PyComplexHeatmap: A Python package to visualize multimodal genomics data

Wubin Ding, David Goldberg, Wanding Zhou

Children's Hospital of Philadelphia
University of Pennsylvania
Salk Institute for Biological Studies

https://github.com/DingWB/PyComplexHeatmap

Ding, W., Goldberg, D. and Zhou, W. (2023), PyComplexHeatmap: A Python package to visualize multimodal genomics data. iMeta e115. https://doi.org/10.1002/imt2.115

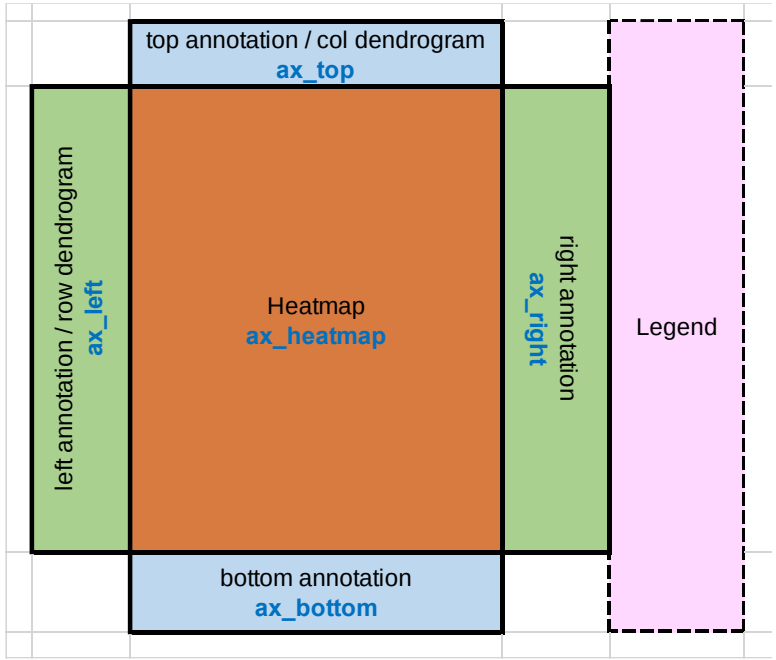
Introduction

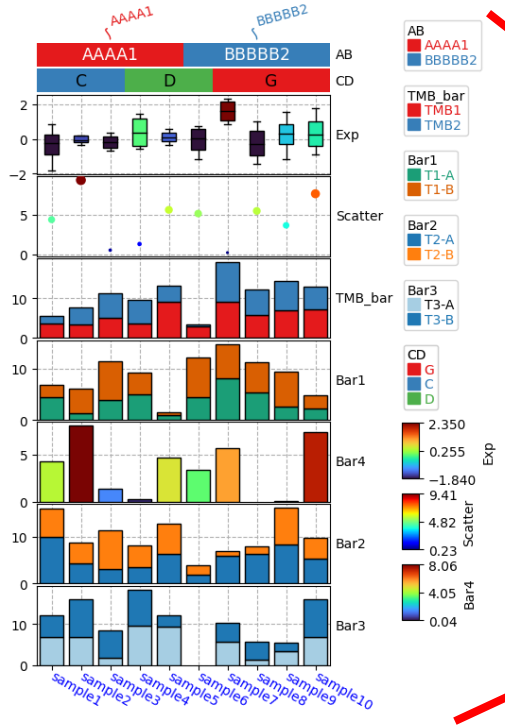
- The popularity of Python in bioinformatic single cell data analysis and **machine learning & deep learning** (SCANPY, EpiScanpy, scVelo, PyDESeq2, GSEApy and so on).

