

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

## Table of Contents

<b>HEMAP Overview .....</b>	<b>1</b>
Home page .....	2
<b>Explore:.....</b>	<b>3</b>
e-staining the gene expression state with <i>FLT3</i> .....	4
e-staining Pathway ROSS_AML_WITH_PML_RARA_FUSION.....	5
e-staining acute precursor B cell vs. chronic lymphocytic leukemia clusters.....	7
<b>Accessing pairwise analysis results.....</b>	<b>8</b>
Search: <i>cancermap cluster CLL and GEXP type</i> .....	9
Search: <i>cancermap cluster aml and DrugSigDB type with GSK subfiltering</i> .....	11
Gene sets of pathway and drug signatures.....	12
<b>TCGA AML Map .....</b>	<b>13</b>
Additional data types for TCGA AML map: GNAB, CNVR, METH, MIRN and CLIN categories.....	14
e-staining <i>mir-125 across METH and MIRNA</i> .....	14
Lookup of copy number aberrations by gene name, vice versa.....	15
<b>Annotation Search, e-staining, and box plotting.....</b>	<b>16</b>
Main Category: <i>Leukemia AND Cytogenetics: crlf2 analysis to IRX3</i> .....	16
Sample groups from map cluster selection (mouse) and plotting combined with annotation search .....	18
<b>Basic use case example: Annotation categories to gene expression e-staining .....</b>	<b>19</b>
<b>Advanced use case example: <i>In silico</i> drug screening using Hemap .....</b>	<b>25</b>
<b>Settings.....</b>	<b>31</b>
<b>Info.....</b>	<b>31</b>
<b>Comparing an independent dataset with Hemap samples using cancer maps.....</b>	<b>31</b>

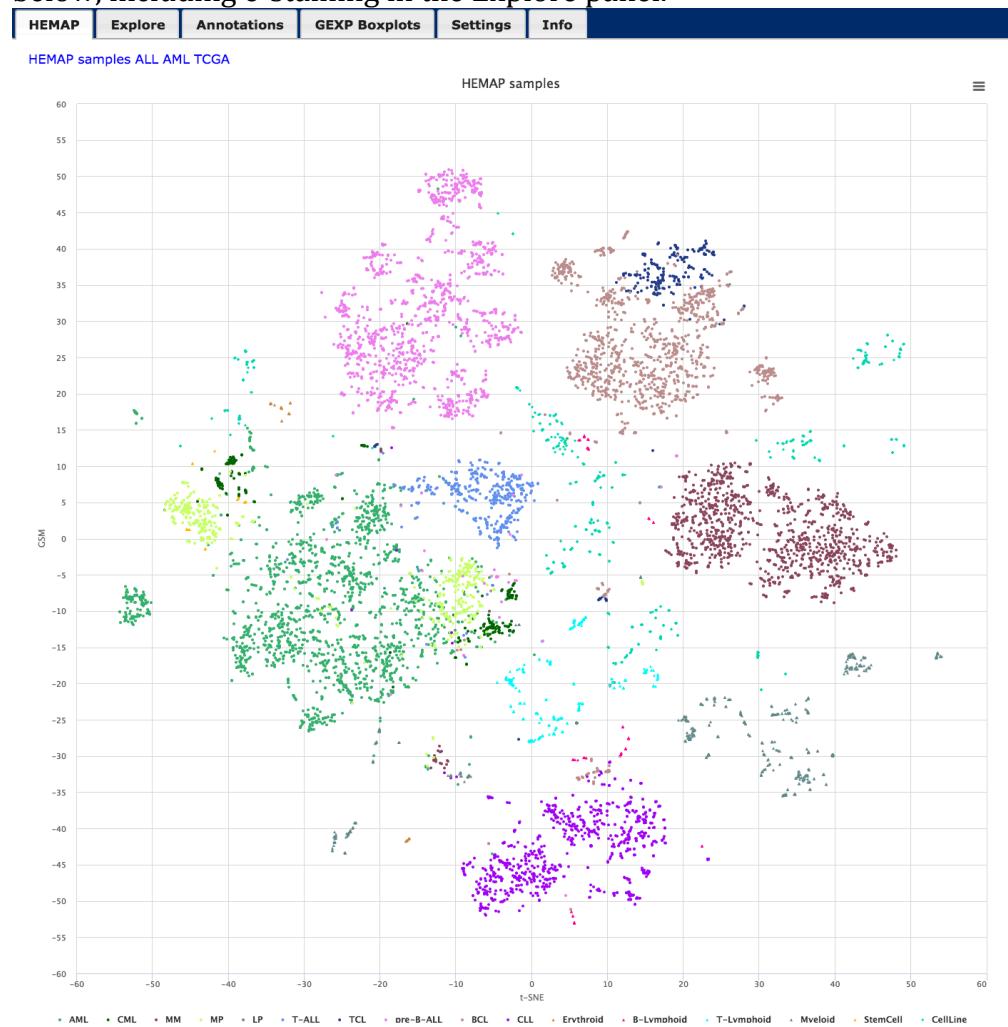
## HEMAP Overview

Hemap (<http://compbio.uta.fi/hemap>) is the interactive online resource component of the project *Hemap: An Interactive Resource for the Molecular Classification and Comparative Analysis of Cancers of Hematopoietic Origin*. Please contact Merja Heinaniemi ([merja.heinaniemi@uef.fi](mailto:merja.heinaniemi@uef.fi)) and Matti Nykter ([matti.nykter@uta.fi](mailto: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 home page shows the main maps (HEMAP samples, ALL, AML and TCGA). The can be selected and toggle by clicking on the provided links on the upper left corner. Map session element sizes can be changed on the Settings pane. Every map supports zooming and direct viewing of annotations, custom group box plots are enabled using mouse drags. Details are provided below, including e-staining in the Explore panel.



**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 samples, 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.

Resource map: HEMAP samples

Type: ✓ Gene Expression

Gene/Pathway/Drug

PW Cluster/Class

Corr. > 0.1

-log10(ePval) 2

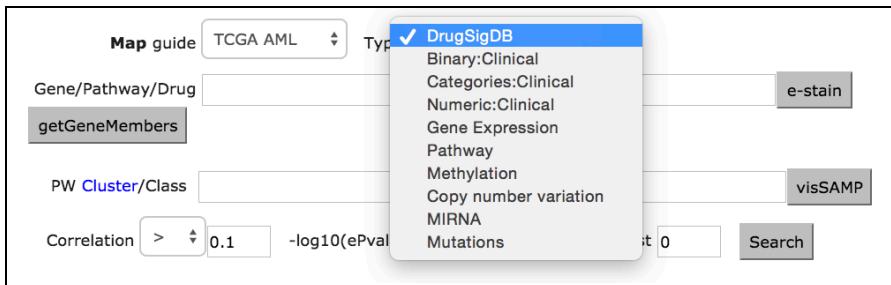
Hypergeometric Test 0

e-stain

viSamp

search

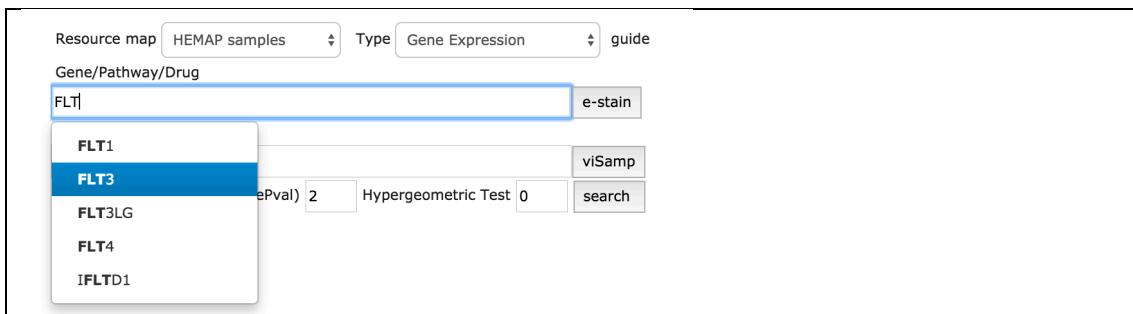
DrugSigDB  
Clinical  
Gene Expression  
Pathway  
guide

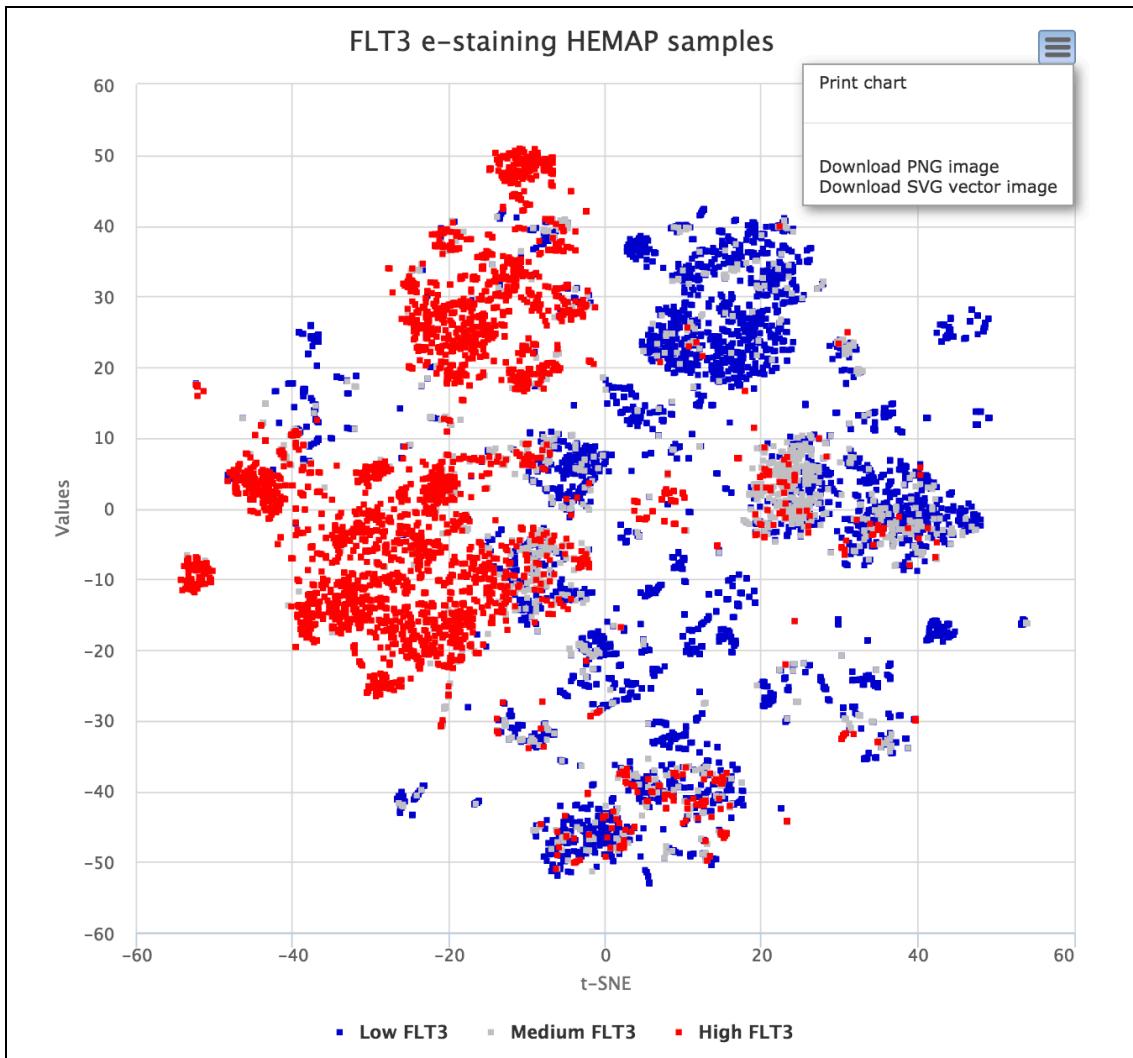


**Figure 2. The explore interface provides intuitive search of pairwise cluster to feature associations and then e-staining. The omic types are DrugSigDB, Clinical, GEXP and GSVA. The bottom figure shows additional TCGA cluster/class types, including copy number and methylation.**

### e-staining the gene expression state with *FLT3*

Auto-suggestions (type aheads) are built into Gene, pathways, and cluster/class fields to ensure valid selections. The example here shows how to visualize *FLT3* e-staining and as well as image export functionality.





