Survival

This is an example of tab with event free survival, dead and gene expression (in this case is rlog)

	Event.Free.Survival.Time.in.Days	Gender	Dead	CDKN1A	GADD45A	MKNK2	HES4
TARGET-10-PARASZ	3087.0	1	0.0	9.606897	9.110655	11.978452	10.382673
TARGET-10-PARAYM	3399.0	2	0.0	5.644741	8.004702	9.967753	8.635177
TARGET-10-PARCVM	2424.0	2	0.0	7.610549	8.393703	10.079970	6.962121
TARGET-10-PAREGZ	3087.0	1	0.0	9.739366	10.607096	10.670771	10.347120
TARGET-10-PARFDL	3075.0	2	0.0	6.246289	8.718152	10.055517	8.054267
...	...	...	...	...	...	...	...
TARGET-10-PAUBPY	1837.0	2	0.0	8.833777	10.198986	10.907244	10.198394
TARGET-10-PAUBTC	119.0	1	0.0	6.933120	8.483856	10.641391	8.906693
TARGET-10-PAUBXP	1747.0	2	0.0	7.411272	8.953100	10.271347	9.853103
TARGET-10-PAUCDC	1867.0	2	0.0	7.386351	8.208279	9.987410	6.247797
TARGET-10-PAUCDY	1793.0	2	0.0	7.827907	8.553985	10.687724	6.868094

262 rows × 7 columns

Convert in z score the value of genes

	CDKN1A	GADD45A	MKNK2	HES4
TARGET-10-PARASZ	1.489123	0.028308	2.493273	1.080664
TARGET-10-PARAYM	-2.612981	-1.519815	-1.609644	-0.049129
TARGET-10-PARCVM	-0.577739	-0.975288	-1.380659	-1.130794
TARGET-10-PAREGZ	1.626270	2.123041	-0.175104	1.057679
TARGET-10-PARFDL	-1.990186	-0.521121	-1.430556	-0.424699
...	...	...	...	...
TARGET-10-PAUBPY	0.688695	1.551764	0.307427	0.961524
TARGET-10-PAUBTC	-1.279096	-0.849091	-0.235056	0.126413
TARGET-10-PAUBXP	-0.784055	-0.192239	-0.990147	0.738286
TARGET-10-PAUCDC	-0.809856	-1.234846	-1.569532	-1.592619
TARGET-10-PAUCDY	-0.352704	-0.750924	-0.140513	-1.191584

convert the value in 0,1,2 respectivly low, med, high expression (quantile method)

	Event.Free.Survival.Time.in.Days	Gender	Dead	CDKN1A	GADD45A	MKNK2	HES4
TARGET-10-PARASZ	3087.0	1	0.0	2.0	1.0	2.0	2.0
TARGET-10-PARAYM	3399.0	2	0.0	0.0	0.0	0.0	1.0
TARGET-10-PARCVM	2424.0	2	0.0	1.0	0.0	0.0	0.0
TARGET-10-PAREGZ	3087.0	1	0.0	2.0	2.0	1.0	2.0
TARGET-10-PARFDL	3075.0	2	0.0	0.0	1.0	0.0	1.0
...	...	...	...	...	...	...	...
TARGET-10-PAUBPY	1837.0	2	0.0	2.0	2.0	1.0	2.0
TARGET-10-PAUBTC	119.0	1	0.0	0.0	0.0	1.0	1.0
TARGET-10-PAUBXP	1747.0	2	0.0	0.0	1.0	0.0	1.0
TARGET-10-PAUCDC	1867.0	2	0.0	0.0	0.0	0.0	0.0
TARGET-10-PAUCDY	1793.0	2	0.0	1.0	0.0	1.0	0.0

262 rows × 7 columns

an example of algorithm in python

```python
for i in  list(genes_apopt.columns):  #list gene

    # convert value of gene expression in 0,1,2 using quantile method
    data_surv_zscore.loc[data_surv_zscore[i] >
float(data_surv_zscore[i].quantile([0.75])),i] = 2 # > quartile 0.75

    data_surv_zscore.loc[((data_surv_zscore[i] >
float(data_surv_zscore[i].quantile([0.25]))) & #0.25 to 0.75 quartile
                (data_surv_zscore[i] <
float(data_surv_zscore[i].quantile([0.75])))),i]=1

    data_surv_zscore.loc[data_surv_zscore[i] <
float(data_surv_zscore[i].quantile([0.25])),i] = 0 # < 0.25
    # Fit Kelpen Meier
    kmf_high = KaplanMeierFitter()
    kmf_med = KaplanMeierFitter()
    kmf_low = KaplanMeierFitter()

    #create variable samples with high, med and low
    High_gene=data_surv_zscore.query(f"{i} == 2")
    Med_gene=data_surv_zscore.query(f"{i} == 1")
    Low_gene=data_surv_zscore.query(f"{i} == 0")

    #Fitting
```

```python
    kmf_high.fit(durations=High_gene["Event.Free.Survival.Time.in.Days"],event_o
bserved=High_HES4["Dead"], label=f"{i}_high")

    kmf_med.fit(durations=Med_gene["Event.Free.Survival.Time.in.Days"],event_obs
erved=Med_HES4["Dead"], label=f"{i}_med")

    kmf_low.fit(durations=Low_gene["Event.Free.Survival.Time.in.Days"],event_obs
erved=Low_HES4["Dead"], label=f"{i}_low")

    #High vs MED #stats significative
    print(f"High vs Med of {i} gene")
    T=High_gene["Event.Free.Survival.Time.in.Days"]
    E=High_gene["Dead"]
    T1=Med_gene["Event.Free.Survival.Time.in.Days"]
    E1=Med_gene["Dead"]
    results= logrank_test(T,T1,event_observed_A=E,event_observed_B=E1)
    results.print_summary()

    #High vs Low #stats significative
    print(f"\n\n High vs Low of {i} gene")
    T=High_gene["Event.Free.Survival.Time.in.Days"]
    E=High_gene["Dead"]
    T1=Low_gene["Event.Free.Survival.Time.in.Days"]
    E1=Low_gene["Dead"]
    results= logrank_test(T,T1,event_observed_A=E,event_observed_B=E1)
    results.print_summary()

    #Med vs Low
    print(f"\n\n Med vs Low of {i} gene")
    T=Med_gene["Event.Free.Survival.Time.in.Days"]
    E=Med_gene["Dead"]
    T1=Low_gene["Event.Free.Survival.Time.in.Days"]
    E1=Low_gene["Dead"]
    results= logrank_test(T,T1,event_observed_A=E,event_observed_B=E1)
    results.print_summary()

    #create and show plot
    kmf_high.plot()
    kmf_med.plot()
    kmf_low.plot()
    plt.gcf().set_dpi(150)
    plt.xlabel("Days passed")
    plt.ylabel("Survival Probability")
```

```
plt.title("KMF")
plt.show()
```