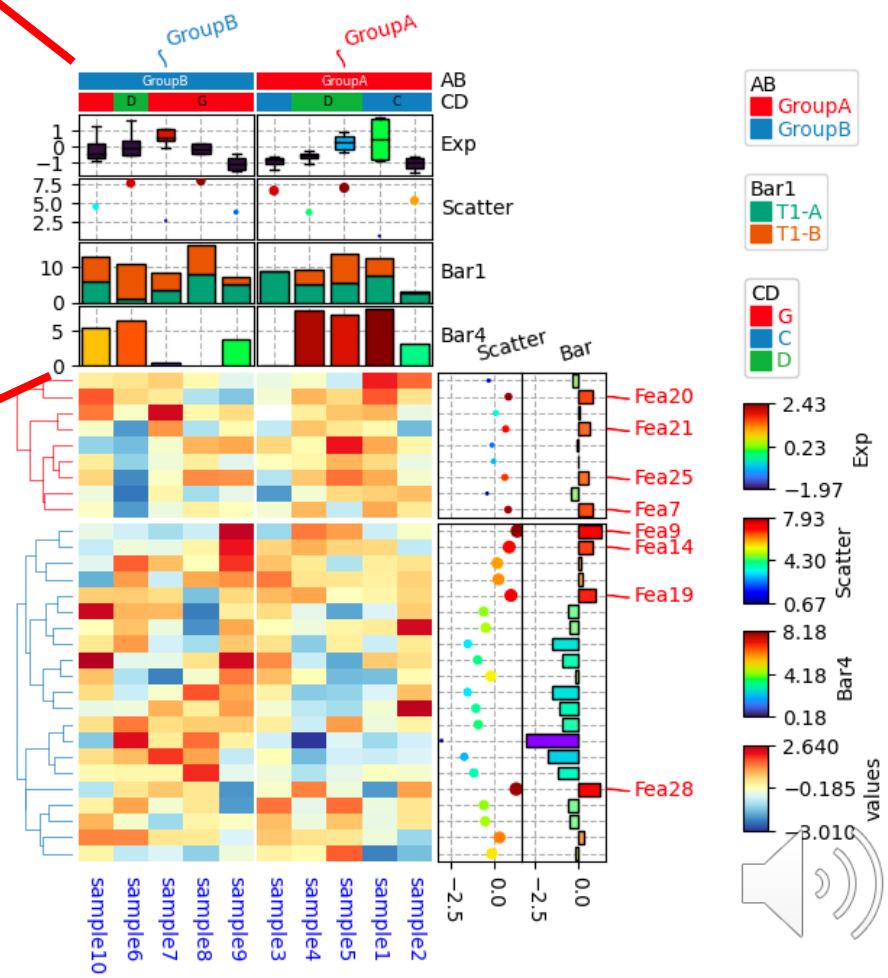
- Compared with R, what hinders the application of python in single-cell data analysis is data visualization, especially for the complex heatmaps.

- To fill this gap in Python, we developed a Python package, PyComplexHeatmap, which allow Python user to visualize **multidimensional biological data** easily.