**Figure 3.** e-staining is performed using FLT3, where low expressing 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 e-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.

Map guide AML Type Pathway

Gene/Pathway/Drug pml\_r

getGeneMembers

PW Cluster/Class

Correlation > 0.001 -log10(ePval) 2 Hypergeometric Test 0 Search

e-stain

ROSS\_AML\_WITH\_PML\_RARA\_FUSION-MsigDB\_c2

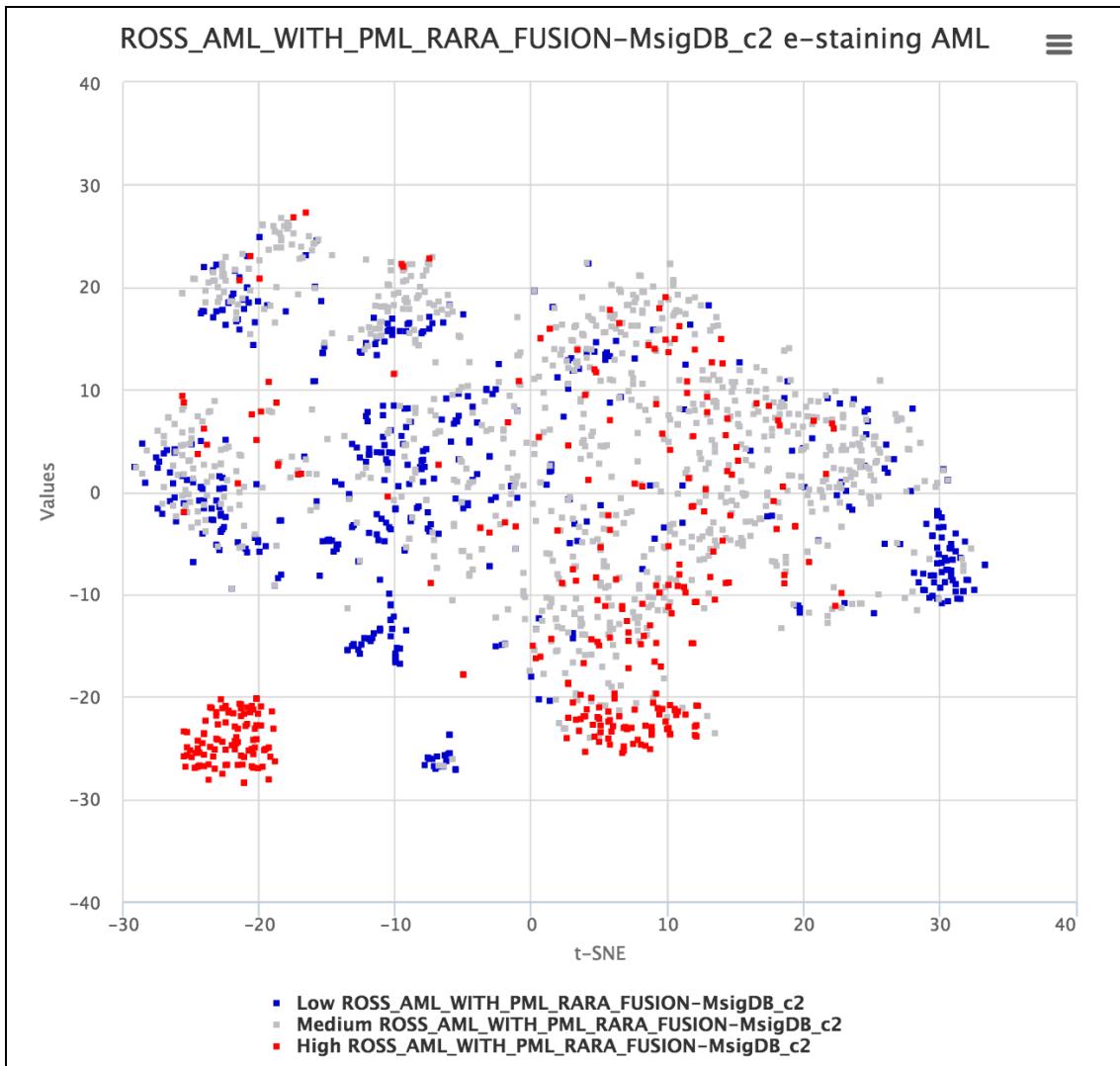
MARTENS\_BOUND\_BY\_PML\_RARA\_FUSION-MsigDB\_c2

ROSS\_AML\_WITH\_PML\_RARA\_FUSION-MsigDB\_c2 info

Source: ROSS\_AML\_WITH\_PML\_RARA\_FUSION-MsigDB\_c2

Gene set:

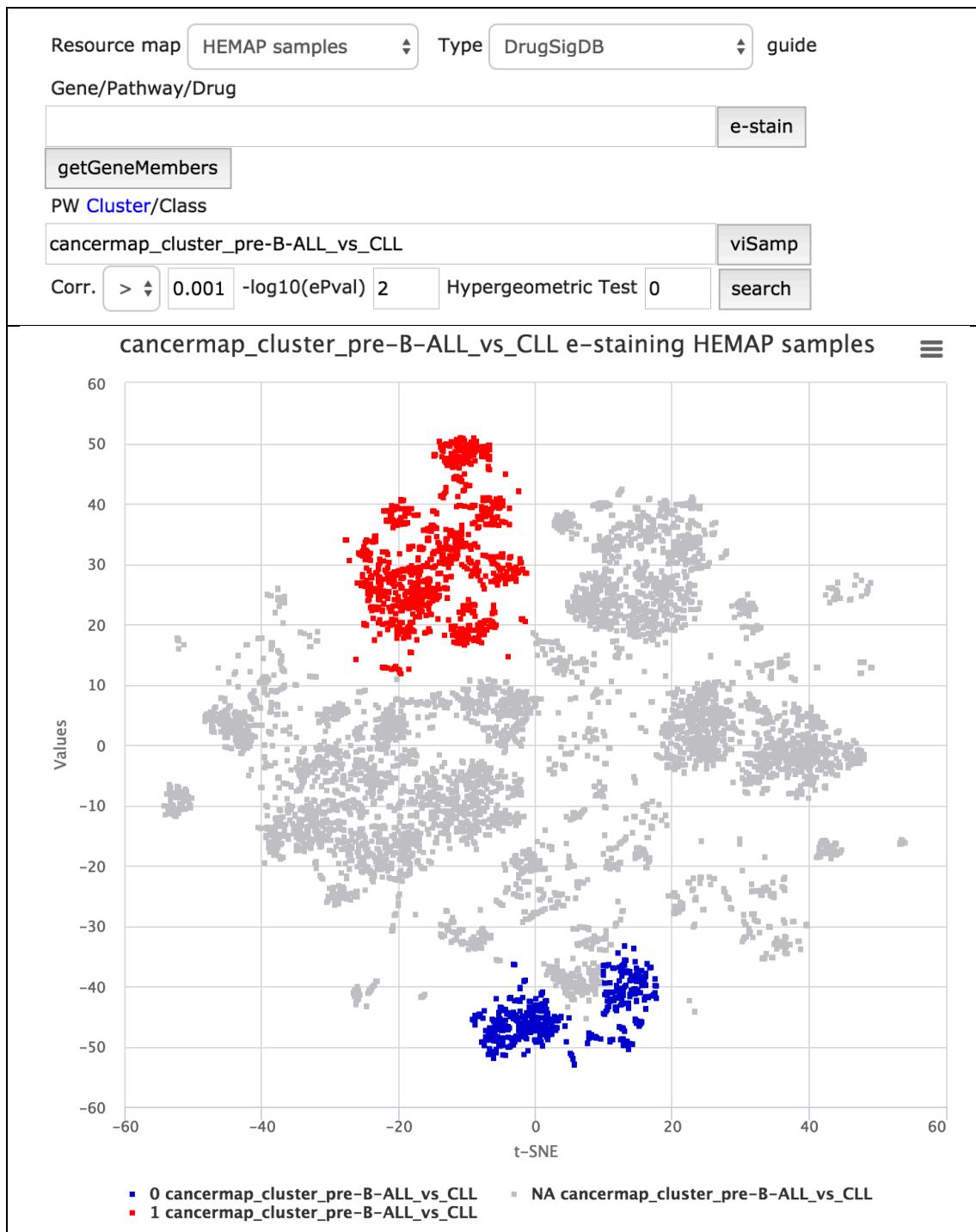
```
PDE3B,KRT18,LRFN4,HGF,SEC31A,CDKN1C,ARHGA  
P4,SERPING1,AUTS2,ALOX5,SMARCD3,ARFGEF1,AG  
RN,PRODH,TMEM87A,MEGF6,FNDC3B,CHN2,STXBP  
1,CERS4,ST3GAL6,NAALADL1,MST1,ITPR2,GRB10,JA  
G1,HYOU1,KIAA0195,PXDNL,ALCAM,CMAHP,RASL12,  
STAB1,ATG13,CALR,MANF,MOSC2,MPO,ANKFY1,GC  
NT1,MAP1A,CFD,GALNS,PRR14,RCN1,SLC25A38,PT  
GDS,MMP2,CST7,C18orf1,NISCH,IGFBP2,CTSW,MRC  
2,FGF13,CPA3,CLMN,ASNS,MAP4,AFF2,MST1P9,MX  
RA7,CHPF2,P4HB,COL2A1,PGBD5,SLC1A4,HSPA5,T
```



**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 2<sup>nd</sup> (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:

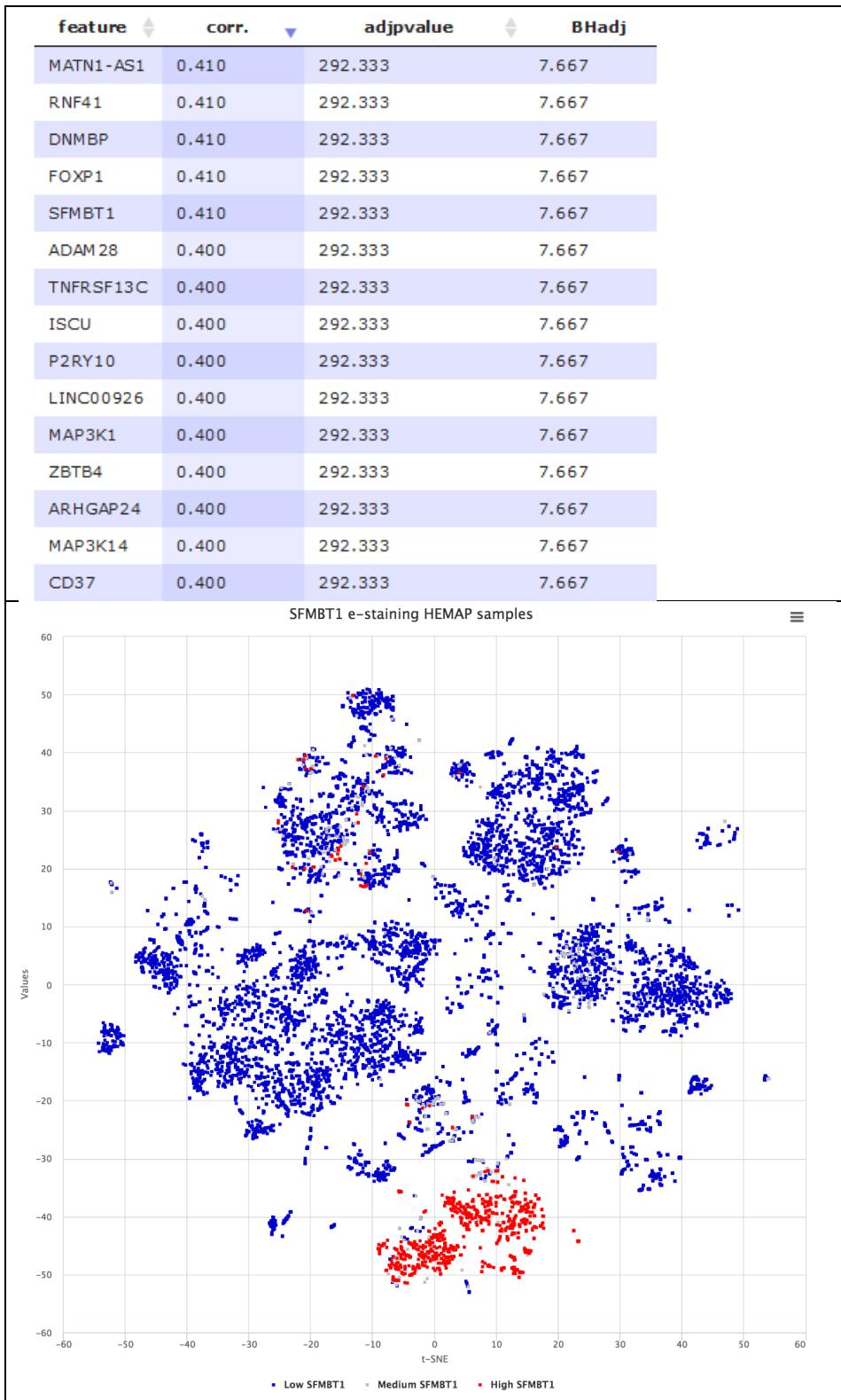
1. Feature (e.g. a Gene or a pathway gene set)
2. Spearman correlation
3. Adjusted  $-\log_{10}(P\text{-value})$  from correlation test
4. BH factor adjustment for correlation test (in multiple testing correction)
5. Adjusted  $-\log_{10}(P\text{-value})$  from cluster enrichment test (hypergeometric test), applicable for GSVA/Drug Signature
6. BH factor adjustment for hypergeometric test
7. Number of samples
8. Diff Samples A: number of samples in A
9. Diff Samples B: number of samples in 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 clicking search where resultant rows (4917) ordered by adjusted P-value are returned and can be downloaded. 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 can be directly stained from results by gene name selection on the row; figure 6 bottom panel is map stained by selection of MATN1-ASN1, the 1<sup>st</sup> record ordered by P-value.

