise search parameters, such as order by, legend placement and visibility, max records and staining category colors. Example usages are listed below.

Interface Overview

The Gene and Pathway/DrugSig form allows users to enter gene/gene set (for pathways and drugs) and cluster/class values for e-staining of selected maps. The Gene/Pathway/Drug Signature input field is implemented with type ahead (auto complete) where the possible choices are determined by **Type** selection. For example, if GEXP is selected, then only Gene names are suggested. Selecting TCGA map with GEXP also dictates the program to have the source type ahead of TCGA genes.

Pairwise results associated with cluster/class values can be explored and downloaded upon searching. The pairwise results are associated with specific Types (DrugSigDB, CLIN, GEXP, GSVA for Leukemia and AML; TCGA those four as well as GNAB, CNVR, METH and MIRN designations). PValue, Correlation and Hypergeometric Test measures are provided for pairwise filtering, other resultant columns are listed in the pairwise search section below.

The figure displays two screenshots of the 'explore' interface. Both screenshots show a search form with various input fields and dropdown menus. The top screenshot shows a dropdown menu for 'Type' with options: DrugSigDB, Clinical, GEXP (selected), and GSVA. The bottom screenshot shows a dropdown menu for 'Type' with options: DrugSigDB (selected), Binary:Clinical, Categories:Clinical, Numeric:Clinical, Gene Expression, Pathway, Methylation, Copy number variation, MIRNA, and Mutations.

Figure 2. The explore interface provides intuitive search of pairwise cluster to feature associations and then e-staining. The omic types are DrugSigDB, Clinical, GEXP and GSVA. The bottom figure shows additional TCGA cluster/class types, including copy number and methylation.

e-staining the gene expression state with *IRX1*

Auto-suggestions (type aheads) are built into Gene, pathways, and cluster/class fields to ensure valid selections. The example here shows how to visualize *IRX1* (from figure 2 of manuscript) e-staining and as well as image export functionality.

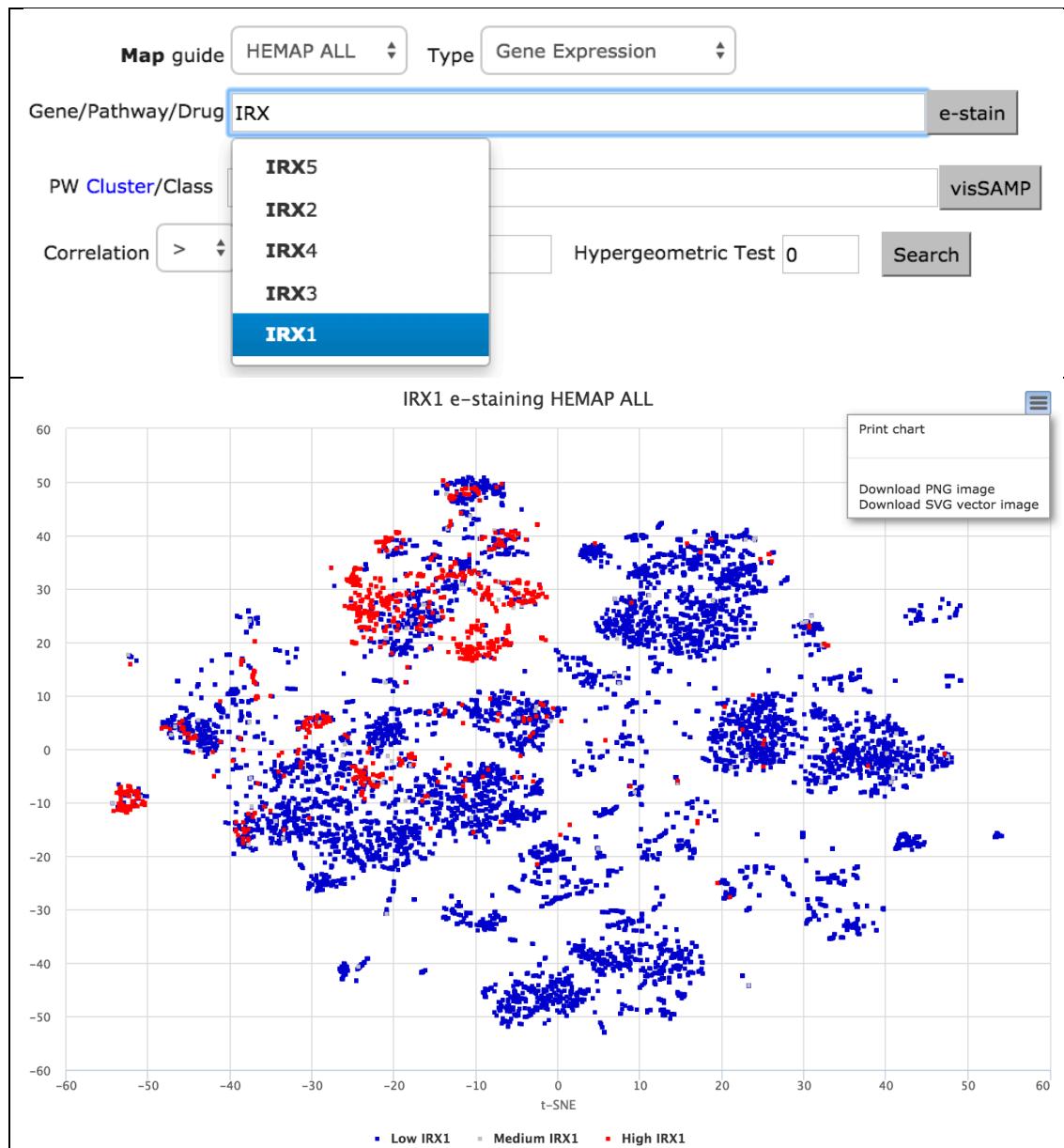


Figure 3. e-staining is performed using *IRX1*, where low profiled samples are blue, medium grey and high red. Samples in the map are selectable and advanced functions include mouse dragging for summary plotting. Details are provided in the use case section.

e-staining Pathway ROSS_AML_WITH_PML_RARA_FUSION

The example here uses the AML map to locate samples with the PML-RARA fusion gene based on a previously published gene set for this subtype. Shown in figure 4, entry of “PML_RAR” automatically brings up suitable matches and the ROSS_AML_WITH_PML_RARA_FUSION from MsigDB is selected and then stained. The graph in the middle lists the gene members of this pathway. To reproduce the result shown below, the FDR cut-off should be set to 0.001 from the Settings tab.

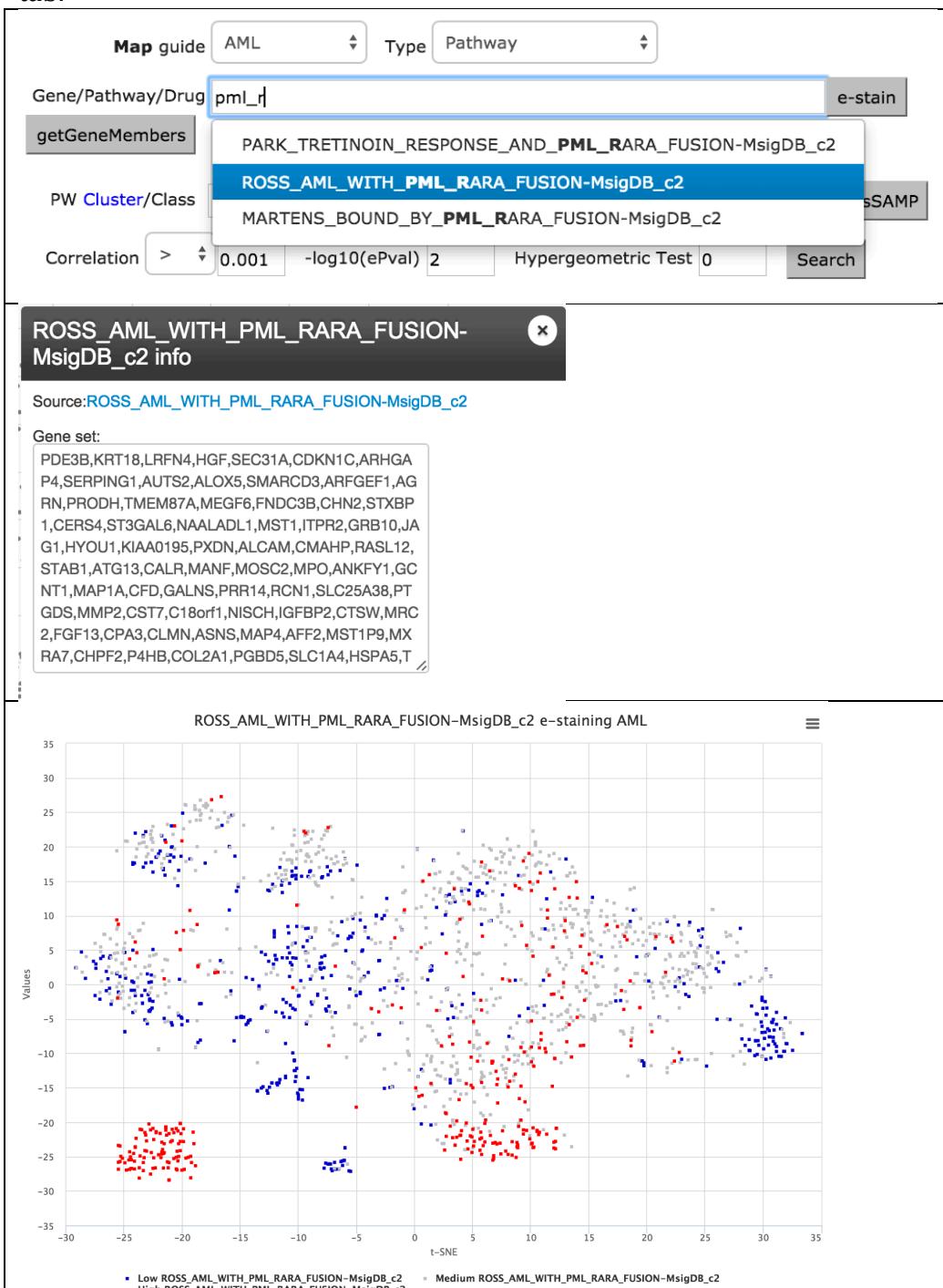


Figure 4. AML map e-stained with ROSS_AML_WITH_PML_RARA fusion event reported in MsigDB. The resource contains type specific autosuggestions and gene membership retrieval functions. The middle figures list gene members of the selected pathway.

e-staining acute precursor B cell vs. chronic lymphocytic leukemia clusters

In order to make the cluster exploration intuitive, it is easy to find out where the samples belonging to a cluster (or clusters being compared) are located on the map as a first step. The scenario below, figure 5, shows the location of acute precursor B cell (pre-B-ALL) vs. chronic lymphocytic leukemia (CLL) using VisSAMP that indicates cluster locations with binary labels vice versa: blue for the 2nd (CLL) category and red for the first (pre-B-ALL), grey nodes are NA. Both diseases have a tight cluster confirming the original map sample placements. VisSAMP stains samples of “_and_” clusters red if both categories applied.

