

Oomicsplayground



**General Notes and Specific Recommendations
for the Design of Data Visualizations and Tables**

February 2, 2023

General Notes:

- For plots I followed these **general style guidelines**:
 - font family is Lato for all text elements (features tabular numbers)
 - is often not the case for many panels (also see comment on tooltips at the end of this list)
 - avoid rotated flipped labels if possible
 - often the case though but handled by plotly based on number of labels and aspect ratio
 - avoid the same color palette to encode different variables
 - a bit inconsistent for now as we have many different variables and also plots that need to be reimplemented
 - use outlines in case colors can become light
 - see for example the scatter plot (c) in the “Data View: Plots”
 - Uppercase title cases for axis titles
 - Include units in axis text instead of axis titles
 - i.e. labels “5% 10% ...” and title “Proportion” instead of “5 10 ...” and “Proportion (%)”
 - Tooltips show both, the variable name and value; the first is styled in regular, the latter in bold
 - also, avoid trace names and provide that information in tooltip
 - for some reason, plotly doesn't sue the theme font (Lato) in the tooltips; needs to be fixed
- **Resolution of text elements** in plots looks pretty bad, doesn't look sharp but pretty pixelated on all my screens
- In general, there are multiple **categorical color palettes**; even though it might become colorful, it would be good to use the same encoding across all charts featured in the app
 - Sx clusters (S1, S2, ...)
 - cx clusters (c1, c2,...)
 - MEx cluster (ME0, ME1, ...)
 - (major) gene types (Kinases, Ribosomal, ...)
 - high-level grouping variables (activity, time, ...)
 - groups used in “Data View: Plots (f)”
 - good/bad encodings (not truly categorical but if only -1 and +1 basically aas categories; most extreme colors from diverging palette, i.e. blue and orange with the later encoding “bad” outcomes)
 - T.cell groups
- There is a reoccurring pattern that **diverging palettes** are used in the “wrong” direction, i.e. with red encoding “good”, high values/correlation/...
 - make sure to use a proper, colorblind-safe diverging palette with the alarming color encoding the lower range
- Sometimes, **text spans the whole screen** (e.g. some captions) which is hard to read and bad in terms of user experience
 - consider to split text in multiple columns in these cases
- The **panel headers** are not consistently formatted
 - sometimes the tag position or styling differs
 - the text formatting switches between upper-case letters for all words versus the first word only
- **Tables** are currently coming in multiple looks, would be good to have one consistent style
 - I prefer the table without the stripped look (rows being alternately colored in white and grey)

Colors:

- The final set of **BootsTrap** colors we settled on are the following:



- The script `ggplot-colors.R` features functions to create **sequential, diverging, and categorical color palettes**
 - palettes are based on the BS colors and a few other colors
 - categorical palettes come in multiple form as `super_light`, `light`, `dark`, `super_dark`, and `muted` variants
 - the sequential and diverging palettes are designed in a way that the lower range and midpoint colors, respectively, don't feature too light colors to avoid "invisible" encodings on the white background
 - in general, color palettes are generated by calling `omics_pal_c()` for sequential and diverging palettes and `omics_pal_d()` for categorical palettes
 - the scale functionality was developed for `ggplot2` output and cannot be used in the current `plotly` setup
 - the color palettes (along with the `ggplot2` functionality) are showcased in the report `showcase_ggplot_theme_scales.Rmd`

Board
“Load”

Load: Load dataset

The screenshot shows the BigOmics Analytics interface for loading datasets. On the left, a sidebar has 'Load dataset' selected. The main area displays a table of loaded datasets and a t-SNE scatter plot.

Data files:

dataset	description	datatype	nsamples	ngenes	conditions	date
example-data	Proteome profiles of activated vs resting human naïve T cells at different times (Geiger et al., Cell 2016).	LCMS proteomics	18	6987	conditiontime	2021-09-30 activated

Dataset explorer: A t-SNE scatter plot showing data points. One point is labeled "example-data".

Action buttons:

- Delete dataset
- Download PGX
- Download ZIP
- Load dataset** (button highlighted with a yellow box)

- Currently it is possible to deselect all data sets and hit the “Load dataset” button
 - button should be inactive in case no data is selected
- The scatter plot on the right looks rather empty
 - use bigger point sizes and label sizes?

Board “Data View”

Data View: Plots



General notes:

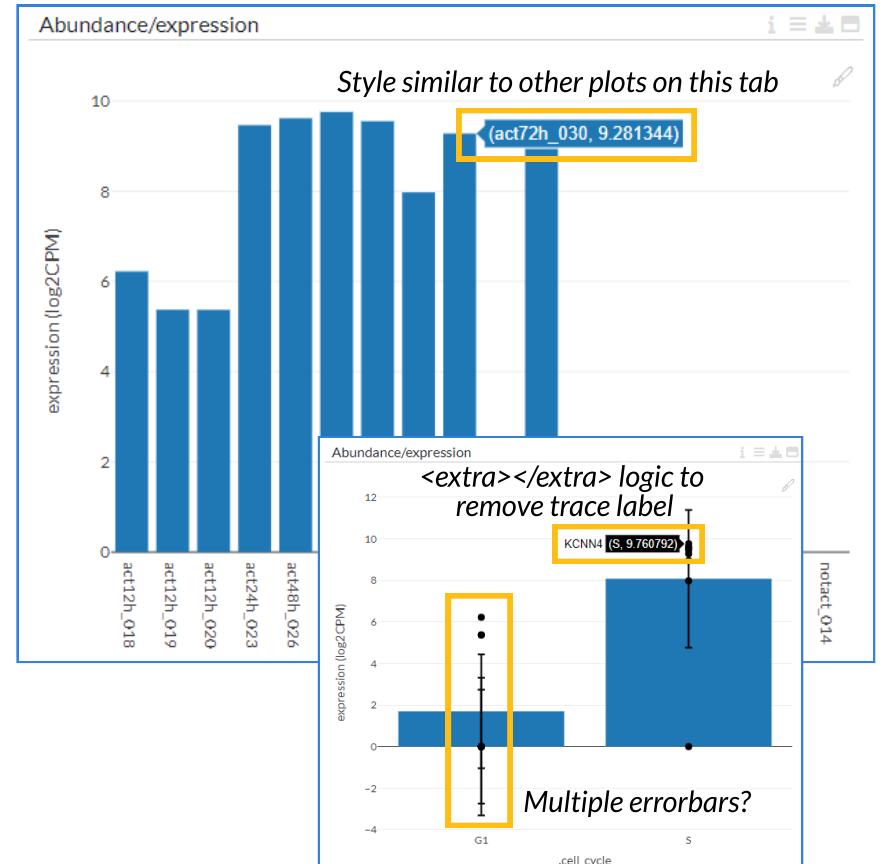
- Tags are missing; not consistent with other tabs and boards plus it is unclear what the caption refers to

(a) Gene Information:

- On my Win machine, the KEGG.db package is still not available thus the gene information is throwing an error

(b) Abundance/expression:

- Does it make any sense to show error bars on a log scale?
Actually, log-transformed values should not be encoded as linear forms.
- Displaying groups in the tooltips don't work currently, as there are multiple groups per bar
 - check if that should really be the case – if so, decide how to handle in tooltips
- Currently showing mean in the tooltips doesn't work as there are multiple averages per bar
 - check if that should really be the case – if so, decide how to handle in tooltips
- I did style the tooltips, however that change seems to be intentionally or unintentionally been overwritten
 - style tooltip similar to the other plots in this tab ("var name + value" and no trace via "<extra></extra>")



(c) Average rank:

- Currently x and y cannot be shown explicitly in the tooltips as the format doesn't comply with the chart
 - check if the format can be adjusted (if information is needed in tooltip at all?)

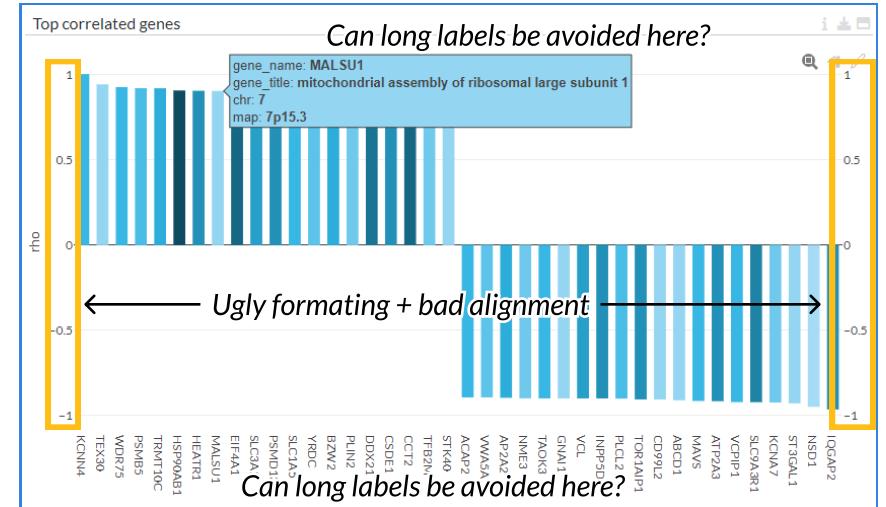
Data View: Plots (continued)

(d) tSNE clustering:

- I have included the grouped plot option with the same logic as the ggplots before but somehow plotly doesn't update the plot in case the grouping is changed; when changing the group in the "Settings" points should be colored by the respective groups
 - check if I messed something up or if plotly needs a different logic

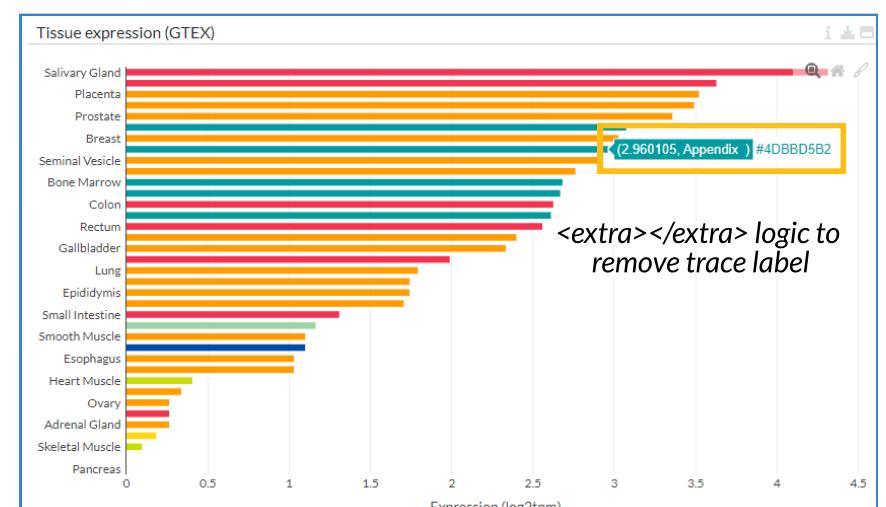
(e) Top correlated genes:

- Currently some x labels are pretty long; also the variable names are cryptic
 - check if names can be shortened and variable names can be formatted nicely
- Numbers on the y axis are badly formatted (i.e. no trailing .0 for integer numbers)
 - it's how it is, according to Carson there is no good solution to fix that
- A second y axis seems useful (even cooler would be a positive on the left and a negative on the right)
 - decide if dual axis should be kept; as Carson mentioned no real solution to the alignment and pos vs neg axes
- Labels in the tooltip are taken from data set as I don't know what they mean (e.g. chr and map)
 - use more meaningful, nicely styled names
- I was always confused by the color encoding
 - as it's encoding another variable (expression), a legend would help to make this clear to everyone
 - fine to use the same color palette as in the tSNE clustering scatter plot but one needs to make sure that the range is the same (currently it's not so the same shade of blue represents different expression values; needs to be fixed)

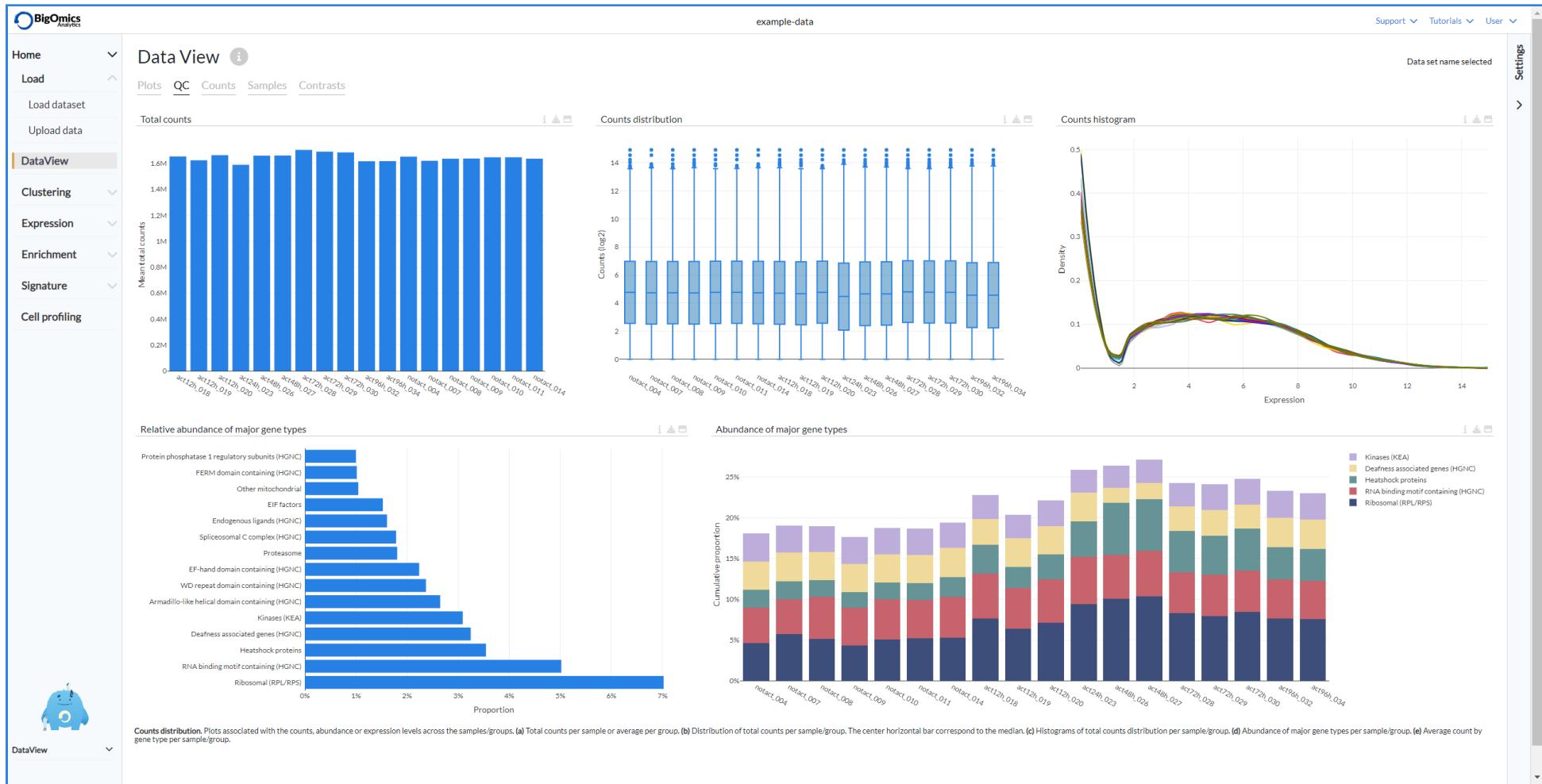


(f) Tissue expression (GTEx)

- Show actual names in legend (as one makes sense here; you could also think about other ways how to encode that information)
 - currently the groups are encoded as color codes, the actual group information seems missing in the current data set passed here
- I have flipped the chart for better readability of the tissue labels (and it's also good for the sake of variety I think)
 - decide if horizontal version should be kept
- Once information is provided, style tooltip similar to the other plots in this tab ("var name + value" and no trace via "<extra></extra>")
- Currently the colors are a nice variety but seem a bit bright as they are based on the BS colors



Data View: QC



(a) Total counts:

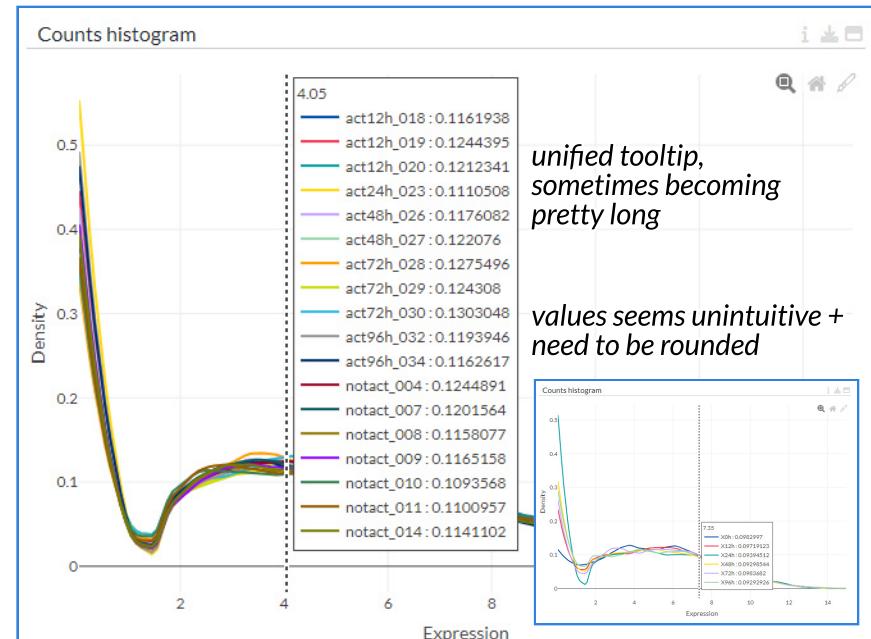
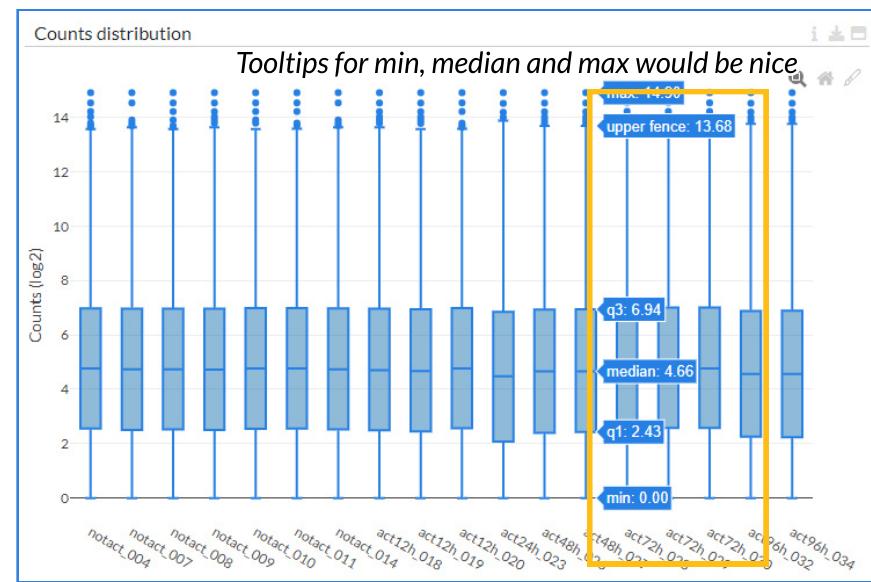
- All fine for now I guess.

