

Update on Anndata Extension: Bioframe

Qi An

2023.02.13

This week: dive into package `Bioframe`

- Bioframe is pandas based operation and acts faster than `PyRanges`
- It returns a filtered item based on given ranges
- Supporting multiple headers
- Easy to work on aggregations

```
import bioframe
bed_column_names = ("chromosome_name", "start_position", "end_position")

query_result = bioframe.select(new_var, "4:0-10000000", cols=bed_column_names)
query_result.head()
```

Problem

- Need to input a data frame with ['chrom', 'start', 'end']
- In AnnData object, coordinates are usually stored in a separate `varm` attribute
- Need to combine the data with coordinate into a temporal new DataFrame

```
new_var = (  
    pbmc.var  
    # Merging  
    .join(gene_positions.set_index("hgnc_symbol"), how="left")  
    # Setting original order  
    .loc[pbmc.var_names]  
)
```

Export

- Export format: DataFrame
- Need to slice on the original AnnData object with DataFrame's index