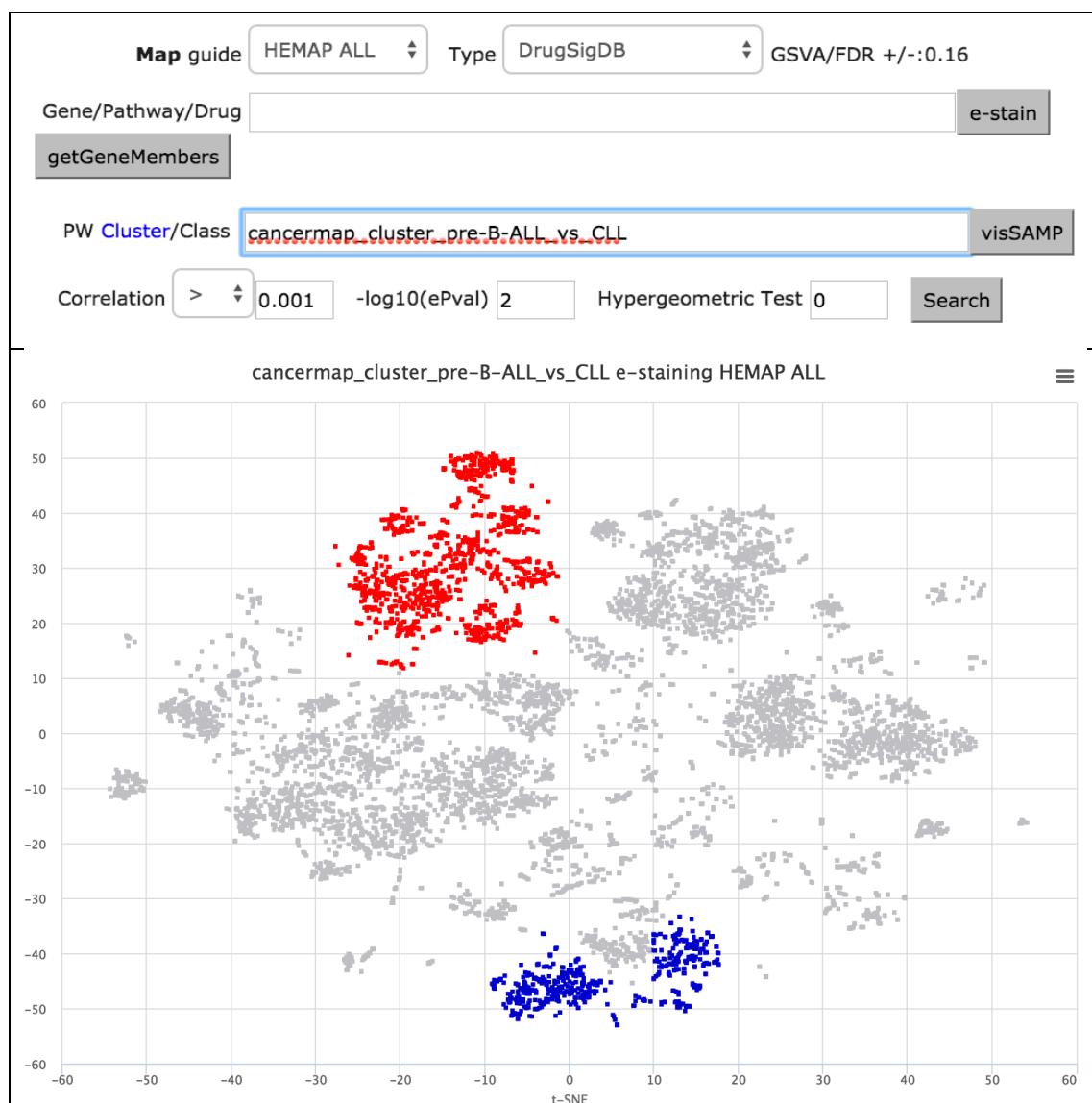


Figure 5. Samples can be directly colored from cluster features. Recalled that pre-B-ALL samples on the original map shared the same location as the red highly expressed samples here. In contrast, the blue samples represent CLL labels.

Accessing pairwise analysis results

In addition to exploration of selected gene features and gene sets (based on expert knowledge on disease biology), the PW Cluster/Class field enables the user to analyze in an unbiased manner the pairwise comparisons between cluster assignment and different sample features. The features of interest are selected from the “Type” drop-down menu and the cluster(s) of interest are specified in the PW Cluster/Class field.

The results (default to max 5000, see Settings at the end of the guide for adjustment) can be downloaded as a table and individual features (rows) can be selected directly for e-staining. Each column is sortable and they are:

1. Feature (e.g. a Gene or a pathway gene set)
2. Spearman correlation
3. Adjusted -log 10(P-value) from correlation test
4. BH factor adjustment for correlation test (in multiple testing correction)
5. Adjusted -log 10(P-value) from cluster enrichment test (hypergeometric test), applicable for GSVA/Drug Signature
6. BH factor adjustment for hypergeometric test
7. Number of samples
8. Diff Samples A
9. Diff Samples B
10. Number of missing values in the first feature
11. Number of missing values in the second feature

Search: cancermap cluster CLL and GEXP type

The figure below describes cancer map cluster CLL selection and then resultant rows ordered by adjusted P-value are returned for the cancermap cluster CLL. All columns are sortable. In the example below, the result was sorted by correlation (highest first) (arrows in the column names are clickable to sort in both directions). The map is stained with SFMBT1 (middle graph with name underlined and mouse over tip) by clicking on the gene name.

The screenshot shows the e-stain search interface with the following settings:

- Map guide: HEMAP ALL
- Type: Gene Expression
- Gene/Pathway/Drug: (empty input field)
- PW Cluster/Class: cancermap_cluster_cl (highlighted in blue)
- Correlation: > (highlighted in blue, showing 'cancermap_cluster_ClL' as the selected value)
- Hypergeometric Test: 0
- Buttons: e-stain, visSAMP, Search

[Download\(500\)](#)

feature	corr.	adjpvalue	BHadj	hypergeom_test	BHadj	nsamples	ndiff_1	ndiff_2	nna_1	nna_2
MATN1-AS1	0.410	292.333	7.667	0		0	9544	0	0	0
RNF41	0.410	292.333	7.667	0		0	9544	0	0	0
DNMBP	0.410	292.333	7.667	0		0	9544	0	0	0
FOXP1	0.410	292.333	7.667	0		0	9544	0	0	0
<u>SFMBT1</u>	0.410	292.333	7.667	0		0	9544	0	0	0
ADAM click to stain	0.410	292.333	7.667	0		0	9544	0	0	0
TNFRSF13C	0.400	292.333	7.667	0		0	9544	0	0	0
ISCU	0.400	292.333	7.667	0		0	9544	0	0	0
P2RY10	0.400	292.333	7.667	0		0	9544	0	0	0
LINC00926	0.400	292.333	7.667	0		0	9544	0	0	0
MAP3K1	0.400	292.333	7.667	0		0	9544	0	0	0
ZBTB4	0.400	292.333	7.667	0		0	9544	0	0	0
ARHGAP24	0.400	292.333	7.667	0		0	9544	0	0	0

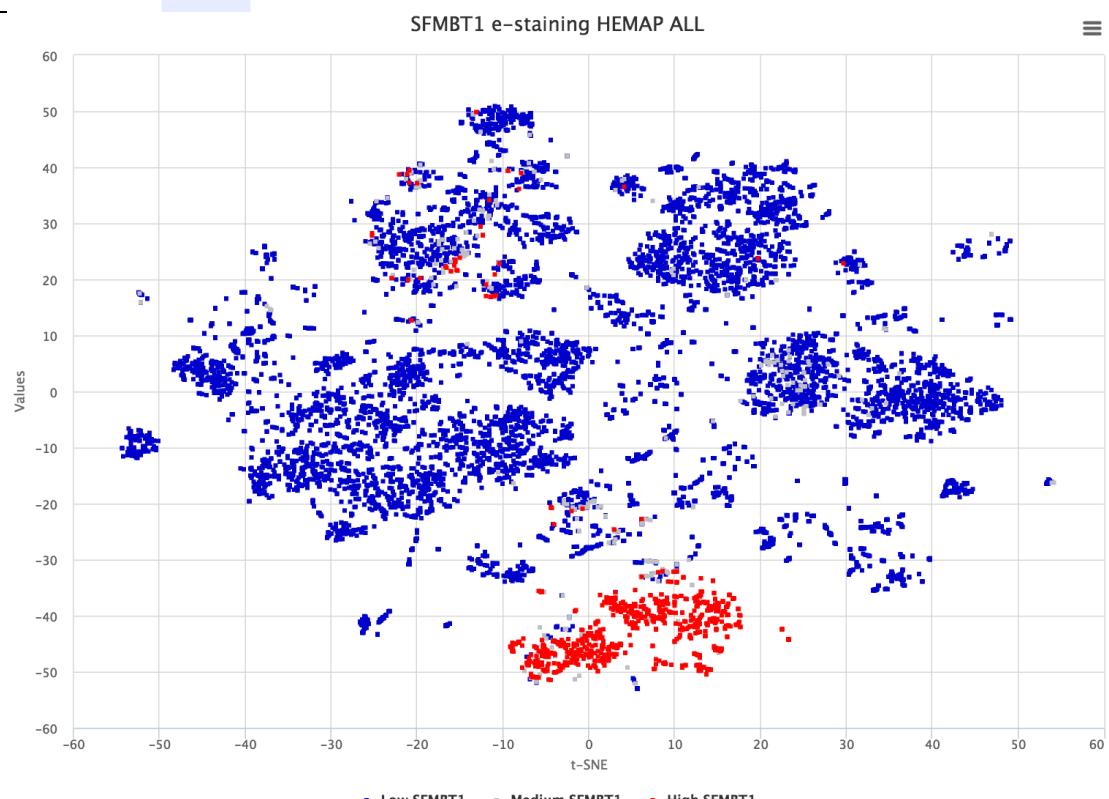


Figure 6. This use case details cancer cluster CLL analysis of finding highly correlated pairwise gene features. The results can be downloaded for advanced statistical studies. Here, SFMBT1 is selected after correlation sorting and then interactively e-stained thus confirming CLL original samples.

