Single-cell Pairwise Relationships Untangled by Composite Topic models

Park Lab

09:42:45 AM, Jun 20, 2022

Cell topic analysis

Multinomial-Dirichlet:

$$p(\mathbf{y}_i|\mathbf{q}_i) = \frac{(\sum_g Y_{ig})!}{\prod_g Y_{ig}!} \prod_g q_{ig}^{Y_{ig}}$$

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

Single-cell generative model:

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

where

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

Bayesian regularization of the model parameters

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

Total Expected log-likelihood Lower-bound (ELBO):

$$\begin{aligned}
\frac{J}{n} &= \frac{1}{n}\sum_{i=1}^{n} \log p(\mathbf{x}_i|\theta_i(\mathbf{z}_i),\beta) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} D_{\mathsf{KL}}\left(q(z_{it})\|p(z_{it})\right) \\
&\quad + \frac{1}{n}\sum_{t=1}^{T}\sum_{g=1}^{G} D_{\mathsf{KL}}\left(q(\beta_{tg})\|p(\beta_{tg})\right) \\
&\approx \frac{1}{B}\sum_{i=1}^{B} \log p(\mathbf{x}_i|\theta_i(\mathbf{z}_i),\beta) \\
&\quad + \frac{1}{B}\sum_{i=1}^{B}\sum_{t=1}^{T} D_{\mathsf{KL}}\left(q(z_{it})\|p(z_{it})\right) \\
&\quad + \frac{1}{n}\sum_{t=1}^{T}\sum_{g=1}^{G} D_{\mathsf{KL}}\left(q(\beta_{tg})\|p(\beta_{tg})\right)
\end{aligned}$$

where B is the mini-batch size.

