

interaction_topic_analysis

June 17, 2022

1 Methods

1.1 Bayesian Topic modelling

Multinomial-Dirichlet:

$$p(\mathbf{y}_i|\mathbf{q}_i) = \frac{(\sum_g Y_{ig})!}{\prod_g Y_{ig}!} \prod_g q_{ig}^{Y_{ig}}$$
$$\mathbf{q}_i \sim \text{Dir}(\mathbf{q}_i|\rho_i) = \frac{\Gamma(\sum_g \rho_{ig})}{\prod_g \Gamma(\rho_{ig})} \prod_g q_{ig}^{\rho_{ig}-1}$$

Single-cell generative model:

$$p(\mathbf{x}_j|\cdot) = \frac{\Gamma(\sum_g \lambda_{jg})\Gamma(1 + \sum_g X_{jg})}{\Gamma(\sum_g \lambda_{jg} + X_{jg})} \prod_{g=1}^G \frac{\Gamma(X_{jg} + \lambda_{jg})}{\Gamma(\lambda_{jg})\Gamma(X_{jg} + 1)}$$

where

$$\lambda_{jg} = \exp\left(\sum_{t=1}^T \theta_{jt}\beta_{tg} + \delta_g\right)$$

Bayesian regularization of the model parameters

$$\beta_{tg} \sim \mathcal{N}(0, 1)$$

```
[ ]: import experiment_interaction_topic as experiment
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
warnings.simplefilter('ignore')

experiment_dir = '/projects/experiments/spruce_topic/
↳3_augmented_lr_multidir_v_beta/'
spruce = experiment.get_experiment_model(experiment_dir)
print(spruce.model_id)
```

```
/home/BCCRC.CA/ssubedi/projects/experiments/spruce_topic/3_augmented_lr_multidir_v_beta/output/interaction_topic/2022061606_ld_25_ep_500/GSE176078mix_202205112335_2022061606
```

```
[ ]: from analysis import _topics
```

```
top_n=5
df_tpgenes = _topics.topic_top_lr_genes(spruce,top_n)
df_tpgenes.to_csv(spruce.model_id+'_itopic_top_'+str(top_n)+'_lr_genes.tsv.
    ↪gz',compression='gzip')
```

```
[ ]: setwd('../scripts/')
library(yaml)
options(warn=-1)
```

```
args_home = "/home/BCCRC.CA/ssubedi/projects/experiments/spruce_topic/
    ↪3_augmented_lr_multidir_v_beta/"
```

```
config = paste(args_home,"config.yaml",sep="")
```

```
args = read_yaml(config)
```

```
model_id =
```

```
    ↪paste(args_home,args$output,args$interaction_topic$out,args$interaction_topic$model_id,sep=
```

```
[ ]: source('fig_2_lr_loss.R')
```

```
plot_loss(paste(model_id,'_itopic_lossm.txt.gz',sep=''))
```