Search: cancermap cluster aml and DrugSigDB type with GSK subfiltering

DrugSigDB and Pathway features are sets of curated gene lists from known published resources, including Drug Signature DataBase (DSigDB), Wikipathways and PWCommons. The figure below shows the 38 results upon sub-filtering (original with 1990) for pathway features labeled 'GSK' (Glaxo-Smith-Kline) developed drugs. Number of rows shown per page is defaulted to 15 and can be adjusted in Settings.

Map guide
HEMAP ALL
Type
DrugSigDB

feature	corr.	adjpvalue	BHadj	hypergeom_test	BHadj	nsamp
GW770249X_GSK-DSigDB_D2	0.490	292.194	7.806	307.017	7.806	9544
GW770249A_GSK-DSigDB_D2	0.420	292.194	7.806	202.716	7.806	9544
GW779439X_GSK-DSigDB_D2	0.380	289.994	7.806	128.264	7.806	9544
GW806290X_GSK-DSigDB_D2	0.370	284.494	7.806	121.114	7.806	9544
GW770220A_GSK-DSigDB_D2	0.350	243.294	7.806	38.194	7.806	9544
GW795493X_GSK-DSigDB_D2	0.340	237.694	7.806	65.342	7.806	9544
GW778894X_GSK-DSigDB_D2	0.340	232.494	7.806	57.194	7.806	9544
GW810576X_GSK-DSigDB_D2	0.330	218.994	7.806	132.08	7.806	9544
GW795486X_GSK-DSigDB_D2	0.320	210.694	7.806	80.239	7.806	9544
GSK_3_INHIBITOR_XIII_RBC-DSigDB_D2	0.270	146.194	7.806	63.614	7.806	9544
GSK_3_INHIBITOR_IX_RBC-DSigDB_D2	0.270	143.794	7.806	55.387	7.806	9544
GW830365A_GSK-DSigDB_D2	0.260	137.994	7.806	0	7.806	9544
GW612286X_GSK-DSigDB_D2	0.240	112.794	7.806	9.342	7.806	9544
GSK2110236A_GSK-DSigDB_D2	0.240	111.994	7.806	43.216	7.806	9544
GW784752X_GSK-DSigDB_D2	0.240	111.194	7.806	17.581	7.806	9544

Showing 1 to 15 of 38 entries (filtered from 1,990 total entries)
 Previous
1
2
3
Next

flip
Filter results

GSK

Figure 7. Sub filtering, using GSK, is demonstrated on AML class features. The number of rows is adjusted to 38 from 1990. Sub-filtering is functional across all columns with character matching.

Gene sets of pathway and drug signatures

Gene members upon pathway/drug gene set selection can be easily looked up. Clicking on the *getGeneMembers* brings up a dialog box showing the source link and gene set.

Map guide HEMAP ALL Type DrugSigDB

Gene/Pathway/Drug
MLN_518_KINOME_SCAN-DSigDB_D2 e-stain

getGeneMembers

MLN_518_KINOME_SCAN-DSigDB_D2 info

Source:MLN_518_KINOME_SCAN-DSigDB_D2

Gene set:

PDGFRB,PDGFRA,MAP4K5,DDR2,CSF1R,EGFR,NTRK1,NTRK2,KIT,CLK1,IRAK3,FLT3

Gene Set: D2 : Kinome Scan - Tandutinib

Collection	D2 : Kinome Scan						
Chemical Name	Tandutinib (From Source : MLN-518)						
FDA	NPC	WHO	Indian	Australia	China	Traditional Herbal	Clinical Trail
Not	Not	Not	Not	Not	Not	Not	Not
Molecular Weight	Hydrogen Bond Donor Count	Hydrogen Bond Acceptor Count	cLogP	Lipinski Rule			
562.703 g/mol	1	10	5.0483	False(2/4)			

Structure

InChI

InChI=1S/C31H42N6O4/c1-23(2)41-25-10-8-24(9-11-25)34-31(38)37-17-15-36(16-18-37)30-26-20-28(39-3)29(21-27(26)32-22-33-30)40-19-7-14-35-12-5-4-6-13-35/h8-11,20-23H,4-7,12-19H2,1-3H3,(H,34,38)

JSmol

Figure 8. The top panel shows MLN-518 (synonym for Tandutinib) drug signature selection and its gene members are shown. The bottom graph shows results of clicking on the source DSigDB (Tan lab) link.

TCGA AML Map

The default map of TCGA colors the samples by cytogenetic subtypes (fusion genes and mutations) Alternatively, the map can be re-plotted with the t-SNE cluster labels by clicking on [Cluster](#) link shown below.

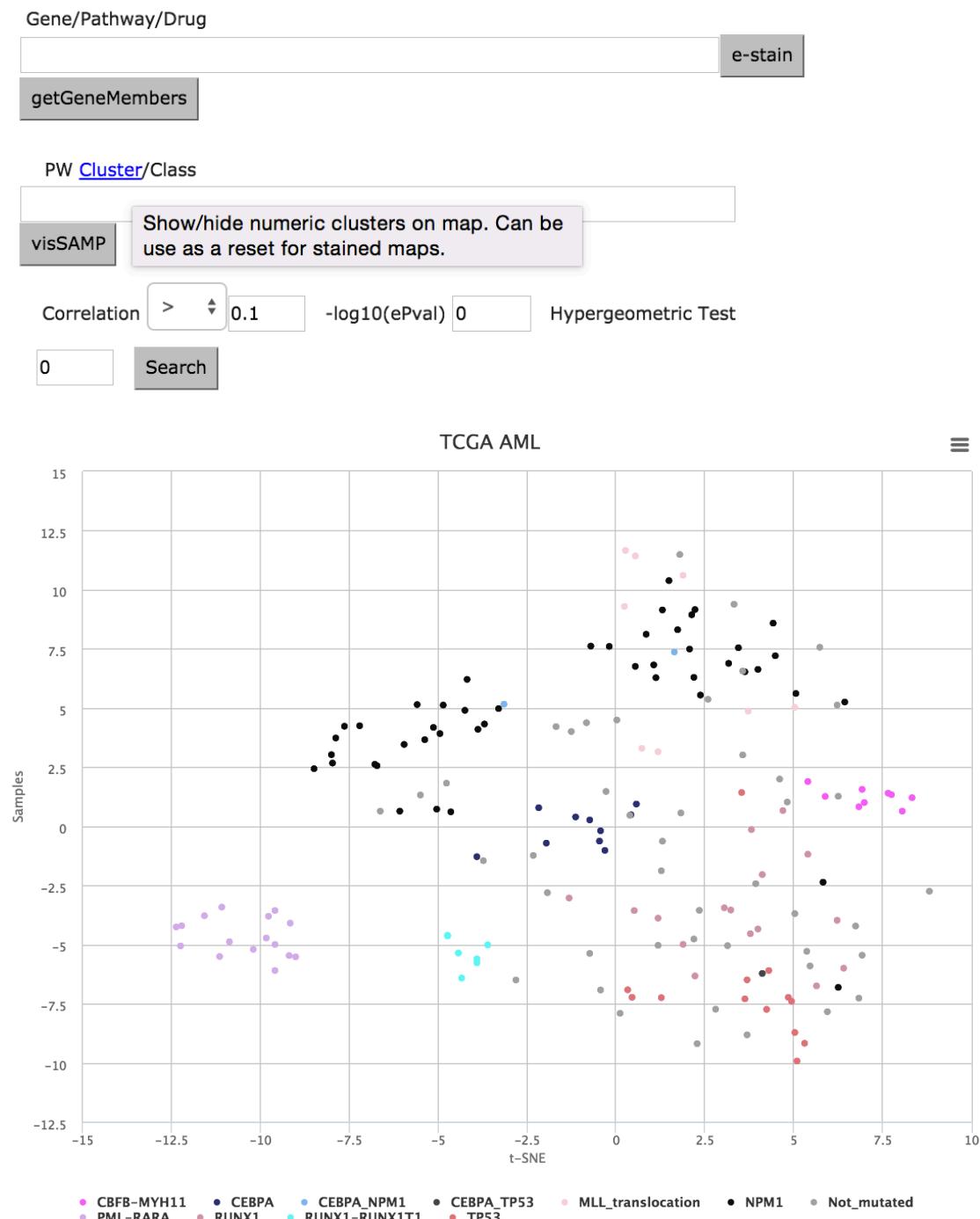


Figure 9. The default view of the TCGA map indicates the location of patients with gene fusion and mutation events in color.

Additional data types for TCGA AML map: GNAB, CNVR, METH, MIRN and CLIN categories

The genome-wide mutation, methylation and copy number profiles are available for the TCGA samples with the feature names GNAB, METH and CNVR. Pairwise exploration of cluster association with these features works the same as described above. Below we will look at a few of the data types and the easy manner they can be explored and e-stained.

Certain TCGA data types are continuous numeric values therefore statistic significance/ranking across each element is rank using z-score, computed as $(x - \text{mean})/\text{standard deviation}$:

$$z = \frac{x - \mu}{\sigma}$$

e-staining *mir-125* across METH and MIRNA

The scenario below finds *mir-125b-1* with high positive correlation (.47 to cancermap cluster 1) after selecting the METH datatype and clicking on cluster 1 (PML-RARA cases) in the map. The quick-selection of an interested cluster for pairwise analysis is also supported for other feature types. As explained earlier, e-staining the methylation data is performed on clicking the feature (gene) name directly on the table (middle figure). It is also possible to check whether the differential methylation status affects the miRNA expression, by choosing Type MIRNA and entering the official identifier to the Gene field (hsa-miR-125b-1) (bottom figure, correlation .48).