(b) Counts distribution:

- In general it's not a good idea to show boxplots on a log scale!
 - Does it really need to show log values? If so, maybe a jitter with indicators for the key summaries might be better?
 - I'd like to show less labels as tooltips; could format them nicely but didn't find any information on how to remove some
 - check if it is possible to reduce the number of tooltips

(c) Counts histogram:

- The chart is not a histogram, title should read “density curves” or something similar
 - As all lines overlap and it is hard to highlight a particular one, I have added an unified tooltip based on the x axis
 - decide if that is the preferred behavior or if single lines should be highlighted
 - maybe two different solutions depending on the grouping (a single tooltip if ungrouped but individual tooltips for each category if grouped)
 - The tooltip needs some further formatting: clearly indicate that the top value is the expression score; round values of groups to x digits
 - are those numbers any meaningful at all?
 - The list in the tooltip can become very long, too long for the overall slide
 - decide how to handle – maybe only show detailed tooltip in a separate window
 - In general way too many colors (but guess no solution to this),
 - maybe use grey only and a overall summary in bright colors



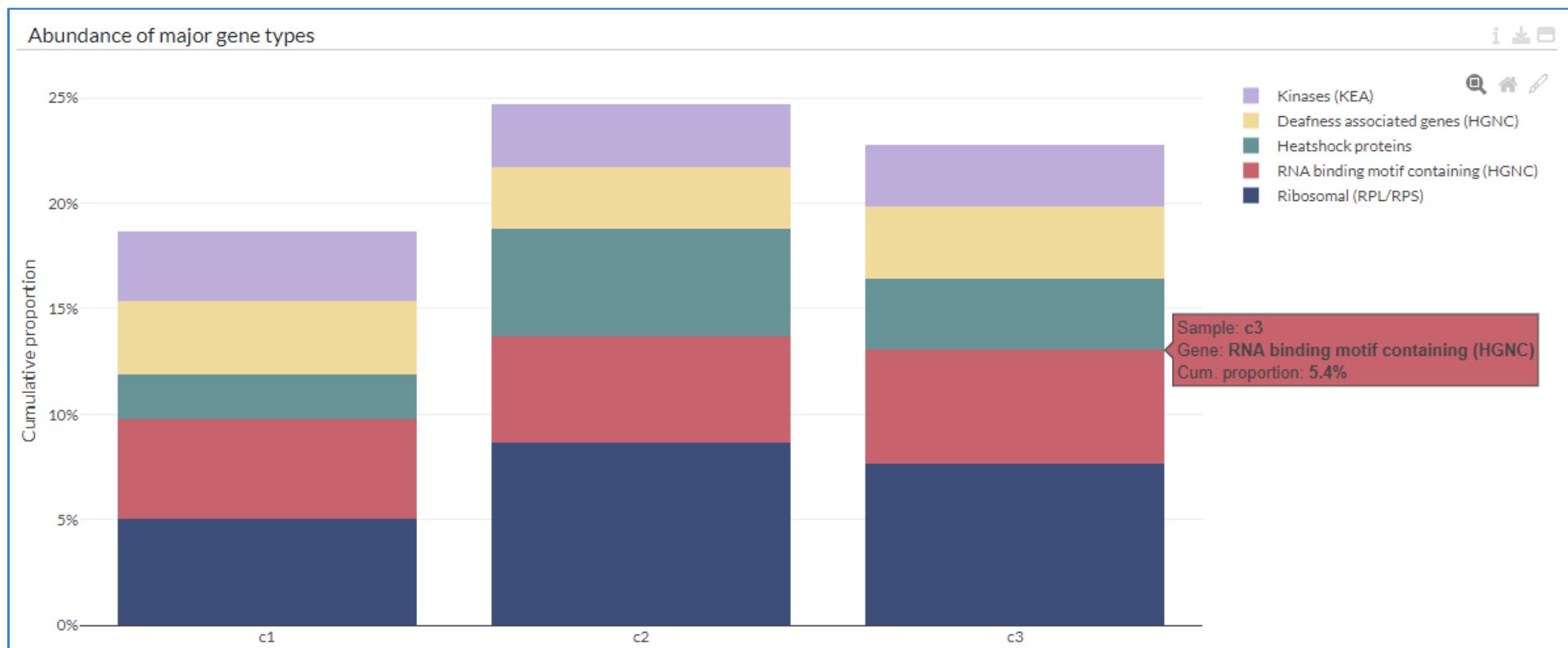
Data View: QC (continued)

(d) Relative abundance of major gene types:

- Uses same color as (a) and (b) which might be okay. In general, something to decide one how to deal with using the same colors for different values (here counts versus proportions, seems okayish)
- As these gene types are the same as the “major gene types” shown in plot (e), it would be great to use the same colors here to allow for an simple recognition of that fact and be consistent in color use
- Why are bars sorted in scending order, shouldn't be those with the highest proportion of most interest?
 - if so, reverse order
- The description says “The center horizontal bar correspond to the median” but that line is not present
- The gene list provided in the tooltip doesn't seem useful
 - can we come up with a different solution here? how about showing number of genes or the top 3?
 - maybe also/instead feature exact proportion value?

(e) Abundance of major gene types:

- I'd love to have a switch to flip from a stacked to a dodged bar chart as it allows to either focus on the totals per group (stacked) or on each gene group across groups (dodged)
 - depending on the main insight provided here, maybe switch to a dodged bar chart as default
- A general thing to check which is obvious here: sometimes the font color used in the tooltip seems to have very low contrast values, i.e. in my opinion the tooltip text in the example below should be white not black
 - not sure if it's possible to control that in an automated way but if it should be adjusted



Data View: Counts

BigOmics Analytics

example-data

Support ▾ Tutorials ▾ User ▾

Home ▾

Load ▾

Load dataset

Upload data

Data View ⓘ

Plots QC Counts Samples Contrasts

Gene expression table

Search:

gene	title	rho	SD	AVG	notact_004	notact_007	notact_008	notact_009	notact_010	notact_011	notact_014	act12h_018	act12h_019	act12h_020	act24h_023	act48h_026	act48h_027	act72h_028	act72h_029	act72h_030
KCNN4	potassium calcium-activated channel subfamily N member 4	1	4.375	4.532	0	0	0	0	0	0	0	6.224	5.374	5.372	9.466	9.619	9.761	9.553	7.977	9.
TEX30	testis expressed 30	0.939	1.168	1.159	0	0	0	0	0	0	0	1.232	2.237	1.231	2.015	1.901	3.033	1.817	1.809	3.
WDR75	WD repeat domain 75	0.925	0.725	6.315	5.342	5.563	5.77	5.3	5.493	5.3	5.917	6.563	6.583	6.225	7.035	7.133	7.267	7.079	7.007	7.
PSMB5	proteasome 20S subunit beta 5	0.918	2.46	4.48	2.272	1.953	1.428	2.448	1.333	1.121	1.382	5.473	4.528	4.261	6.795	7.718	7.531	6.749	6.833	6.
TRMT1OC	tRNA methylation transferase 10C, mitochondrial RNase P	0.917	0.772	6.817	6.102	5.901	6.121	5.813	5.49	6.12	6.191	7.34	7.242	6.857	7.723	7.662	8.053	7.469	7.306	7.
HSP90AB1	heat shock protein 90 alpha family class B member 1	0.904	0.848	12.73	11.878	11.98	11.698	11.906	11.577	11.407	12.126	13.171	12.284	12.645	13.514	13.791	13.764	13.773	13.	
MALS1U	mitochondrial assembly of ribosomal large subunit	0.902	1.176	1.291	0	0	0	0	0	0	0	1.725	1.762	1.767	2.482	3.332	2.76	2.027	2.533	2.
HEATR1	HEAT repeat containing 1	0.902	0.712	8.161	7.235	7.797	7.677	7.083	7.532	7.526	6.955	8.202	8.278	8.216	8.847	9.126	9.233	8.834	8.822	8.
EIF4A1	eukaryotic translation initiation factor 4A1	0.9	1.079	10.754	9.589	9.629	9.639	9.365	9.595	9.708	9.106	11.406	10.715	10.873	11.843	12.033	12.127	11.681	11.734	11.
SLC3A2	solute carrier family 3 member 2	0.899	1.44	7.92	6.041	6.399	6.151	5.931	6.61	6.434	6.086	9.096	8.013	8.631	9.673	9.646	9.466	8.977	8.989	9.
SLC1A5	solute carrier family 1 member 5	0.894	2.338	5.184	2.342	1.902	2.04	3.499	2.918	2.969	2.758	6.618	3.763	5.307	7.291	7.776	8.196	7.657	7.376	1.
YRDC	yrDC N6-threonylcarbamoyltransferase domain contain	0.894	1.794	4.033	2.105	1.593	1.896	2.463	2.473	2.188	2.218	5.903	2.763	4.622	6.356	6.321	6.712	5.445	5.271	5.
PSMD13	proteasome 26S subunit, non-ATPase 13	0.894	0.407	8.545	8.36	8.346	8.104	8.225	7.983	7.881	8.075	8.456	8.453	8.401	9.013	9.019	9.034	9.076	8.799	9.
PUN2	perilipin 2	0.893	1.79	2.793	2.565	0.726	0.286	0.636	0.715	1.533	1.486	4.466	2.248	2.21	5.161	5.83	5.591	3.543	3.891	3.
BZW2	basic leucine zipper and W2 domains 2	0.893	1.086	6.916	6.159	6.014	5.223	5.794	5.904	4.875	6.169	7.078	6.984	7.07	8.154	8.272	8.217	7.851	7.702	8.
DDX21	DEAD-box helicase 21	0.892	0.97	10.096	9.125	8.906	0.945	8.928	9.156	9.036	9.348	10.992	9.982	10.634	11.609	11.423	11.596	10.687	10.605	10.
TFB2M	transcription factor B2, mitochondrial	0.891	1.613	3.374	1.276	1.545	2.567	0.868	1.637	2.129	1.718	4.108	3.056	3.238	5.234	5.362	5.912	4.953	4.87	5.
CSEDE1	cold shock domain containing E1	0.891	0.668	8.291	7.128	7.552	7.467	7.535	7.224	7.291	7.046	8.768	8.271	8.381	9.325	9.42	9.434	9.133	9.045	9.
CCT2	chaperonin containing TCP1 subunit 2	0.891	0.415	10.815	10.509	10.713	10.234	10.53	10.245	10.12	10.482	11.045	10.708	10.539	11.265	11.343	11.197	11.257	11.127	11.
FABP5	fatty acid binding protein 5	0.89	1.742	8.869	6.667	6.663	6.752	7.016	6.6	7.391	6.733	9.965	9.054	9.136	10.242	10.849	10.492	10.576	10.421	10.
STK40	serine/threonine kinase 40	0.89	1.393	1.195	0	0	0	0	0	0	0	2.174	0	2.747	2.989	2.555	2.802	3.19	2.	
MAP1LC3B	microtubule associated protein 1 light chain 3 bet	0.889	1.809	2.395	1.262	1.241	0	0	0.442	0	0	3.254	2.167	1.991	3.464	4.519	4.159	4.674	4.522	4.
NAMPT	nicotinamide phosphoribosyltransferase	0.889	1.319	8.753	7.768	7.757	7.343	7.442	7.554	7.256	6.363	8.975	8.285	8.749	10.21	10.639	10.632	9.879	9.935	9.
RRP12	ribosomal RNA processing 12 homolog	0.889	0.805	7.87	7.207	7.063	6.426	6.99	7.367	7.029	7.009	8.533	8.1	8.518	8.985	8.896	9.126	8.221	8.198	8.
MTHFD2	methylenetetrahydrofolate dehydrogenase (NADP+ dep)	0.885	3.108	4.909	0.97	1.507	1.246	0.603	1.45	1.305	2.732	6.469	3.773	5.402	7.275	8.41	8.104	8.124	8.003	8.
RBM19	RNA binding motif protein 19	0.885	0.82	4.915	4.293	4.122	3.932	3.691	3.39	4.241	5.01	5.597	4.174	5.062	6.137	5.774	5.899	5.453	5.651	5.
LRRC14	leucine rich repeat containing 14	0.885	0.572	0.63	0	0	0	0	0	0	0	0.838	0.91	0.909	0.603	0.84	1.239	1.218	1.312	1.
IPO4	importin 4	0.884	1.431	7.349	6.103	5.453	5.466	5.349	6.431	6.3	5.623	7.893	6.597	7.249	8.342	9.22	9.215	8.748	8.827	8.
COA7	cytochrome c oxidase assembly factor 7 (putative)	0.883	1.494	1.734	0	0	0	0	0	0	0	2.337	2.064	1.996	2.536	3.38	3.43	3.116	3.42	3.
TRAF1	TNF receptor associated factor 1	0.882	1.467	6.065	3.819	3.86	3.86	4.538	4.844	4.959	4.868	7.79	6.343	7.111	7.952	7.811	7.614	7.24	6.87	1.
ZCHC3	zinc finger CCHC-type containing 3	0.882	0.928	1.08	0	0	1.055	0	0	0	0	1.975	0.76	1.009	2.266	2.576	2.317	1.738	1.12	2.
4	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	

Showing 1 to 33 of 6,987 entries

Gene table: The table shows the gene expression values per sample, or average expression values across the groups. The column 'rho' reports the correlation with the gene selected in 'Search gene' in the left side bar.

Gene table:

- Number of digits should be consistent so that numbers align nicely
 - trailing zeros can be achieved by using `sprintf("%1.3f", x)`
 - decide how to work with zero values: add trailing zeros? show at all? maybe color-encode in case they are of special interest?

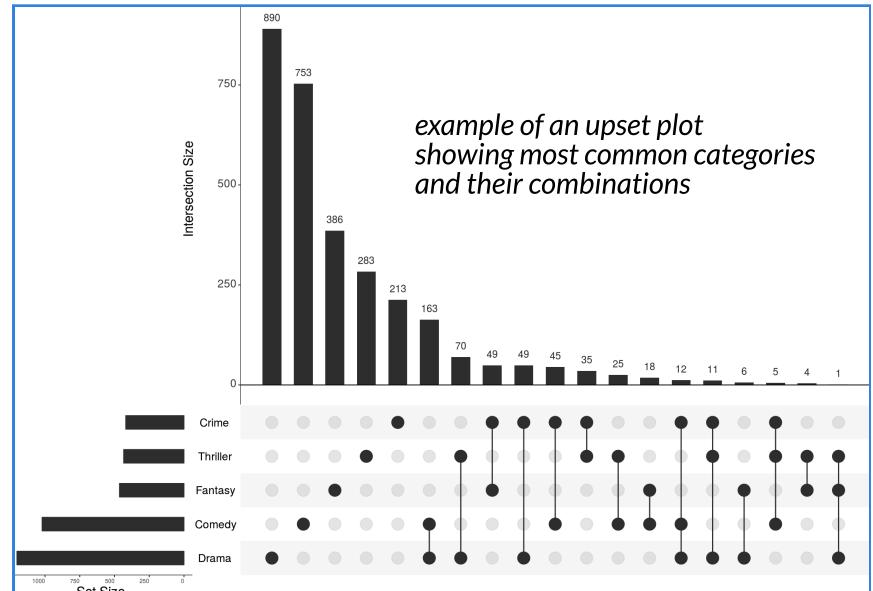
gene	↑↓ title	↑↓ rho	↑↓ SD	↑↓ AVG	↑↓ notact_004	↑↓ notact_007	↑↓ notact_008
KCNN4	potassium calcium-activated channel subfamily N member 4	1	4.375	4.532	0	how to deal with zeros?	0
TEX30	testis expressed 30	0.939	1.168	1.159	0	0	0
WDR75	WD repeat domain 75	0.925	0.725	6.315	5.342	5.563	5.77
PSMB5	proteasome 20S subunit beta 5	0.918	2.46	4.48	2.272	1.953	1.428
TRMT10C	tRNA methyltransferase 10C, mitochondrial RNase P	0.917	0.772	6.817	6.102	5.901	6.121
HSP90AB1	heat shock protein 90 alpha family class B member	0.904	0.848	12.73	11.878	11.98	11.698
MALSU1	mitochondrial assembly of ribosomal large subunit	0.902	1.176	1.291	0	0	0
HEATR1	HEAT repeat containing 1	0.902	0.712	8.161	7.235	7.797	7.677
EIF4A1	eukaryotic translation initiation factor 4A1	number of digits should be consistent	0.9	1.079	10.754	9.589	9.629
SLC3A2	solute carrier family 3 member 2		1.44	7.92	6.041	6.399	6.151
SLC1A5	solute carrier family 1 member 5	0.894	2.338	5.184	2.342	1.902	2.04

Data View: Samples

The screenshot shows the BigOmics Analytics Data View interface. The top navigation bar includes 'example-data' and 'Support ▾ Tutorials ▾ User ▾'. The left sidebar has sections for Home, Load, DataView (selected), Clustering, Expression, Enrichment, Signature, and Cell profiling. The main area displays three panels: 'Phenotype clustering' (a heatmap of sample information), 'Phenotype association' (a correlation matrix heatmap), and 'Sample information' (a table of phenotype details). A legend on the right lists various phenotype variables with their corresponding color codes.