The screenshot shows a search interface with the following fields and controls:

- Resource map: HEMAP samples
- Type: Gene Expression
- guide
- Gene/Pathway/Drug: (empty input field)
- e-stain button
- PW Cluster/Class:
  - cancermap\_cluster\_cl
  - cancermap\_cluster\_CLL (highlighted in blue)
- viSamp button
- Hypergeometric Test: 0
- search button



**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.

Resource map HEMAP samples Type DrugSigDB guide							
Gene/Pathway/Drug							
e-stain							
getGeneMembers							
PW Cluster/Class							
annotated_class_AML							
viSamp							
Corr.	> 0.1	-log10(ePval)	2	Hypergeometric Test	0	search	
<a href="#">Download(1990)</a>							
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)							
<a href="#">Previous</a> <a href="#">1</a> <a href="#">2</a> <a href="#">3</a> <a href="#">Next</a>							
Filter results							
<input type="text" value="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.

Resource map HEMAP samples Type DrugSigDB guide

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:

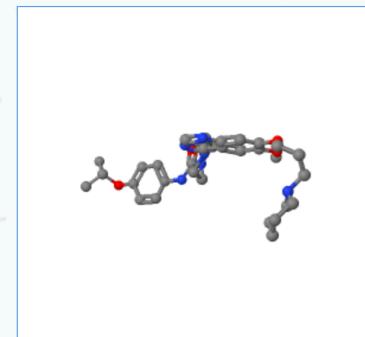
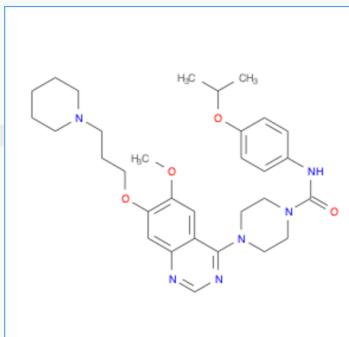
PDGFRB,PDGFRA,MAP4K5,DDR2,CSF1R,EGFR,NTRK1,NTRK2,KIT,CLK1,IRAK3,FLT3

### Gene Set: D2 : Kinome Scan - Tandutinib

Collection	D2 : Kinome Scan
Chemical Name	Tandutinib (From Source : MLN-518)
FDA	Not
NPC	Not
WHO	Not
Indian	Not
Australia	Not
China	Not
Traditional Herbal	Not
Clinical Trail	Not

Molecular Weight: 562.703 g/mol | Hydrogen Bond Donor Count: 1 | Hydrogen Bond Acceptor Count: 10 | cLogP: 5.0483 | Lipinski Rule: False(2/4)

Structure

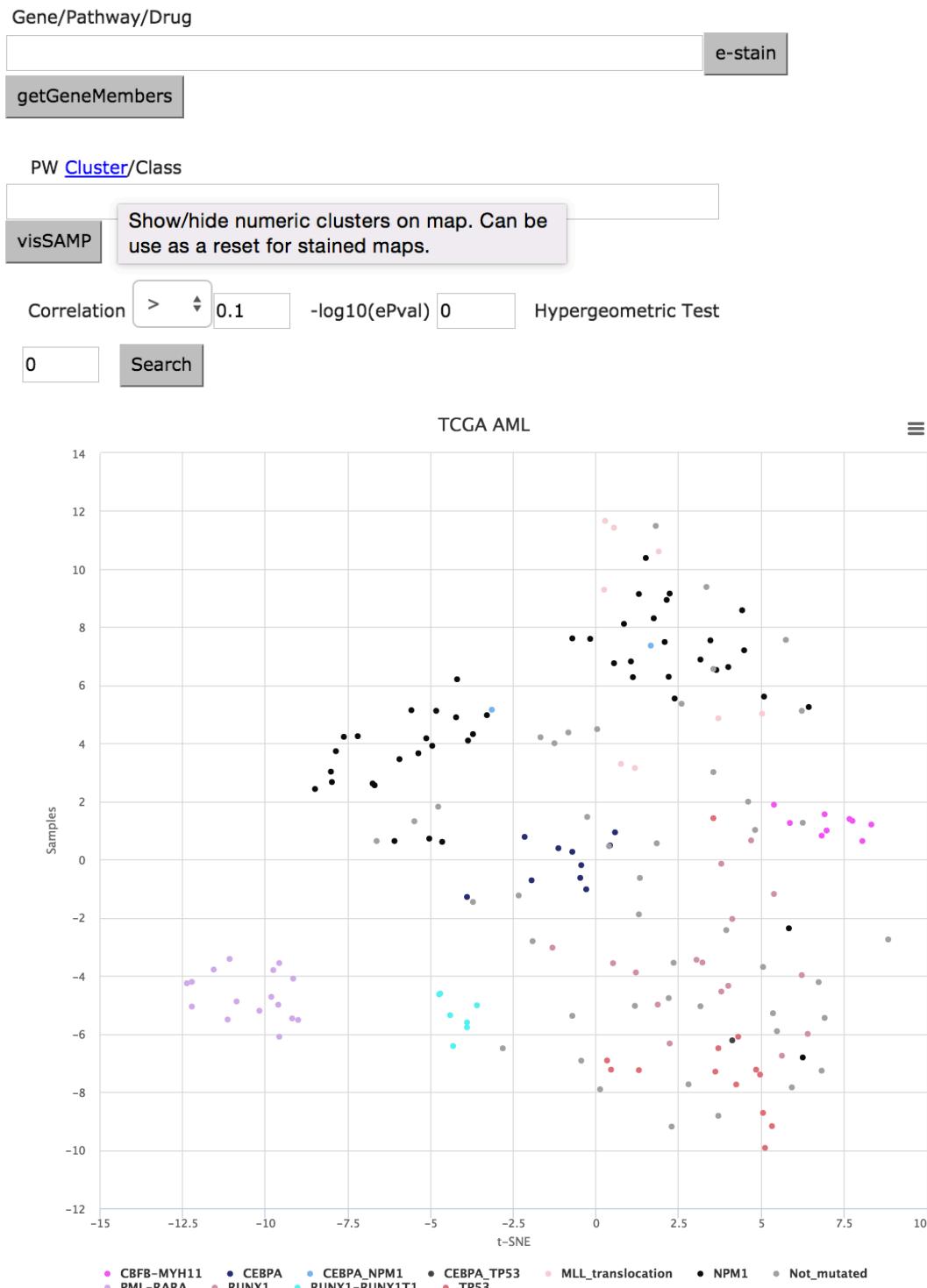


InChI: InChI=1S/C31H42N6O4/c1-23(2)41-25-10-8-24(9-11-25)34-31(38)37-17-15-36(16-18-37)30-26-20-28(39-3)29(21-27(26)32-22-33-30)40-19-7-14-35-12-5-4-6-13-35/h8-11,20-23H,4-7,12-19H2,1-3H3,(H,34,38)

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

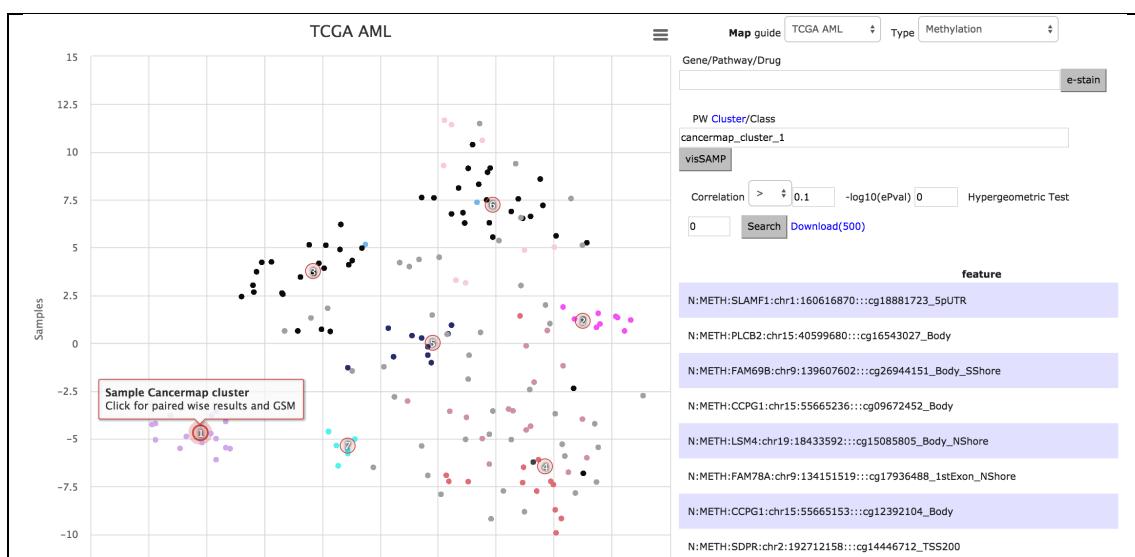
The genome-wide mutation, methylation and copy number profiles are available for the TCGA samples with the feature names GNAB, METH and CNVR. Pairwise exploration of cluster association with these features works the same as described above. Below we will look at a few of the data types and the easy manner 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 2<sup>nd</sup> 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