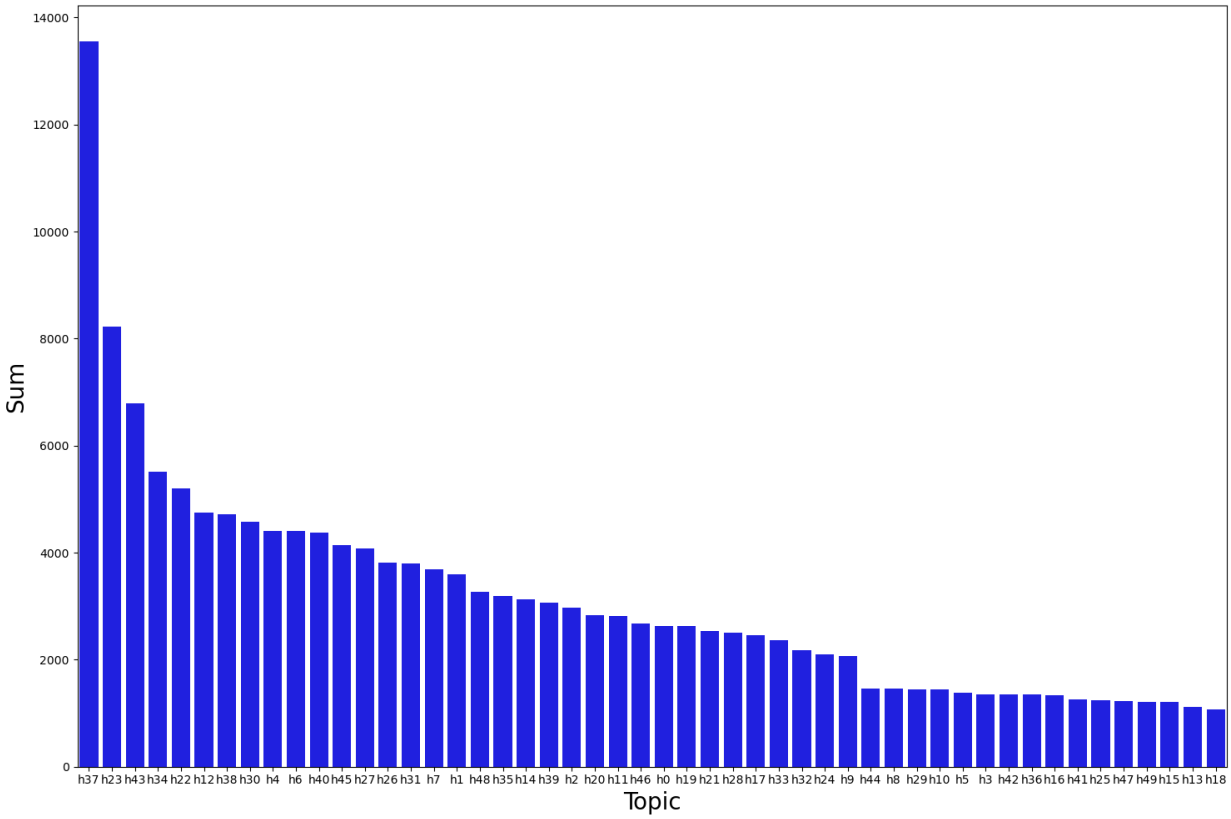

Figure 1: Topic sum

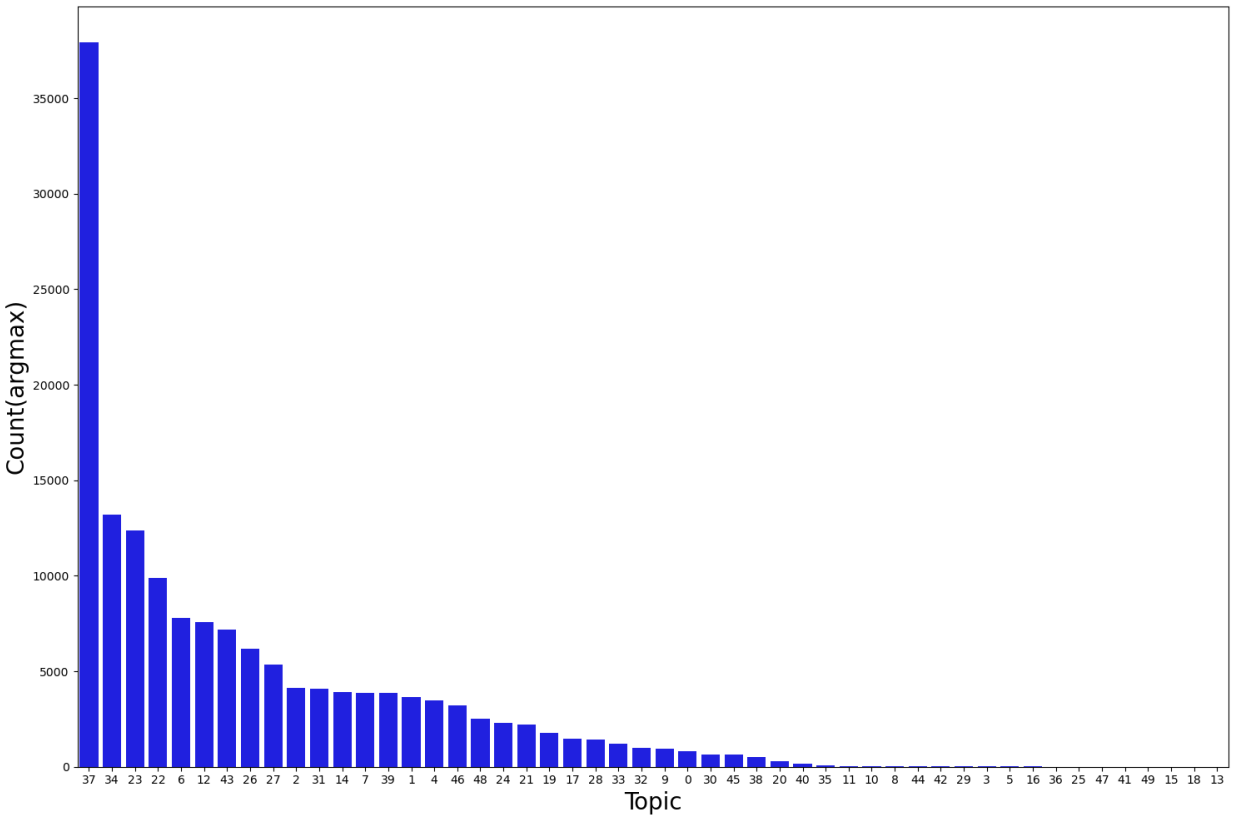