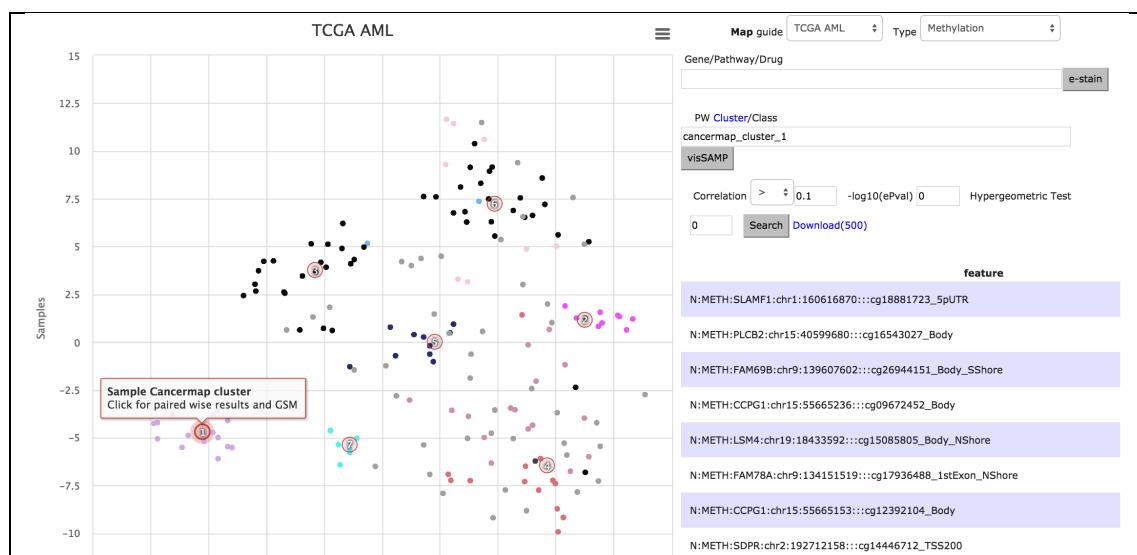




Figure 10. Data analysis is performed visually for mir-125 via e-staining across methylation and miRNA. Pairwise results are interactively retrieved with cluster label mouse click on map.

Lookup of copy number aberrations by gene name, vice versa

Copy number aberrations often include large chromosomal region gains or losses that houses multiple genes. As such, it is useful to search by gene name and then look up the corresponding copy number labels, often in cytoband terms. The figure below illustrates the situation of looking up *TP53*, 20q13.2 is displayed upon clicking *Gene2Cyto*, *Cyto2Genes* click returns gene members.

Map guide TCGA AML **Type** Copy number variation

Gene/Pathway/Drug TP53

e-stain Gene2Cyto Cyto2Genes

Map guide TCGA AML **Type** Copy number variation

Gene/Pathway/Drug 20q13.2

e-stain Gene2Cyto Cyto2Genes

20q13.2 Gene members

RPS4XP3, CDH22, PARD6B, RPL35AP, TOMM34, PLTP, TMSB4XP6, PTPRT, FAS1, RPL12P4, EIF4EBP2P, TFAP2C, STK4-AS1, TMEM189, UBE2V1, SALL4, ZNFX1, AURKA, CCNB1IP1P2, ZHX3, TMEM189, GAPDHP54, 52P, MIR3646, TP53RK, MYBL2, SRMP1, PIGT, HNF4A, RN7SKP33, ZMYND8, WFDC6, SNX21, ZSWIM1, ZSWIM3, SNORD12, SNRPF1, SPINT5P, RN7SL24, DBNDD2, RN7SL197P, SNAP23P, PPIAP10, RPL13P2, MIR4756, MIR3194, KCN5, TRERNA1, RPS2P7, HSPEP1, PKIG, ZNF335, SPATA25, L3MBTL1, DOK5, NEU-miR-4756-5p, hsa-miR-645, hsa-miR-3646, FKSG56, SPINLW1-WFDC6, LOC100505783, LOC100131496, C20orf43, Mir_147, LOC284751, RNU6-5p, RNU6ATAC, LOC79015, hsa-mir-3616, hsa-mir-3617, hsa-mir-1302-5, ZNFX1-AS1, hsa-miR-3194-5p, hsa-mir-645, hsa-miR-3616-3p, SCARNA15, SNORD112, hsa-mir-3646, hsa-miR-3617-5p, hsa-miR-4756-3p, hsa-mir-4756, hsa-mir-3194, hsa-miR-3617-3p

Figure 11. Copy number features are often named as cytobands inclusive of many genes. Hemap allows for lookups in both directions. Example depicts looking up TP53 (top) for 20q13.2(middle) and then all members within the cytoband.

Annotation Search, e-staining, and box plotting

Main Category: *Leukemia AND Cytogenetics: crlf2 analysis to IRX3*

The web application allows for flexible searches across all annotations supporting type ahead and wild card search. For instance and detailed in figure 12, updating the column dropdown to *Main Category* and typing in Leukemia (or use type ahead) returns 4778 records (not shown). The advanced interface is turned on (or hide) by clicking on [Advanced](#) link and setting the 2nd column dropdown to *Cytogenetics* with *crlf2* entry with AND clause, 26 records are returned as depicted in figure 11 (top with partial number of rows). Using the button “See Results on Map” on upper right corner, the middle figure shows the 26 samples (including all *crlf2* cytogenetic deletions and fish) in blue (color picker available as shown). From the annotation results next to the Search button is AddResults2BoxPlot. This adds the sample group to Hemap GEXP box

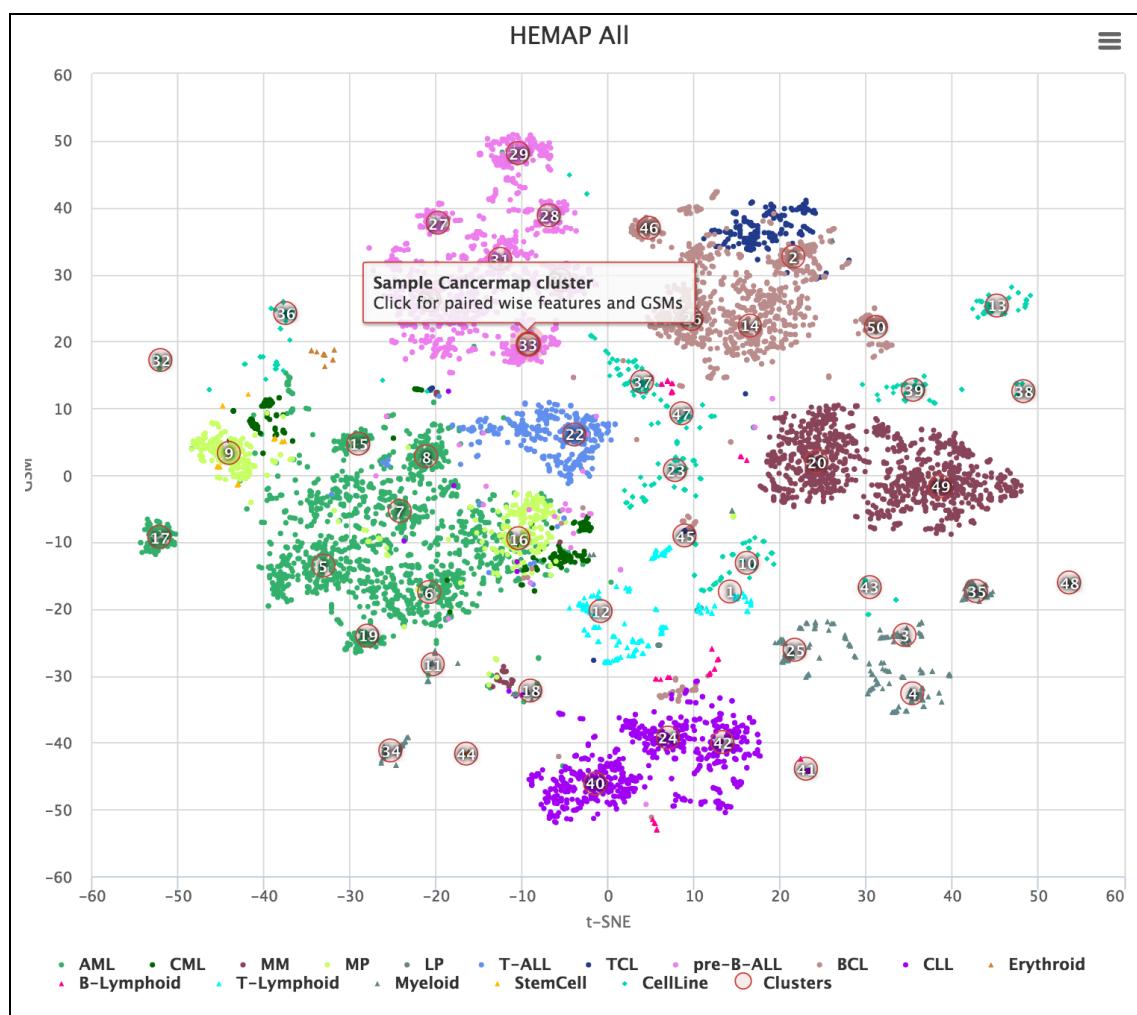
plots; more custom groups can be added across user selected gene(s); in this case *IRX3*.



Figure 12. Annotation search allows for intuitive basic search but also provides powerful advanced functions. Almost 5000 records are labeled Leukemia and shown here is Leukemia samples annotated also with Cytogenetics *crlf2*. The 26 results are then visualized on the map and then gene *FLT3*'s expression summary statistics are plotted and contrasted, revealing substantially higher mean expression.

Sample groups from map cluster selection (mouse) and plotting combined with annotation search

From the main map, the 9,544 samples have been classified into 50 clusters. Prior, we described toggling the map labels using Explore interface 'Clusters' link. For the main map and using the selection for pairwise TCGA searches. Cluster selection on the main map can also add the clustered samples for box plotting. Figure 13 shows selection of Cluster 33 (82 samples) and then box plotted with IRX3 expression values. User need to select the cluster label (red enlarged circles), not sample nodes (selection on sample nodes opens up GEO), the 2nd panel shows a zoomed in area of cluster 33 for easier selection. Upon clicking, the AddCluster33_2Boxplot button needs to be clicked, also in the Explore interface, above the tabular results and shown in same panel. The IRX1 box plot for the 82 samples (from prompt after button) is shown in 3rd panel of figure 13. Hemap is integrated across maps, annotations and plotting. The bottom panel includes cluster 33 IRX3 together with annotation search of subtype pre-B-ALL. Other annotation searches or cluster selections are also compatible.