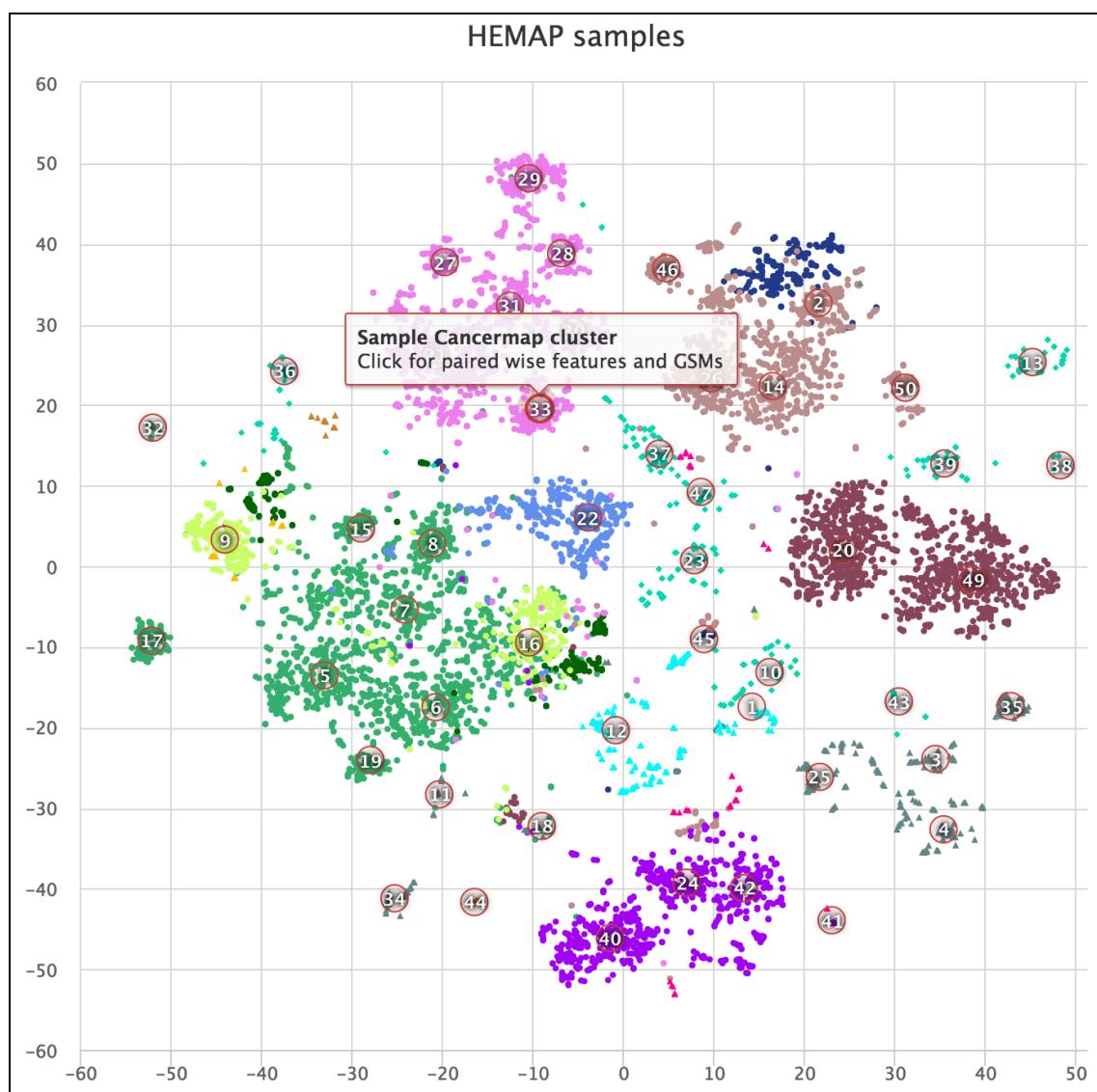
plots; more custom groups can be added across user selected gene(s); in this case *IRX3*.

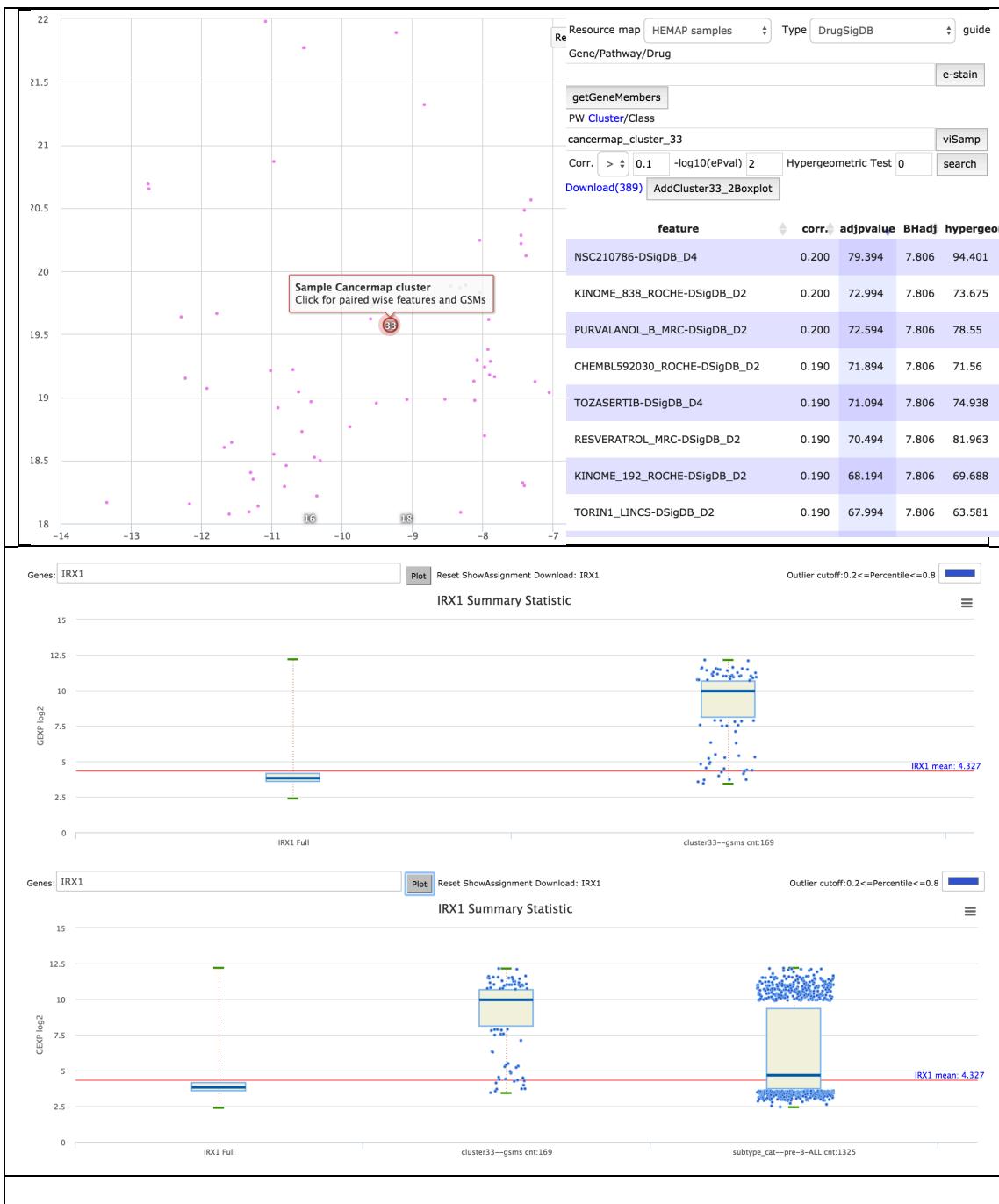


**Figure 12. Annotation search allows for intuitive basic search but also provides powerful advanced functions. Almost 5000 records are labeled Leukemia and shown here is Leukemia samples annotated also with Cytogenetics *crlf2*. The 26 results are then visualized on the map and then gene *FLT3*'s expression 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 (169 samples) and then the box plot result 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 2<sup>nd</sup> 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 169 samples (from prompt after button) is shown in 3<sup>rd</sup> 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

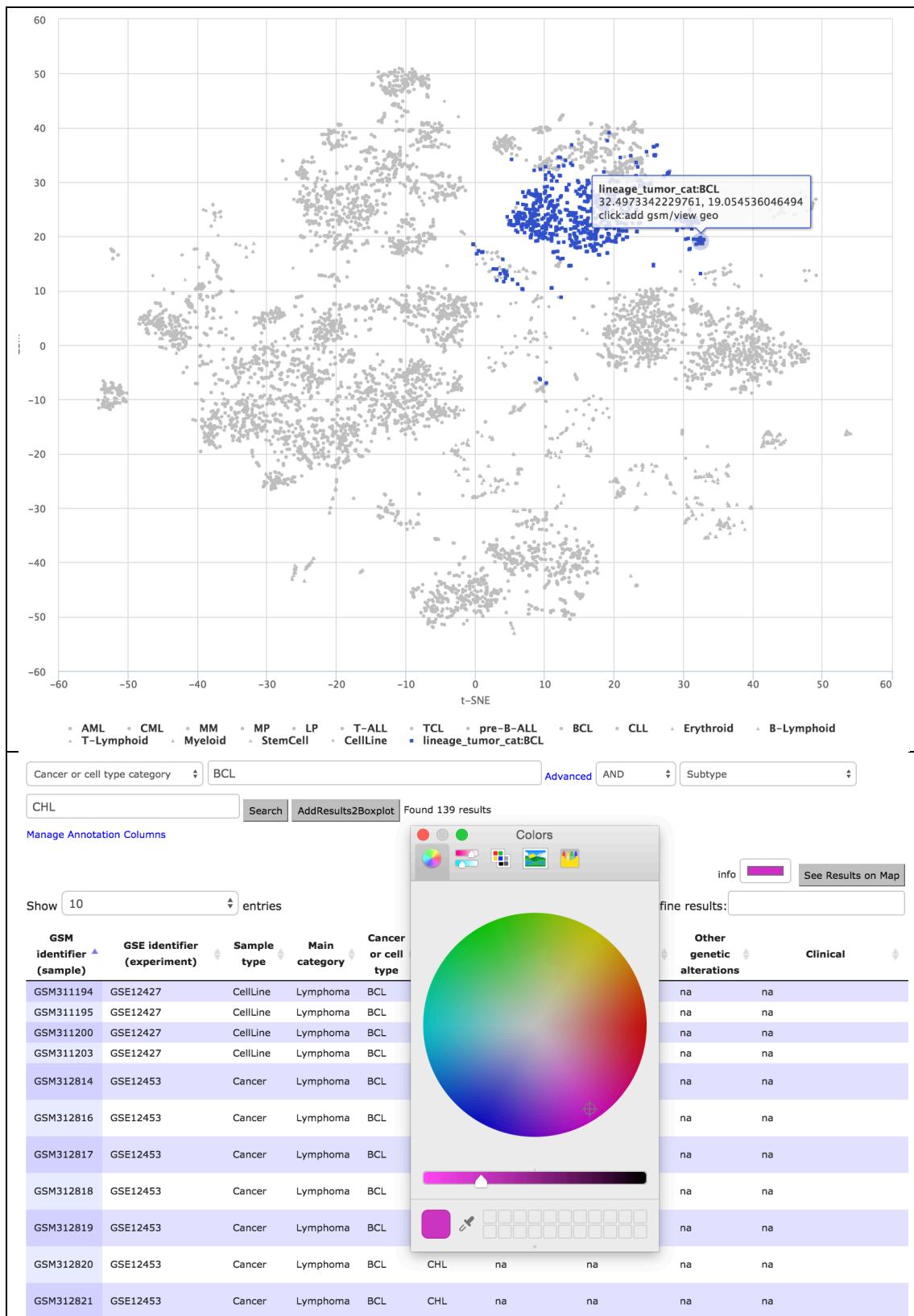
In parallel with Fig. S5 presented in the manuscript, this example illustrates the usage of combining filtering values to locate samples from three cancer types.