Figure 2: Topic argmax

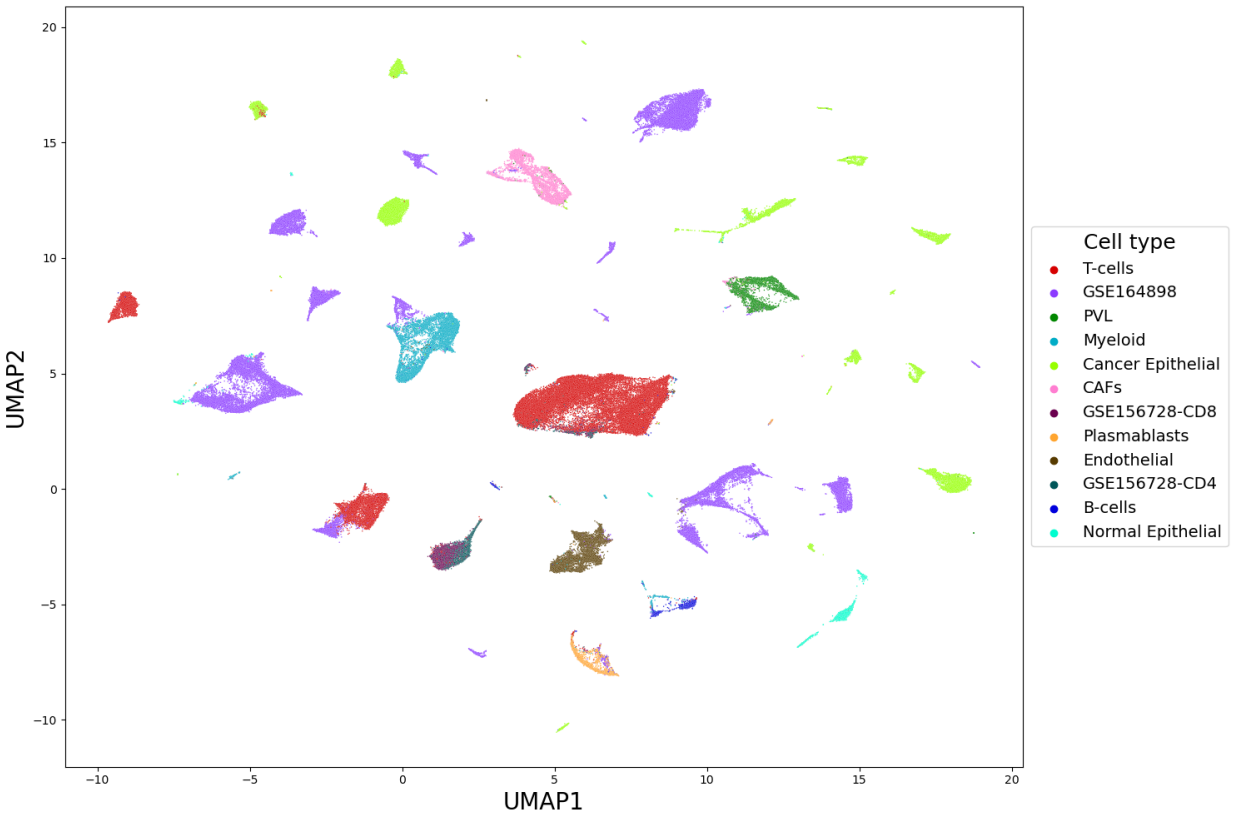

Figure 3: UMAP cell-type

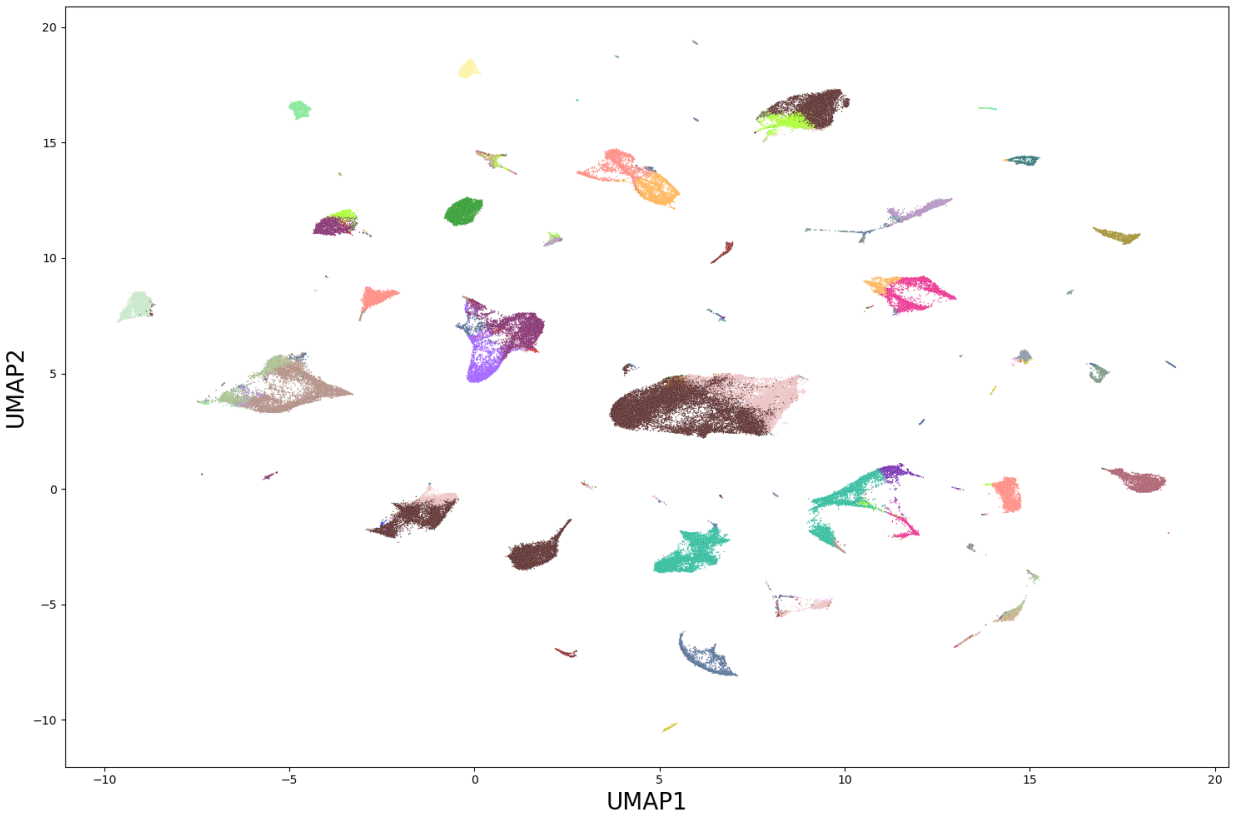