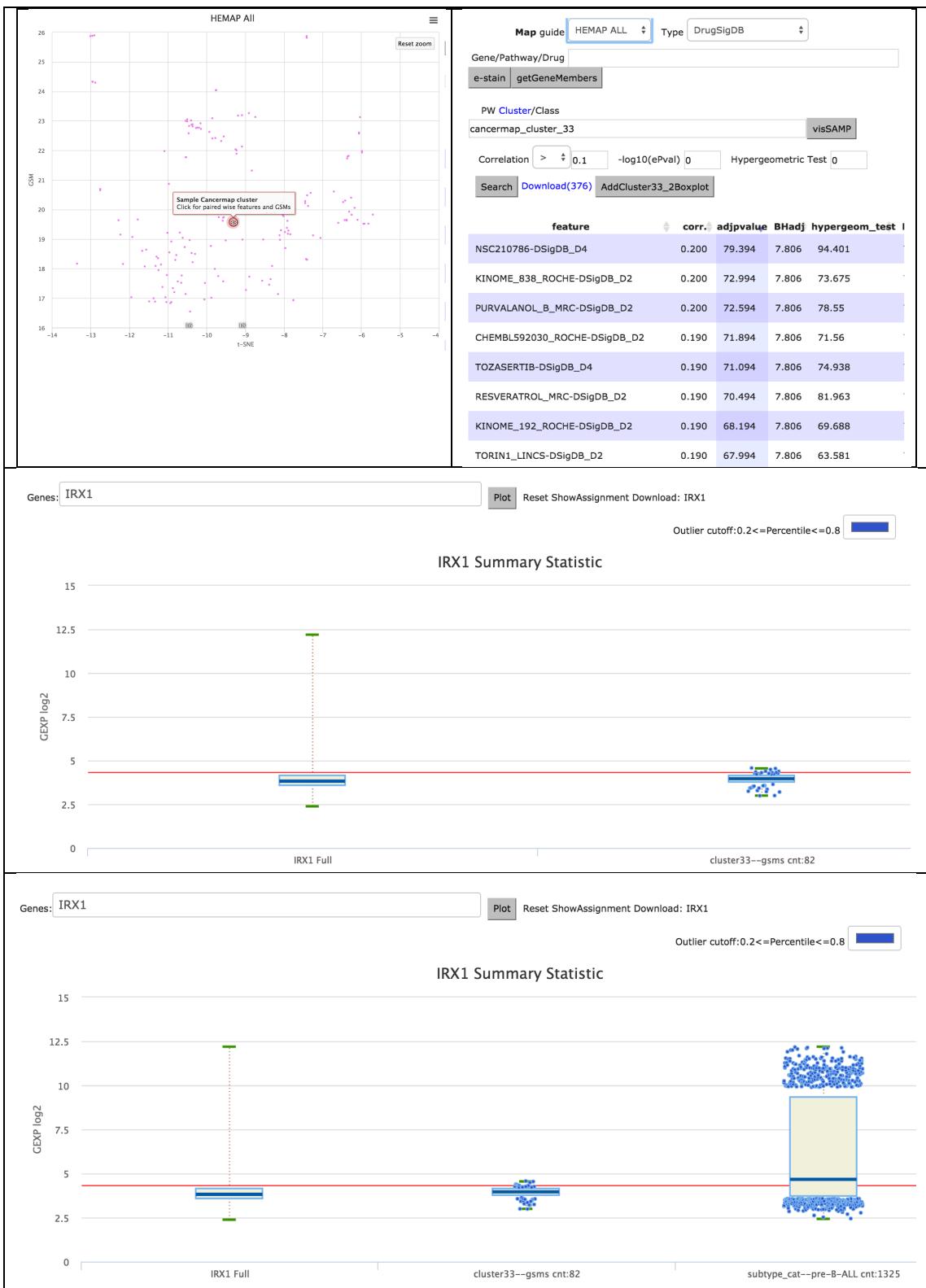
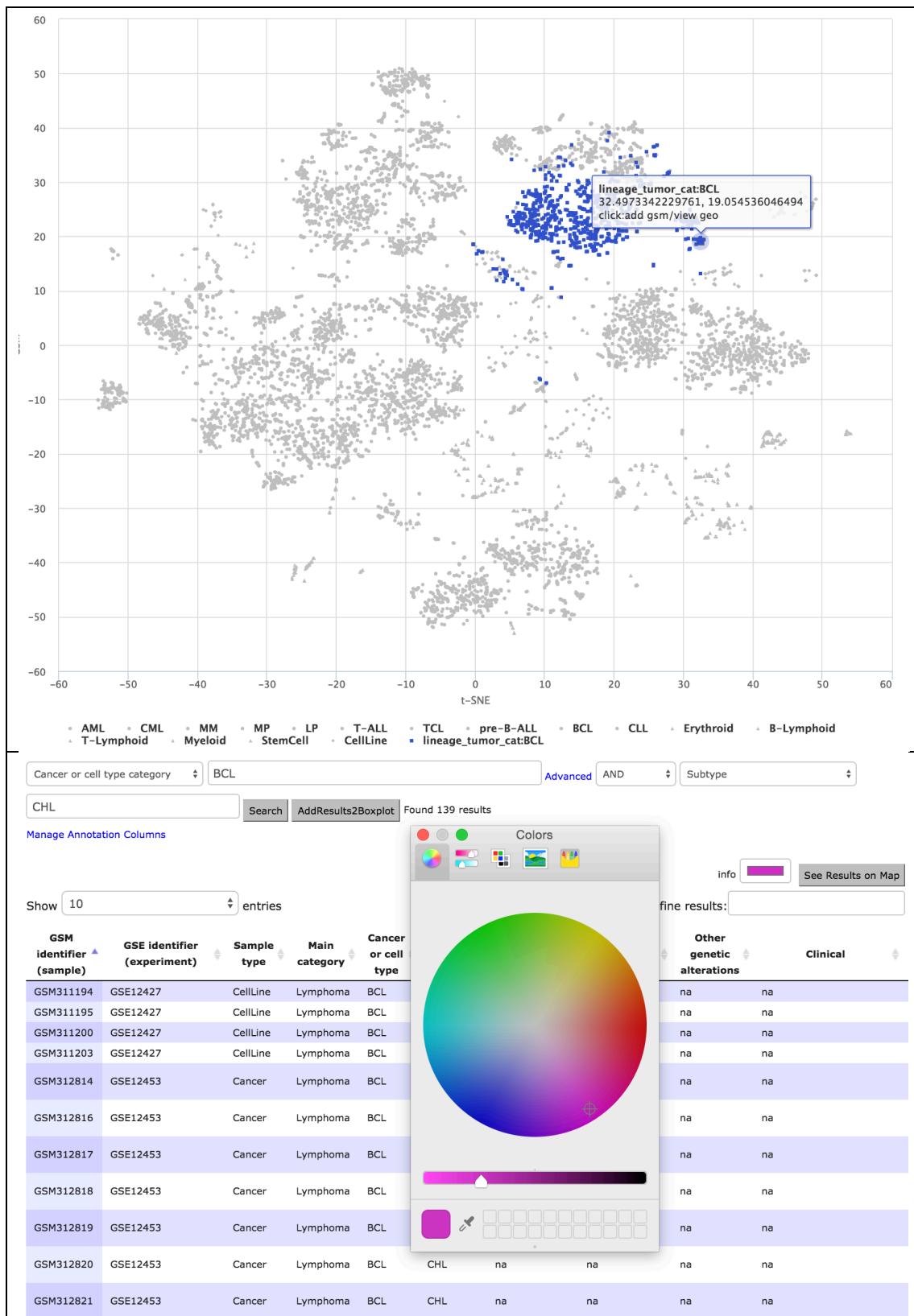


Figure 13. Labeled cluster 33 selection is triggers pairwise selection and the samples classified to cluster 33 can also be prepared for box plot. This sample groups is plotted to IRX1 and the bottom most figure demonstrates combining annotation search on subtype pre-B-ALL, contrasting with cluster selection and also revealing higher IRX1 expression on pre-B-ALL samples.

Basic use case example: Annotation categories to gene expression e-staining

In parallel with the manuscript's Supplement Fig. 5, this example illustrates the usage of combining filtering values, using BCL cancer/cell type across subtypes DLBCL (792 samples stained blue (Fig. 14 rows 1 and 2)) and CHL (139 samples stained purple, see row 3 usage of color picker (Fig. 14 rows 3 and 4)). Without the advanced AND subtype filtering, there are 1280 BCL labeled samples. Furthermore, subtype pre-B-ALL and MLL cytogenetic labeled samples (238) are stained pink in the same figure.

Annotations									
GEXP Boxplots									
Settings									
Info									
<input type="text" value="Cancer or cell type category"/> <input type="text" value="BCL"/> Advanced AND <input type="text" value="Subtype"/>									
<input type="text" value="DLBCL"/> <input type="button" value="Search"/> AddResults2Boxplot Found 792 results									
DLBCL <input type="button" value="columns"/> DLBCL-ABC DLBCL-GCB DLBCL-PMBL									
GSE Sample type Main category Cancer or cell type Subtype Further specification Cytogenetics Other genetic alterations *Clinical Sample source Sample Purity/Cancer cell content Patient Gender Race Age Phenotype SurvivalStatus SurvivalTime *Recurrence *RecurrenceTime *Stage *InvivoTreatment Cultured *ExvivoTreatment *ExvivoTreatmentTime *Additional notes *Alternate GSMS *Alternate GSEs *Cluster *hidden									
Info <input type="button" value="See Results on Map"/>									
Show 10 entries <input type="button" value="Refine results:"/>									
GSM identifier (sample)	GSE identifier (experiment)	Sample type	Main category	Cancer or cell type	Subtype	Further specification	Cytogenetics	Other genetic alterations	Sampl sour
GSM102466	GSE2109	Cancer	Lymphoma	BCL	DLBCL	na	na	na	lymph i
GSM152566	GSE2109	Cancer	Lymphoma	BCL	DLBCL	na	na	na	lymph i
GSM152639	GSE2109	Cancer	Lymphoma	BCL	DLBCL	na	na	na	ovary
GSM179927	GSE2109	Cancer	Lymphoma	BCL	DLBCL	na	na	na	lymph i
GSM188663	GSE7788	Cancer	Lymphoma	BCL	DLBCL	THRBC	na	na	lymph i
GSM188665	GSE7788	Cancer	Lymphoma	BCL	DLBCL	THRBC	na	na	lymph i
GSM188674	GSE7788	Cancer	Lymphoma	BCL	DLBCL	THRBC	na	na	lymph i