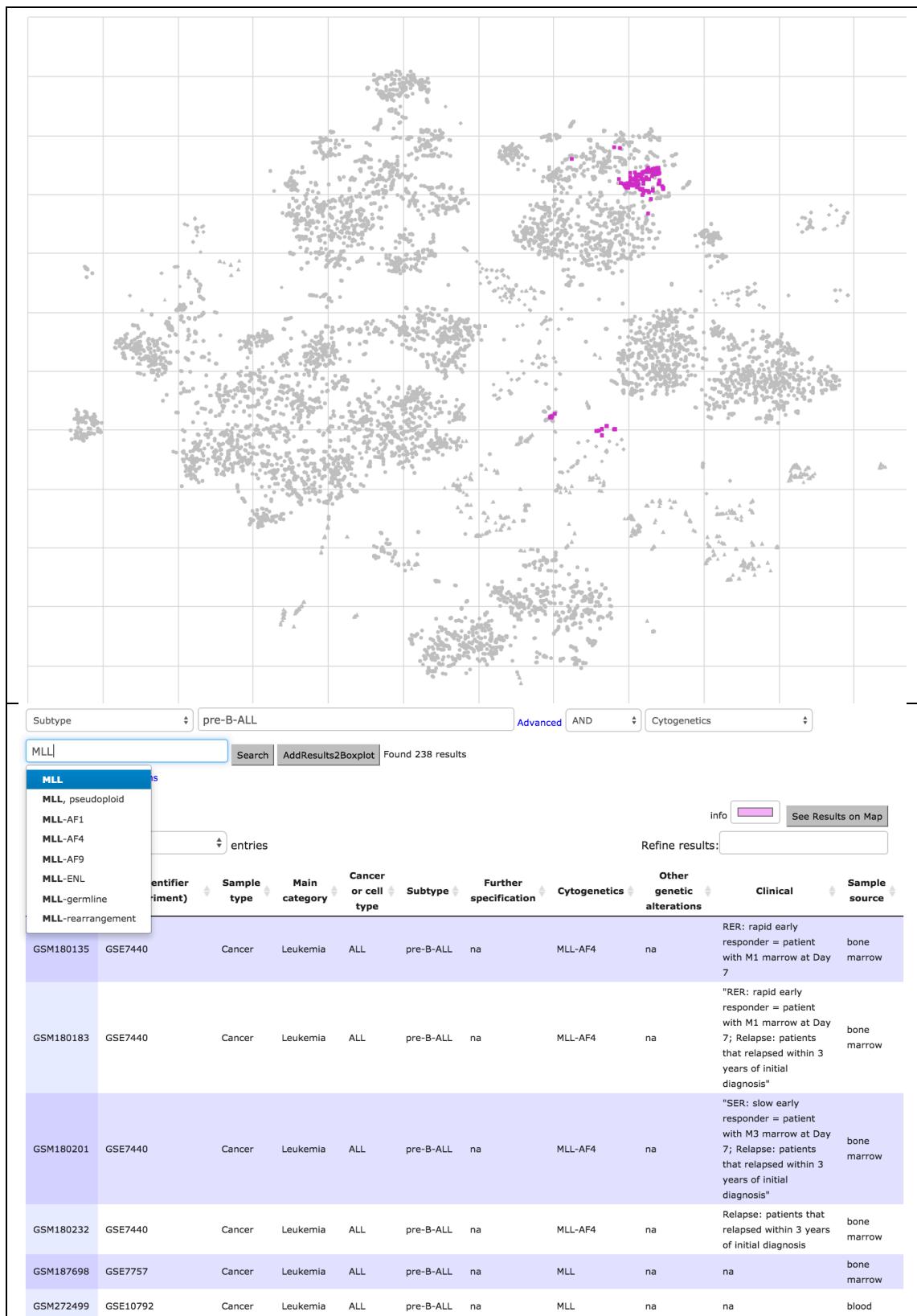
The first filter is BCL cancer/cell type and next we select across subtypes DLBCL (792 samples stained blue (Fig. 14 rows 1 and 2)) or CHL (139 samples stained purple, see row 3 usage of color picker (Fig. 14 rows 3 and 4)). In a similar manner, more BCL types can be stained on the map (not shown). Without the advanced AND subtype filtering, there are 1280 BCL labeled samples. Furthermore, the subtype selection of pre-B-ALL with MLL cytogenetic labeled samples (238) are stained pink in the same figure.

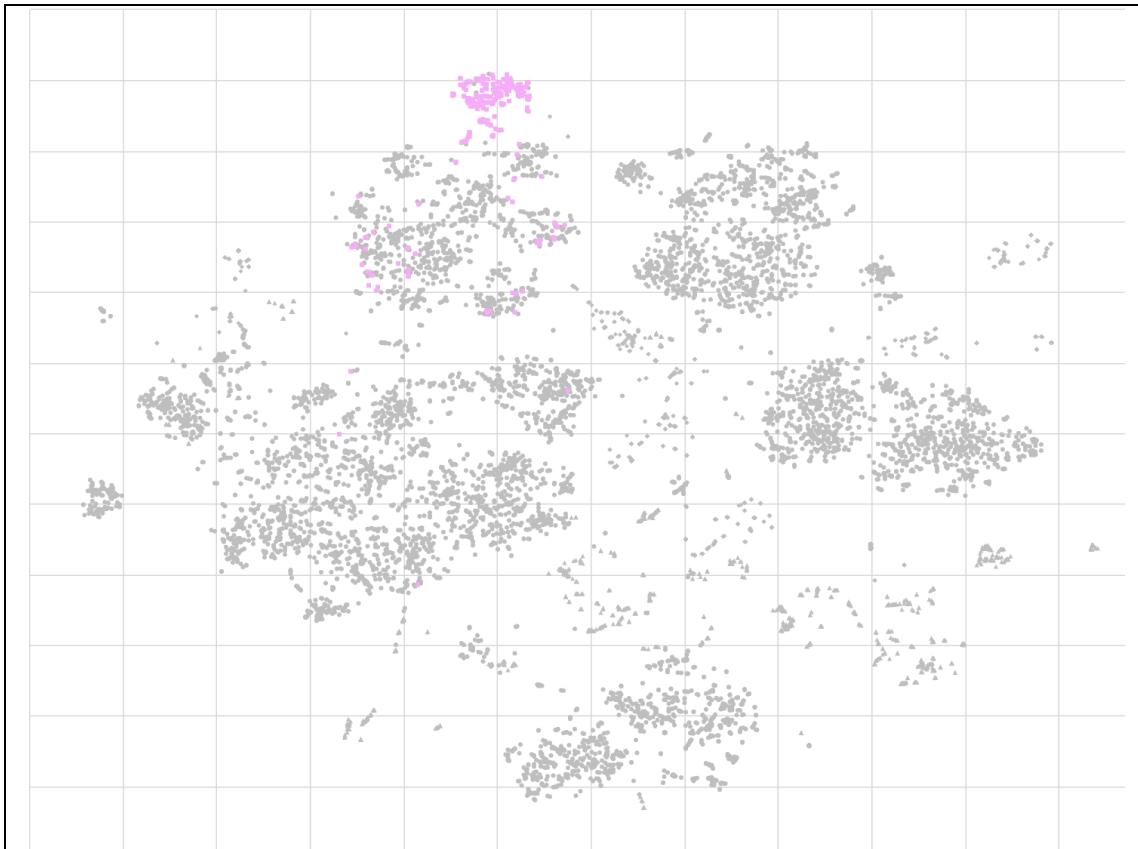
The screenshot shows the HEMAP software interface with the following details:

- Header:** HEMAP, Explore, Annotations, GEXP Boxplots, Settings, Info.
- Search Bar:** Cancer or cell type category dropdown set to BCL, Advanced AND dropdown set to Subtype. Search button and AddResults2Boxplot button. Result count: Found 792 results.
- Filter Panel:** DLBCL selected in the dropdown. Columns dropdown is open, showing DLBCL-ABC, DLBCL-GCB, and DLBCL-PMBL.
- Table Headers:** GSM identifier (sample), GSE identifier (experiment), Sample type, Main category, Cancer or cell type, Subtype, Further specification, Cytogenetics, Other genetic alterations, Sample source.
- Table Data:** A list of 7 entries matching the search criteria. The first entry is highlighted in blue.

GSM identifier (sample)	GSE identifier (experiment)	Sample type	Main category	Cancer or cell type	Subtype	Further specification	Cytogenetics	Other genetic alterations	Sample source
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



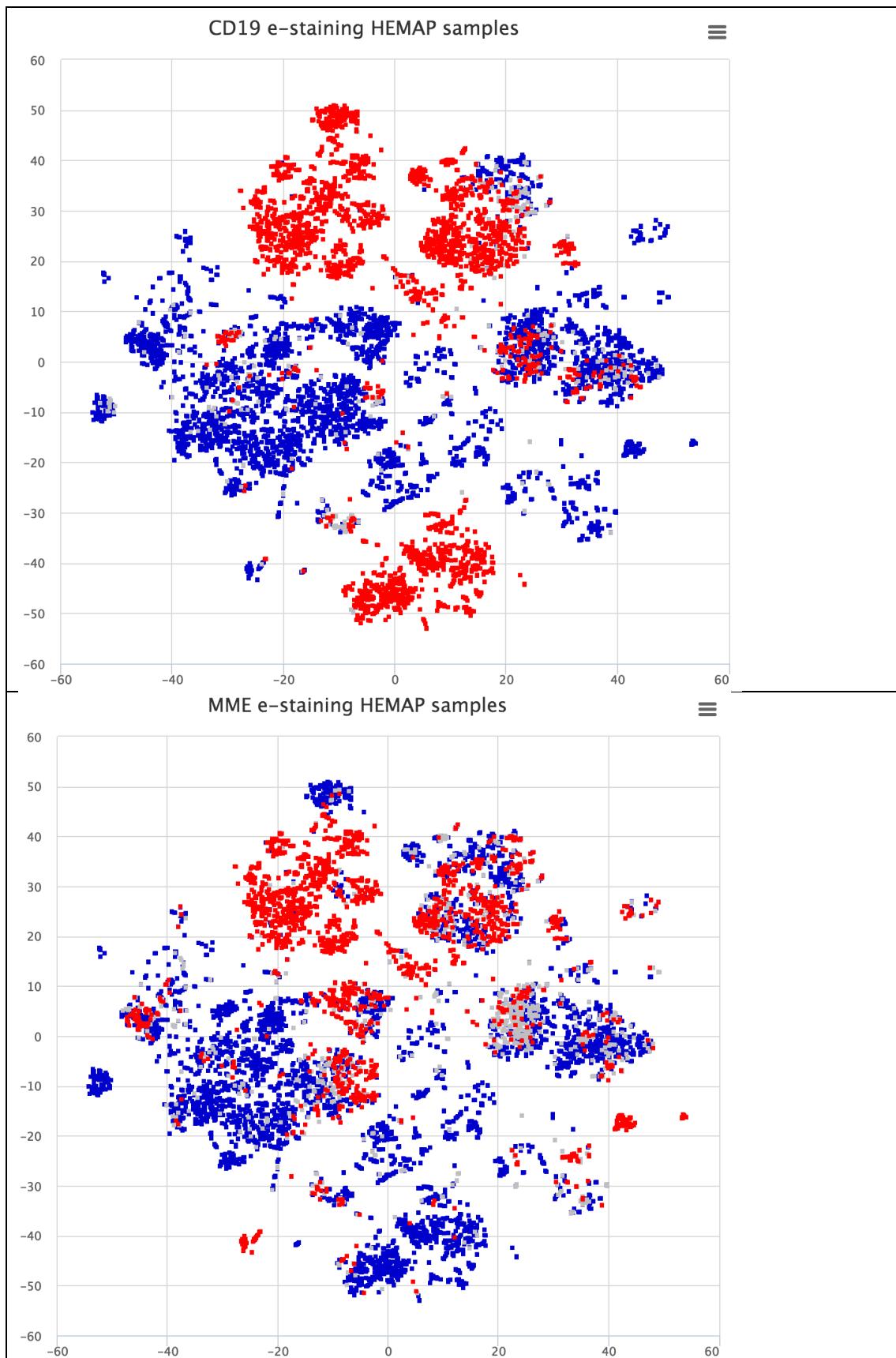




**Figure 14.** HEMAP samples are further classified and e-stained based on cancer type BCL subtypes DLBCL (1<sup>st</sup> 2<sup>nd</sup> rows) and CHL (middle 3<sup>rd</sup>,4<sup>th</sup> rows). The results are obtained via combining cancer/cell type with subtype. Filtering and its combinations are 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, the cell surface 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.

Resource map	HEMAP samples	Type	Gene Expression	guide
Gene/Pathway/Drug				
CD19				e-stain



**Figure 15** Gene based e-staining is accomplished for cell surface protein 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.