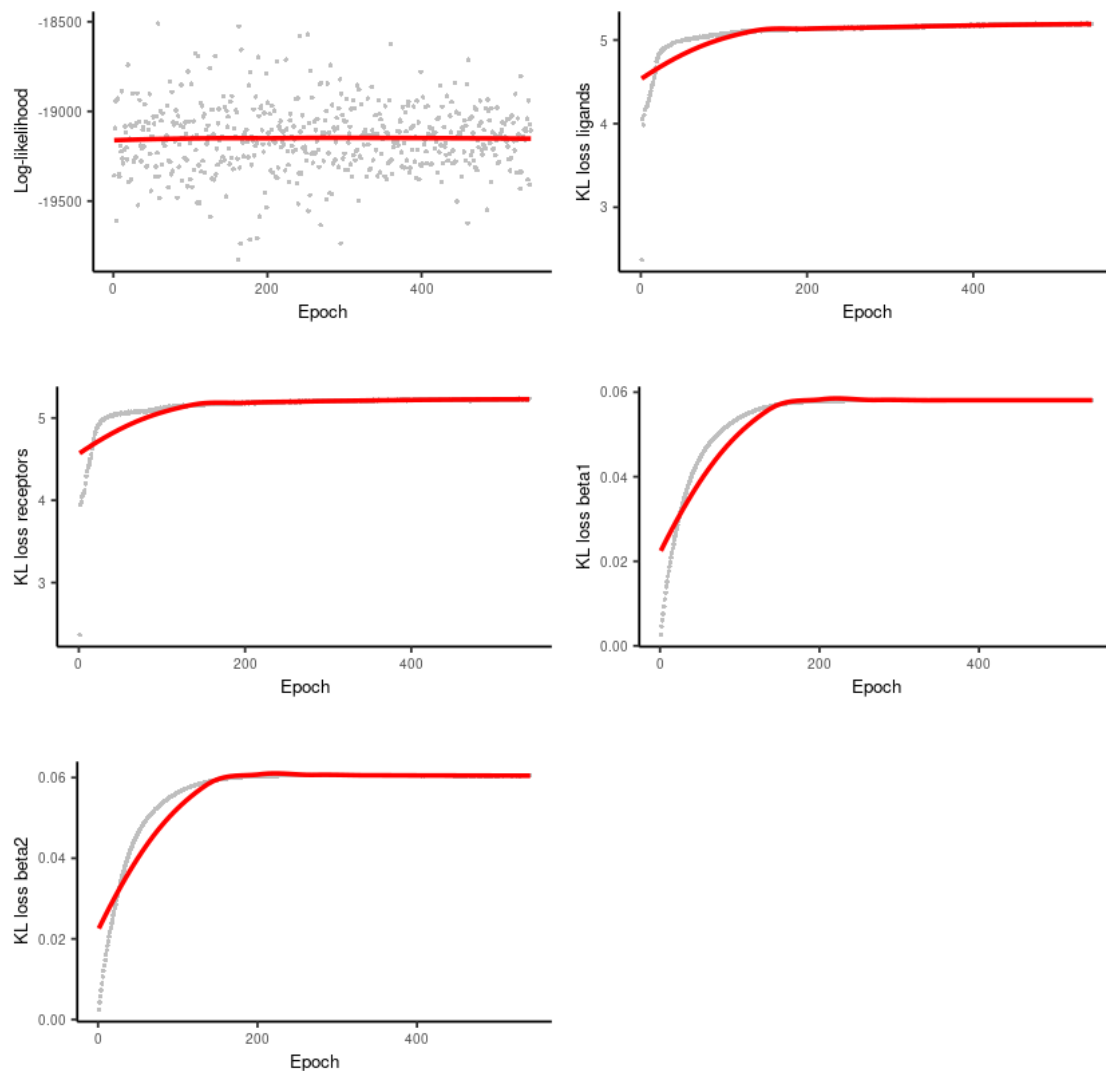
```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
[ ]: source('fig_2_lr_hmap_v2.R')
topgenes_file = paste(model_id, "_itopic_top_5_lr_genes.tsv.gz", sep="")

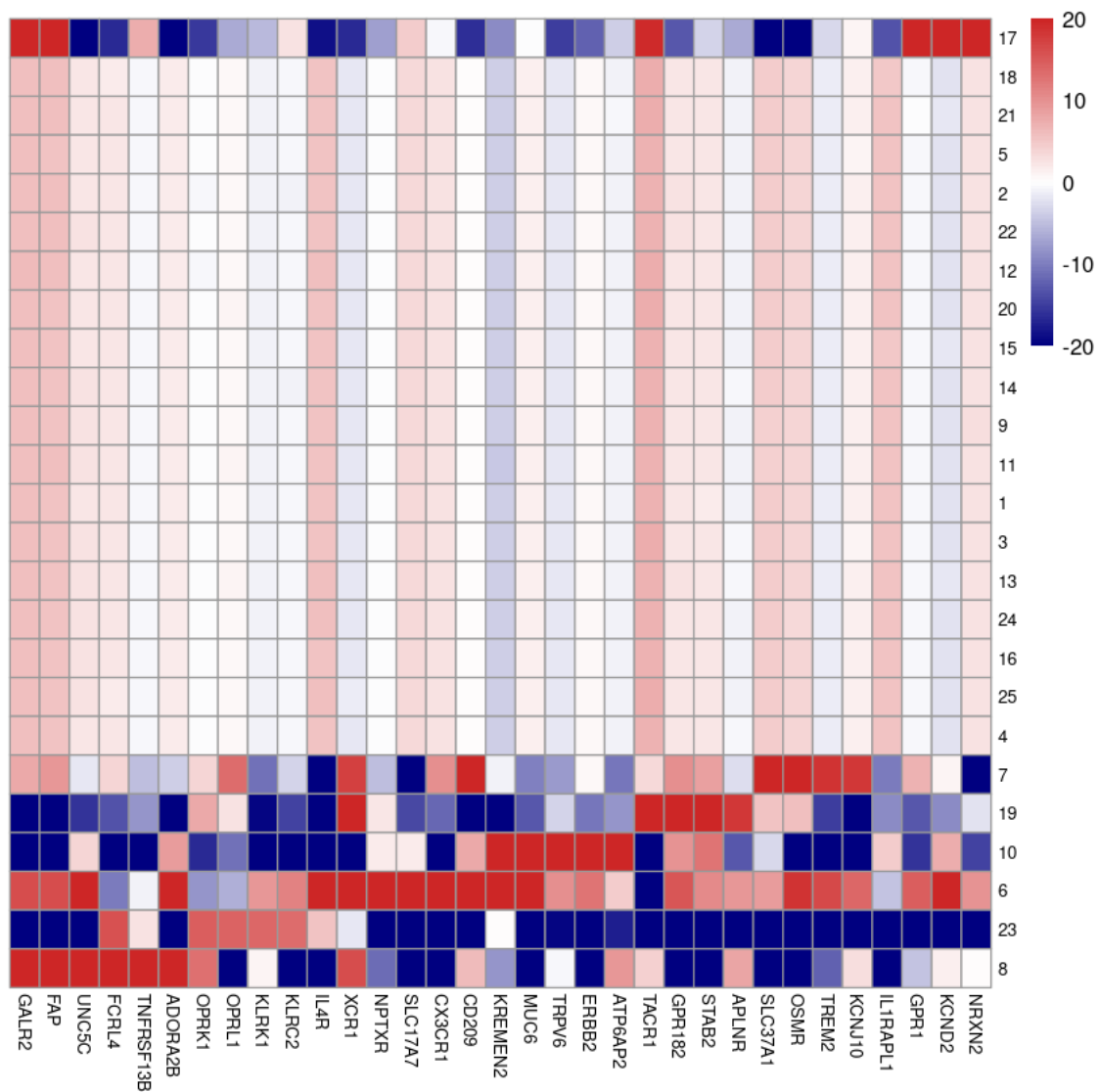
beta = paste(model_id, "_itopic_beta_1.tsv.gz", sep="")
beta_cols = read.table(paste(args$home, args$data, args$sample_id, 'receptors.csv.
  ↪gz', sep=' '), header=TRUE)
