

main

July 10, 2021

1 Metrics summary for splicing

```
[1]: import numpy as np
import pandas as pd
```

```
[2]: df = pd.read_csv('../_m/cluster_ds_results_annotated.txt', sep='\t')
df['chr'] = df.coord.str.split(":", expand=True)[0]
df.head(2)
```

```
[2]:
```

	clusterID	N	coord	gene	annotation	FDR	chr
0	clu_755_-	3	chrX:53217966-53220839	KDM5C	annotated	2.270000e-43	chrX
1	clu_744_-	10	chrX:53176622-53193437	KDM5C	cryptic	3.640000e-40	chrX

```
[3]: print("There are %d unique DS!" % len(df.clusterID.unique()))
print("There are %d unique DS (gene name)!" % len(df.gene.unique()))
```

There are 352 unique DS!

There are 306 unique DS (gene name)!

```
[4]: dfx = df[(df['chr']=='chrX')].copy()
print("There are %d unique DS for allosomes!" % len(dfx.clusterID.unique()))
print("There are %d unique DS (gene name) for allosomes!" % len(dfx.gene.
→unique()))
```

There are 22 unique DS for allosomes!

There are 16 unique DS (gene name) for allosomes!

```
[5]: print("There are %d unique DS for autosomes!" % (len(df.clusterID.unique()) -
→len(dfx.clusterID.unique())))
print("There are %d unique DS (gene name) for autosomes!" % (len(df.gene.
→unique()) - len(dfx.gene.unique())))
```

There are 330 unique DS for autosomes!

There are 290 unique DS (gene name) for autosomes!

```
[ ]:
```