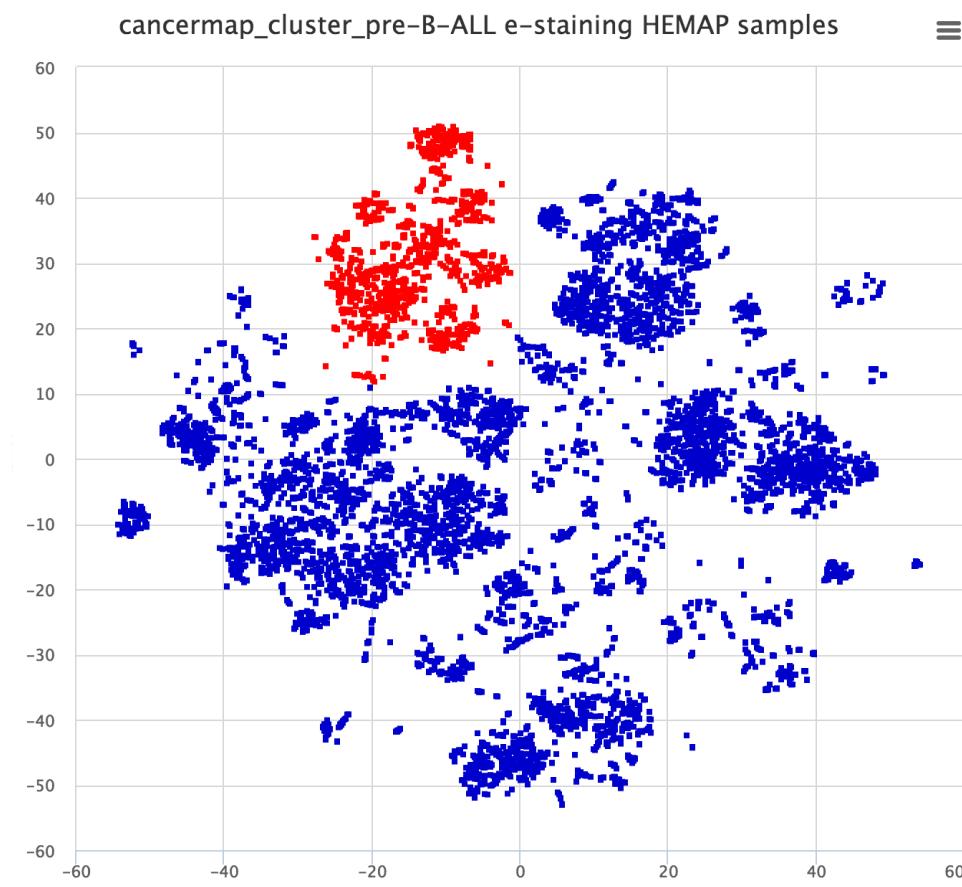
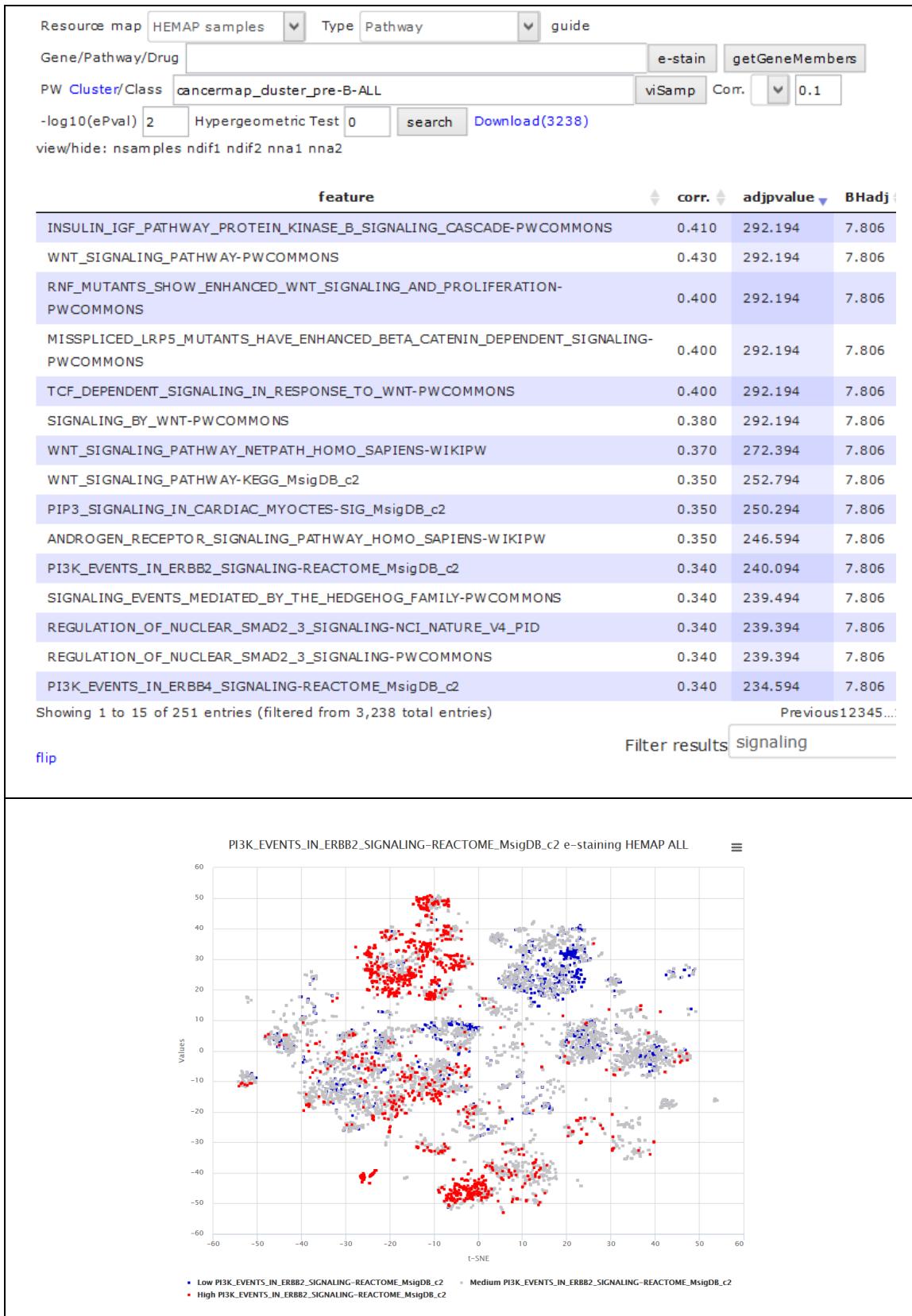


Figure 16. pre-B-ALL clusters stained using visSAMP.

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

Resource map: HEMAP samples | Type: DrugSigDB | guideGSVA/FDR

+/-.49  
Gene/Pathway/Drug: BEZ235\_LINCS-DSigDB\_D2

e-stain | getGeneMembers

PW Cluster/Class: cancermap\_cluster\_pre-B-ALL

viSamp | Corr. > 0.1 | -log10(ePval) 2 | Hypergeometric Test 0 | search

[Download\(1192\)](#)

feature	corr.	adjpvalue	BHadj	hypergeom_test	BHadj	nsam
BEZ235_LINCS-DSigDB_D2	0.470	292.194	7.806	316	7.806	954
GALLAMINE_TRIETHIODIDE-DSigDB_D4	0.450	292.194	7.806	316	7.806	954
HC_TOXIN_ALL_DOWN-DSigDB_D3	0.400	292.194	7.806	306.259	7.806	954
REPAGLINIDE-DSigDB_D4	0.400	292.194	7.806	269.54	7.806	954
KINOME_192_ROCHE-DSigDB_D2	0.440	292.194	7.806	263.938	7.806	954

BEZ235\_LINCS-DSigDB\_D2 e-staining HEMAP samples

**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**

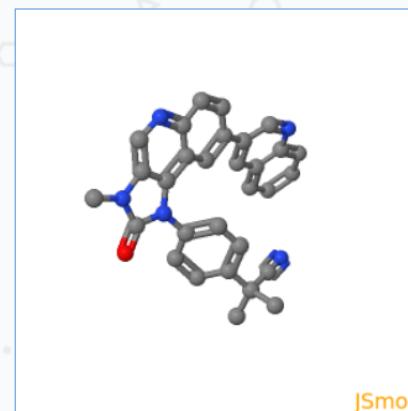
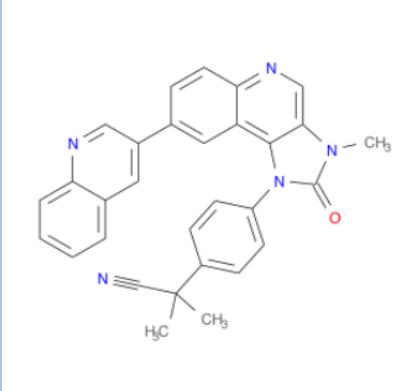
## Gene Set: D2 : HMS LINCS - Nvp-Bez235

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

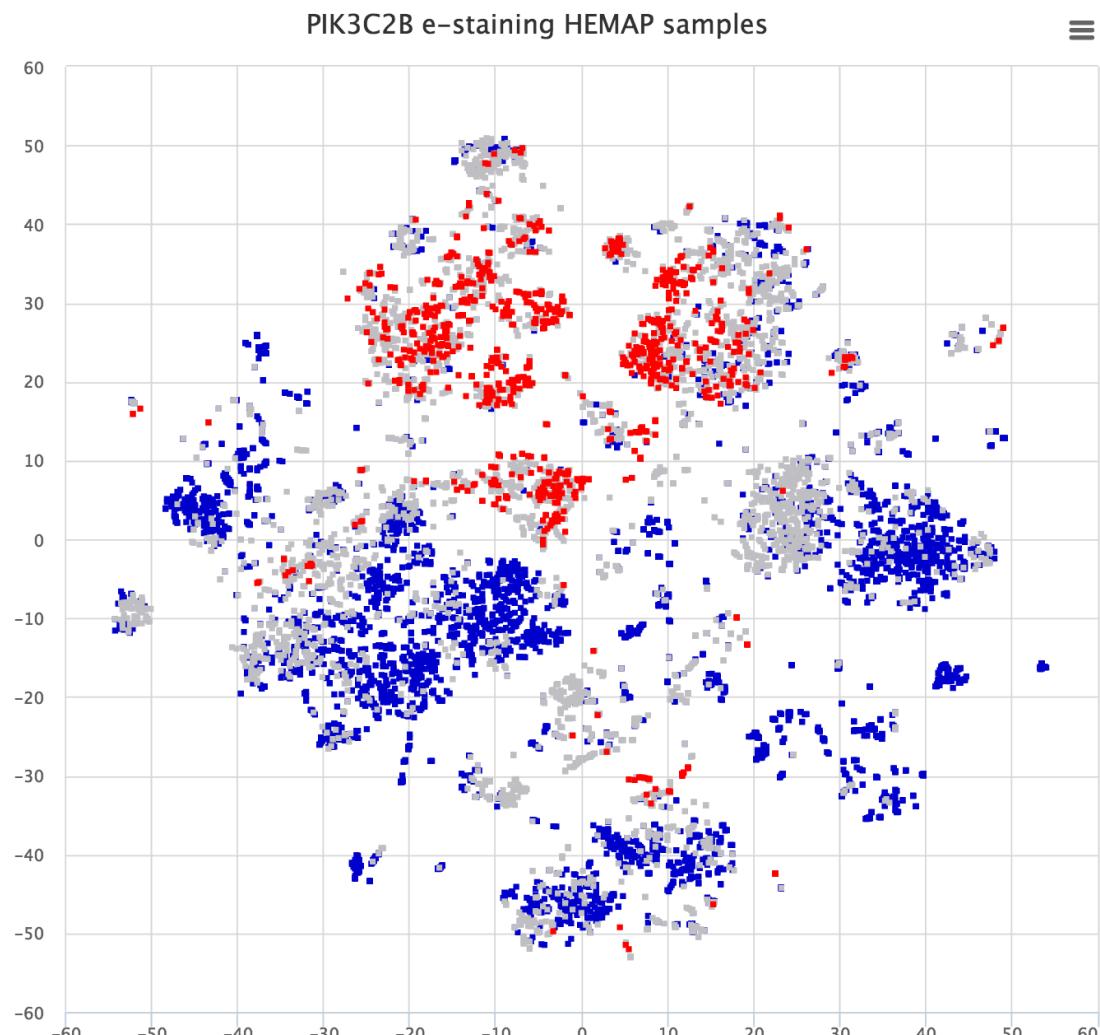
CAS Num : 915019-65-7

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

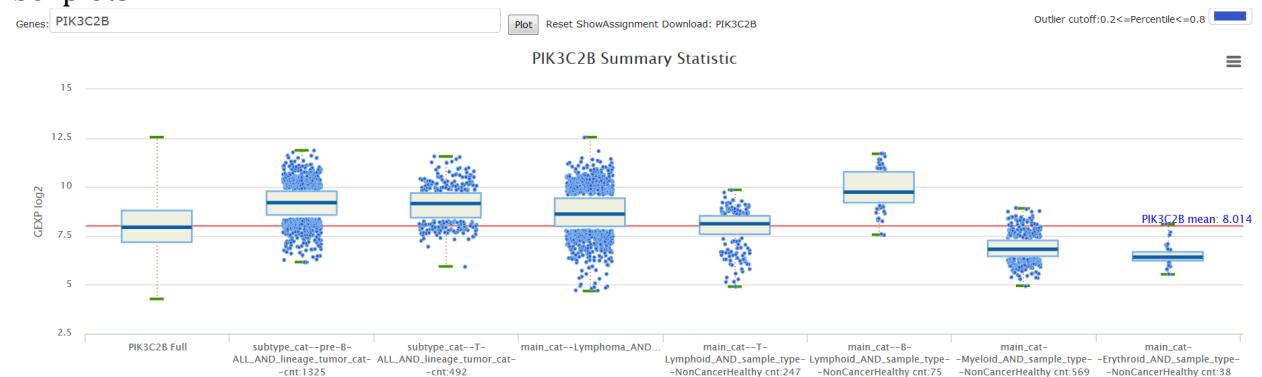
Download gmt, text, Detailed text  
gene sets