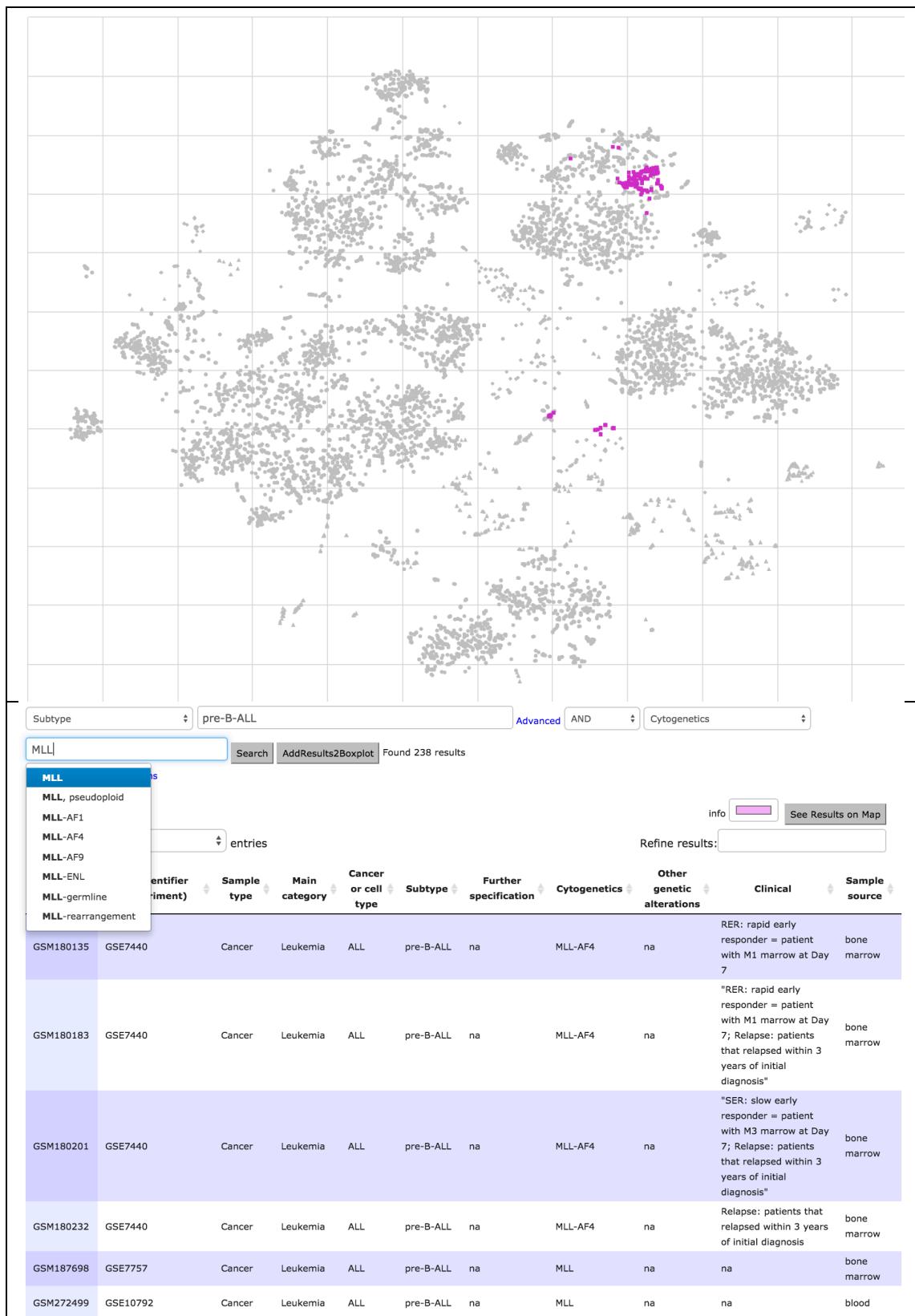
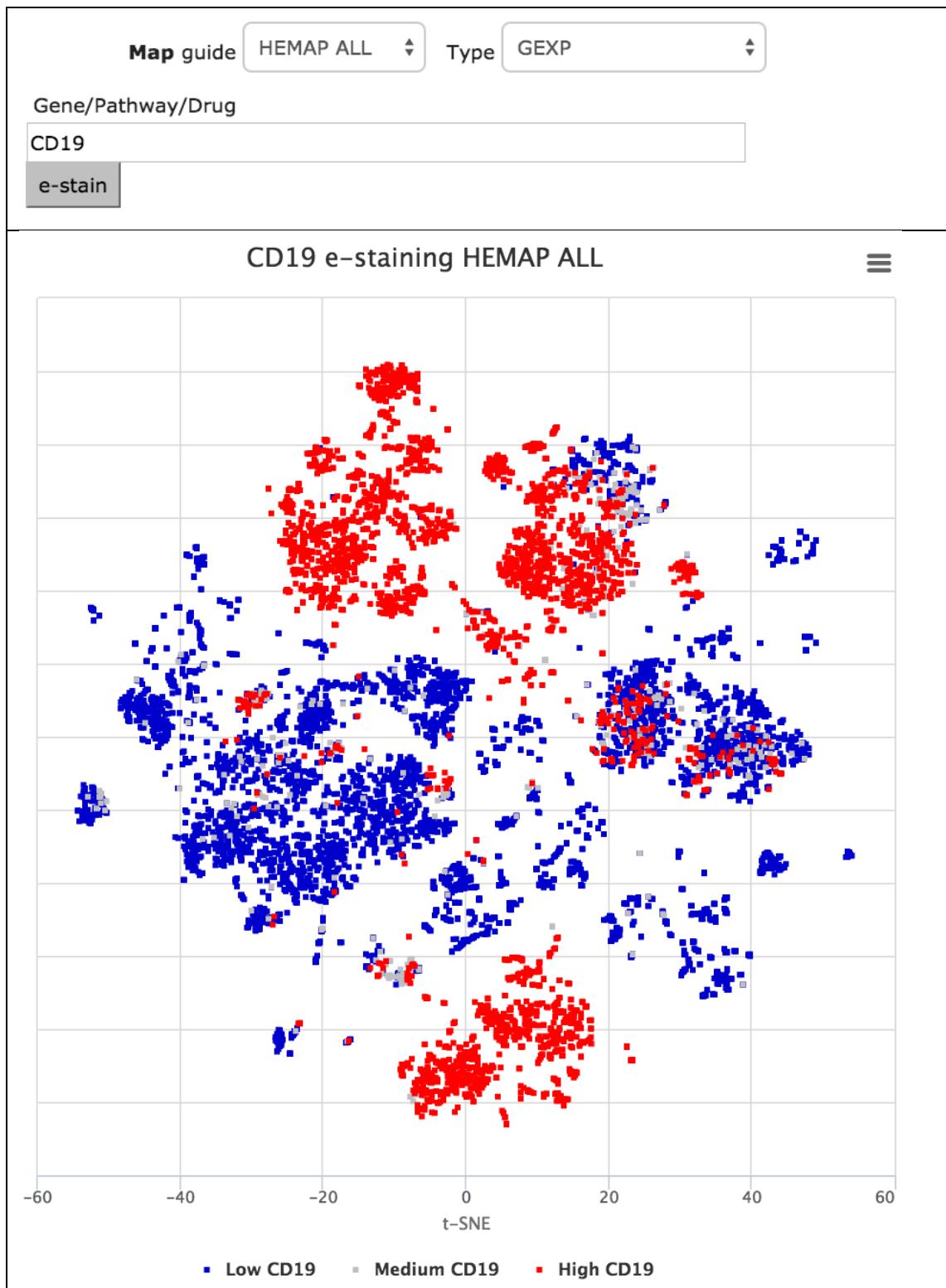




Figure 14. HEMAP samples are further classified and e-stain based on cancer type BCL subtypes DLBCL (1st 2nd rows) and CHL (middle 3rd,4th rows). The results are obtained via combining cancer/cell type with subtype. Filtering and its combinations are highly flexible, as demonstrated in rows 5 and 6 of 238 samples via using subtype pre-B-ALL and MLL cytogenetic event.

The Explore tab allows for easy gene and pathway e-staining as well as pairwise statistical data analysis, highlighted in the next Advanced use case section. Here in Fig. 15, antigen genes CD19 and MME (aka CD10) e-staining maps are plotted. The gene input, as well as pathway and drug signature fields, provides type ahead.



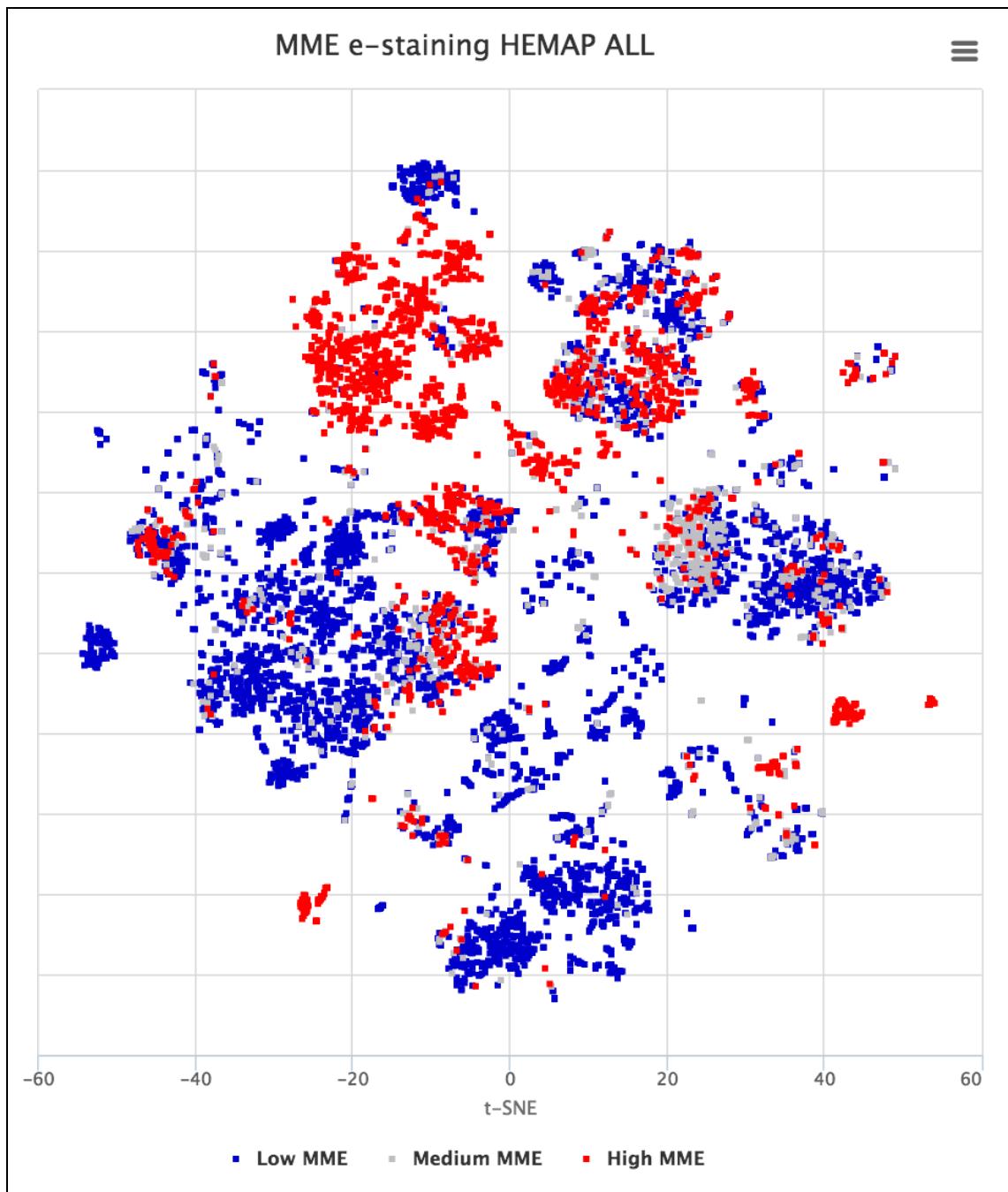


Figure 15 Gene based e-staining is accomplished for antigen coding CD19 and MME (aka CD10) genes. The maps are exportable. Low, medium and high colors can be customized using the settings, described in the last section of this guide.

