GMQL

Simone Mosciatti

April 2, 2017

1 Exercise 1

```
# Among the different epigenetic signals, H3K27ac and H3K27me3 are generally
# associated with active and repressed chromatin regions, respectively.
# Considering H3K4me1 in cell line A549 and the aforementioned signals
# (H3K27ac and H3K27me3) in broadPeak format, under ethanol treatment (EtOH)

K4me1EtOH  = SELECT(cell == 'A549' AND treatment == 'EtOH_0.02pct' AND antibod
K27acEtOH  = SELECT(cell == 'A549' AND treatment == 'EtOH_0.02pct' AND antibod
K27me3EtOH = SELECT(cell == 'A549' AND treatment == 'EtOH_0.02pct' AND antibod

Promoters  = SELECT(annotation_type == 'promoter') HG19_BED_ANNOTATION;

ActiveUnion = UNION() K4me1EtOH K27acEtOH;
ActiveOverlapping = COVER(2, ANY) ActiveUnion;
Active = MAP() ActiveUnion ActiveOverlapping;

RepressedUnion = UNION() K4me1EtOH K27me3EtOH;
RepressedOverlapping = COVER(2, ANY) RepressedUnion;
Repressed = MAP() RepressedUnion RepressedOverlapping;

PoisedUnion = UNION() Repressed Active;
PoisedOverlapping = COVER(2, ANY) PoisedUnion;
Poised = MAP() PoisedUnion PoisedOverlapping;

ActiveInPromotersUnion = UNION() Active Promoters;
ActiveInPromotersOverlapping = COVER(2, ANY) ActiveInPromotersUnion;
ActiveInPromoters = MAP() ActiveInPromotersUnion ActiveInPromotersOverlapping

ClosestK4me1FartherThan10k = JOIN(DGE(10000), MD(1); output: cat) ActiveInPro

MATERIALIZE ClosestK4me1FartherThan10k INTO Result;
```

This job runned in about 2 minutes (113 sec), I obtained three sample, without duplicate. The join is expliciting asking for one element, father than 10000 from the anchor in "ActiveInPromoter" from the "K4me1EtOH".

2 Exercise 2

```
Sample = SELECT(antibody_target == 'JUN' AND (treatment == 'IFNa30' OR treatme
Promoters  = SELECT(annotation_type == 'promoter') HG19_BED_ANNOTATION;

SampleOverlapping = COVER(2, ANY) Sample;
SampleIntersect = MAP() Sample SampleOverlapping;

SampleAndPromotersUnion = UNION() Promoters SampleIntersect;
SampleAndPromotersOverlapping = COVER(2, ANY) SampleAndPromotersUnion;
SampleAndPromoters = MAP() Promoters SampleAndPromotersOverlapping;

MATERIALIZE SampleAndPromoters INTO Result;
```