Figure 4: UMAP topic

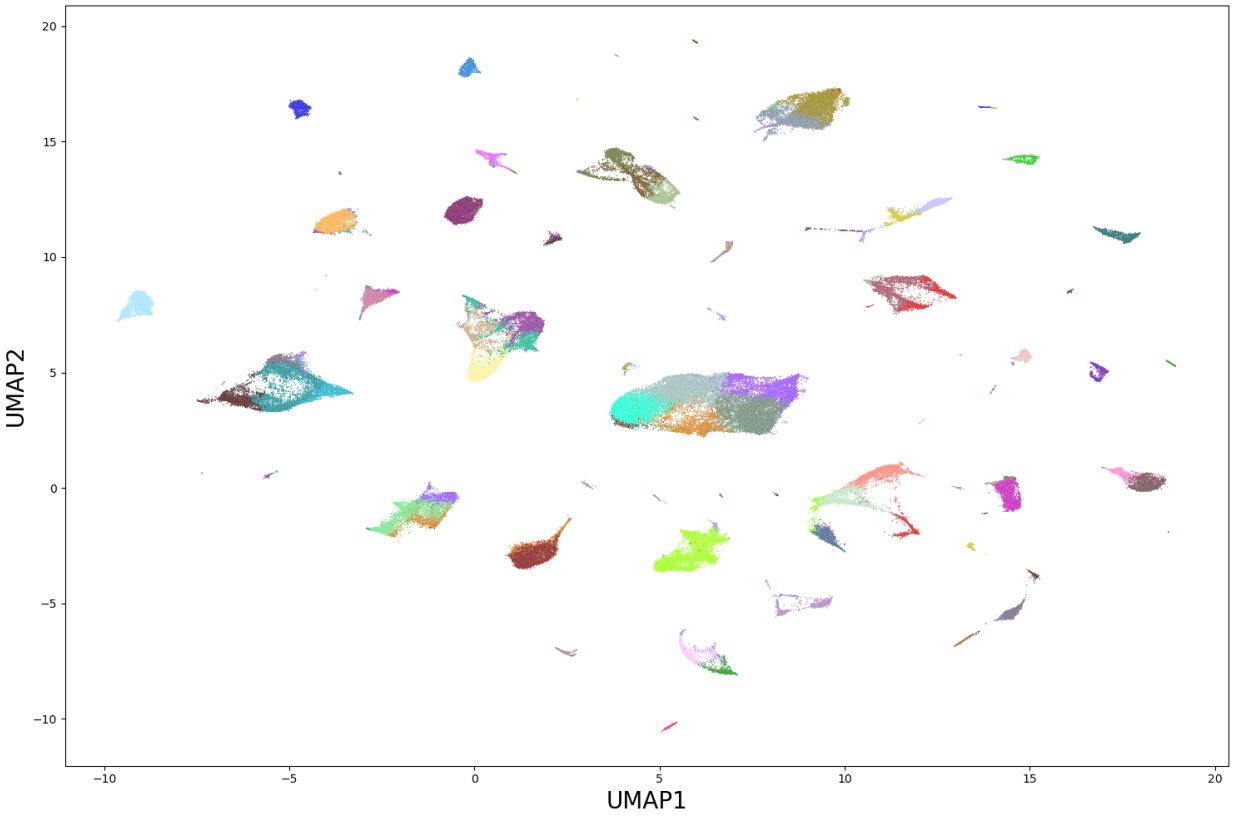

Figure 5: UMAP-kmeans (k=50)