(a) Phenotype clustering:

- Uses some other library for interactive heatmaps (but the chart is not interactive)
 - should be reimplemented in plotly as a truly interactive chart
- Color palettes should be unique; the same categorical color palettes (grey scale palette and multi-hue categorical palette) are used multiple times, should be avoided
 - becomes very colorful if many groups though
 - maybe a different plot type would be better suited for the expected insights? how about an upset plot?
- The text is not set in Lato, please update



(b) Phenotype association:

- Uses some other library to generate a correlation plot, not interactive
 - should be reimplemented in plotly as a truly interactive chart
- Is it possible to have/fit horizontal text for the upper labels?
- Legend text should be bigger
- The text is not set in Lato, please update

(c) Sample information table:

- Basic table, no recommendation except potential use of color for encoding cell entries

Data View: Contrasts

BigOmics Analytics example-data

Support ▾ Tutorials ▾ User ▾ Data set name selected Settings >

Home ▾ Load ▾ Load dataset Upload data DataView ▾ Clustering ▾ Expression ▾ Enrichment ▾ Signature ▾ Cell profiling

Data View ⓘ

Plots QC Counts Samples Contrasts

Contrast table

Search:

	act vs notact <small>T1</small>	act12h vs notact <small>T1</small>	act24h vs notact <small>T1</small>	act48h vs notact <small>T1</small>	act72h vs notact <small>T1</small>	act96h vs notact <small>T1</small>
notact_004	-1	-1	-1	-1	-1	-1
notact_007	-1	-1	-1	-1	-1	-1
notact_008	-1	-1	-1	-1	-1	-1
notact_009	-1	-1	-1	-1	-1	-1
notact_010	-1	-1	-1	-1	-1	-1
notact_011	-1	-1	-1	-1	-1	-1
notact_014	-1	-1	-1	-1	-1	-1
act12h_018	1	1				
act12h_019	1	1				
act12h_020	1	1				
act24h_023	1		1			
act48h_026	1			1		
act48h_027	1			1		
act72h_028	1				1	
act72h_029	1				1	
act72h_030	1				1	
act96h_032	1					1
act96h_034	1					1

Showing 1 to 18 of 18 entries

Contrast table. summarizing the contrasts of all comparisons. Non-zero entries '+1' and '-1' correspond to the group of interest and control group, respectively. Zero or empty entries denote samples not use for that comparison.



DataView

Contrast table:

- Uses corporate colors with the alarming color for negative contrasts
- Currently the table uses bars to color cells; as there are only two values, +1 and -1 (plus zero), one could color the cells directly
 - if cells are colored, all numbers and color encoding fits nicely in one narrow column with labels consistently being placed either inside the color or next to it (maybe the numbers aren't even needed here?)
 - if text is placed inside color chunks, it should be of white color on blue background (i.e. stating "1", if needed at all)

+1	+1
-1	+1
+1	+1
-1	-1
-1	+1

+1	+1
-1	+1
+1	+1
-1	-1
-1	+1

examples of how table could be encoded with filled cells or aligned "bars" next to the text

General notes:

- Overall, this tab is rather empty (at least using the example data set)
 - maybe it can be joined with another tab? if not, maybe make it bigger covering more of the tab area?

Board “Clustering”

Clustering Samples: Heatmap + Annotate clusters

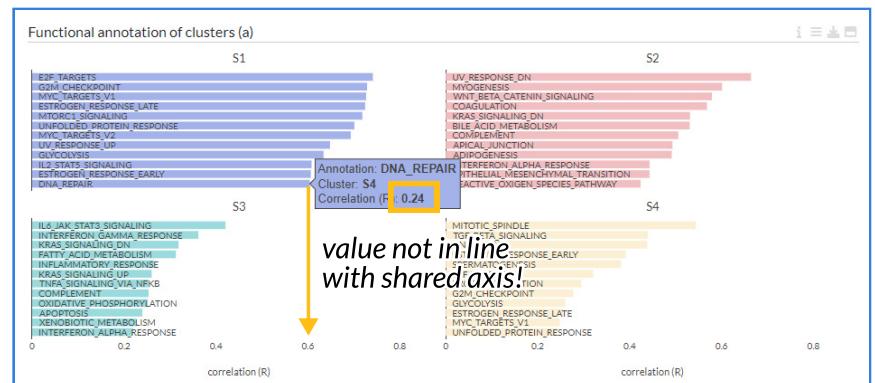


Clustered heatmap:

- Uses some other library for interactive heatmaps (but the chart is not interactive)
 - should be reimplemented in plotly as a truly interactive chart
- As in the “Phenotype clustering” heatmap, unique categorical palettes should be used for different categories
- The currently used (default ?) sequential palette is pointing in the wrong direction with red encoding “good” which is very unintuitive as red usually represents bad outcomes
- The diverging palette should be replaced by the blue-orange one (with orange indicating negative values)
- The text is not set in Lato, please update

(a) Functional annotation of clusters:

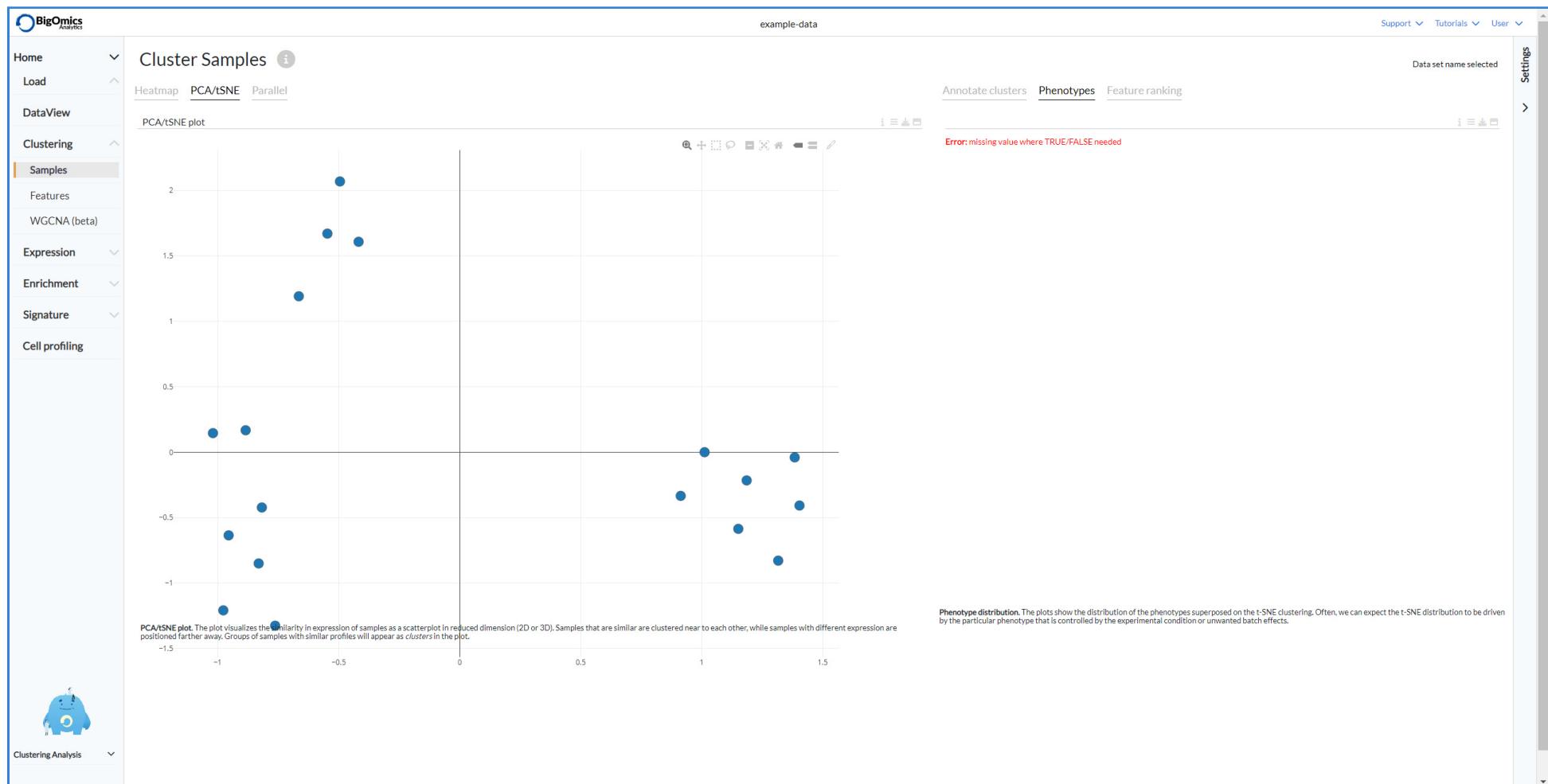
- The cluster ID in the tooltip is assigned wrongly (it's always S4) even though I refer to `colnames(rho[i])` as in all other charts - no clue what's wrong here
 - needs to be fixed (or that information to be removed)
- Uses the new colors; not overall happy with the outcome as either the bars are hardly visible or the text is hardly readable; however, are the colors needed at all?
 - I suggest to not use a categorical palette for the different clusters; the panels alone highlight the different groups and a single color would allow for a fair comparison (in terms of visual weight), solve all readability problems and would make the page much more calm
 - if you wish to use colors, maybe darker, intense colors in combination with white text are a better solution? (however, it would also require a solution for overlapping text)
 - if you use colors, use the same colors to encode the Sx functional clusters in any chart
- while in the lower row x is scaled from 0 to .9, in the upper it's ranging free (kinda; when you plot the axis, the axis range is the same but the tooltip and axis are out of sync; illustrated in previous message and screenshot)
 - **check x axis ranges!**



(b) Annotation scores:

- cluster colors (if any) could be re-used in this table

Clustering Samples: PCA/tSNE + Phenotypes



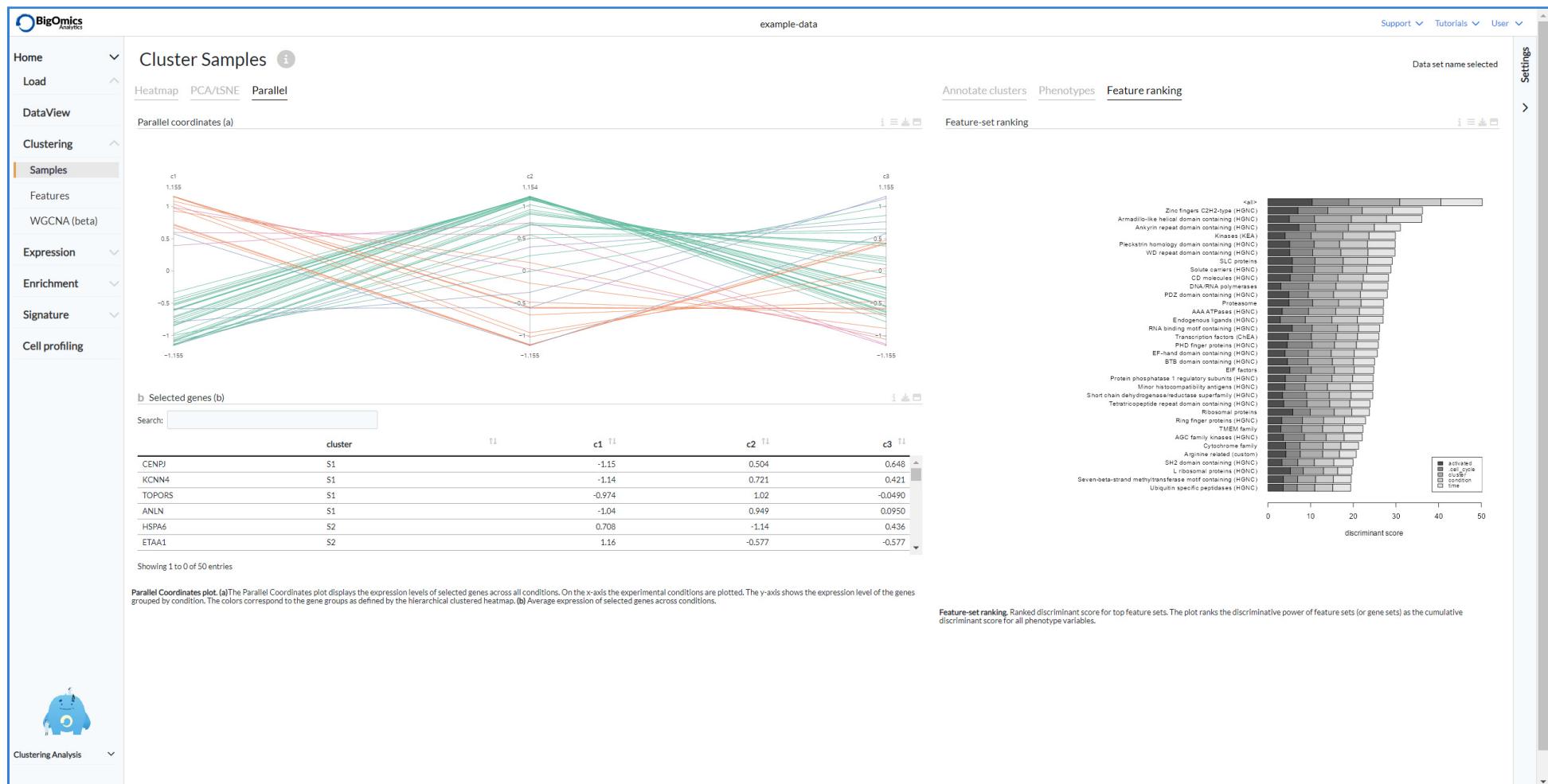
PCA/tSNE:

- Use same style as for the “tSNE clustering” (d) in “Data View” with outlines around points
- Needs to use corporate blue (or a sequential palette to encode value ranges)
- Caption is overlapping with the plot in the current version
- Again, it would be nice to have a consistent number of digits on the axes, especially on the y axis as it would allow for a clean alignment (but not possible in plotly)
- Plot doesn't change when grouping is applied
 - guess it should color the points accordingly or add some annotation for the different groups
- Style tooltips similar to polished plotly charts in the previous tabs
- The text is not set in Lato, please update

Phenotype distribution:

- Throws an error in the current version

Clustering Samples: PCA/tSNE + Phenotypes



(a) Parallel Coordinates plot:

- Use the color palette that is used for the gene groups in the hierarchical clustered heatmap (use same colors for these across the app)
- Again, would be nice to have a decent number format on the axes
- A bit more highlight on the cluster labels on the top would be good (bigger font size + bold?)
- The chart does not feature any tooltips (on purpose?)
- The text is not set in Lato, please update

(b) Average expression:

- Basic table; again color could be reused here to encode cx cluster columns

Feature-set ranking:

- Base plot, needs to be re-implemented in plotly
- Grey scale palette not recommended here; use a categorical color palette (in best case the same across all charts)
- Maybe a switch from stacked to dodged bars is useful here?
- The text is not set in Lato, please update

Clustering Features: Gene

BigOmics Analytics

example-data

Support ▾ Tutorials ▾ User ▾

Data set name selected

Settings

Home ▾

Load

DataView

Clustering

Samples

Features

WGCNA (beta)

Expression

Enrichment

Signature

Cell profiling

Cluster features ⓘ

Gene Geneset

(a) Gene Map

sd.X

UMAP-y

UMAP-x

1
0.22

(b) Gene Signatures

act

UMAP-y

UMAP-x

-2.12
2.12

notact

UMAP-y

UMAP-x

-4.000
4.645

(c) Gene table

Search:

gene	T1	title	T1	sd.X T1	act T1	notact T1
ETAA1		ETAA1 activator of ATR kinase		3.893	-2.956	4.645
KCNN4		potassium calcium-activated channel subfamily N member 4		3.799	2.884	-4.532
POTEE		POTE ankyrin domain family member E		3.707	-2.815	4.423
CENPJ		centromere protein J		3.681	2.795	-4.392
ANLN		anillin actin binding protein		3.438	2.61	-4.102
CTNNBIP1		catenin beta interacting protein 1		3.418	-2.595	4.078
KPN2A		kinase/nuclear ribonucleoprotein subunit alpha 2		3.354	2.544	-4.000

Showing 1 to 0 of 6,987 entries

(a) Gene Map:

- Actually, a diverging palette might be a bad choice in general as the midpoint is not meaningful
 - what is the intended insight here? identifying points below and above the mean score?
 - If you stick to a diverging palette:
 - again, the palette should be reversed (red == bad)
 - consider using a midpoint color that is a bit darker; alternatively point outlines could solve the issue
 - The labeling needs to be adjusted as some annotations are placed outside the panel
 - Maybe tooltips are good enough and no labeling is needed at all?
 - The text is not set in Lato, please update

(b) Gene Signatures:

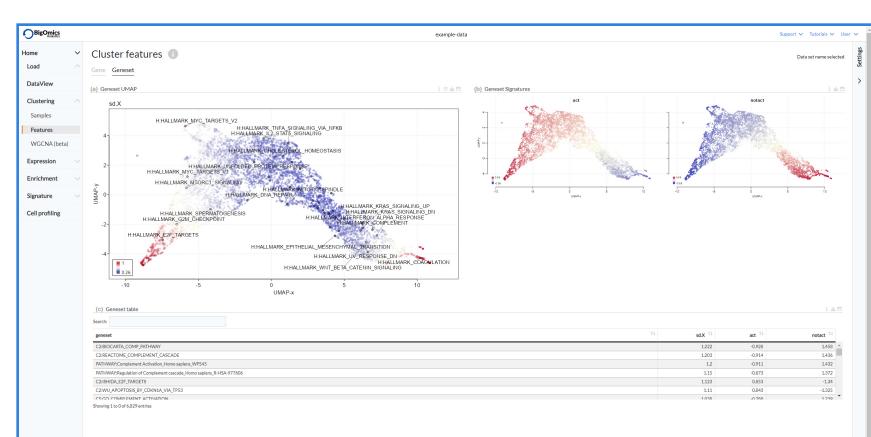
- In general same recommendations as in (a)
 - Here, the values actually range from negative to positive; make sure that the zero baseline is the midpoint (currently also the case)
 - Is the variable encoded by color the same as in (a)? If not, use a different palette here!
 - The text is not set in Lato, please update

(c) Gene table:

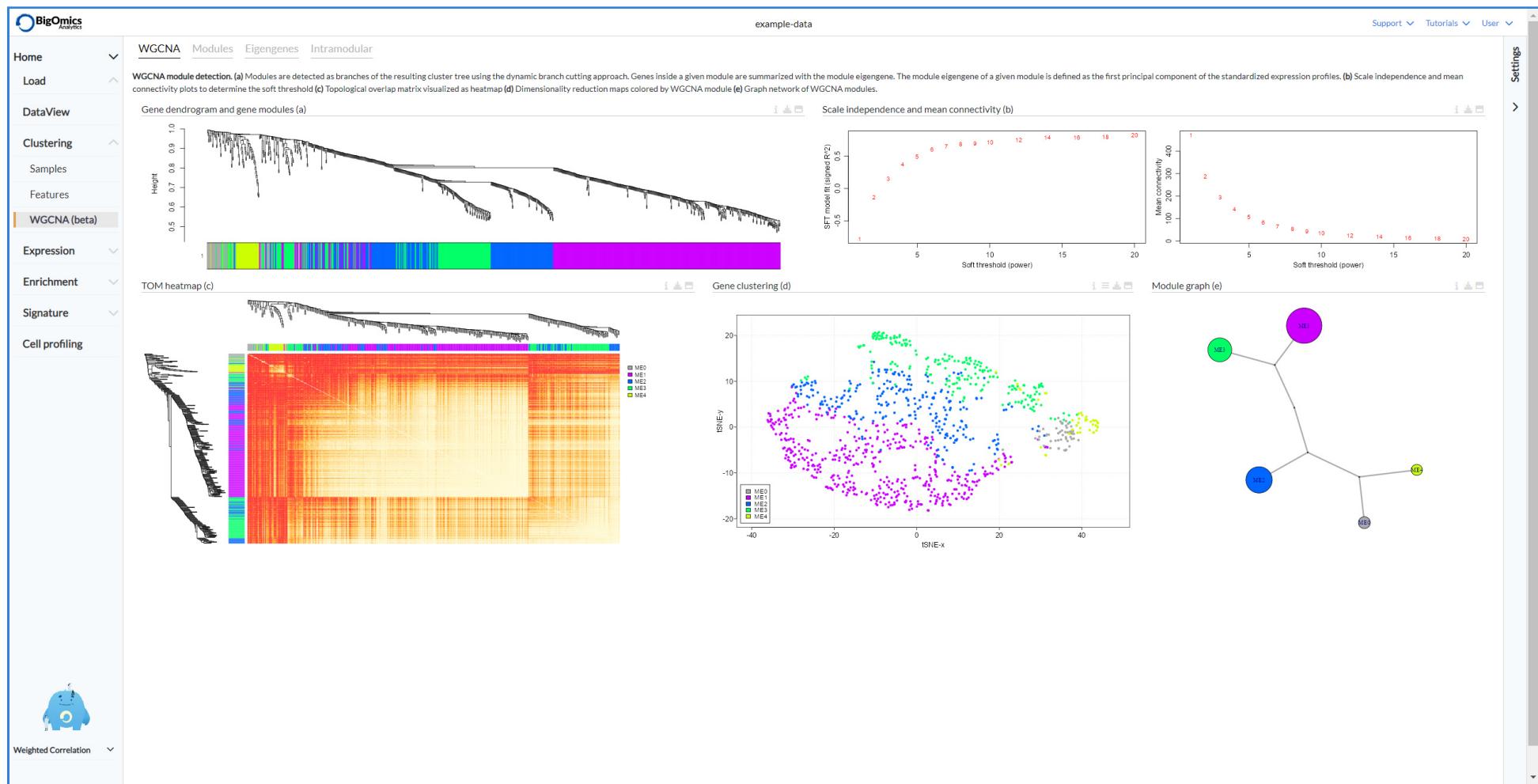
- Basic table, no further recommendations
 - Just noticed that the table styles differ between tabs – would be nice to have a consistent style for tables as well!
 - As a side note: also make sure to consistently use title or upper case for the headers
 - e.g. “Gene Map” versus “Gene table” (likely the same style should be applied to all kind of headers across the app)

Gene Set tab:

- Same recommendations as plots and tables are basically the same as in the “Gene” tab discussed above



Clustering WGCNA: WGCNA



General notes:

- Colors should be based on corporate color palettes; the colors here feel very bright and neon
- The colors are the same across all plots, maybe a general, overarching legend makes more sense than the same legend again and again?
- Charts need to be converted to plotly (if possible at all?)
- A bit larger spacing between upper and lower rows of charts would be good
- General layout of tab differs from all others: no tab title, caption at the top, plot tag after chart title
- The text used in the plots is not set in Lato, please update

(b) Scale independence and mean connectivity:

- It is unclear what the numbers encode
- Why are the numbers colored red? Does it somehow relate to the big heatmap in (c)?
- Does it make sense to connect the points (numbers) by lines?

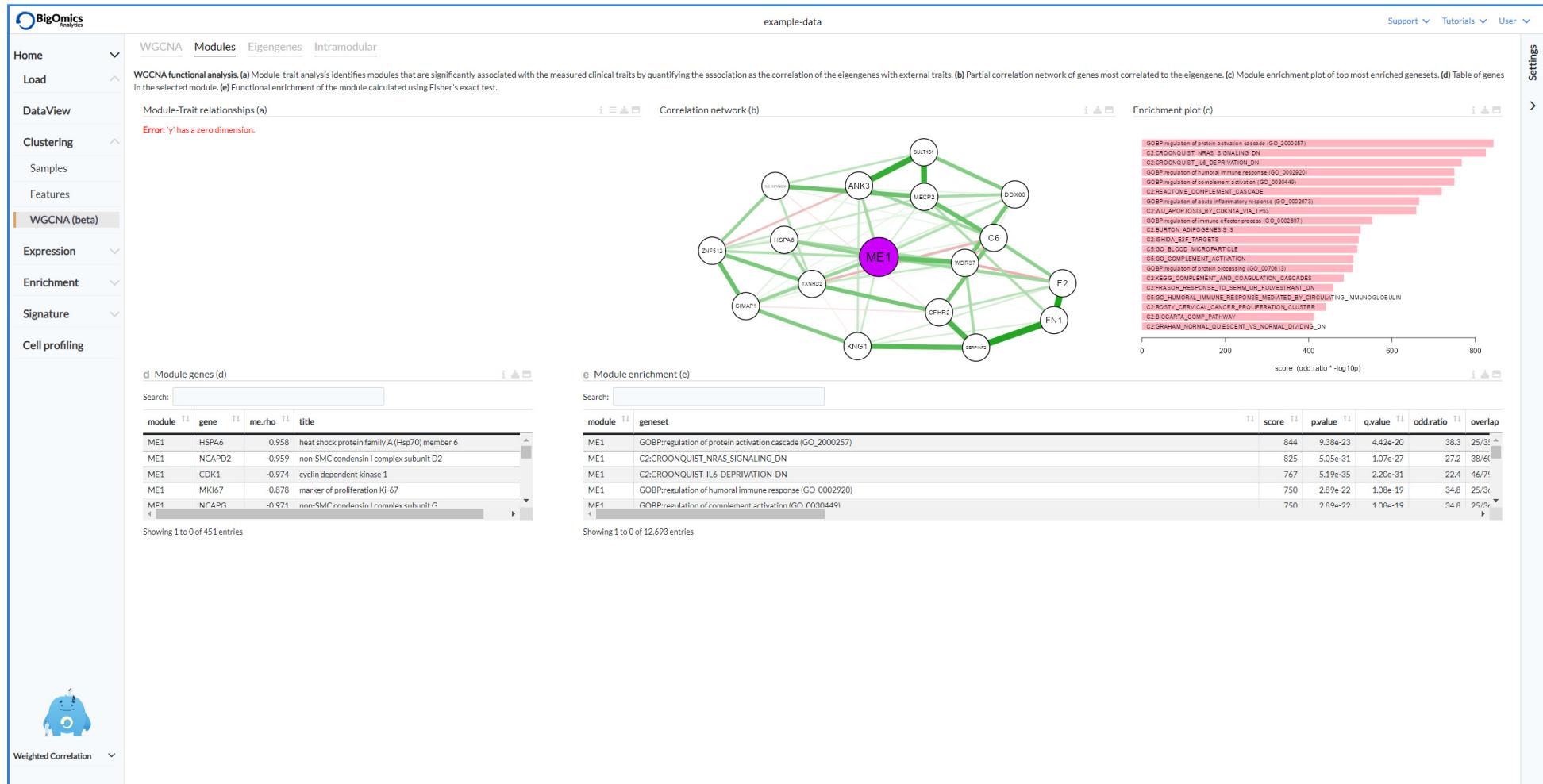
(c) TOM heatmap:

- Depending on what is shown and the desired input, a diverging color palette might make sense (i.e. if the focus is on the extreme outcomes not only one end of the value range)
- It is unclear which color is encoding which end of the values
 - either add legend or at least a note to the caption

(e) Module graph:

- Labels are hardly readable and actually not needed as it uses the same color encodings as before
 - if you wish to use labels, consider to place them outside the nodes

Clustering WGCNA: Modules



General notes:

- General layout of tab differs from all others: no tab title, caption at the top, plot tag after chart title
- Charts are not interactive and need to be implemented in plotly

(a) Module-trait analysis:

- Throws an error in the current version.

(b) Correlation network:

- It is unclear which colors encode positive and negative correlation
 - given the bad red-green color range I assume red represents negative correlation?
 - in any case: don't use a color palette that is known to be problematic for people suffering from color-vision deficiencies
 - instead, use one of the corporate diverging palettes and make sure that colors associated with "bad" are used for the lower end (e.g. orange for negative and blue for positive)
- The text is not set in Lato, please update

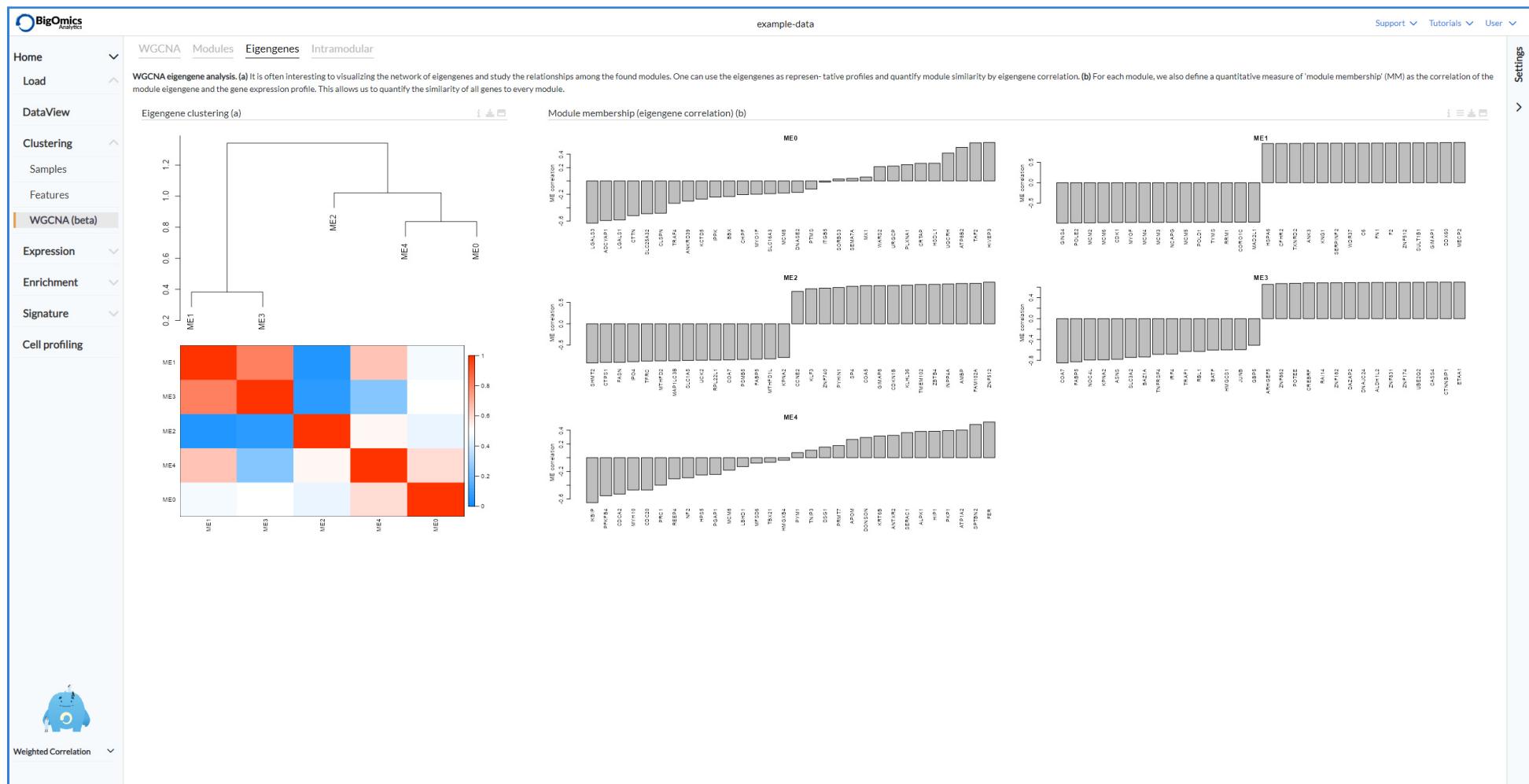
(c) Enrichment plot:

- Why are the bars colored? Is it somehow connected to the red colors in plot (b)?
 - I don't assume so but it makes it feel like
- Could one clean and standardize (and shorten?) the labels
- The text is not set in Lato, please update

(d) + (e) Module genes + enrichment:

- basic tables, no further recommendations besides using a consistent table style across the app

Clustering WGCNA: Eigengenes



General notes:

- General layout of tab differs from all others: no tab title, caption at the top, plot tag after chart title
- Charts are not interactive and need to be implemented in plotly

(a) Eigengene clustering:

- As before, the currently used (default ?) sequential palette is pointing in the wrong direction with red encoding “good” which is very unintuitive as red usually represents bad outcomes

(b) Module membership:

- A zero-baseline plus some color encoding might help to see general positive/negative patterns
- The different axes ranges seem a bit weird and potentially misleading
 - are the exact numbers relevant? is it important to compare values across groups?
 - if yes: ensure same axes ranges across panels
 - if not: maybe use no axes labels or standardize them within ME clusters?
- If the intended insight is about “to quantify the similarity of all genes to every module”, I’d expect/prefer genes to be sorted in the same order across panels? How do I compare genes otherwise without reading out all rotated labels?

Clustering WGCNA: Intramodular

BigOmics Analytics

example-data

Support ▾ Tutorials ▾ User ▾

Settings >

Home ▾

Load ▾

DataView ▾

Clustering ▾

Samples

Features

WGCNA (beta)

Expression ▾

Enrichment ▾

Signature ▾

Cell profiling

Weighted Correlation ▾

WGCNA Modules Eigengenes Intramodular

WGNCNA intramodular analysis. We quantify associations of individual genes with our trait of interest (weight) by defining Gene Significance GS as (the absolute value of) the correlation between the gene and the trait. For each module, we also define a quantitative measure of module membership MM as the correlation of the module eigengene and the gene expression profile. Using the GS and MM measures, we can identify genes that have a high significance for weight as well as high module membership in interesting modules.

Membership-trait heatmap (a) Membership vs. trait correlation (b)

Error: X has a zero dimension.

Error: X has a zero dimension.

General notes:

- General layout of tab differs from all others: no tab title, caption at the top, plot tag after chart title

(a) Membership-trait heatmap + Membership vs. trait correlation:

- Both plots are not working in the current version

Board “Expression”

Differential Expression: Plot + Table

The screenshot shows the BigOmics Analytics web application interface. The left sidebar has a tree view with categories like Home, Load, DataView, Clustering, Expression (selected), Differential expression, Correlation analysis, Enrichment, Signature, and Cell profiling. A small blue cartoon character icon is in the bottom-left corner.

The main area has tabs at the top: Plot, Top genes, Volcano (all), and Volcano (methods). Below are four panels: (A) Volcano plot, (B) MA plot, (C) Differential expression, and (D) Gene in contrasts. Panel (A) shows a scatter plot of significance (log10) vs effect size (log2:FC) with points for ETAA1. Panel (B) shows a scatter plot of effect size (log2:FC) vs average expression (log2:CPM) with points for ETAA1. Panel (C) is a barplot of top differentially expressed genes. Panel (D) is a barplot of gene sets. Below the panels is a table of differential expression analysis results and another table of gene sets.

I Differential expression analysis (I)

gene_name	TL	gene_title	TL	chr	TL	logFC	TL	stars	TL	meta.q	TL	AveExpr0	TL	AveExpr1	TL
ETAA1		ETAA1 activator of ATR kinase		2		-7.601		***		7.054e-9		7.601		0.000	
KCNN4		potassium calcium-activated channel subfamily N member 4		19		7.493		***		0.0003096		-0.03814		7.455	
POTEE		POTE ankyrin domain family member E		2		-7.239		***		1.177e-9		8.588		1.349	
CENPJ		centromere protein J		13		7.159		***		0.1922		2.596		9.756	
ANLN		anillin actin binding protein		7		6.820		***		0.00004724		1.653		8.473	
CTNNBIP1		catenin beta interacting protein 1		1		-6.673		***		5.576e-9		6.673		0.000	
ANLN		anillin actin binding protein		17		4.674		***		7.479e-8		1.017		0.200	

Showing 1 to 16 of 4,000 entries

II Gene sets with gene (II)

geneset	TL	rho	TL	fx	TL
C2:ZHANG_TLX_TARGETS_60HR_DN		-0.696		0.8647	
DRUG:vitamin c DB00126 mouse GSE32994 sample 3510 (up)		-0.604		0.5144	
GOC:replication fork (GO_0005657)		-0.636		0.4992	
C2:GOBERT_OLIGODENDROCYTE_DIFFERENTIATION_UP		-0.605		0.4823	
GENE:BST2 OE human GSE50002 sample 1651 (down)		-0.619		0.4236	
GOM:protein kinase activator activity (GO_0030295)		0.959		-0.4166	
GO:translational_stress_response_vs_TREATED_ACT_TREATED		0.740		0.3000	

Showing 1 to 16 of 38 entries

General notes:

- Panel headers differ with the tag being grey, bold and in upper case (i.e. (A) and not (a)) plus the capital roman tags for tables (i.e. (I) and (II)) which haven't been used before
- The header "Differential expression" in the sidebar is wrapped

(a) Volcano plot + (b) MA plot:

- Tooltips and axis titles need to be styled
- It would be nice if the selected point when hovering would be colored
- Why is only one gene annotated?