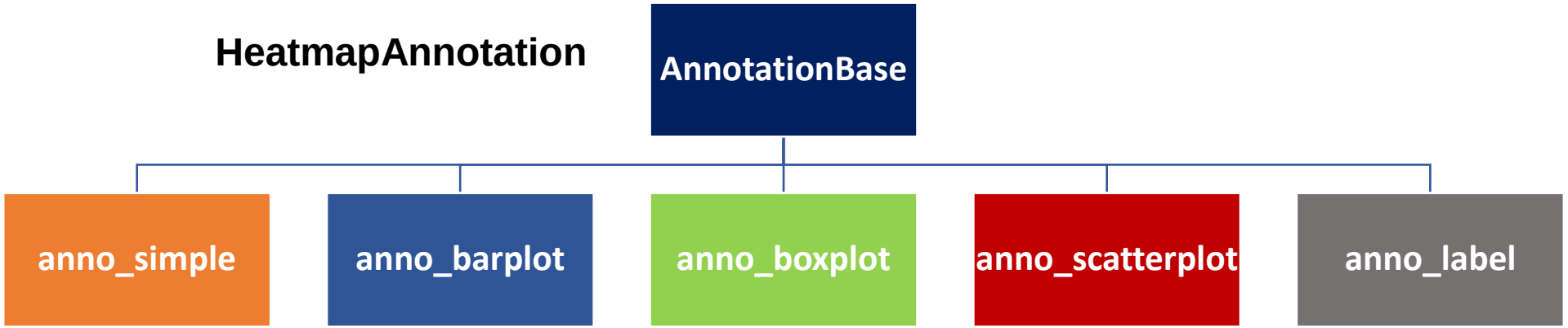
Layout

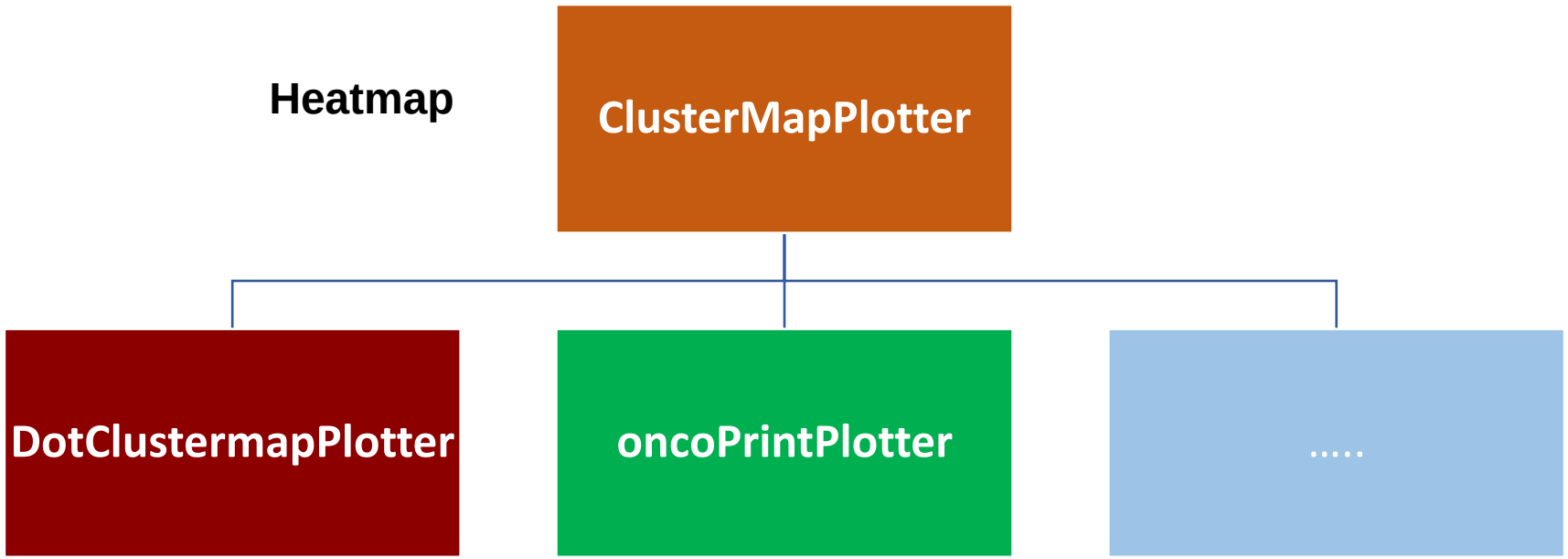

top annotation / col dendrogram
ax_top
left annotation / row dendrogram
ax_left
Heatmap
ax_heatmap
right annotation
ax_right
Legend
bottom annotation
ax_bottom


AAAA1
BBBBB2
AB
CD
Exp
Scatter
TMB_bar
Bar1
Bar4
Bar2
Bar3
sample1
sample2
sample3
sample4
sample5
sample6
sample7
sample8
sample9
sample10


GroupB
GroupA
AB
CD
Exp
Scatter
Bar1
Bar4
Bar
Fea20
Fea21
Fea25
Fea7
Fea9
Fea14
Fea19
Fea28
sample10
sample6
sample7
sample8
sample9
sample3
sample4
sample5
sample1
sample2
values