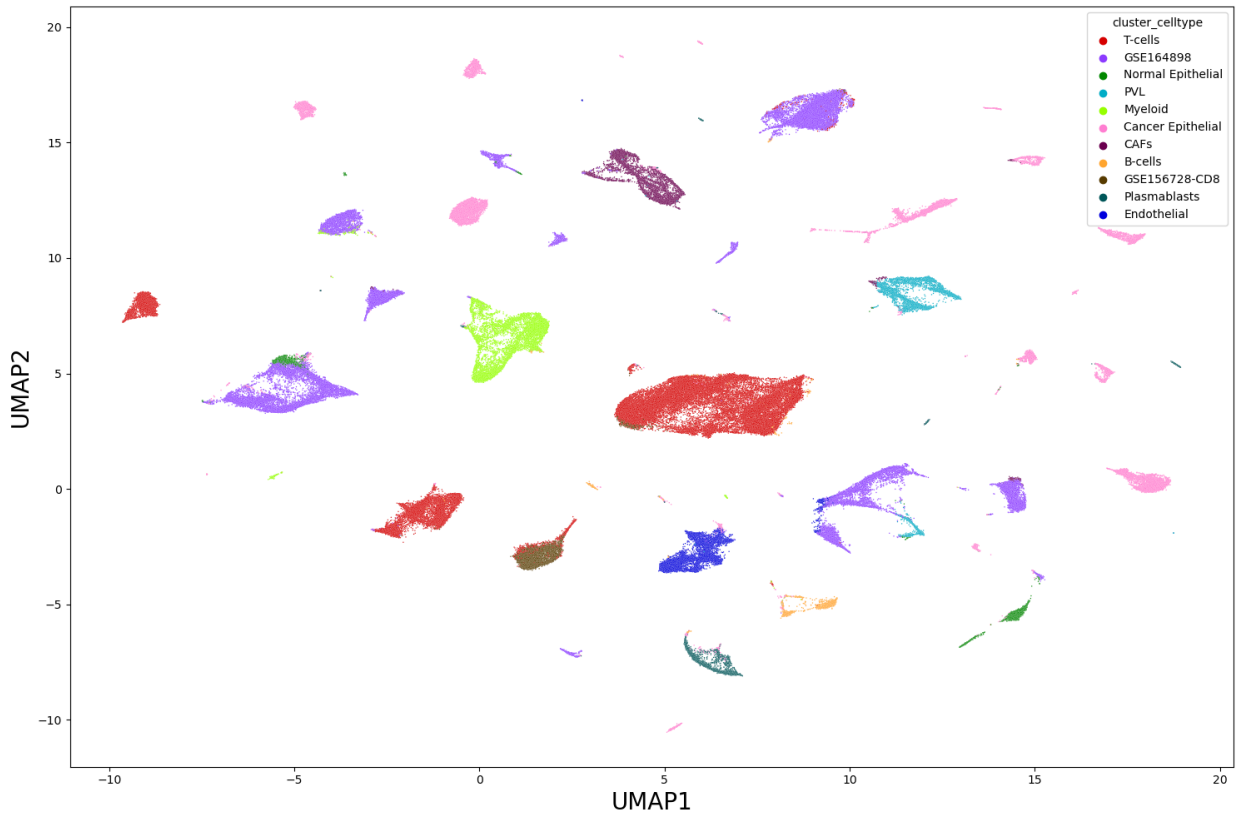

Figure 6: UMAP-kmeans(k=50) cell-type

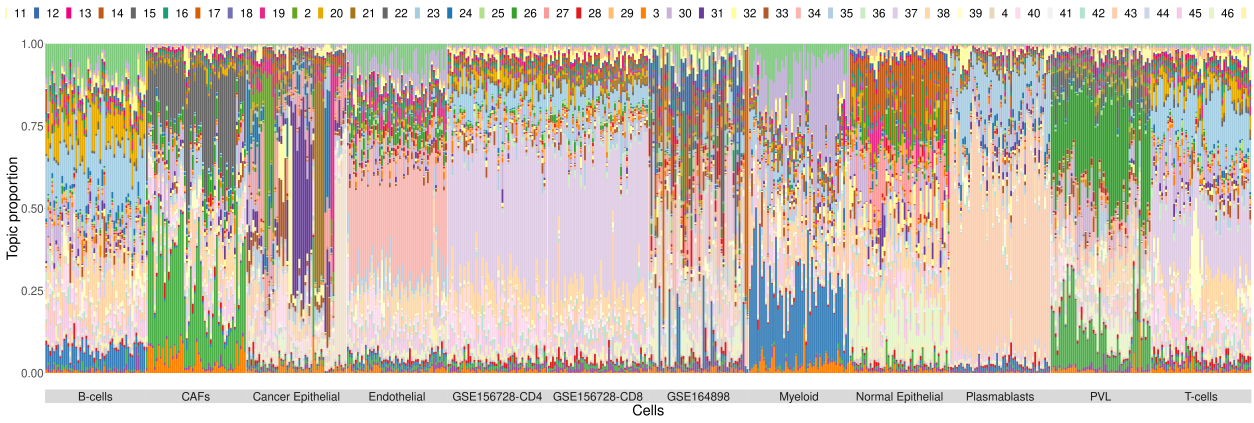

Figure 7: Structure plot cell-type

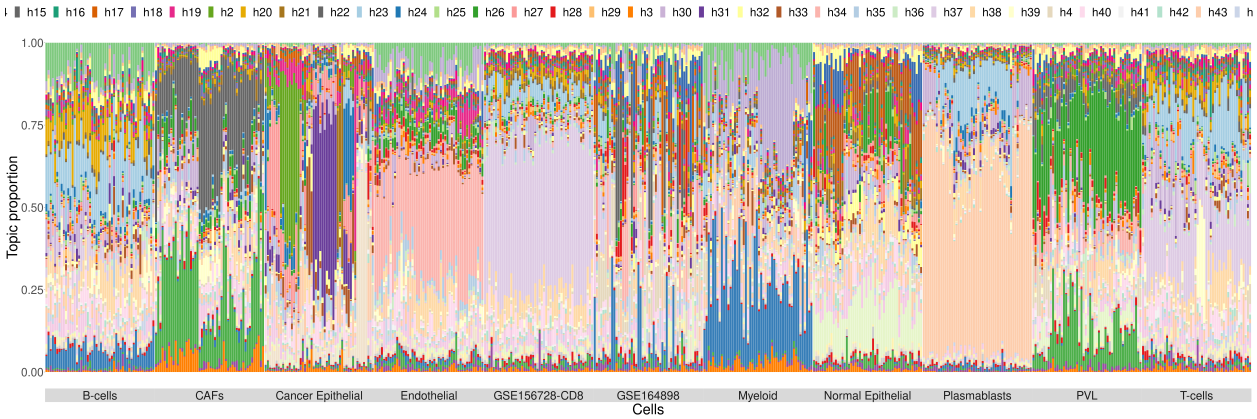

Figure 8: Structure plot kmeans majority cell-type

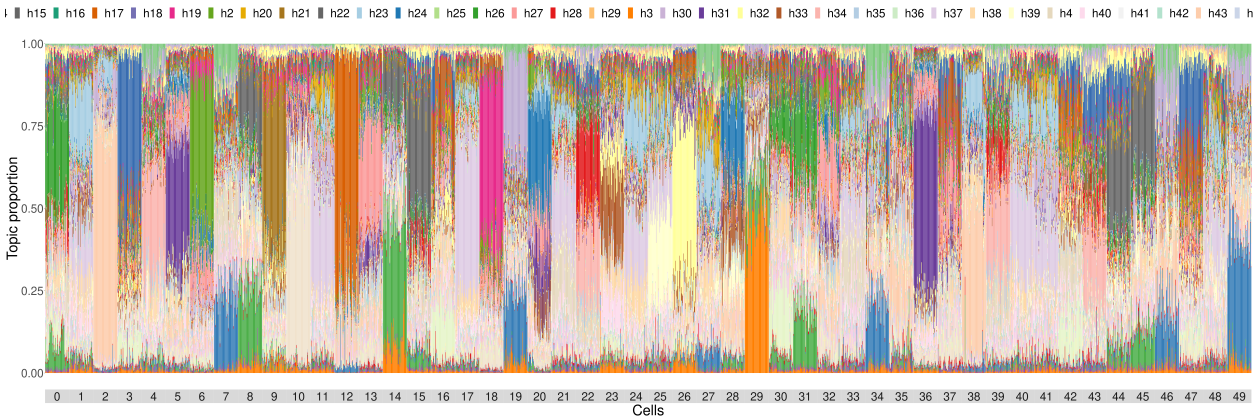

Figure 9: Structure plot kmeans cluster

Interaction topic analysis

Multinomial-Dirichlet:

$$p(\mathbf{y}_i|\mathbf{q}_i) = \frac{(\sum_g Y_{ig})!}{\prod_g Y_{ig}!} \prod_g q_{ig}^{Y_{ig}}$$

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

Single-cell generative model:

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

where

$$\lambda_{jg} = \lambda_0 \exp\left(\sum_{t=1}^{T} \theta_{jt}(\beta_{tg} + \delta_g)\right)$$

$$\lambda_0 = \exp(\tilde{\lambda}_0)$$

$\sum_t \theta_{jt} = 1$

Bayesian regularization of the model parameters

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

Total Expected log-likelihood Lower-bound (ELBO):

$$
\begin{aligned}
\frac{J}{n} &= \frac{1}{n}\sum_{i=1}^{n}\log p(\mathbf{x}_i|\theta_i(\mathbf{z}_i),\beta) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} D_{\mathsf{KL}}\left(q(z_{it})\|p(z_{it})\right) \\
&\quad + \frac{1}{n}\sum_{t=1}^{T}\sum_{l=1}^{L} D_{\mathsf{KL}}\left(q(\beta_{tl})\|p(\beta_{tl})\right) \\
&\quad + \frac{1}{n}\sum_{t=1}^{T}\sum_{r=1}^{R} D_{\mathsf{KL}}\left(q(\beta_{tr})\|p(\beta_{tr})\right)
\end{aligned}
$$

Neighbour cells calculation

- removed self neighbour pair
- 18 topics with cell count less than 100 were removed during generating annoy model list. Topics were not removed during generating neighbours, only selected topics were used to create a model list.

```
h35      57
h11      50
h10      39
h8       36
h44      34
h42      20
h29      18
h3       16
h5       13
h16      12
h36       7
h25       6
h47       6
h41       5
h49       4
```

```
h15        3
h18        2
h13        2
```

- Neighbours are calculated from the remaining 32 topics and 5 neighbours from each topic - 159 neighbours for each cell.