(c) Differential expression + (d) Gene in contrasts:

- Plots load forever, I never managed to see them before the app was timed-out

(I) Differential expression analysis:

- The tag is shown twice in the header (i.e. "(I) Differential expression analysis (I)"), remove one of them
 - also, decide on a consistent nomenclature for those tags; either use roman numbers for tables or the general (a), (b), (c) annotation style
 - "Gene sets with gene" sounds strange, maybe reword
- Numbers in the chr column should be right-aligned
- The values in the logFC column could be placed next to the bars, it looks a bit strange that they are always placed inside the positive range
- The diverging color palette needs to be updated; again, make sure that the alarming color is used for negative values
- Consider to reword the column titles to be cleaner, less "code-looking" and more descriptive
- A consistent formatting of the numeric values would be nice to have
 - I can see how this may become problematic for large values, e.g. in the meta.q column, but for many others it should work? I.e. I would suggest to always use scientific notation in case the numbers do not fit the x-digit rule (and x digits otherwise)

The table from section (I) is shown again with several annotations:

- Yellow boxes highlight the first two rows (ETAA1 and KCNN4).
- A callout bubble says "shorten to 3 digits (or change the other numbers to show 4)" pointing to the meta.q column.
- A callout bubble says "use scientific notation if more than 4 digits are needed" pointing to the AveExpr0 and AveExpr1 columns.
- A callout bubble says "align numbers to the right" pointing to the logFC column.

Differential Expression: Plot + Table (continued)

(II) Gene sets with genes:

- The tag is shown twice in the header (i.e. "(II) Gene sets with genes (II)"), remove one of them
 - also, decide on a consistent nomenclature for those tags; either use roman numbers for tables or the general (a), (b), (c) annotation style
 - "Gene sets with gene" sounds strange, maybe reword?
- The number placement inside the fx column looks strange with not being truly centered but going into both areas, positive and negative
 - maybe place next to the bars?

Differential Expression: Top genes + Foldchange (all)

BigOmics Analytics example-data Support ▾ Tutorials ▾ User ▾ Settings >

Home ▾ Load DataView Clustering Expression Differential expression Correlation analysis Enrichment Signature Cell profiling

Plot Top genes Volcano (all) Volcano (methods)

(A) Expression of top differentially expressed genes

Top differentially expressed genes. Expression barplots of the top most differentially (both positively and negatively) expressed genes for the selected contrast.

Table Foldchange (all) FDR table Gene fold changes for all contrasts Differential expression (fold-change) across all contrasts. The column 'rms.FC' corresponds to the root-mean-square fold-change across all contrasts.

Search:

gene	rms.FC ↑↑	FC.act vs notact ↑↑	q.act vs notact ↑↑	FC.act12h vs notact ↑↑	q.act12h vs notact ↑↑	FC.act24h vs notact ↑↑	q.act24h vs notact ↑↑	FC.act48h vs notact ↑↑	q.act48h vs notact ↑↑	FC.act72h vs notact ↑↑	q.act72h vs notact ↑↑	FC.act96h vs notact ↑↑	q.act96h vs notact ↑↑
KCNN4	7.86	7.46	0.000310	5.66	0.0147	9.47	0.00216	9.69	0.000371	8.94	0.000422	4.48	0.0126
ETAA1	7.60	-7.60	7.05e-9	-7.60	0.0000203	-7.60	0.0000641	-7.60	8.76e-7	-7.60	1.71e-7	-7.60	0.00000153
POTEE	7.26	-7.24	1.18e-9	-7.36	7.08e-7	-7.19	0.000149	-8.17	4.64e-7	-7.05	9.96e-8	-6.44	0.0000261
ANLN	7.20	6.78	0.0000472	3.64	0.141	7.27	0.0471	9.34	0.0000211	7.11	0.000312	7.82	0.000294
CENPJ	7.13	7.17	0.192	7.81	0.362	6.86	0.733	6.56	0.592	7.05	0.408	7.25	0.483
CTNNBIP1	6.67	-6.67	5.58e-9	-6.67	0.0000165	-6.67	0.000104	-6.67	6.84e-7	-6.67	1.31e-7	-6.67	0.00000126
KPNA2	6.66	6.56	7.67e-10	5.41	2.38e-7	6.53	0.00000765	8.07	4.17e-9	6.96	2.15e-9	6.13	1.25e-7
CCNE2	6.57	-6.50	0.00000303	-5.72	0.00236	-6.78	0.0150	-6.78	0.000486	-6.78	0.0000916	-6.78	0.000816

Showing 1 to 0 of 4,080 entries

(a) Expression of top differentially expressed genes:

- The plot loads forever, I never managed to make it appear before the app was timed-out

Gene fold changes for all contrasts:

- The header doesn't feature a tag (while the only plot in this tab does); as there is no tag the panel header has a weird spacing in front, thus it's misaligned with rest of the tab
- Strange placement of the caption below the second header but above the table; not consistent with previous boards and the previous "Table" tab
- Style recommendations similar to the tables before:
 - consistent styling of numbers
 - corporate diverging color palette with alarming color encoding negative values
 - if possible better styled column names
- With the example data set being loaded, the last column shows only one tiny bar which looks a bit strange
 - The custom maximum for the bars seems to be off, the bar should cover the full range in case it's 1 but it doesn't



for some columns the range for the bars seems to be off

Differential Expression: Volcano (all) + FDR table

The screenshot shows the BigOmics Analytics web application interface. The left sidebar has a tree view of features: Home, Load, DataView, Clustering, Expression, Differential expression (which is selected), Correlation analysis, Enrichment, Signature, and Cell profiling. The main area has tabs: Plot, Top genes, Volcano (all) (selected), and Volcano (methods). Below these are six volcano plots for contrasts: act12h_vs_notact, act12h_vs_nodat, act24h_vs_notact, act24h_vs_nodat, ad48h_vs_notact, and ad72h_vs_notact. A caption below the plots says: "Volcano plot for all contrasts. Simultaneous visualisation of volcano plots of genes for all contrasts. Experimental contrasts with better statistical significance will show volcano plots with 'higher' wings." Below the plots is a table header: "Number of significant genes" and "Number of significant genes versus FDR". The table body contains data for various methods (deseq2.wald, edger.qif, trend.limma, deseq2wald, act12h_vs_notact, edger.qif, trend.limma, deseq2wald, act24h_vs_notact, edger.qif) across different contrasts and FDR thresholds (1e-16 to 1). The right side of the table shows counts for downregulated genes (blue) and overexpressed genes (red). A search bar and a note "Showing 1 to 0 of 18 entries" are also present.

Volcano plots for all contrasts:

- No tag in the panel title (which might be fine but isn't in line with the previous tabs "Plot" and "Top genes"), panel header consequently misaligned with rest of the tab
- Not interactive (likely due to performance issues?)
- All text labels are way too small
- Axis titles are redundant
 - can it be turned into a true grid of plots with shared axes and axis titles?
- Annotations in the plot are overlapping
- Strange empty row below the row of volcano plots
 - can the grid placement be automated in a more clever way?
- What is the main insight from these plots?
 - maybe one can adjust the styling to the main purpose, e.g. (made-up example as I don't know what these plots mean) in case the difference in "lava height" is important, report this maximum and/or indicate the difference to the highest volcano, either with text or maybe even using a sequential palette (i.e. the closer to the global maximum the darker); another option could be a reference line across all panels

Number of significant genes:

- Strange placement of the caption below the second header but above the table; not consistent with previous boards and the "Table" tab
- It's strange that the same colors as before are used to now indicate different types of FDR and now all point to the left (negative) side while being colored differently
 - use a different palette here!
 - instead of "just" indicating the upper and lower groups, one could use two sequential palettes, one for the upper and one for the lower, to actually encode the counts
- Why are the bars pointing towards the left while values are positive?
 - I would recommend to make them point to the right as it is common for bar graphs and also more intuitive in my opinion (if there is not a valid reason why they point to the left)
- Maybe use a big mark to make large numbers more readable and to follow the American formatting rules? (I just noticed it here; no matter if you decide to use or omit big marks, make sure numbers are styled consistently across the app)

Differential Expression: Volcano (methods)

BigOmics Analytics example-data Support Tutorials User Settings >

Home ▾ Load DataView Clustering Expression Differential expression Correlation analysis Enrichment Signature Cell profiling

Plot Top genes Volcano (all) Volcano (methods)

Volcano plots for all methods

Volcano plot for all statistical methods. Simultaneous visualisation of volcano plots of genes by multiple differential expression methods for the selected contrast. Methods showing better statistical significance will show volcano plots with higher wings.

Table Foldchange (all) FDR table

Number of significant genes

Number of significant genes versus FDR. This table reports the number of significant genes at different FDR thresholds for all contrasts and methods. This enables to quickly see which methods are more sensitive. The left part of the table (in blue) correspond to the number of significant down-regulated genes, the right part (in red) correspond to the number of significant overexpressed genes.

Search:

method	contrast	DOWN FDR = 1e-16	1e-08	1e-06	1e-04	0.01	0.05	0.1	0.2	0.5	1	UP FDR = 1e-16	1e-08	1e-06	1e-04	0.01	0.05	0.1	0.2	0.5	1
deseq2.wald	act_vs_notact	128	402	570	866	1199	1272	1294	1310	1325	1328	124	290	376	524	745	855	908	963	1046	1095
edger.qif	act_vs_notact	0	23	123	495	1144	1292	1317	1325	1326	1328	0	25	119	336	659	792	836	866	912	1095
trend.limma	act_vs_notact	0	35	207	630	1221	1341	1382	1394	1400	1400	0	35	161	459	887	1024	1053	1064	1069	1069
deseq2.wald	act12h_vs_notact	2	9	16	54	236	400	493	591	702	747	9	36	56	120	265	387	446	529	694	806
edger.qif	act12h_vs_notact	0	0	1	5	130	323	436	606	725	747	0	0	6	32	222	430	524	623	756	806
trend.limma	act12h_vs_notact	0	4	10	53	309	493	587	669	746	747	0	3	16	66	332	522	619	718	805	906
deseq2.wald	act24h_vs_notact	0	8	23	50	213	462	753	1143	1575	1650	11	53	88	150	326	465	577	704	947	1128
edger.qif	act24h_vs_notact	0	0	0	7	113	398	678	1103	1605	1650	0	0	15	77	349	540	664	821	1030	1128

Showing 1 to 0 of 18 entries

Expression Analysis

Volcano plots for all methods:

- No tag in the panel title (which might be fine but isn't in line with the previous tabs "Plot" and "Top genes"), panel header consequently misaligned with rest of the tab
- Not interactive (likely due to performance issues?)
- All text labels are way too small
- Yet another style of these volcano plots with no redundant axes but also no spacing and no axis titles at all
 - add some spacing between the single panels if possible
 - add axis titles
- Annotations in the plot are overlapping and partly crossing the panel border
- Strange empty row below the row of volcano plots
 - can the grid placement be automated in a more clever way?
- What is the main insight from these plots?
 - maybe one can adjust the styling to the main purpose, e.g. (made-up example as I don't know what these plots mean) in case the difference in "lava height" is important, report this maximum and/or indicate the difference to the highest volcano, either with text or maybe even using a sequential palette (i.e. the closer to the global maximum the darker); another option could be a reference line across all panels

Correlation Analysis: Correlation

The screenshot shows the BigOmics Analytics interface with the 'Correlation analysis' tab selected. The main panel displays three sections: (a) Top correlated genes, (b) Correlation table, and (c) Correlation scatter plots.

(a) Top correlated genes: A bar chart showing correlation values for various genes. The y-axis ranges from -1.0 to 1.0. A legend indicates that grey bars represent correlation and black bars represent partial correlation. Gene names are listed along the x-axis, including KCNN4, WDPR75, PSMB5, TEX30, TRMT10C, HSP90AB1, HEATR1, PSMD13, EIF4A1, and SLC1A5.

(b) Correlation table: A table showing correlations between KCNN4 and other genes. The table includes columns for gene, title, cor (correlation), and pcor (partial correlation). The top row shows the correlation of 1.000 for KCNN4 itself.

gene	title	cor	pcor
KCNN4	potassium calcium-activated channel subfamily N member 4	1.000	1.000
WDPR75	WD repeat domain 75	0.9255	0.1240
PSMB5	proteasome 20S subunit beta 5	0.9230	0.000
TEX30	testis expressed 30	0.9163	0.2973
TRMT10C	tRNA methyltransferase 10C, mitochondrial RNase P subunit	0.9145	0.07668
HSP90AB1	heat shock protein 90 alpha family class B member 1	0.9077	0.03617
HEATR1	HEAT repeat containing 1	0.9043	0.07665
PSMD13	proteasome 26S subunit, non-ATPase 13	0.9021	0.06849
EIF4A1	eukaryotic translation initiation factor 4A1	0.8997	0.000
SLC1A5	solute carrier family 1 member 5	0.8890	-0.01870

(c) Correlation scatter plots: A grid of 20 scatter plots showing gene expression of top correlated genes. Each plot has 'KCNN4' on the x-axis. The y-axis labels for each plot include ACTB, WDPR75, PSMB5, TEX30, TRMT10C, HSP90AB1, HEATR1, PSMD13, EIF4A1, SLC1A5, KCNN4, WDPR75, PSMB5, TEX30, TRMT10C, HSP90AB1, HEATR1, PSMD13, EIF4A1, and SLC1A5. Red dots represent KNN1, and blue dots represent KNN4.

General notes:

- Yet another style of tags in bold and grey but lower case
 - make sure to use a consistent style for the panel headers and tags
- The text used in the plots is not set in Lato, please update

(a) Top correlated genes:

- Needs to be implemented in plotly
- To me it is unclear, if the bars pf correlation and partial correlation are stacked or overlapping
- Overall not a very attractive plot with lots of bars that are mostly of same height and all these rotated labels
 - maybe the plot can be flipped to allow for horizontal labels?
- Difficult to match labels to bars (but will be solved when using an interactive version with tooltips)
 - one could add dotted lines or something similar from the labels to the zero baseline

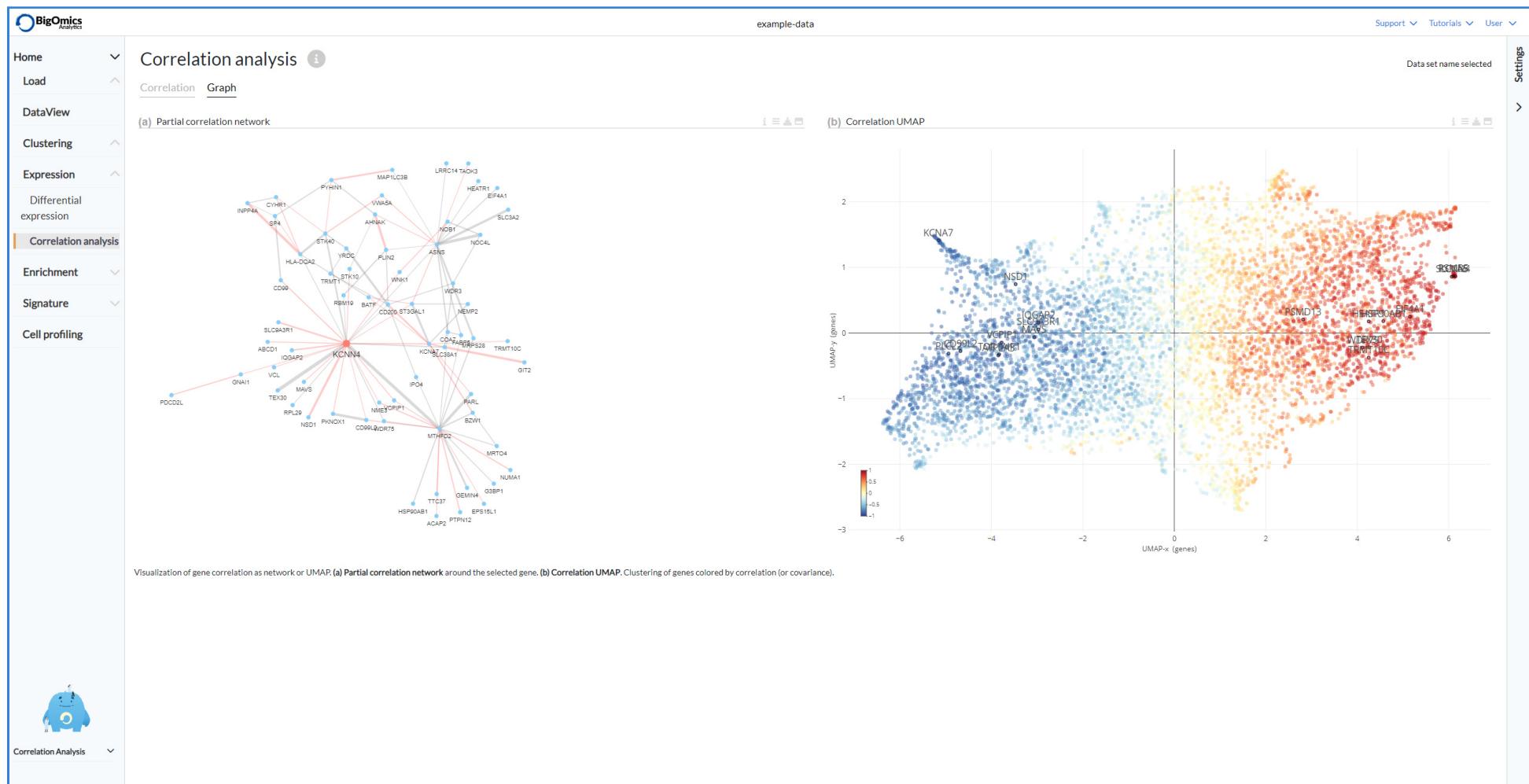
(b) Correlation table:

- The tag levels seem a bit strange to me, I would expect the table to be labelled with (c) and the grid of scatter plots to be labelled with (b)
- Inconsistent table styling, make sure to streamline the table looks (I prefer the other look without grey rows)
- As in several tables before, the number of digits should be consistent to allow for properly aligned numbers and the diverging palette for bars needs to replaced and the alarming color should encode negative/low values
- For me it is strange that the cor column is in addition encoded with bars while the pcor column isn't
 - The column names could be more descriptive and less "code-like"
 - also, decide if you want to use upper case letter for the column headers and use the style consistently across all tables featured in the app

(c) Correlation scatter plots:

- Needs to be implemented in plotly
- The legend is repetitive
 - only show one legend but more prominently (on the top or right of the grid) instead of placing it inside of one/ several of the panels
- All text elements are too small
- Consider to add some transparency to the points as there are likely multiple points on top of each other?