Implementation



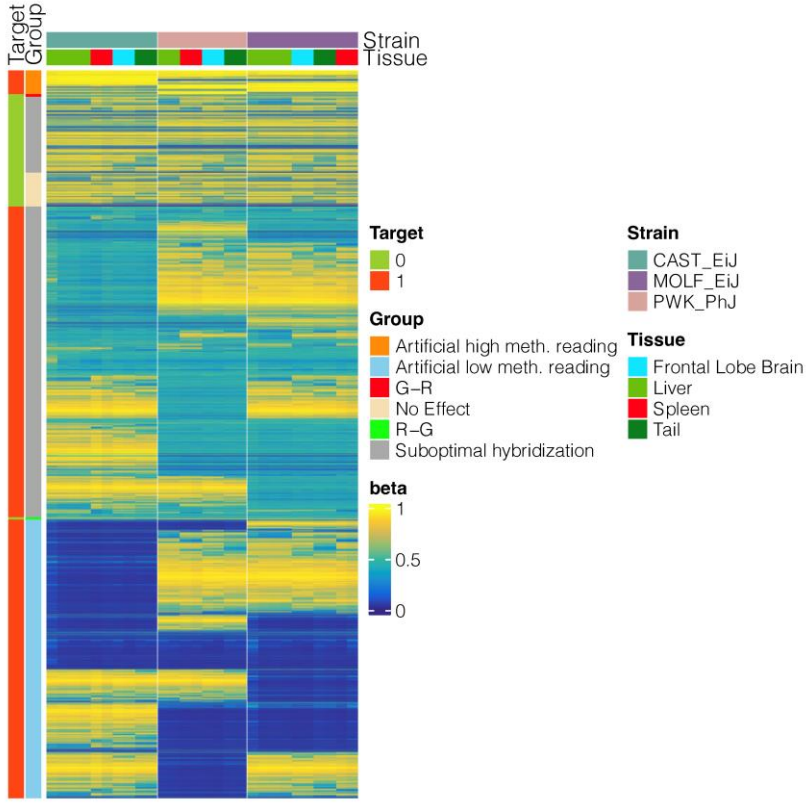


Compared with other Python packages

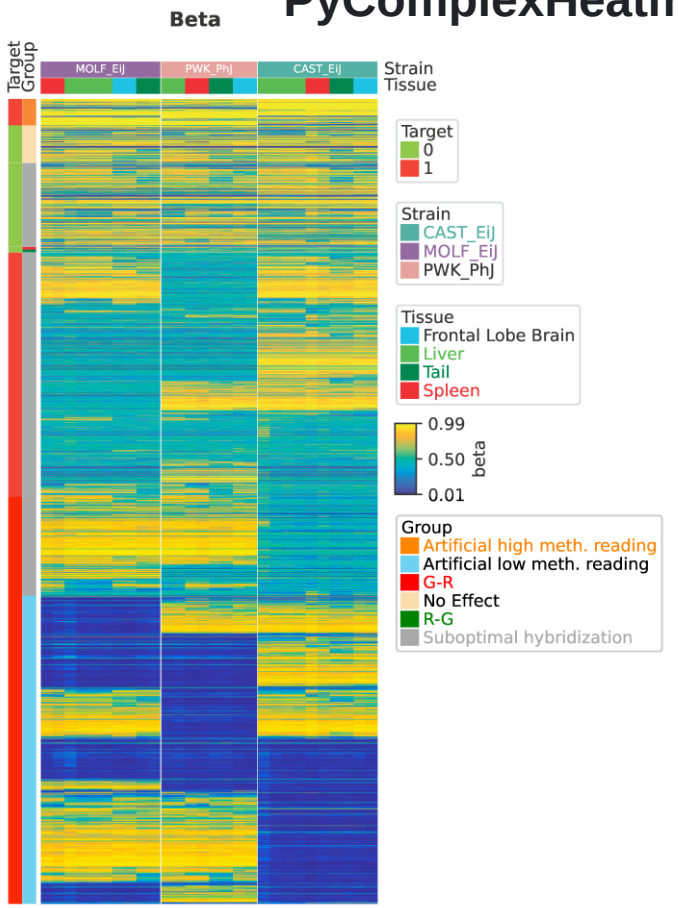
	PyComplexHeatmap	matplotlib	seaborn	ComplexHeatmap	Description
Version	1.3.8	3.7.1	0.12	2.13.1	Test version
URL	https://github.com/DingWB/PyComplexHeatmap	https://github.com/matplotlib/matplotlib	https://github.com/mwaskom/seaborn	https://github.com/jokergoo/ComplexHeatmap	Github link
Language	Python	Python	Python	R	Programming Language
Row / Columns scale	✓	×	✓	×	Scale rows or columns using standard scale or z score
Clustering	✓	×	✓	✓	Clustering rows and columns using different clustering methods and metrics
Heatmap Split	✓	×	×	✓	Split rows or columns using int number or given annotations
Support Missing Values	✓	×	×	✓	Whether support matrix containing NaN (missing values)
Heatmap Annotations	✓	×	custom function	✓	Add different kinds of annotations, including simple annotation, barplot, boxplot and scatterplot annotations
Text Annotations	Automatically	×	×	anno_block	Add text on the top of simple annotations
Float Annotated Heatmaps	✓	custom function	✓	custom function	Add float values on the top of heatmap cells
Heatmap Splicing	✓	×	×	✓	Combine multiple heatmap vertically or horizontal to generate a joint figure
Support Dot Heatmap	✓	×	✓	custom function	Plot dot heatmap using different size, colormap and marker for different categories
Dot Heatmap Annotations	✓	×	custom function	✓	Add rows/cols annotations to scatterplot heatmap
Plot Legends Separately	✓	×	×	packLegend	Plot the figure legends together or separately in another figure
Plot Annotations Separately	✓	×	×	✓	Plot the heatmap annotations together or separately in another figure
OncoPrint	✓	×	×	✓	Support oncoPrint
Label Annotations	Automatically Distributed	×	×	Manually Distributed	Add labels annotations to show the interested rows or columns and merge the labels of the rows/cols belonging to the same group and distribute them evenly throughout the axis without overlapping