Advanced use case example: *In silico* drug screening using Hemap

This example illustrates how the Hemap resource can be used for *in silico* drug screening. First, the pairwise results are used to find signaling pathways that could be targeted in pre-B-ALL and as a next step to find candidate drugs that specifically target this signaling pathway. Furthermore, as a third step, drug gene set details of drug chemical screening are used to evaluate drug target specificity. As a final step, gene e-staining and boxplot functions are used to compare gene expression in disease and normal cells to assess drug safety and potential side effects.

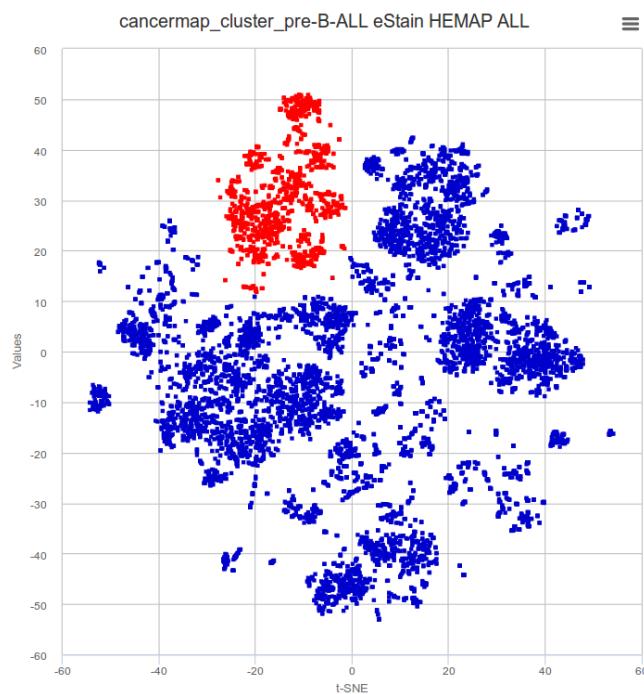


Figure 16. pre-B-ALL clusters stained using visSAMP.

Step 1: Identification of candidate pathways

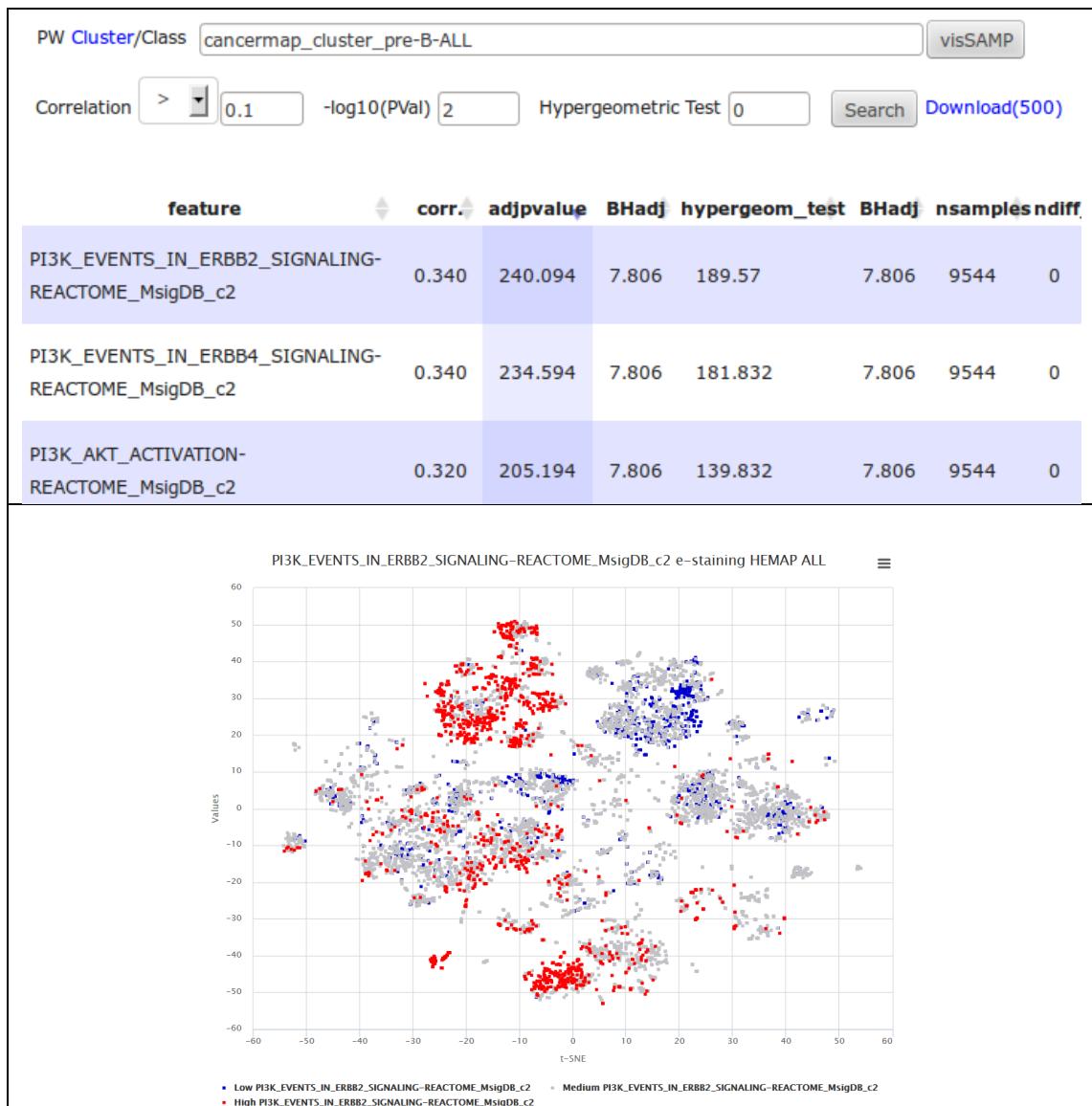
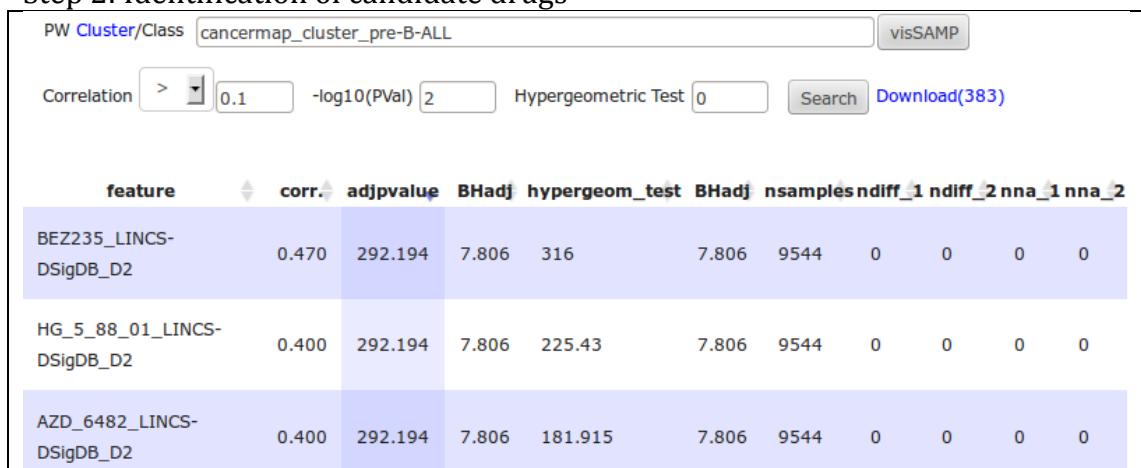


Figure 17. Searching for pre-B-ALL cluster correlated gene sets in pairwise results and filtering using term signaling. PI3K pathway is enriched in pre-B-ALL samples as shown in e-staining.