Correlation Analysis: Graph



General notes:

- Make sure to use a consistent style for the panel headers and tags

(a) Partial correlation network:

- Unclear what the colors (red and grey) represent
 - add a legend or at least a note to the caption
- Why are nodes colored in light blue?
 - if color is not used for encoding, replace with corporate blue
- Text labels seem to be too small
- The text is not set in Lato, please update

(b) Correlation UMAP:

- A new diverging color palette (with yellow as midpoint color), not consistent with the previous palettes for similar plots
 - should be replaced by one of the corporate color palettes that is used consistently across the app
 - again, the red color should likely point towards the negative range of values
- Text labels are hard to read on top of the dark colors
 - either lighten and/or desaturate the point colors or add a white stroke to the text labels
- Tooltips need to be polished accordingly to the general style

Board
“Enrichment”

Geneset Enrichment: Top enriched + Table

BigOmics Analytics example-data Support Tutorials User Settings Data set name selected

Home Load DataView Clustering Expression Enrichment Geneset enrichment Pathway analysis Word cloud Drug connectivity Signature Cell profiling Geneset Enrichment

Geneset enrichment

Top enriched Plots Compare Volcano (all) Volcano (methods)

(a) Top enriched gene sets

(b) Frequency in top gene sets

(a) Top enriched gene sets. Enrichment plots of the top differentially enriched gene sets. Black vertical bars indicate the rank of genes in the gene set in the sorted list metric. The green curve corresponds to the 'running statistics' of the enrichment score. (b) Gene frequency. The plot shows the number of times a gene is present in the top-N genesets sorted by frequency.

Table Foldchange (all) FDR table

Enrichment tables. (I) Table summarizing the statistical results of the gene set enrichment analysis for selected contrast. The number of stars indicate how many methods identified the geneset significant. (II) Table showing the fold-change, statistics and correlation of the genes in the selected gene set.

(I) Enrichment analysis

Search:

geneset	size	logFC	meta.q	stars	AveExpr	AveExpr1
C2:ISHIDA_E2F_TARGETS	46	2.007	0.0005000	***	3.527	5.533
C2:REACTOME_DNA_STRAND_ELONGATION	29	1.899	0.0005000	***	4.585	6.484
PATHWAY:DNA strand elongation_Homo sapiens_R-HSA-69190	30	1.839	0.0005000	***	4.610	6.449
C2:WU_APOTOSIS_BY_CDKN1A_VIA_TP53	43	1.720	0.0005000	***	3.433	5.153
PATHWAY:Activation of the pre-replicative complex_Homo sapiens_R-HSA-689...	26	1.611	0.0005000	***	3.828	5.439

Show 1 to 15 of 4,920 entries Previous 1 2 3 4 5 ... 328 Next

(II) Genes in gene set

warning: no genes.

General notes:

- Make sure to use a consistent style for the panel headers and tags
- Tab looks crowded and unstructured, mostly due to missing spacing between the plots and the table but also due to multiple captions and their placements
- The header “Gene enrichment” in the sidebar is wrapped
- Table recommendations are the same as in the previous “Differential Expression” notes
 - note that the stars in the stars column are not working here
- None of the plots in these tabs use Lato as typeface, please update

(a) Top enriched gene sets:

- Needs to be reimplemented in plotly
- Awful colors, replace with corporate color palettes
- Here, the diverging palette is pointing in the right direction (but not consistent with all previous plots and tables using a blue-red gradient)
 - most importantly, it is not in line with the table below the plot!
 - should be solved as you are replacing all diverging palettes anyway
- All text elements are way too small
- Redundant axes labels, turn into a true grid if possible
- Pretty busy plots (but seems to be a standard chart for bioinformatics so I guess no chance to replace it?)

(b) Frequency in top gene sets:

- Needs to be reimplemented in plotly
- Grey-scale palette not suitable here, also not meaningful with a legend and/or tooltips
- I have no clue how this chart is providing insights due to the many groups (shades of grey), the missing information what grey represents and all the stacked chunks
 - what is the main insight that should be transmitted? Is the overall height of bars more important than the share of each of the grey shades and which grey dominates?
 - in case the grey shades are of high interest, maybe allow a regrouping into these different groups?
 - maybe add a switch to quickly change stacked to grouped bars?
- How are the stacks sorted?
 - to me it would make sense to rank them by overall occurrence, currently it seems a bit random
- As before, a flipped version would increase readability of the gene names

Geneset Enrichment: Plots + Foldchange (all)

example-data

Support ▾ Tutorials ▾ User ▾

Data set name selected

Settings >

Geneset enrichment

- Home
- Load
- DataView
- Clustering
- Expression
- Enrichment
 - Geneset enrichment
 - Pathway analysis
 - Word cloud
 - Drug connectivity
- Signature
- Cell profiling

Enrichment plots associated with the gene set (selected in Table I) and gene (selected in Table II). (a) Volcano-plot showing significance versus fold-change on the y and x axes, respectively. Genes in the gene set are highlighted in blue. (b) Barplot of the gene set enrichment in the groups. (c) Barplot of selected gene in the groups. (d) Scatter plot of the enrichment versus the expression of the selected geneset and gene, on the y and x axes, respectively.

Table Foldchange (all) FDR table

Gene set enrichment for all contrasts

Enrichment for all contrasts. Table summarizing the enrichment for all gene sets across all contrasts. The column "fcvar" corresponds to the variance of the gene set across all contrasts.

Search:

geneset	fcvar	act vs notact	act12h vs notact	act24h vs notact	act48h vs notact	act72h vs notact	act96h vs notact
C2:ISHIDA_E2F_TARGETS	5.134	2.007	0.266	0.678	3.091	2.99	2.783
C2:REACTOME_DNA_STRAND_ELONGATION	4.734	1.899	-0.005	1.197	2.818	2.965	2.575
PATHWAY:DNA strand elongation_Homo sapiens_R-HSA-69190	4.472	1.839	-0.032	1.111	2.741	2.889	2.521
C2:WU_APOPTOSIS_BY_CDKN1A_VIA_TPS3	3.993	1.72	0.001	0.62	2.749	2.649	2.458
PATHWAY:Activation of the pre-replicative complex_Homo sapiie	3.367	1.611	0.028	1.232	2.37	2.484	2.075
PATHWAY:DNA Replication_Mus musculus_WP150	3.207	1.599	0.154	1.142	2.239	2.365	2.179
PATHWAY:DNA replication_Homo sapiens_hsa03030	3.18	1.597	0.181	1.055	2.234	2.369	2.186

Showing 1 to 0 of 4,920 entries

General notes:

- Make sure to use a consistent style for the panel headers and tags
- “Plots” as tab name is not very descriptive
- Tab looks misaligned and unstructured, mostly due to missing spacing between the plots and the table but also due to multiple captions and their placements
- The header “Gene enrichment” in the sidebar is wrapped

(a) to (d):

- None of the plots loads, not even a spinner shows up
- Same for the plot(s) “Enrichment of geneset across multiple contrasts” in the “Compare” tab

Gene set enrichment for all contrasts:

- No tag but panel header with weird spacing in front, thus misaligned with rest of the tab
- Recommendations are the same as in the previous “Differential Expression” notes on tables
 - ensure consistent number of digits
 - diverging palette needs to be replaced and the alarming color should point towards negative values
 - note that the maximum ranges seem to be off again with bars not reaching the maximum values at all

Geneset Enrichment: Volcano (all) + FDR table

BigOmics Analytics example-data Support ▾ Tutorials ▾ User ▾ Data set name selected Settings >

Home Load DataView Clustering Expression Enrichment Geneset enrichment Pathway analysis Word cloud Drug connectivity Signature Cell profiling

Geneset enrichment ⓘ

Top enriched Plots Compare Volcano (all) Volcano (methods)

Volcano plots for all contrasts

Volcano plots for all contrasts. Simultaneous visualisation of volcano plots of gene set enrichment across all contrasts. Volcano-plot are plotting enrichment score versus significance on the x and y axes, respectively. Experimental contrasts showing better statistical significance will show volcano plots with 'higher' wings.

Table Foldchange (all) FDR table

Number of significant gene sets

FDR table. Number of significant gene sets versus different FDR thresholds, for all contrasts and all methods. The blue color denote the number of downregulated genes, the red color for upregulated genes.

Search: ⌂

method	TU	contrast	TU	DOWN FDR = 1e-16 TU	1e-08 TU	1e-06 TU	1e-04 TU	0.01 TU	0.05 TU	0.1 TU	0.2 TU	0.5 TU	1 TU	UP FDR = 1e-16 TU	1e-08 TU	1e-06 TU	1e-04 TU	0.01 TU	0.05 TU	0.1 TU	0.2 TU	0.5 TU	1 TU
camera		act_vs_notact		0	0	0	24	327	607	758	937	1329	1762	8	247	560	1715	3491	3973	4141	4325	4646	5067
fgsea		act_vs_notact		0	0	0	0	688	842	942	1067	1333	1758	0	0	0	0	4153	4329	4412	4491	4710	5071
fisher		act_vs_notact		0	2	12	68	379	573	662	777	986	1763	450	1839	2665	3441	4066	4295	4386	4495	4759	5066
camera		act12h_vs_notact		0	0	0	0	8	64	135	269	704	1770	0	12	49	167	1245	2203	2654	3102	3827	5059
fgsea		act12h_vs_notact		0	0	0	0	38	112	194	358	809	1655	0	0	0	0	2191	2745	3044	3408	4114	5174
fisher		act12h_vs_notact		0	0	4	7	26	54	76	115	249	1796	22	459	946	1695	2612	2971	3196	3488	4067	5030
camera		act24h_vs_notact		0	0	1	6	204	459	627	848	1271	1918	0	170	412	1130	2841	3372	3609	3835	4308	4911

Showing 1 to 0 of 18 entries

Geneset Enrichment ⌂

Volcano plots for all contrasts:

- No tag but panel header with weird spacing in front, thus misaligned with rest of the tab
- Not interactive (likely due to performance issues?)
- All text labels are way too small
- Axis titles are missing
- Spacing between panels seems a bit too low, consider to increase it
- Strange empty row below the row of volcano plots
 - can the grid placement be automated in a more clever way?
- Plot tile inside panel while they are placed on top (as it should be) in the “Volcano plots for all contrast” grid in the expression board (why are you using a different plot function and/or style here?)
- Suggestions for adding more insights as for the other volcano plots

Number of significant gene sets:

- No tag but panel header with weird spacing in front, thus misaligned with rest of the tab
- Recommendations are the same as in the previous “Differential Expression” notes on tables:
 - most importantly the two colors need to be changed as they are confusing as they are now used in a different context
 - consider to use two sequential palettes to actually encode the values, not only the upper and lower FDR groups
 - if you decide to include a big mark, do so here as well

Geneset Enrichment: Volcano (methods)

BigOmics Analytics

example-data

Support ▾ Tutorials ▾ User ▾

Home ▾ Load ▾ DataView ▾ Clustering ▾ Expression ▾ Enrichment ▾ Geneset enrichment ▾ Pathway analysis ▾ Word cloud ▾ Drug connectivity ▾ Signature ▾ Cell profiling ▾

Geneset enrichment ⓘ

Top enriched Plots Compare Volcano (all) Volcano (methods)

Volcano plots for all methods

Spearmann

sigsea

fisher

tgsea

camera

0 DOWN 0 UP

Volcano plots for all methods. Simultaneous visualisation of volcano plots of gene sets for different enrichment methods. Methods showing better statistical significance will show volcano plots with 'higher' wings.

Table Foldchange (all) FDR table

Gene set enrichment for all contrasts

Enrichment for all contrasts. Table summarizing the enrichment for all gene sets across all contrasts. The column "fcvar" corresponds to the variance of the gene set across all contrasts.

Search:

geneset	fcvar	act vs notact	act12h vs notact	act24h vs notact	act48h vs notact	act72h vs notact	act96h vs notact
GOBP:regulation of DNA recombination (GO_0000018)	0.218	0.442	0.189	0.322	0.652	0.609	0.423
GOBPrribosomal large subunit assembly (GO_0000027)	0.695	0.779	0.554	0.999	1.134	0.813	0.558
GOBPrribosomal small subunit assembly (GO_0000028)	0.904	0.894	0.493	1.051	1.241	1.026	0.829
GOMF:tRNA binding (GO_0000049)	0.073	0.267	0.263	0.214	0.31	0.384	0.098
GOBP:G1/S transition of mitotic cell cycle (GO_0000082)	0.391	0.548	0.016	0.299	0.859	0.881	0.664
GOBP:regulation of transcription involved in G1/S transition	1.399	1.046	0.128	0.459	1.731	1.558	1.284
GOBPrnuclear-transcribed mRNA catabolic process, nonsense-me	0.509	0.679	0.472	0.814	0.91	0.725	0.591

Showing 1 to 0 of 4,920 entries

Geneset Enrichment

Settings

Volcano plots for all methods:

- Needs to be reimplemented in plotly
- All text labels are way too small
- Spacing between panels could be a bit larger
- Axis titles are missing; also, missing x axis labels for all plots 2-5
- Strange grid behavior, could be optimized
- Title should be placed outside of the panel as in the expression board; in general, use uniform style for all volcano plots

Pathset analysis: KEGG

The screenshot shows the BigOmics Analytics interface for Pathway Analysis. The left sidebar has a 'Pathway analysis' section selected, containing options like 'Word cloud', 'Drug connectivity', 'Activation matrix', and 'Cell profiling'. A small blue cartoon character icon is also present. The main area displays two panels: 'Kegg pathway map' and 'Enrichment table', both of which show error messages: 'Error: Fehler bei der Auswertung des Argumentes 'X' bei der Methodenauswahl für Funktion 'as.list': es gibt kein Paket namens 'KEGG.db''.

General notes:

- None of the plots work due to the missing KEGG.db package

Pathset analysis: GO graph

The screenshot shows the BigOmics Analytics platform's Pathway Analysis section. The left sidebar includes links for Home, Load, DataView, Clustering, Expression, Enrichment, Geneset enrichment, Pathway analysis (selected), Word cloud, Drug connectivity, Signature, and Cell profiling. A small blue cartoon character icon is also present.

The main area has tabs for KEGG and GO graph (selected). Below these are sections for Gene Ontology graph, Activation matrix, and example-data. The Activation matrix tab is currently active, displaying a heatmap of GO term enrichment across multiple contrast profiles. The GO score table and Gene Ontology graph tabs are also visible.

Activation matrix: This heatmap visualizes the enrichment of GO terms across multiple contrast profiles. The y-axis lists GO terms, and the x-axis lists contrast profiles. Red dots indicate positive enrichment, while blue dots indicate negative enrichment.

GO score table: This table provides cumulative scores for GO terms. The columns include id, T1, term, T1, score, T1, logFC, T1, and meta.q. The first few rows are:

id	T1	term	T1	score	T1	logFC	T1	meta.q
GO:0070126		mitochondrial translational termination		7.500		0.6042		0.0005000
GO:0000466		maturity of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA,		6.975		0.5027		0.01880
GO:0002181		cytoplasmic translation		6.883		0.9166		0.0005000
GO:0070125		mitochondrial translational elongation		6.126		0.6512		0.0005000
GO:006364		rRNA processing		6.109		0.6971		0.0005000
GO:0042255		ribosome assembly		5.814		0.7383		0.0005000

Gene Ontology graph: This tree diagram visualizes the enrichment of GO terms. Nodes represent GO terms, and edges represent hierarchical relationships. Nodes are colored based on their enrichment score.

General notes:

- Headers don't feature tags but you refer to them in the caption
- As before, none of the plots features Lato as typeface, please update

Gene Ontology graph:

- Plot looks lost due to too much padding on the left and right and the tiny text labels
- Text hardly readable due to the size but also overlapping boxes and low contrast in many cases
 - make plot wider to allow for larger font sizes and no overlap of boxes
- Replace diverging palette with one of the corporate ones
- Unclear what colors encode, add legend (or at least a note to the caption)
 - please reverse palette if, as in almost all cases so far, red is used for "good"

GO score table:

- As with most tables before:
 - update diverging palette plus alarming color for low/negative values
 - ensure consistent number formatting, i.e. same number of digits (especially "0.005000" looks very strange – why the trailing zeros here?? same for the single trailing zero for "0.01880")

Activation matrix:

- Needs to be reimplemented in plotly
- Plot looks lost due to too much padding on the left and right and the tiny text labels
 - flip the chart?
 - maybe limit to a lower number of GO terms as the visual is very text-heavy
- Replace diverging palette with one of the corporate ones
- Unclear what colors encode, add legend (or at least a note to the caption)
 - please reverse palette if, as in almost all cases so far, red is used for "good"

Word cloud



General notes:

- Inconsistent tag style as discussed before

(a) Enrichment plots:

- Recommendations as for the other set of enrichment plots in the “Geneset enrichment” tab
 - reimplement in plotly
 - replace colors and color palettes
 - increase label sizes
 - consider to remove redundant axes
 - use Lato as typeface

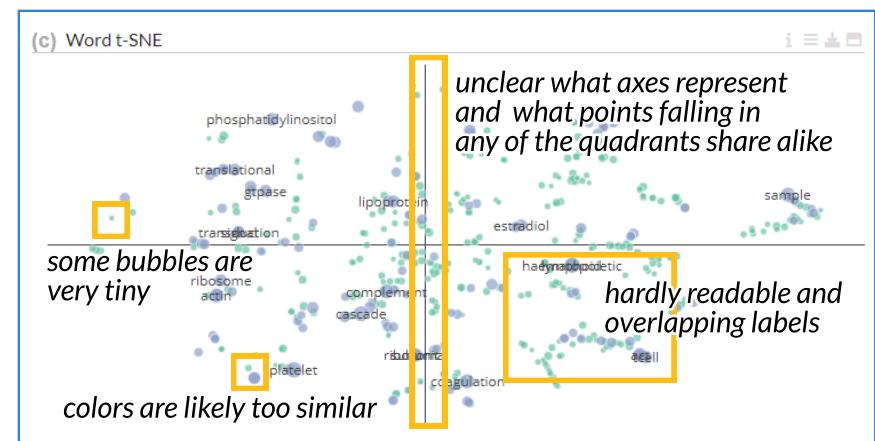
(b) Word cloud:

- Padding to the left and right too large, plot cloud lost
- Text is not set in Lato, please update
- The sequential palette is not very suited
 - replace wioth a limited sequential palette with a darker colors for low values
- Word clouds in general are problematic as they tend to be messy, make one focus one especially on the longer words, and are less insightful than ranked bar plots or tables of single words (at the same time, many researchers love to show word clouds, so that might be reason enough to feature it)

Word cloud (continued)

(c) Word t-SNE:

- Unclear what the axes encode
 - add titles and numbers or words indicating what's shown (e.g. "low interest" or whatever)
- Unclear what the colors represent
 - add a legend or at least a note in the caption
 - as I assume that these are somehow connected to the other panels, consider to use the same color to encode groups
 - I would recommend two colors that are less similar and more easily distinguishable
- Some points look a bit too tiny to make it easy enough to hover over them
- Words that are crossing points are hard to read
 - consider to add a white stroke to the labels
 - are they needed at all as the chart features tooltips?