Compared with R ComplexHeatmap

ComplexHeatmap



PyComplexHeatmap

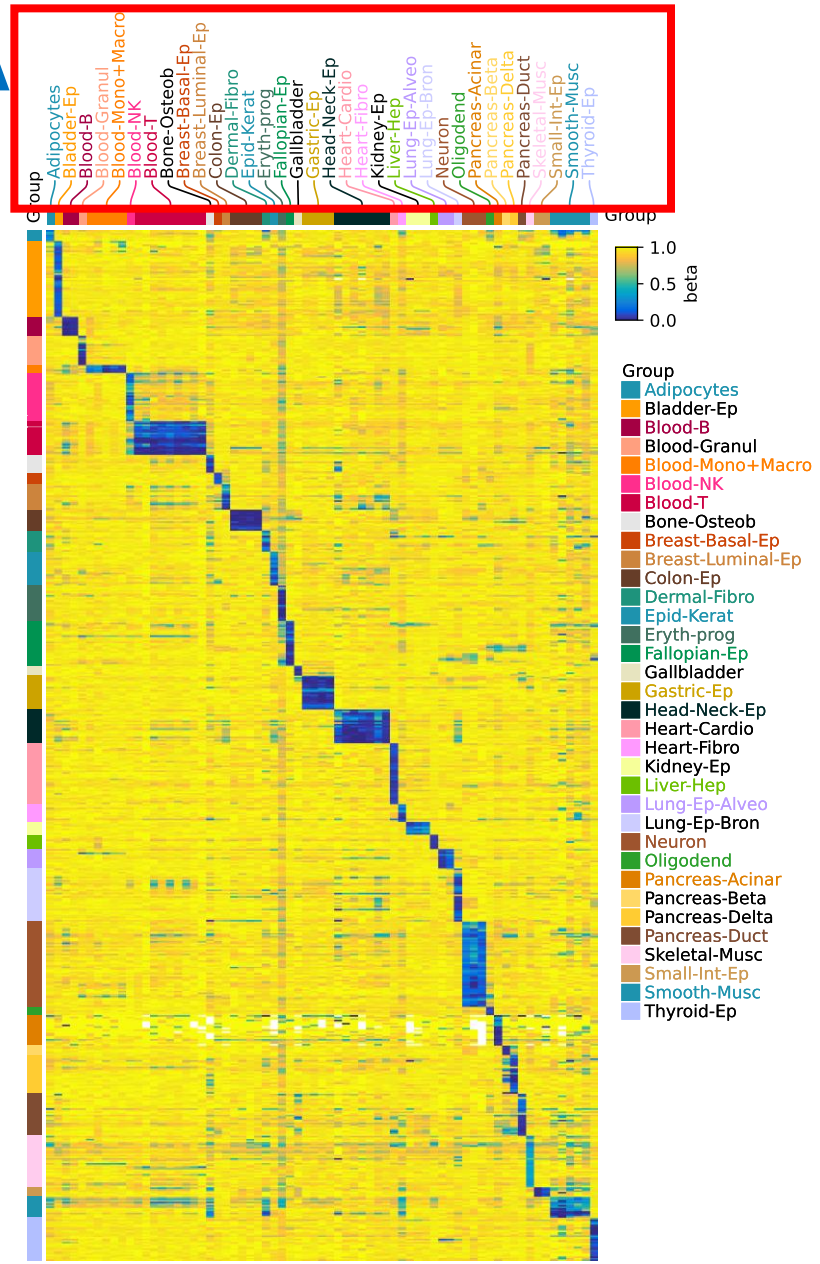


Package Name	Processing Time (s)	Memory (kb)
ComplexHeatmap	40.21	3,366,768
PyComplexHeatmap	22.57	1,037,944