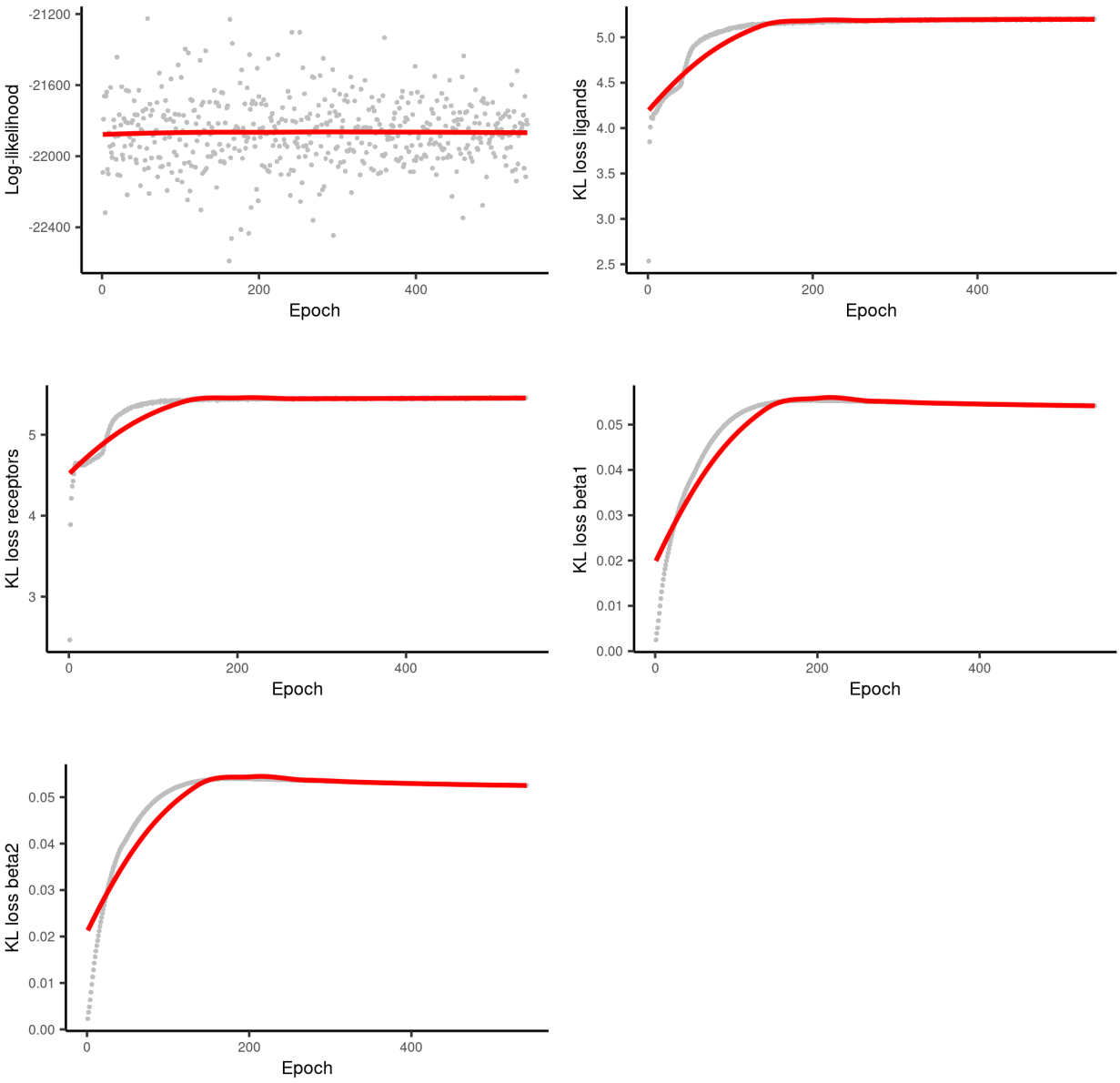

Figure 10: Loss plot

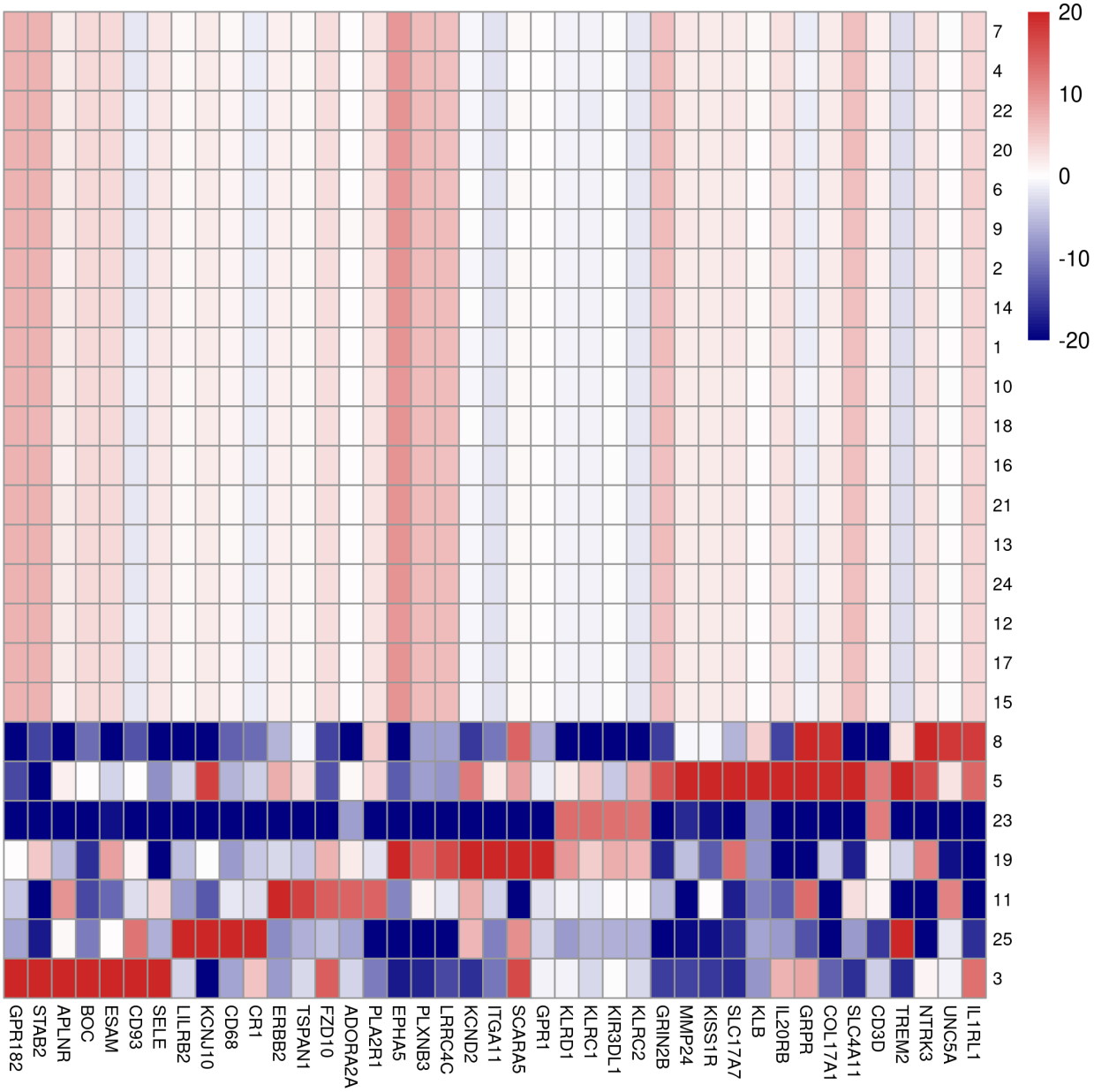

Figure 11: Beta - top receptor genes