df_beta = read.table(beta, sep = "\t", header=TRUE)
colnames(df_beta) = beta_cols$X0

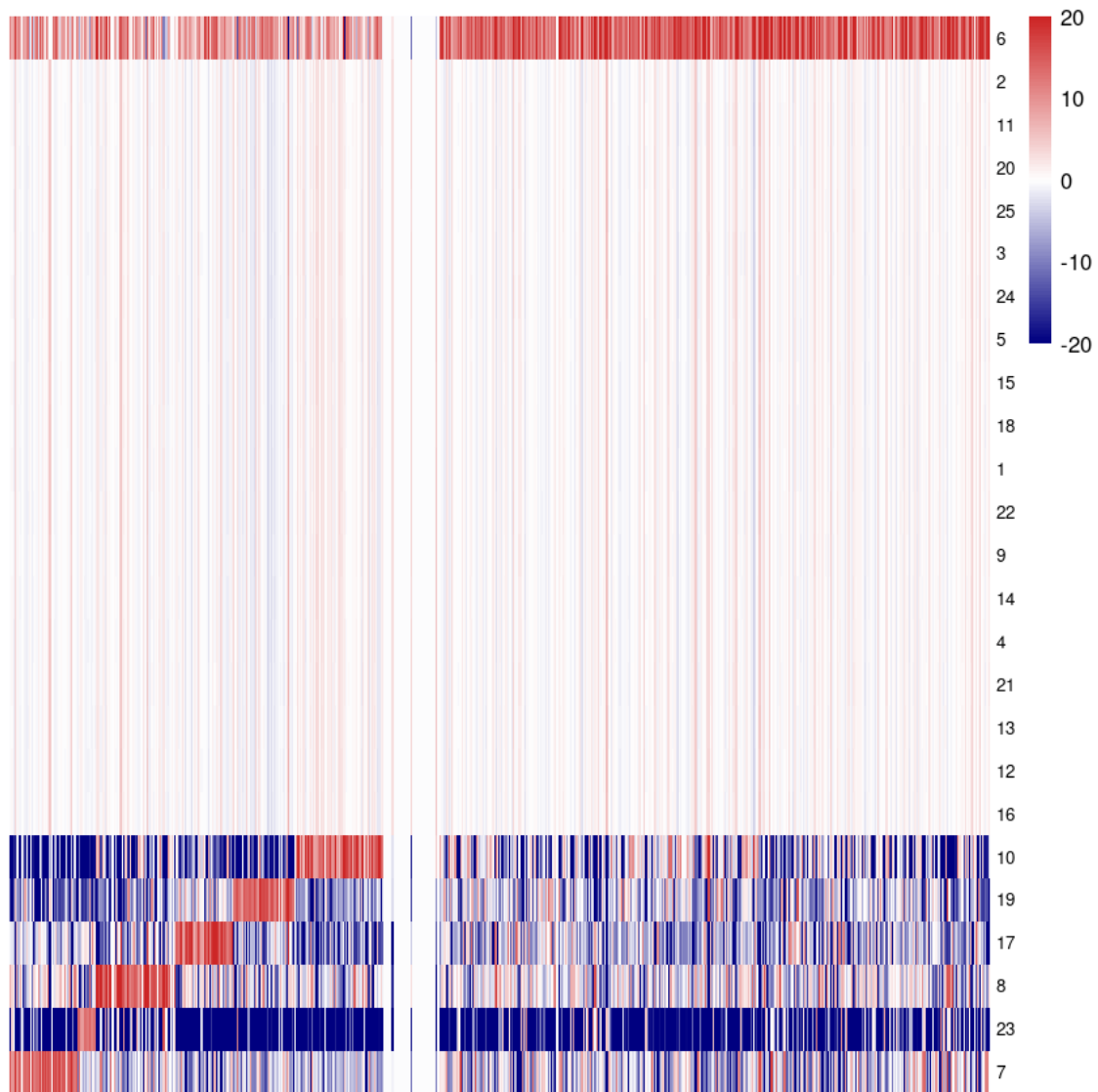
tag='receptors'
top_genes=TRUE
weightmat_lr_plot(df_beta, tag, top_genes, topgenes_file)
```

```
top_genes=FALSE
weightmat_lr_plot(df_beta,tag,top_genes,topgenes_file)
```

Using topic as id variables

Using topic as id variables





```
[ ]: beta = paste(model_id,"_itopic_beta_r.tsv.gz",sep="")
beta_cols = read.table(paste(args_home,args$data,args$sample_id,'ligands.csv.
→gz',sep=' '),header=TRUE)
df_beta = read.table(beta, sep = "\t", header=TRUE)
colnames(df_beta) = beta_cols$X0

tag='ligands'
top_genes=TRUE
weightmat_lr_plot(df_beta,tag,top_genes,topgenes_file)

top_genes=FALSE
weightmat_lr_plot(df_beta,tag,top_genes,topgenes_file)
```

Using topic as id variables

Using topic as id variables