Applications

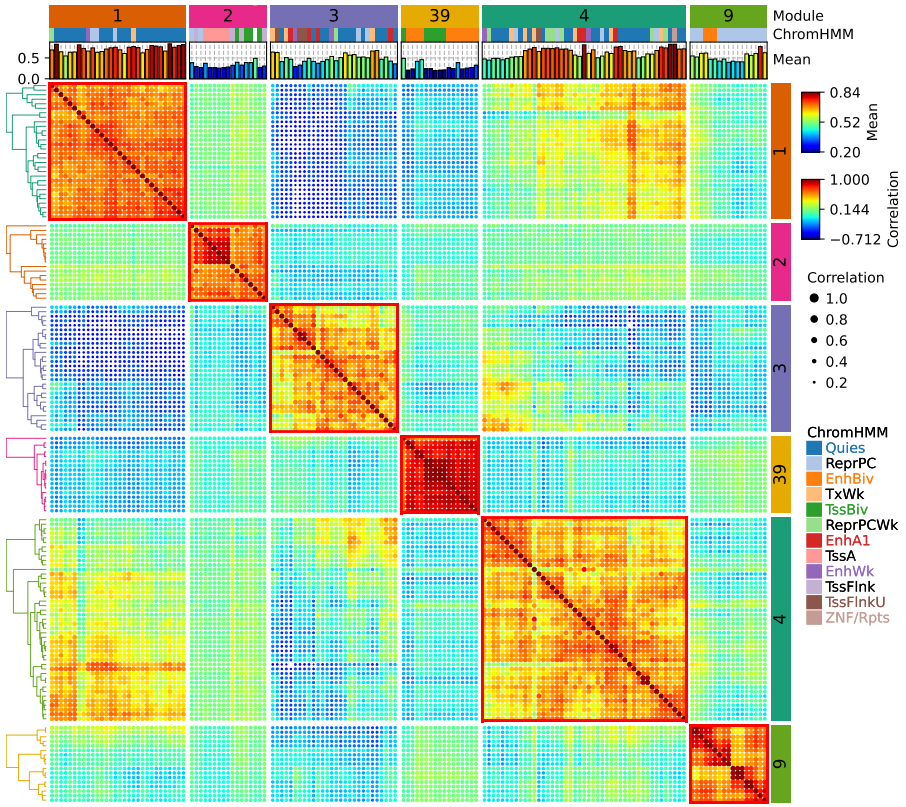
- **Visualization of cell type specific DNA methylation markers**

```
col_colors_dict={
                'Adipocytes':'#1E93AE','Bladder-Ep':'#FF9C00','Blood-B':'#A40043','Blood-Granul':'#FF9F
                'Blood-Mono+Macro':'#FF7F00','Blood-NK':'#FF2E8D','Blood-T':'#CC0043','Bone-Osteob':'#E
                'Breast-Basal-Ep':'#CC4407','Breast-Luminal-Ep':'#CC843D','Colon-Ep':'#663D28','Dermal-
                'Epid-Kerat':'#1E93AE','Eryth-prog':'#40705F','Fallopian-Ep':'#009351','Gallbladder':'#
                'Gastric-Ep':'#CCA300','Head-Neck-Ep':'#002929','Heart-Cardio':'#FF99AA','Heart-Fibro':
                'Kidney-Ep':'#F6FF99','Liver-Hep':'#6CBF00','Lung-Ep-Alveo':'#BA99FF','Lung-Ep-Bron':'#
                'Neuron':'#9e542e','Oligodend':'#2ca02c','Pancreas-Acinar':'#DF7F00','Pancreas-Beta':'#
                'Pancreas-Delta':'#FFCC32','Pancreas-Duct':'#7F4C33','Skeletal-Musc':'#FFCCEE','Small-I
                'Smooth-Musc':'#1E93AE','Thyroid-Ep':'#B2BFFF'}

col_ha = HeatmapAnnotation(label=anno_label(df_col['Group'],merge=True,rotation=90,extend=True,
                                            colors=col_colors_dict,adjust_color=True,luminance=0.75,
                                            relpos=(0.5,0)), #fontsize=10
                           Group=anno_simple(df_col['Group'],colors=col_colors_dict), #legend_kws={'fon
                           verbose=0,axis=1)
row_ha = HeatmapAnnotation(
                           Group=anno_simple(df_row['Group'],legend=True,
                                             colors=col_ha.annotations[1].color_dict),
                           verbose=0,axis=0,plot_legend=False) #label_kws={'rotation':90,'rotation_mode

plt.figure(figsize=(6, 10))
cm = ClusterMapPlotter(data=data.loc[df_row.index.tolist(),df_col.index.tolist()],
                       top_annotation=col_ha, left_annotation=row_ha,
                       row_cluster=False,col_cluster=False,
                       label='beta', row_dendrogram=False,legend_gap=7,
                       cmap='parula',rasterized=True)
plt.savefig("Loyfer2023_heatmap.pdf",bbox_inches='tight')
plt.show()
```