Step 2: Identification of candidate drugs



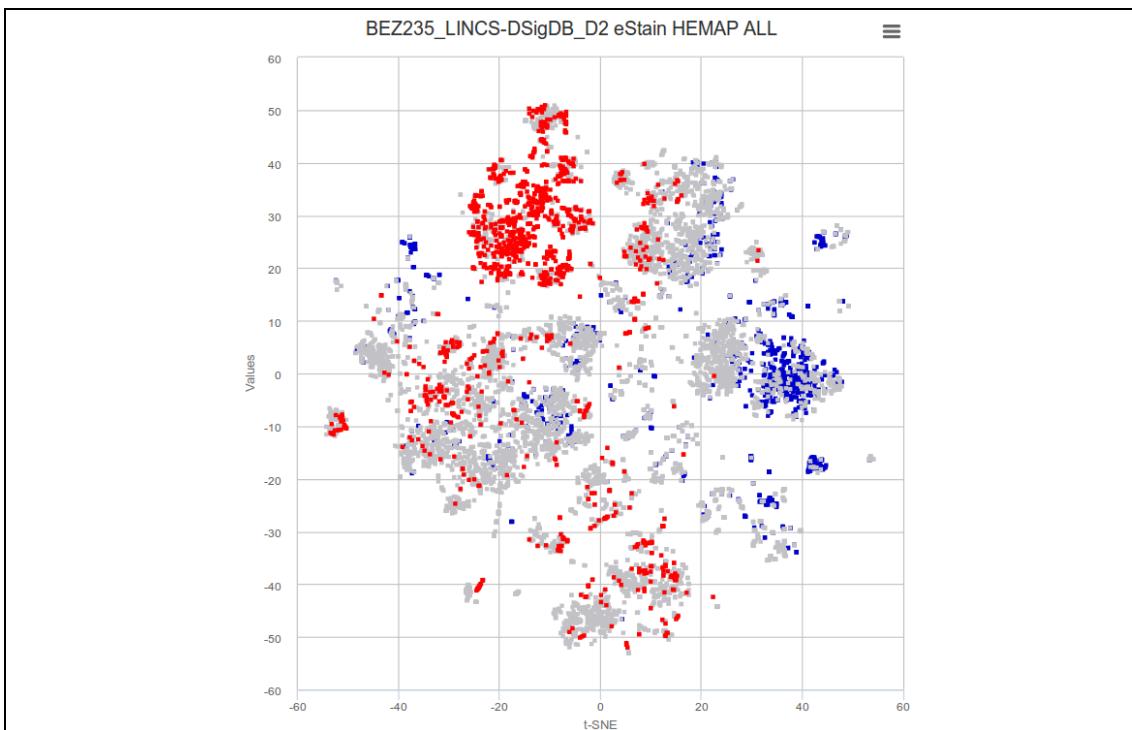


Figure 18. Searching for pre-B-ALL cluster correlated drug gene sets in pairwise results. Search is filtered to contain only LINCS chemical screen drugs. Two PI3K inhibitors, BEZ235 and AZD_6482, are among top correlated drug gene sets for pre-B-ALL. BEZ235 e-staining reveals high specificity for pre-B-ALL

Step 3: Examining the drug gene set and accessing drug target information

[BEZ235_LINCS-DSigDB_D2 info](#)

Source:[BEZ235_LINCS-DSigDB_D2](#)

Gene set:

MAP4K2,PIK3CA,PIK3CD,PIK3C2B,PIK3C2G,FLT3

Figure 19. BEZ235 gene set composition and link to drug details can be accessed by clicking [GetGeneMembers](#)

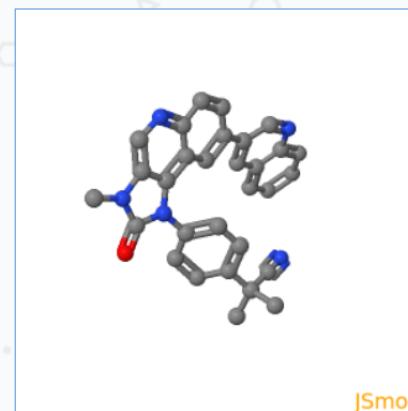
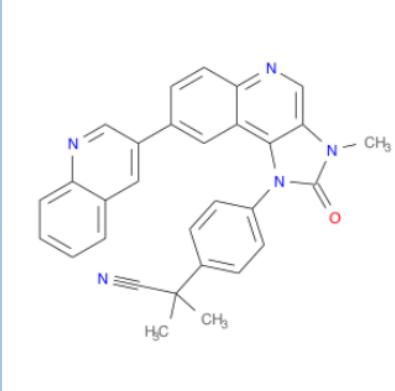
Gene Set: D2 : HMS LINCS - Nvp-Bez235

Collection D2 : HMS LINCS

Chemical Nvp-Bez235 (From Source : BEZ235)
Name

FDA	NPC	WHO	Indian	Australia	China	Traditional Herbal	Clinical Trail
Not	Not	Not	Not	Not	Not	Not	Not
Molecular Weight 469.537 g/mol	Hydrogen Bond Donor Count 0	Hydrogen Bond Acceptor Count 6	cLogP 5.89378	Lipinski Rule False(3/4)			

Structure



JSmol

InChI InChI=1S/C30H23N5O
/c1-30(2,18-31)22-9-11-23(12-10-22)35-28-24-15-19(21-14-20-6-4-5-7-25(20)32-16-21)8-13-26(24)33-17-27(28)34(3)29(35)36
/h4-17H,1-3H3

InChIKey JOGKUKXHTYWWRGZ-UHFFFAOYSA-N

Links

CAS Num : 915019-65-7

CAS Num : 1146702-52-4

Gene (14 / 14)	Value Type	Value↑	Concentration	Gene	PMID / Source
θ Less	POC	2.800	1uM	PIK3C2B	HMS LINCS
	POC	2.800	1uM	PIK3CA	HMS LINCS
	POC	2.800	1uM	PIK3CA(E542K)	HMS LINCS
	POC	3.000	1uM	PIK3CA(E545K)	HMS LINCS
	POC	3.800	1uM	PIK3CA(Q546K)	HMS LINCS
	POC	4.400	1uM	PIK3C2G	HMS LINCS
	POC	4.600	1uM	PIK3CA(C420R)	HMS LINCS
	POC	4.800	1uM	PIK3CA(H1047L)	HMS LINCS
	POC	5.100	1uM	PIK3CA(E545A)	HMS LINCS
	POC	6.200	1uM	PIK3CA(H1047Y)	HMS LINCS
	POC	7.200	1uM	PIK3CD	HMS LINCS
	POC	11.000	1uM	MAP4K2	HMS LINCS
	POC	11.000	1uM	PIK3CA(M1043I)	HMS LINCS
	POC	12.000	1uM	FLT3(D835Y)	HMS LINCS

Download gmt, text, [Detailed text](#)
gene sets

Figure 20. Drug target details for BEZ235 reveal that PIK3C2B and PIK3CA have the highest specificity for BEZ235.

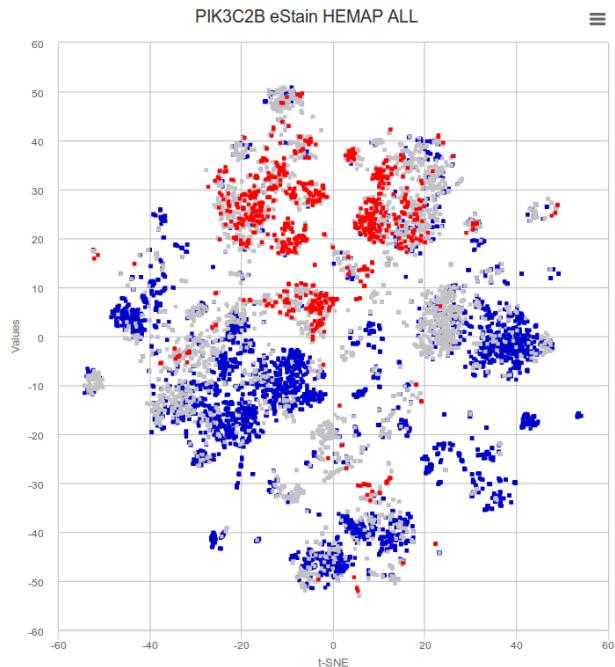


Figure 21. e-staining PIK3C2B shows that high expression of this drug target is prevalent in pre-B-ALL. BEZ235 is therefore an interesting drug candidate for pre-B-ALL treatment.

Final step: Comparing the expression in cancer to normal cell types using boxplots

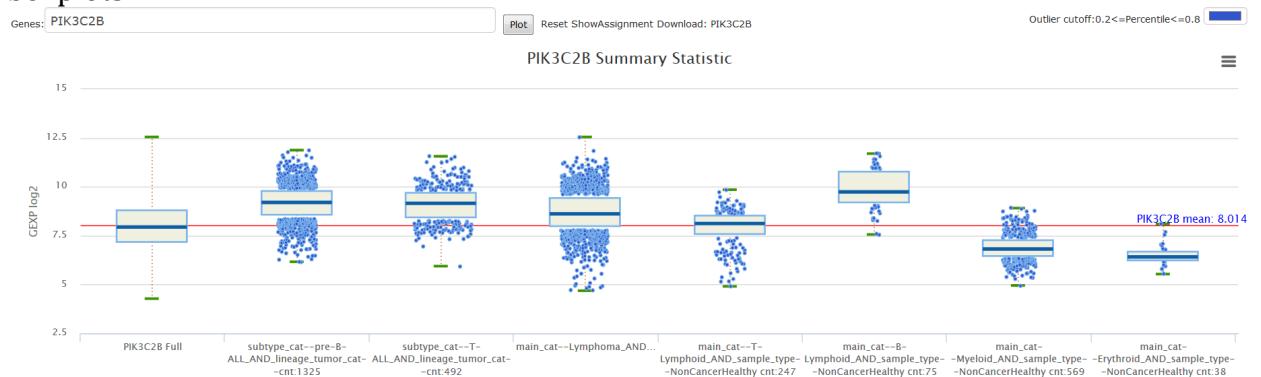


Figure 22. Sample sets of interest (pre-B-ALL, T-ALL and lymphoma vs T-Lymphoid, B-lymphoid, Erythroid and Myeloid) were defined using the Annotation table and selected for box plotting. The result shows that PIK3C2B is highly expressed in pre-B-ALL and T-ALL but is also highly expressed in normal B-lymphoid cells, which could indicate potential side effects.

Settings

Users can specify custom plotting and database search settings. It should be noted that plotting and search performance have large dependencies on network and browser RAM memory capacity. Brief descriptions are listed below.

The screenshot shows a user interface for setting parameters. At the top, there's a dropdown for 'Max GSM mouse selection' set to 500, a dropdown for 'Symbol size' set to 'Default', and a dropdown for 'Outlier Percentiles' set to 0.2. Below these are two dropdowns for 'Pairwise rows limit' and 'shown on page' both set to 500, and a dropdown for 'Map-PWTable WidthRatio' set to 60:40. There's also a dropdown for 'ePval' set to .05 and one for 'Order by' set to 'pval'. Underneath these are three color bars labeled 'Low' (blue), 'Medium' (grey), and 'High' (red). At the bottom right are 'Save' and 'Cancel' buttons.

Figure 23. Hemap allows for session customization of max number of selections, rows returned and colors. These settings are stored at the session and will revert to defaults on closing of the browse.

The default settings can be individually updated using the Settings interface.

Max mouse sample map selections – default 500

Map Symbol size, can increase or decrease using drop down.

Outlier Percentiles for Boxplots <20% and >80%

Max number of pairwise rows – default 5000, values 500 to 50,000

Rows shown per page – default 15

Map:Table Explore Screen Ratio – default 60:40 (flip function in Explore interface)

ePval default .05, empirical Pvalue used for pairwise cutoff – drop down

Custom e-staining colors

Info

This section provides contact information and project issue tracker. Hemap is open sourced and free for non-profit usage. The software is use as it is and does not offer any warranty or guarantee. Please contact matti.nykter@uta.fi and jake.lin@uta.fi for commercial usage permissions.