(e) Enrichment table:

- Ensure same number of digits per column so that numbers align
- As before, update the diverging color palette make sure that the alarming color encodes low values
- The table uses the striped style, use one consistent style across the app
- The column headers could be more descriptive instead of using "codes"

(f) Leading-edge table:

- Same comments on the colored bars, column headers, and table style as for table (e) before

size ↑↓
950.0
31.00
32.00
21.00
72.00
12.00

Word cloud



General notes:

- The tab “Drug enrichment” is superfluous as there are no other tabs
- No tags for individual panels but reference to those in the caption
- Table recommendations as before

(a) Drug connectivity:

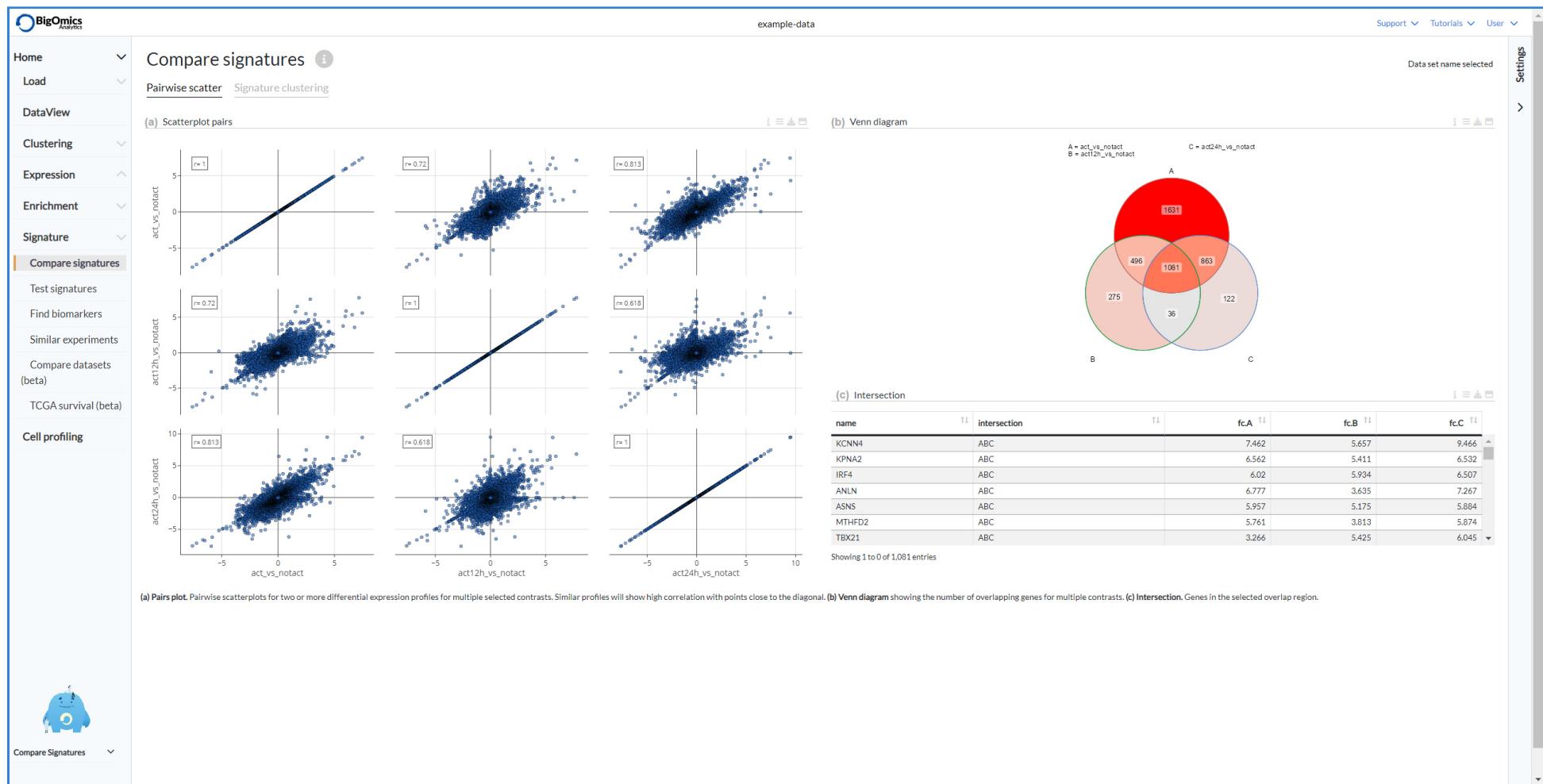
- Recommendations as for the other set of enrichment plots in the “Geneset enrichment” tab
 - reimplement in plotly
 - replace colors and color palettes
 - increase label sizes
 - consider to remove redundant axes
 - use Lato as typeface

(b) Mechanisms of action:

- Needs to be reimplemented in plotly
- Text is not set in Lato, please update
- The rotated labels along the x axis are not great, some too long
 - consider to flip the chart and/or shorten/break the labels
- The labels are hard to match to bars
 - consider to add dotted lines or something similar to the zero baseline to simplify matching labels with (especially positive) bars

Board
“Signature”

Compare Signatures: Pairwise scatter



General notes:

- Make sure to use a consistent style for the panel headers and tags

(a) Scatterplot pairs:

- Text seems a bit too small
- As there are lots of overlapping points, one could consider one of the following:
 - instead of showing single points, a grid binning could be used with color encoding number of points
 - similarly, one could still show single points but color them based on the density (i.e. the number of neighbors in a given distance)
- Tooltips need to be polished
- In my opinion adding true facet headers instead of using the axis titles would make it easier to read the grid (i.e. placing the titles for each column at the top not the bottom)

(b) Venn diagram:

- Well, the area of overlap does not represent the numbers.... How is this visualization any helpful??
 - make sure to scale the area accordingly – otherwise a table is more efficient
- Strange colors, why red? The outline colors are awful
 - replace fill colors with corporate color palettes and use simply a black/grey outline for the circles
- Why is there a legend/glossary?
 - there is enough space to add those labels directly instead of encoding them with letters
- Text is not set in Lato, please update

(c) Intersection:

- Basic table, no further suggestions besides using a consistent style and more descriptive column names

Compare Signatures: Signature clustering



General notes:

- Make sure to use a consistent style for the panel headers and tags
- The naming sounds redundant and a bit ridiculous: Signatures > Compare signatures > Signature clustering
- Somehow the layout looks off with all the padding around the plots, the caption is floating around in the nowhere

(a) Foldchange heatmap:

- Needs to be reimplemented in plotly
- Text seems a bit too small
- Text is not set in Lato, please update
- A zero baseline could help to make the upper part easier readable
 - also, consider to color the boxplots, e.g. based on the median, min/max difference from zero, ...
- As in most cases, the diverging palette should be reversed
 - when updating the palette, make sure the alarming color encodes negative values

(b) Contrast correlation:

- Text is not set in Lato, please update
- As in most cases, the diverging palette should be reversed
 - when updating the palette, make sure the alarming color encodes negative values
 - it feels a bit strange that the legend ranges from -1 to +1 but only red values are featured in the heatmap
- Style tooltips consistently

Test Signatures: Enrichment + Enrichment table

BigOmics Analytics example-data Support Tutorials User Settings Data set name selected

Home Load DataView Clustering Expression Enrichment Signature Compare signatures Test signatures Find biomarkers Similar experiments Compare datasets (beta) TCGA survival (beta) Cell profiling

Test signatures

Enrichment Volcano plots Overlap/similarity Markers

Enrichment plots

act48h_vs_notact q=0.004 act96h_vs_notact q=0.004 act72h_vs_notact q=0.004
act_vs_notact q=0.034 act24h_vs_notact q=0.007 act12h_vs_notact q=0.199

rank metric

Enrichment plots. Enrichment of the query signature in all contrasts. Positive enrichment means that this particular contrast shows similar expression changes as the query signature.

Enrichment table

(a) Enrichment by contrasts

contrast	T1	NES	T1	q
act48h_vs_notact		3.055		0.004200
act96h_vs_notact		3.044		0.004200
act72h_vs_notact		2.940		0.004200
act_vs_notact		2.933		0.004200
act24h_vs_notact		1.538		0.007100
act12h_vs_notact		-1.228		0.1086

Show 1 to 6 of 6 entries

(b) Genes in signature

Enrichment of query signature across all contrasts. (a) Enrichment scores across all contrasts for the selected query signature. The NES corresponds to the normalized enrichment score of the GSEA analysis. (b) Genes in the query signature sorted by decreasing (absolute) fold-change corresponding to the selected contrast.

Enrichment plots:

- Needs to be reimplemented in plotly
- Same recommendations as for all other enrichment plots before
- Too much empty space between grid of plots and caption
- Text is not set in Lato, please update

Enrichment table:

- The tab is superfluous as there is only one entry

(a) Enrichment by contrast:

- As before, make sure to use the alarming color for negative scores when updating the diverging palette
- As before, make sure that the number of digits is consistent within columns (why trailing zeros in the q column?)

(b) Gene in signature:

- Nothing loading here in the current version, also no spinner showing up

All other tabs:

The plots are not updated and all tabs show the same grid of enrichment plots

BigOmics Analytics example-data Support Tutorials User Settings Data set name selected

Home Load DataView Clustering Expression Enrichment Signature Compare signatures Test signatures Find biomarkers Similar experiments Compare datasets (beta) TCGA survival (beta) Cell profiling

Test signatures

Enrichment Volcano plots Overlap/similarity Markers

Enrichment plots

act48h_vs_notact q=0.004 act96h_vs_notact q=0.004 act72h_vs_notact q=0.004
act_vs_notact q=0.034 act24h_vs_notact q=0.007 act12h_vs_notact q=0.199

rank metric

Enrichment plots. Enrichment of the query signature in all contrasts. Positive enrichment means that this particular contrast shows similar expression changes as the query signature.

Enrichment table

(a) Enrichment by contrasts

contrast	T1	NES	T1	q
act48h_vs_notact		3.055		0.004200
act96h_vs_notact		3.044		0.004200
act72h_vs_notact		2.940		0.004200
act_vs_notact		2.933		0.004200
act24h_vs_notact		1.538		0.007100
act12h_vs_notact		-1.228		0.1086

Show 1 to 6 of 6 entries

(b) Genes in signature

Enrichment of query signature across all contrasts. (a) Enrichment scores across all contrasts for the selected query signature. The NES corresponds to the normalized enrichment score of the GSEA analysis. (b) Genes in the query signature sorted by decreasing (absolute) fold-change corresponding to the selected contrast.

BigOmics Analytics example-data Support Tutorials User Settings Data set name selected

Home Load DataView Clustering Expression Enrichment Signature Compare signatures Test signatures Find biomarkers Similar experiments Compare datasets (beta) TCGA survival (beta) Cell profiling

Test signatures

Enrichment Volcano plots Overlap/similarity Markers

Enrichment plots

act48h_vs_notact q=0.004 act96h_vs_notact q=0.004 act72h_vs_notact q=0.004
act_vs_notact q=0.034 act24h_vs_notact q=0.007 act12h_vs_notact q=0.199

rank metric

Enrichment plots. Enrichment of the query signature in all contrasts. Positive enrichment means that this particular contrast shows similar expression changes as the query signature.

Enrichment table

(a) Enrichment by contrasts

contrast	T1	NES	T1	q
act48h_vs_notact		3.055		0.004200
act96h_vs_notact		3.044		0.004200
act72h_vs_notact		2.940		0.004200
act_vs_notact		2.933		0.004200
act24h_vs_notact		1.538		0.007100
act12h_vs_notact		-1.228		0.1086

Show 1 to 6 of 6 entries

(b) Genes in signature

Enrichment of query signature across all contrasts. (a) Enrichment scores across all contrasts for the selected query signature. The NES corresponds to the normalized enrichment score of the GSEA analysis. (b) Genes in the query signature sorted by decreasing (absolute) fold-change corresponding to the selected contrast.

Biomarker Selection

The screenshot shows the Biomarker Selection module of the BigOmics Analytics platform. The left sidebar contains a navigation menu with various analytical tools. The 'Find biomarkers' option is currently selected. The main workspace is divided into four panels: 'Variable importance', 'Decision tree', 'Heatmap', and 'Biomarker expression'. A status bar at the top right shows 'Data set name selected'. A note at the bottom of the page provides a brief explanation of biomarker selection.

Biomarker selection. The expression of certain genes may be used as *markers* to predict a certain phenotype such as response to a therapy. Finding such *biomarkers* are of high importance in clinical applications. (a) An importance score for each feature is calculated using multiple machine learning algorithms, including LASSO, elastic nets, random forests, and extreme gradient boosting. The top features are plotted according to cumulative ranking by the algorithms. (b) The heatmap shows the expression distribution for the top most important features. (c) The decision tree shows (one) tree solution for classification based on the top most important features. (d) Boxplots show the expression of biomarker genes across the groups.

General notes:

- No plots load in the current version, not even spinners indicating that plots are loading
- Overall, this tab looks as it's too long (the caption is not visible without scrolling)
- No tags in panel header but references in the caption

Similar Experiments: FC correlation

General notes:

- No plots load in the current version, not even spinners indicating that plots are loading
- Overall, this tab looks as it's too long (the caption is not visible without scrolling)
- Tags after the panel header, inconsistent with other boards and tabs
- Caption at the top for some reason, inconsistent with most other boards and tabs
- Caption is misaligned with the rest of the tab

→ same applies to all the other tabs featured here (“FC heatmap”, “Meta-graph”, and “Experiment clustering”)

Compare Datasets: Compare



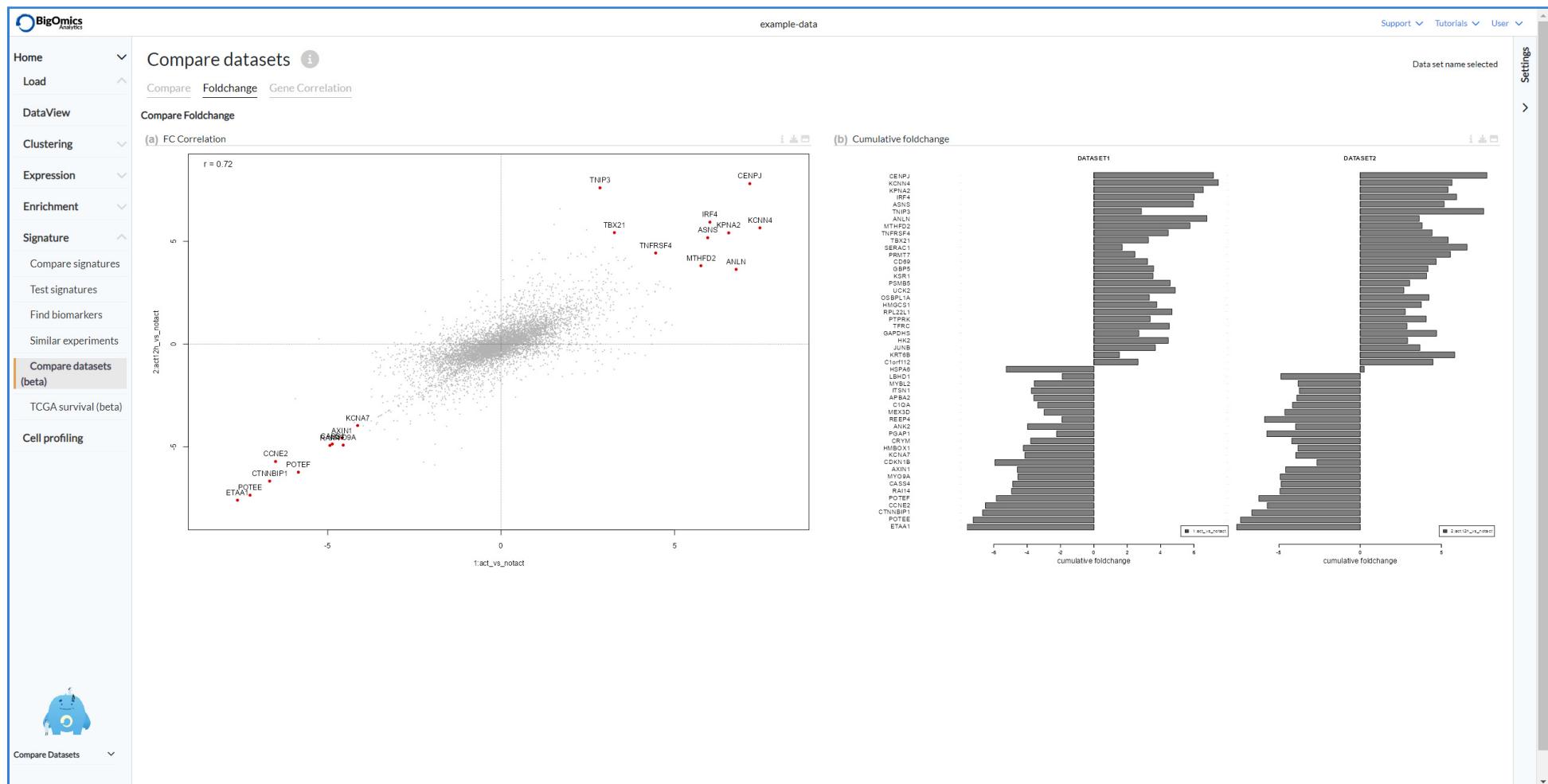
General notes:

- The middle-header (?) is not aligned with the rest
 - not sure why it's there in the first place – maybe a placeholder for the caption?
 - again, it feels a bit redundant: Compare datasets > Compare > Compare expressions
- Make sure to consistently style the tags and headers in general
- The header "Compare datasets (beta)" in the sidebar is wrapped

(a) Dataset 1 + (b) Dataset 2:

- Again, the diverging palette should be reversed (red == bad)
 - consider using a midpoint color that is a bit darker; alternatively point outlines could solve the issue
- The labeling needs to be adjusted as some annotations are crossing the panel border
 - Maybe tooltips are good enough and no labeling is needed at all?
- The text is not set in Lato, please update
- Overall, all text labels seem a bit too small
- Legend could be placed much more prominently and in bigger size

Compare Datasets: Foldchange



General notes:

- The middle-header (?) is not aligned with the rest
 - not sure why it's there in the first place – maybe a placeholder for the caption?
 - again, it feels a bit redundant: Compare datasets > Foldchange > Compare foldchange
- Make sure to consistently style the tags and headers in general

(a) FC Correlation:

- Needs to be reimplemented in plotly
- Avoid overlapping labels
 - maybe labels are not needed at all when turning this into an interactive visualization?
- It is unclear why some points are / which points have been chosen to be colored in red

(b) Cumulative foldchange:

- Needs to be reimplemented in plotly
- Difficult to match y-axis labels to bars
 - consider to add some helping lines going connecting the labels with the bars
- Legend not helpful at all as there is only one color anyway
 - if that information is important, it should be part of the title or added as subtitle
- Axis breaks on the right bar chart is not really helpful
 - add further breaks
 - if comparison between the two plots is of high interest, make sure to use the same axis limits for both plots

Compare Datasets: Gene Correlation

The screenshot shows the BigOmics Analytics interface with the 'Compare datasets' section selected. The main area is divided into three sections: (a) Expression, (b) Correlation score, and (c) Gene correlation.