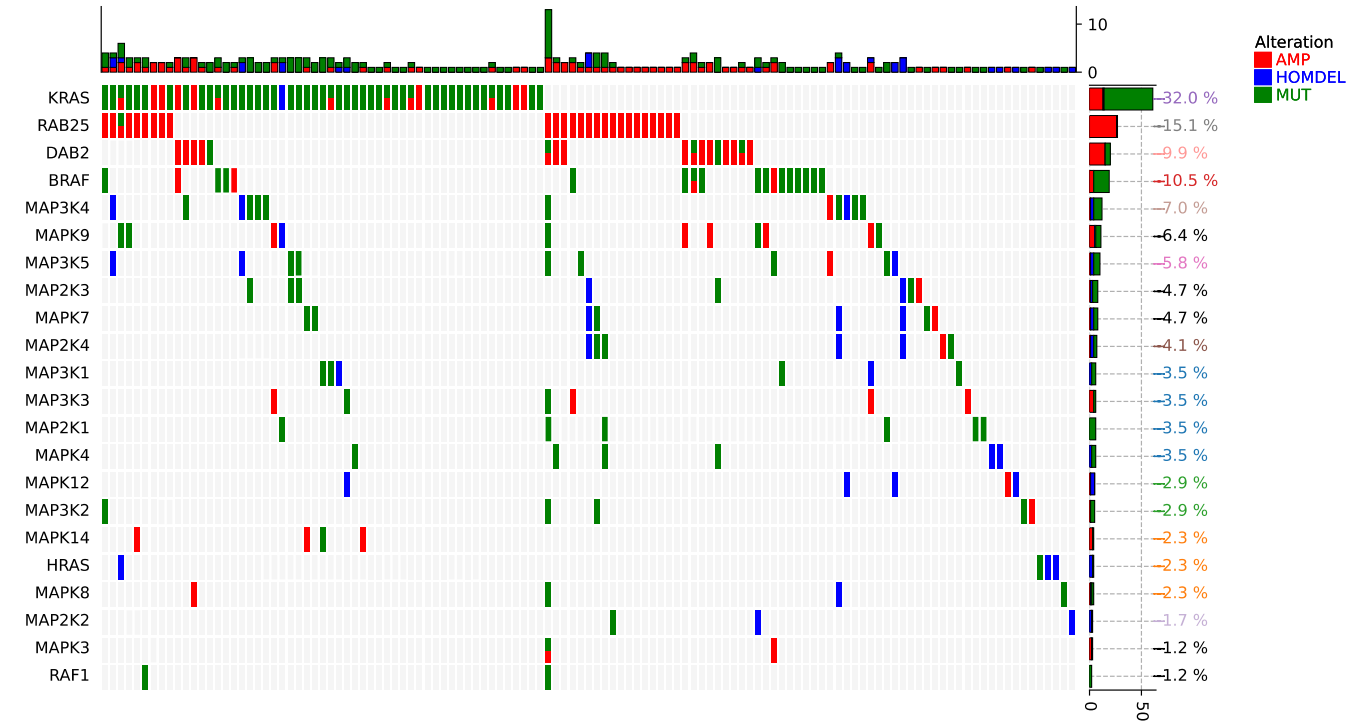


Applications

- **Dot heatmap**



- **OncoPrint**



Applications

- **Combine two heatmaps horizontally**

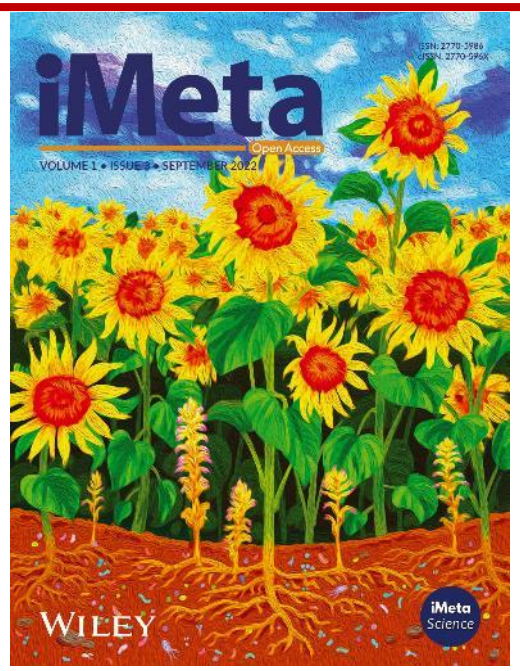
Conclusions

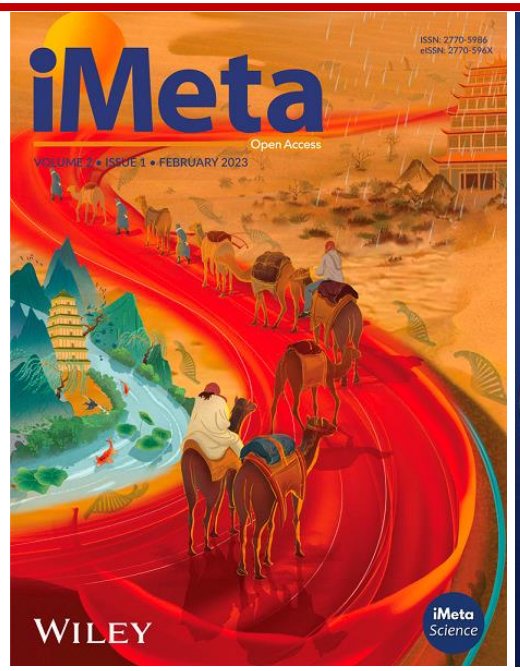
- We introduced PyComplexHeatmap, a versatile and user-friendly Python package, to fill the multidimensional data visualization gap in the Python-based data science ecosystem.
- We showcased the main features of PyComplexHeatmap in rendering complex biological datasets with rich annotations. Our benchmark indicated improved computational efficiency over the R implementation.
- We demonstrated its power in advanced genomics data analysis, including rendering the OncoPrint and dot heatmap.

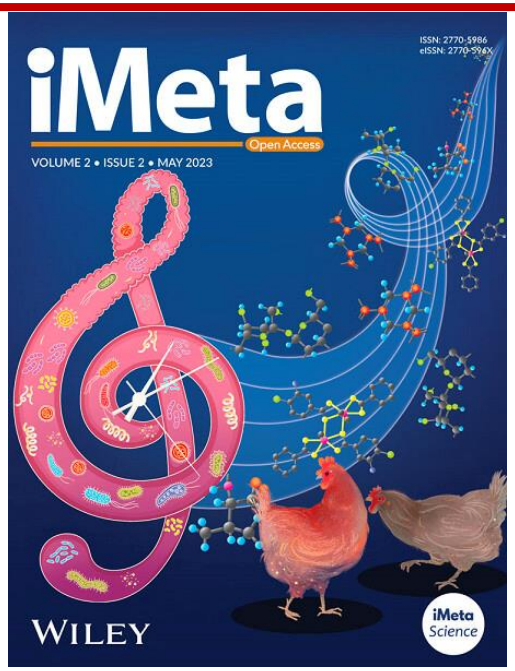
Ding, W., Goldberg, D. and Zhou, W. (2023), PyComplexHeatmap: A Python package to visualize multimodal genomics data. iMeta e115. https://doi.org/10.1002/imt2.115

iMeta: Integrated meta-omics to change the understanding of the biology and environment

WILEY







"*iMeta*" is an open-access Wiley partner journal launched by iMeta Science Society consist of scientists in bioinformatics and metagenomics world-wide. iMeta aims to promote microbiome, and bioinformatics research by publishing research, methods/protocols, and reviews. The goal is to publish high-quality papers (top 10%, IF > 15) targeting a broad audience. Unique features include video submission, reproducible analysis, figure polishing, APC waiver, and promotion by social media with 500,000 followers. Four issues were released in March, June, September, and December 2022. Index by Google Scholar, Crossref, Dimensions, PubMed(partial), DOAJ and Scopus.

Society: http://www.imeta.science
Publisher: https://wileyonlinelibrary.com/journal/imeta
Submission: https://mc.manuscriptcentral.com/imeta

office@imeta.science
Promotion Video

iMetaScience
iMetaScience