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 are dropdown menus for 'Max GSM mouse selection' (set to 500), 'Symbol size' (set to Default), and 'Outlier Percentiles' (<= 0.2, >= 0.8). Below these are more dropdowns for 'Pairwise rows limit' (500), 'Rows shown on page' (15), 'Map-PWTable WidthRatio' (60:40), 'ePval' (.05), and 'Order by' (pval). Further down are three color bars labeled 'Low' (blue), 'Medium' (grey), and 'High' (red), followed by '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 browser.**

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 used as it is and does not offer any warranty or guarantee. Please contact [matti.nykter@uta.fi](mailto:matti.nykter@uta.fi) and [merja.heinaniemi@uef.fi](mailto:merja.heinaniemi@uef.fi) for commercial usage permissions.

## Comparing an independent dataset with Hemap samples using cancer maps

### Background

The normalized data matrix is available for download from the web resource for further integrative analysis using standard R/Bioconductor software. The required objects are available as RData files and facilitate joint analysis of new datasets.

### Description

In this example, the Ross acute lymphoblastic leukemia (referred to as Ross ALL) study will be jointly analyzed with Hemap samples.

We will compare this new study to Hemap samples by generating a cancer-map, similar to those provided in the Hemap resource. In this case, mean-shift clustering using bandwidth parameter 0.9 results in 9 clusters with high correspondence to cytogenetic annotations.

Next, the similar sample groups will be identified from the Hemap sample collection based on gene set scores. For this purpose genes that correlate with the cluster assignment will be identified, followed by visualization of the gene set scores on the map using e-staining.

The analysis available from Explore can then used to characterize these sample groups by correlating the cluster assignment with gene set, pathway and annotation features (refer to the other User guide examples for instructions).

### Usage

The provided R-script allows users to create a cancer-maps from a new dataset (see input data format below) and Hemap samples. The gene set scoring is performed using functions implemented in the R/Bioconductor package GSVA. Cytogenetic, mutation, copy number or DNA methylation data available for the samples can be used as input for the e-staining function to ensure that the clusters resemble genetically similar sample groups. A wrapper script that runs all steps for the example is included ([Rscript\\_find\\_clusters\\_TSNE\\_map.R](#)).

## Input data

This section describes what is needed to run similar analysis for any new dataset for which comparable samples are found in Hemap. For the example ALL dataset, the ready input data is available as an RData object.

- 1) The new dataset to be compared should be uploaded to R for analysis. The following format is expected if you follow the example code given below:

SYMBOL	JD-ALD428-v5-U133A	JD-ALD011-v5-U133A	JD-ALD035-v5-U133A	JD-ALD386-v5-U133A	JD-ALD387-v5-U133A	JD-ALD388-v5-U133A	JD-ALD389-v5-U133A	JD-ALD391-v5-U133A
NAT2	3.90166023474923	4.57436193087391	3.92277207590922	4.08680065887691	4.06379164803862	3.93637972132651	4.03170843475255	4.0834907646788
ADA	8.24254014225021	7.07010932649795	7.06755348234404	7.55517885848556	7.08785336409747	6.83781901229187	7.53580432692934	6.5599291300822
CDH2	3.82164945384477	3.93542032197657	3.78126498325648	3.85013124097616	4.00660971479218	3.98865899143423	3.90673942451677	3.99828135926627
AKT3	2.90641186851881	3.37670619898111	3.10952156012042	3.07324336114044	2.95517160705336	3.00761798980276	3.47935088918691	3.0354172038487
MED6	4.27791058213262	4.46966844020427	4.4020276743519	4.09260880775334	4.52785941142264	4.17960012480704	4.66802475984958	4.48481440058593
NR2E3	3.06201500286937	3.55366394681636	3.16177284431474	3.13311275991516	3.15882925086513	3.17392790329423	3.17126236673153	3.13359454070503
NAALAD2	2.23737722829235	2.17489686470953	2.17882349192806	2.22059138791208	2.23896340526372	2.10476890248036	2.12295490713235	2.10876948464123
NAALAD1	2.73444736384272	3.24394276682265	2.98518230223398	3.63526541241986	3.06462541241986	2.97003942291121	3.12359370478477	2.955852744715
CDKN2B-AS1	4.44473148571337	4.765324405599714	4.33983410074841	4.80351812905632	4.80148261626042	4.9131881510479	4.51417680818917	4.817096303175
ACOT8	5.28476350278237	5.17849161617626	5.21916418263883	4.72872260889102	4.99752280001247	5.18928505340086	4.99062096078961	5.2921687605282
AB11	6.56151705173912	6.15567221443852	6.22062860614184	6.4856304787852	7.18156434753867	6.23777332527874	6.32527062861189	6.59269761321854
GNPDA1	3.50951193949713	3.30802111078989	3.62083160938696	3.67137269151345	3.54356055781845	3.46587485251744	3.80097010453393	3.37275957308621
ZBTB33	2.55111273084133	2.72632604154482	2.7052448841611	2.48744126918428	2.63265327060339	2.54864063496103	2.23351745199273	2.74601909020327
GS1-600G8.3	2.99819580082003	3.18476009142842	2.88755379311625	2.9409366453268	3.12220318077858	3.09211862860817	3.15895127191349	2.9126305765382
CDH3	3.75336522867252	4.14131075504685	3.516899241581	3.7030268104871	3.56639213598694	3.90106436967326	3.57286389295936	3.849203239333
TANK	3.04612294754131	2.95308253353195	3.18029497251378	3.1002090653741	3.1014545456157	3.17579411732521	3.1903729854312	3.85421970071311
EGOT	3.31271790440567	3.57681610419996	2.99453372025441	3.44881537486818	3.1366107695785	3.31458943218616	3.0341612742087	3.0562520590026
HGC6.3	3.0529367183507	3.06833551104816	3.21010401603803	3.45653732310925	3.4418733265517	3.40162907618177	3.59990613749718	3.21074890389713
ACVR2B-AS1	3.21122203428237	2.86191970778242	3.11606238041263	2.8371077160453	2.83220602734411	3.13392435462571	3.06502342253644	2.840652420675
KHDC1L	3.30835687916482	2.90372016069118	2.84256323716618	3.5894815024902	2.98463232443134	3.00713585748736	3.16118441281854	3.13482452947633
TOPORS-AS1	4.7075628608164	4.61976118307796	4.51989578484985	4.62218101268014	4.99256077093214	4.65712205145365	4.72424132298132	4.688325458415
C1orf68	4.2269368174902	4.0634097143006	3.7120315404914	3.76555911547762	3.97624943632885	3.98814955420688	3.97628632024202	4.13831091626194
SMIM10L1	5.15209462422466	4.37072115197994	5.65068191076254	4.27562062090971	4.53277741318928	4.555822090494	4.25091783715449	3.565865751465
DPY19L1P1	2.66032821106212	3.20770013670623	2.63190208037877	3.02554682330667	2.87540924033935	2.69348843744764	2.67490493268737	2.819598917965
ZNF378P	2.56842787372607	3.29964391906447	2.7339869870329	3.3478546099432	2.8981399542636	3.28184373618672	2.95994594575905	2.89428592622704
PP13	5.29891278910915	5.77703724873662	5.14590651926768	5.61180339991909	5.43188487800759	5.27908245222969	5.42390995955149	5.36770750567261
MUC8	6.93088500100671	7.4479484377518	7.06628170733906	7.01444430695525	7.20131169389749	7.54224666561607	7.60294924226109	7.32812693075868

If using Excel to generate the file, save your data as "tab delimited text"

- 2) The map can be used for e-staining different sample features (cytogenetic type, clinical features, mutation data, etc). Use the sample annotations available from your study to create vectors that specify the colors to be used in the visualization.

You may find e.g. this page helpful

<http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf>

For example, if you have the following sample annotations

```
sample1      MLL
sample2      other
sample3      MLL
```

Your color vector could be specified

```
myColors=c("pink", "orange", "pink")
```

- 3) A subset of samples from Hemap that matches the new dataset is next retrieved.

Any disease subset can be retrieved by loading the full data matrix and taking a subset based on the vector *subtypes*. The annotation table can be used to specify other selections. For the example below that compares ALL studies, the subset has been generated in the following manner:

```
load("HEMAP_data.Rdata")
subset=subtypes=="T-ALL"|subtypes=="pre-B-ALL"
matrix=matrix[,subset]
```

To generate a cancer map from the subset and required objects for the next steps use the provided function CancerMap. For parameters in capital letters refer to the end of the document. Output is a dataframe (*X*) with columns id, x, y and cluster.

```
CancerMap(t(matrix), NAME, VAR, BW, PATH_OUTPUT)
clusters=unique(X[,4])
```

4) Quality control. At this step it is wise to verify that the separation of the samples on the map reflects biology and not the data provider. We recommend to use established molecular subtypes for the disease in question. The following example code does the quality control to the Hemap subset map. Obtain the relevant phenotype information for the sample subset using the Annotation table and read to R for analysis.

Calculate the NMI metrics  
(NMI function found in file infoMeasures.R)

Mixing of data series (low value is better)  
NMIm=NMI(clusters, dataseries[subset])

Separation of phenotypes (high value is better)  
phenotypes=scan("phenotypes.txt", what="character")  
NMIp=NMI(clusters, phenotypes)

The parameter VAR for the CancerMap function specifies what selection of genes is used (default 15% most variable). Different values can be tested to improve the result.