- (a) Expression:** Contains bar charts for genes ETAA1, CENPJ, POTE, CTNNBIP1, KCNN4, CCNE2, POTEF, and IRF4. Each chart compares expression levels between two groups: 1:act_vs_notact (blue) and 2:act12h_vs_notact (green).
- (b) Correlation score:** A table showing correlation statistics for the same genes. The table includes columns for title, score, rho, and two correlation coefficients (1:act_vs_notact and 2:act12h_vs_notact). The data is as follows:

title	score	rho	1:act_vs_notact	2:act12h_vs_notact
ETAA1	ETAA1 activator of ATR kinase	57.8	1.00	-7.60
CENPJ	centromere protein J	56.0	1.00	7.17
POTE	POTE ankyrin domain family member E	53.3	1.00	-7.24
CTNNBIP1	catenin beta interacting protein 1	44.5	1.00	-6.67
KCNN4	potassium calcium-activated channel subfamily N member 4	42.2	1.00	7.46

(c) Gene correlation: A grid of scatter plots comparing expression levels between two datasets (1:act_vs_notact and 2:act12h_vs_notact) for various genes. The genes include ETAA1, CENPJ, POTE, CTNNBIP1, KCNN4, CCNE2, POTEF, IRF4, KPNA2, ASNS, ANLN, RAI14, CASS4, MYO3A, MTHFD2, and TNIP3. The plots show a strong positive correlation for most genes.

General notes:

- The middle-header (?) is not aligned with the rest
 - not sure why it's there in the first place – maybe a placeholder for the caption?
 - again, it feels a bit redundant: Compare datasets > Gene Correlation > Compare Correlation (also, upper case formatting differs between tabs)
- Make sure to consistently style the tags and headers in general
- The order of tags is confusing to me, I expect (b) and (c) to be switched
- The

(a) Expression:

- Needs to be reimplemented in plotly
- How do these colors relate to other tabs and boards?
 - make sure to use a consistent color palette for these groups
- The color shades are a bit confusing at a first glance
 - took a moment to understand that intensity is not mapped to the value but represents groups
 - suboptimal also as the different intensities make viewers focus on the right groups as their visual weight is higher than that of the left bars
- Flipping the charts could avoid rotated, space-consuming x labels
 - why are these rotated in the first place? it seems there is enough space to place them horizontally

(b) Correlation score:

- Basic table, no further recommendations besides deciding on a single, consistent stable style

(c) Gene correlation:

- The plots feel very empty due to the clustering of the points but also the choice of the color palette (might be specific to this example data set?)
 - maybe add a diagonal to the plot?
- As the plot is showing categorical data, the colors should be categorical as well (colors of same weight, not shades of blue that make you focus one one group and “hides” the other group)
- All text elements are way too small
- Legend is redundant, should be a single, bigger legend on the top or the right of the grid

TGCA: TGCA Survival

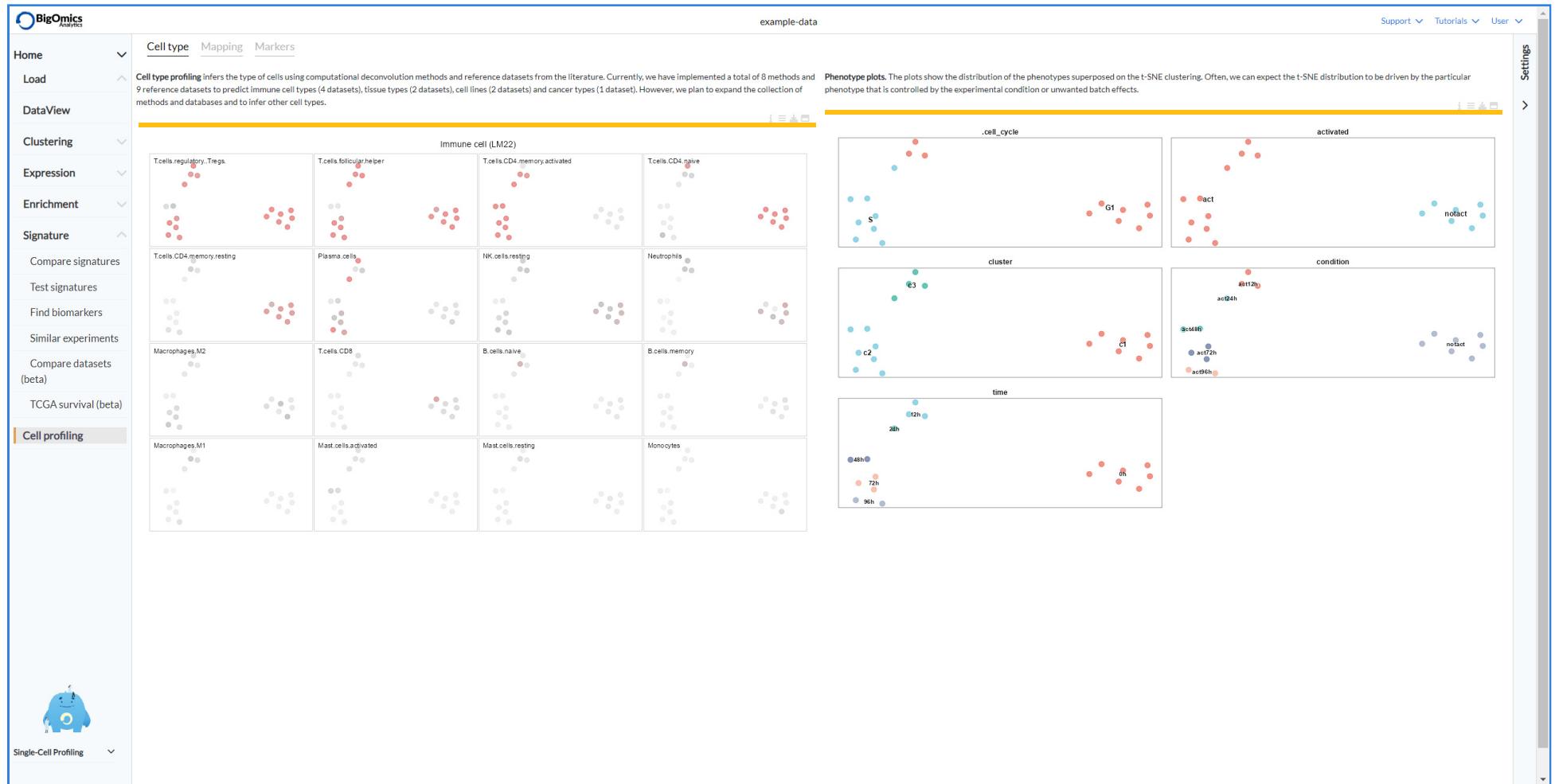
The screenshot shows the BigOmics Analytics web application interface. The left sidebar has a tree view with categories like Home, Load, DataView, Clustering, Expression, Enrichment, Signature, and Cell profiling. Under Signature, 'TCGA survival (beta)' is selected and highlighted with an orange border. The main content area is titled 'TCGA' and contains a sub-section 'TCGA survival'. Below it, a link '(a) TCGA survival analysis' is visible. At the bottom of the main content area, there is a small note: 'TCGA survival analysis. Survival probability of cancer patients in 32 TCGA cancer types. Each cohort is dichotomized into positively and negatively correlated with your signature. The survival probabilities are computed and tested using the Kaplan-Meier method.' On the far left, there is a small blue cartoon character icon. The top right corner shows 'example-data' and 'Data set name selected'. The top right also has 'Support', 'Tutorials', 'User', and 'Settings' buttons.

General notes:

- The tab is superfluous as there are no other tabs
- Plot doesn't load in the current version, not even a spinners is showing up
- Make sure to use a consistent header + tag style across all tabs and boards

Board
“Cell Profiling”

Cell Profiling: Cell type



General notes:

- The left caption is not aligned with the rest of the tab
- It would be nice if the plot grids would be aligned, too

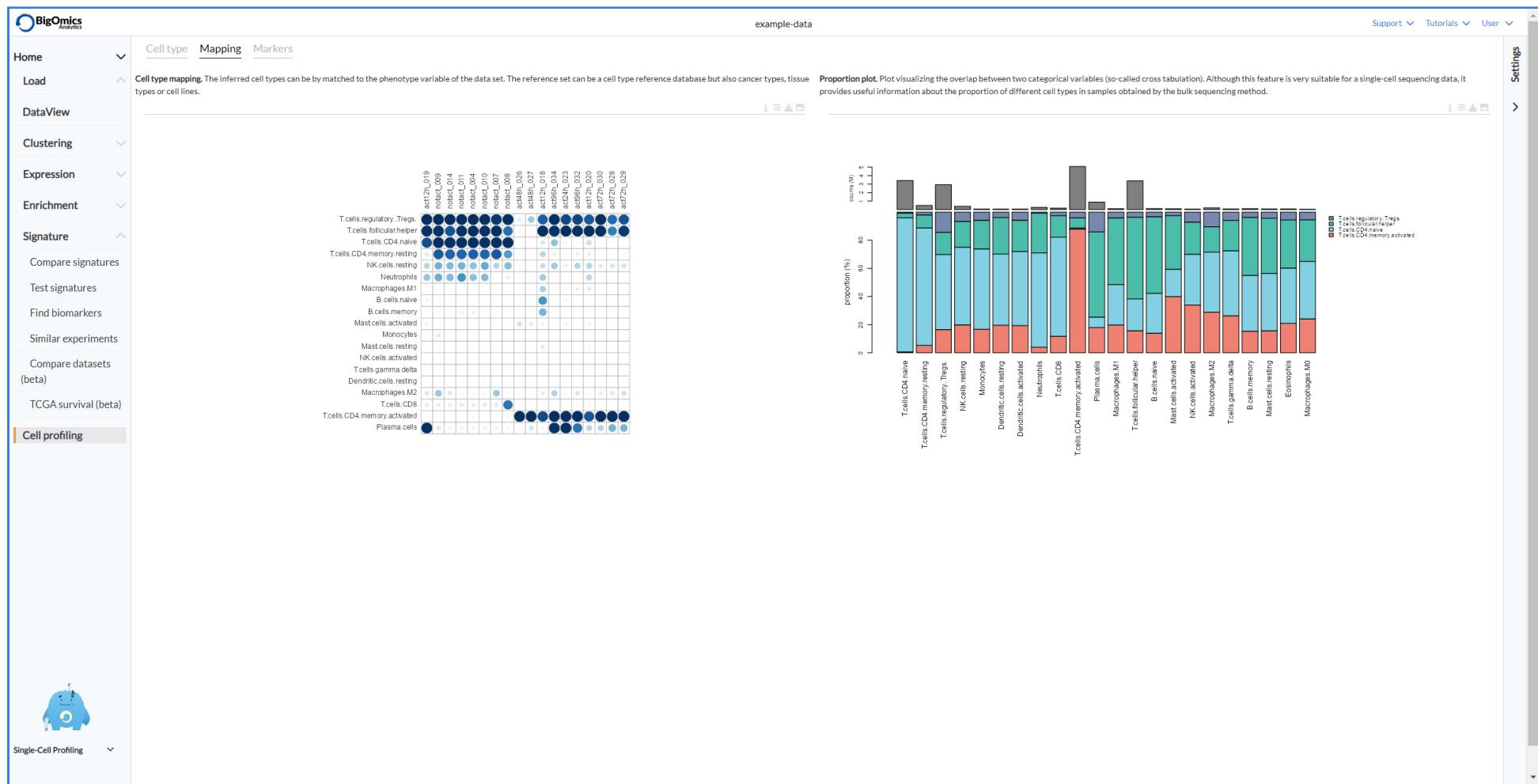
Cell type profiling:

- Needs to be reimplemented in plotly
- All text seems too small
- Text is not set in Lato, please update
- The color palette is suboptimal
 - limit the lower range of colors
 - or add outlines to points which allows to handle the issues of too light colors on a white background
- Unclear what is shown along the axes (might be fine but unusual to show no title and ticks)

Phenotype plots:

- Needs to be reimplemented in plotly
- All text seems too small
- Text is not set in Lato, please update
- The color palette is suboptimal
 - use a true categorical color palette with darker/bright colors for the different colors
 - consider to add an outline to the points if using light colors as well (or as a general detail / corporate style)
 - maybe encircle the groups

Cell Profiling: Cell type



General notes:

- The left caption is not aligned with the rest of the tab
- The plots look a bit lost with all the padding

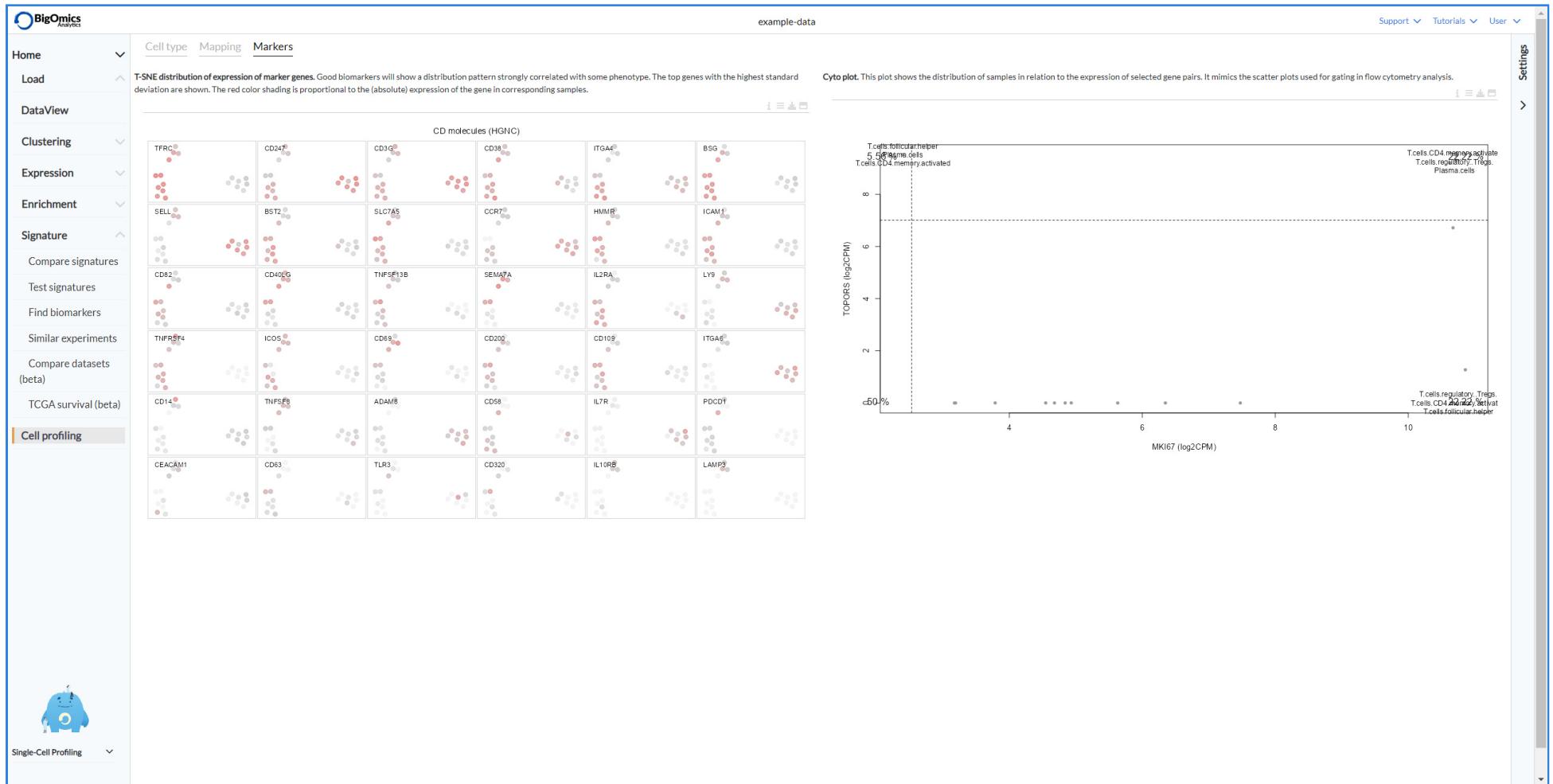
Cell type mapping:

- Needs to be reimplemented in plotly
- Text elements are too small
- Text is not set in Lato, please update
- Unclear what bubble size and color encode
 - limit lower range of sequential palettes to allow to see all bubbles
 - alternatively add outlines (but remove grid, otherwise likely to many lines)

Proportion plot:

- Needs to be reimplemented in plotly
- How are the bars sorted along the x axis?
 - consider ranking them based on one of the variables of interest (e.g. the dark grey amount)
- Polish x axis labels if possible
- Use unique categorical color palette for T.cell groups

Cell Profiling: Markers



General notes:

- The left caption is not aligned with the rest of the tab

T-SNE distribution of expression of marker genes:

- Needs to be reimplemented in plotly
 - Usually, it is “tSNE” but here “T-SNE”
 - All text seems too small
 - Text is not set in Lato, please update
 - The color palette is suboptimal
 - limit the lower range of colors
 - or add outlines to points which allows to handle the issues of too light colors on a white background
 - Unclear what is shown along the axes (might be fine but unusual to show no title and ticks)

Proportion plot:

- Needs to be reimplemented in plotly
 - Text seems a bit too small
 - Overlapping text labels in the corners
 - adjust position of percentage labels
 - align annotations with panel border
 - Points are hardly visible
 - increase point size, also to make plot feel less empty (I honestly had to search for the points, thought the plot is empty at first glance)
 - use corporate blue (or another color with a high visual weight)