Save the objects for later analysis as an RData object.  
save.image("myData.RData")

## Example analysis

Tested with R version 3.2.2 and Bioconductor version 3.2. For list of parameters, optional pre-calculated RData and wrapper script, refer to end of document.

```
# Load data, set parameters and load libraries. Installs missing R packages
# automatically.
source("load_default_parameters.R")

# Modify accordingly or use ROSS and Hemap ALL example data

source("load_data.R")
# Objects loaded:
# newdata: New dataset as data.frame for analysis
# newdata_color_vector: color vector for the new dataset
# matrix: Hemap data matrix as data.frame
# subtypes: Vector defining subtypes for Hemap data
# dataseries: Vector defining dataseries (i.e. data provider) for Hemap data
# hemap_color_vector: Vector defining colors for plotting

# Step 1: Generate cancer-maps using BH-SNE and cluster samples on the map
# using the mean-shift algorithm.

# new data cancermap
clust=CancerMap(t(newdata), NAME, VAR, BW, PATH_OUTPUT)

# Hemap cancermap
clust_hemap=CancerMap(t(matrix), HEMAP, VAR, BW_HEMAP, PATH_OUTPUT)
```

### Output

```
cancermap_Ross_rma_u133a_15pct_genes_BH-SNE_mean-shift_BW0.9.txt
cancermap_Ross_rma_u133a_15pct_genes_BH-SNE_mean-
shift_BW0.9_cluster_centroids.txt
cancermap_HEMAP_ALL_15pct_genes_BH-SNE_mean-shift_BW1.5.txt
cancermap_HEMAP_ALL_15pct_genes_BH-SNE_mean-
shift BW cluster centroids.txt
```

```
# Read in new cancer-map coordinate data
X=read.delim(paste0(PATH_OUTPUT, "cancermap_", NAME, "_", VAR,
"pct_genes_", "BH-SNE_mean-shift_BW", BW, ".txt"), header=T,
stringsAsFactors=F)

# Read in new cancer-map cluster centroid data
```

```

peaks=read.delim(paste0(PATH_OUTPUT, "cancermap_", NAME, "_", VAR,
"pct_genes_", "BH-SNE_mean-shift_BW", BW, "_cluster_centroids.txt"), header=T,
stringsAsFactors=F)

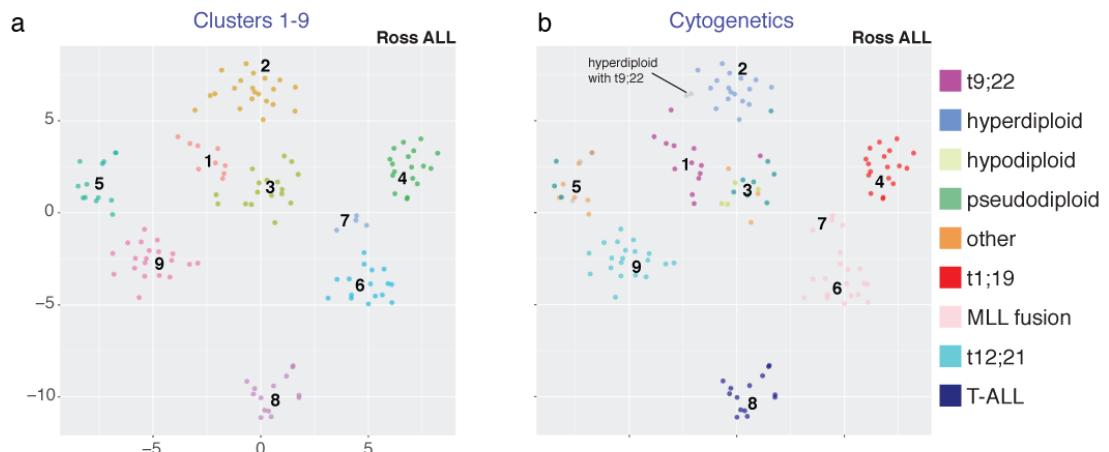
# Plot clusters with different colors
Plot_cancermap_clusters(X, peaks, CLUSTER_CENTRE, NAME, SIZE, VAR, BW,
NAME, PATH_OUTPUT)

# Plot color vector
Plot_color_vector(X, NAME, SIZE, newdata_color_vector, NAME, PATH_OUTPUT,
peaks)

```

### Output

cancermap\_Ross\_rma\_u133a\_15pct\_genes\_BH-SNE\_mean-shift\_BW0.9\_singlepage.pdf  
Ross\_rma\_u133a\_singlepage.pdf



**Figure 23. Rscript step 1.** a. The data-driven cluster assignment (left, colors indicate different clusters) can be compared to annotated cytogenetic types (indicated in color in b) and correlated with sample molecular features.

# Step 2: make gene sets that are identified from Ross cancermap, run GSVA.

```

# Obtain cluster number vector
clusters=unique(X[,4])

# Make gene sets to identify similar clusters from Hemap
genesets=unlist(lapply(clusters, Find_correlated_genes, newdata, X),
recursive=F)

# Run GSVA to get gene set scores
gsva_es <- gsva(as.matrix(matrix), method="gsva", genesets, mx.diff=F, tau=0.25,
verbose=T, rnaseq=F, min.sz=5, max.sz=500)
# Obtain GSVA score matrix

```

```

gsva_es=gsva_es$es.obs
feats=rownames(gsva_es)

# Read in cancer-map coordinate data
X_hemap=read.delim(paste0(PATH_OUTPUT, "cancermap_", HEMAP, "_", VAR,
"pct_genes_", "BH-SNE_mean-shift_BW", BW_HEMAP, ".txt"), header=T,
stringsAsFactors=F)

# Read in cancer-map cluster centroid data
peaks_hemap=read.delim(paste0(PATH_OUTPUT, "cancermap_", HEMAP, "_",
VAR, "pct_genes_", "BH-SNE_mean-shift_BW", BW_HEMAP,
"_cluster_centroids.txt"), header=T, stringsAsFactors=F)

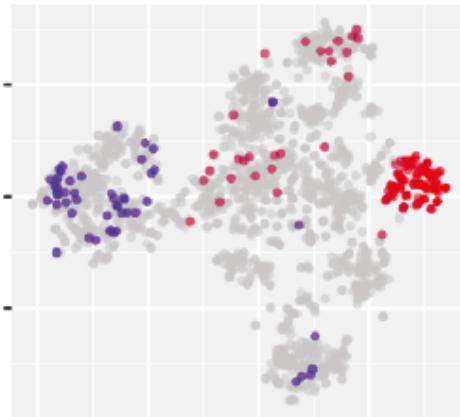
# Generate plots for each cluster with gsva score colored
p.all=lapply(feats, Plot_GSVA_scores, gsva_es, VALUE, SIZE, CLUSTER_CENTRE,
X_hemap, peaks_hemap)

# Make an A4 size figure with multiple panels
ggsave(paste0(PATH_OUTPUT, HEMAP, "_", NAME, "_", PATHW,
"_multipage.pdf"),
do.call(marrangeGrob, list(grobs=p.all, nrow=4, ncol=3)), width = 210,
height = 297, units = "mm", dpi=150)

```

### Output

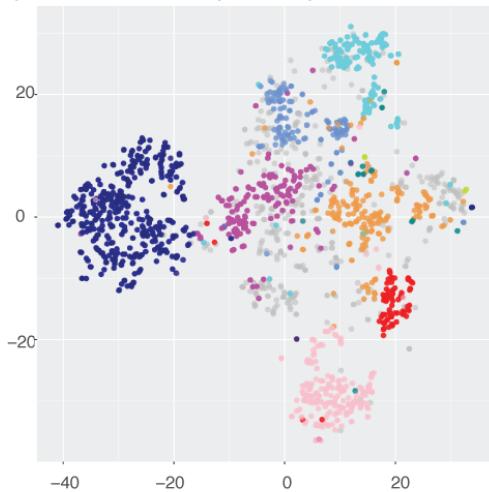
HEMAP\_ALL\_Ross\_rma\_u133a\_clusters\_top20\_GSVA.Rdata  
HEMAP\_ALL\_Ross\_rma\_u133a\_clusters\_top20\_multipage.pdf



**Figure 24.** The cluster-specific gene sets distinguish similar samples from Hemap ALL cancer map. Ross cluster 5 gene set score is shown as an example. Red color corresponds to GSVA score above 0.4 and blue below 0.4.

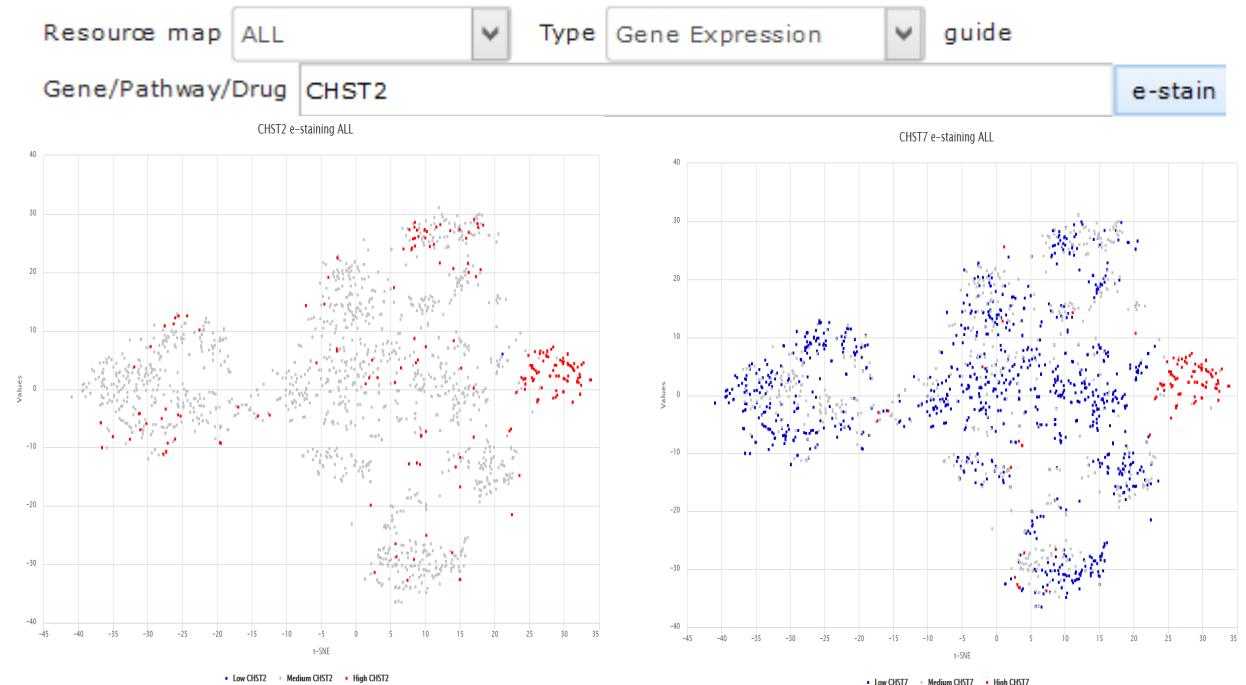
# Step 3: Plot cytogenetics in hemap.

```
Plot_color_vector(X_hemap, HEMAP, SIZE, hemap_color_vector, HEMAP,
PATH_OUTPUT, peaks_hemap)
```



**Figure 25.** The annotated cytogenetic sample type is colored as in Figure 23. The correspondence of gene set score based matching (in Figure 24) with the annotated sample category can be evaluated.

**Step 4.** Characterizing the new cancer-map clusters can then be carried out by utilizing the Hemap Explore analysis. This is exemplified below taking a biomarker example from Harvey et al. Blood 2010, where a good prognosis subtype was identified that has high expression of the genes *CHST2* and *CHST7*.



**Figure 26.** E-staining marker gene expression for a good prognosis subtype.  
Parameters for the analysis

### Parameters used in generating gene sets per cluster

PVAL	P-value threshold for correlation (default 0.05)
NUM_GENES	Number of genes to select per gene set (default 20)

### Parameters used in plotting

VALUE	GSVA score cutoff (default is 0.4 when NUM_GENES is 20)
CLUSTER_CENTRE	Boolean variable whether to plot cluster centroids to cancer maps (default=T)
SIZE	Variable defining point size in cancer maps (default 0.7)

### Parameters for cancer map generation

VAR	Percentage of most variable genes to be included in the dataset when generating cancer map (default 15)
BW	Bandwidth parameter for mean-shift clustering for new data, which typically has less samples and smaller bandwidth (default 0.9)
BW_HEMAP	Bandwidth parameter for mean-shift clustering for hemap. Hemap bandwidth was 2.5 and Hemap AML/ALL bandwidth was 1.5

### Set output names for new cancer-map and gene set results

NAME	New dataset name
PATHW	Gene set name

### Set output name for Hemap cancer-map

HEMAP	HEMAP cancer-map used
-------	-----------------------

### Additional data objects and wrapper script

Precomputed data and GSVA scores are provided as an RData object which can be used to **plot data directly**.

```
load("Example_data_ROSS_HEMAP_ALL.Rdata")
# Contains objects:
# feats: Gene set names.
# gsva_es: GSVA scores for genesets found from Ross ALL clusters.
# X: Ross ALL cancermap coordinates and clustering.
# peaks: Cluster centroids for Ross ALL cancermap.
# X_hemap: Hemap ALL cancermap coordinates and clustering.
# peaks_hemap: Cluster centroids for Hemap ALL cancermap.
```

All analysis steps presented above can be run using the wrapper script  
`source("Rscript_find_clusters_TSNE_map.R")`

### **Adding your data to Hemap**

If you would like to upload your data to Hemap resource, please contact  
[matti.nykter@uta.fi](mailto:matti.nykter@uta.fi)  
and [merja.heinaniemi@uef.fi](mailto:merja.heinaniemi@uef.fi)