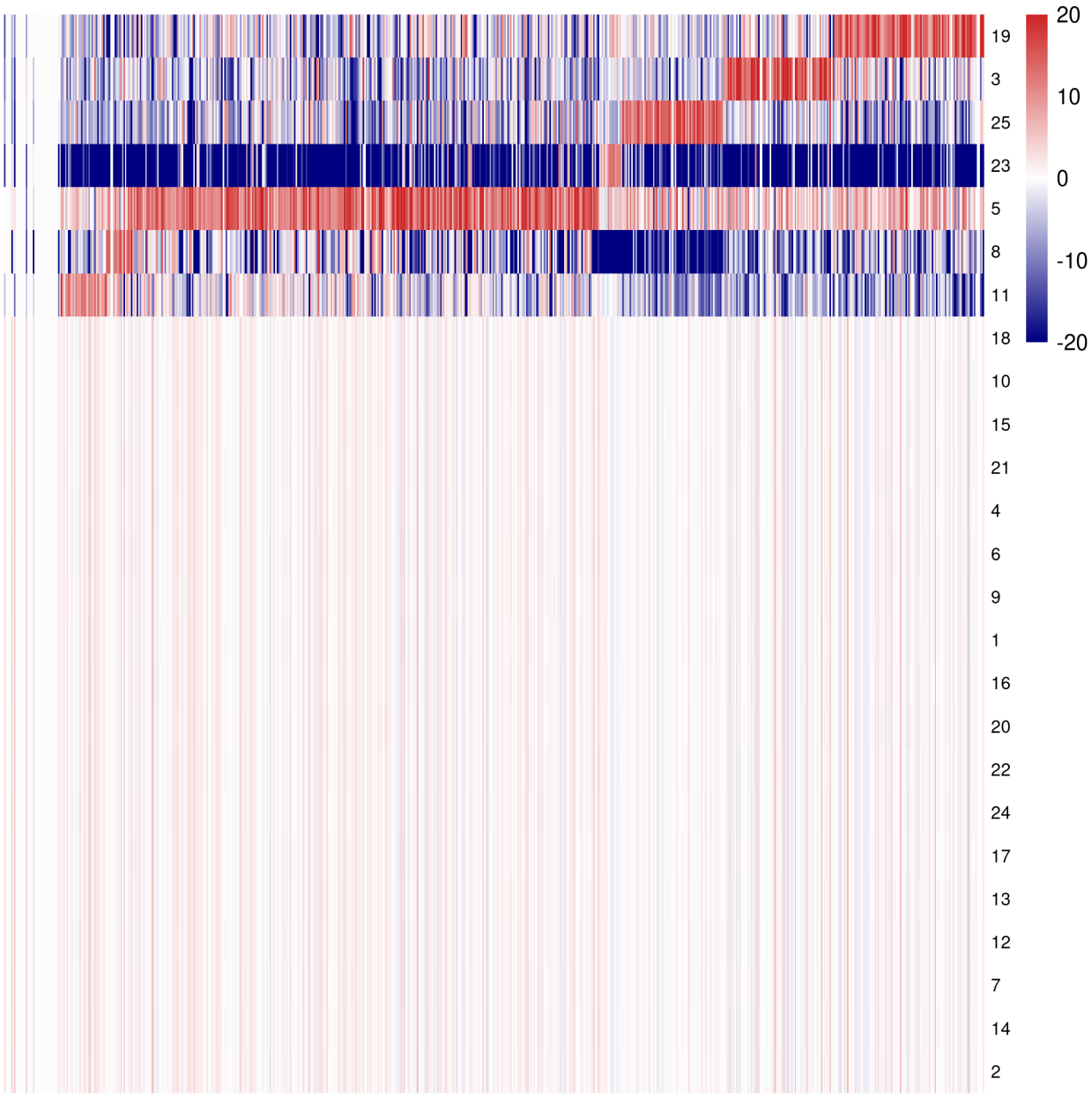

Figure 12: Beta - all receptor genes

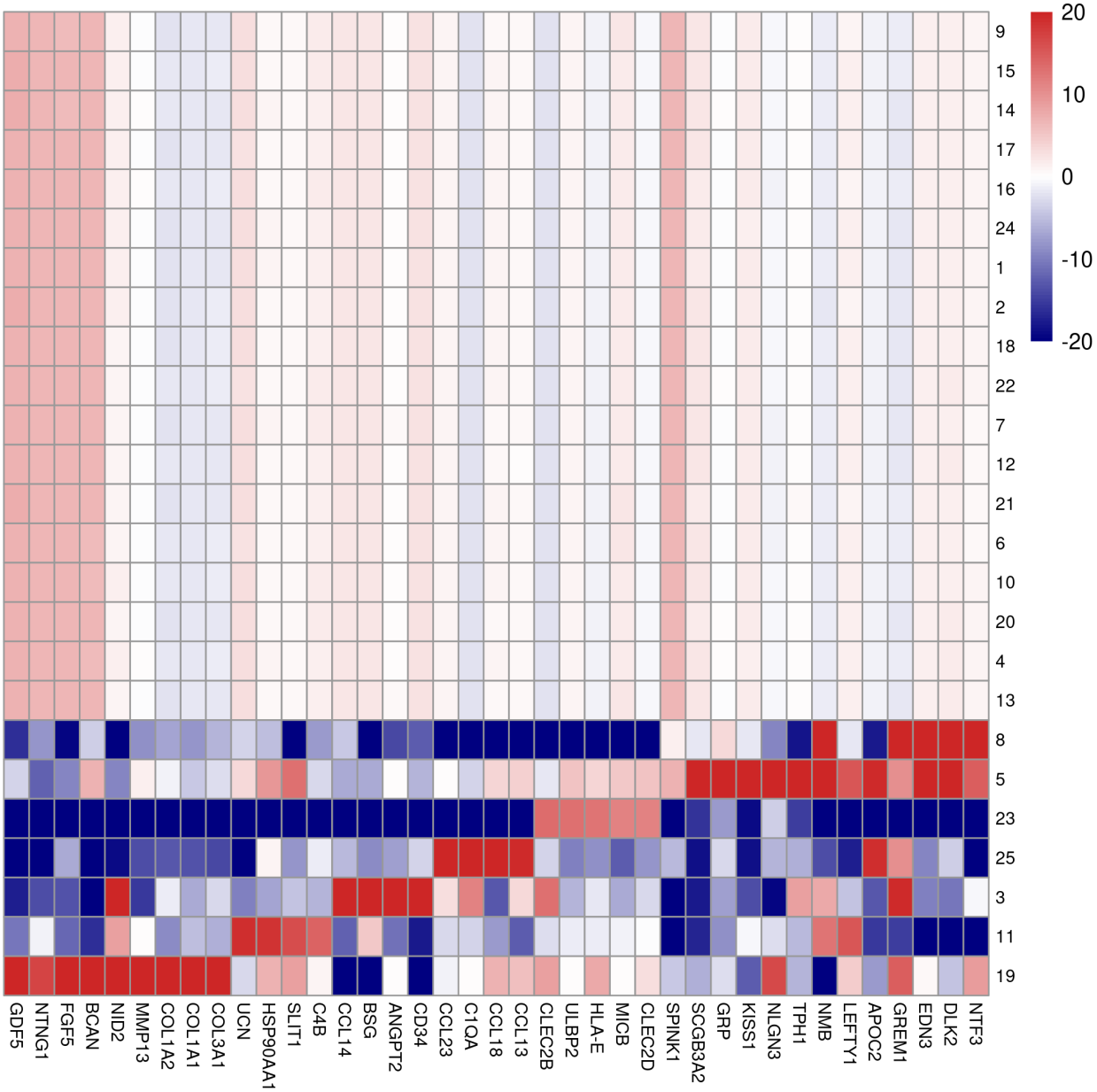

Figure 13: Beta - top ligand genes

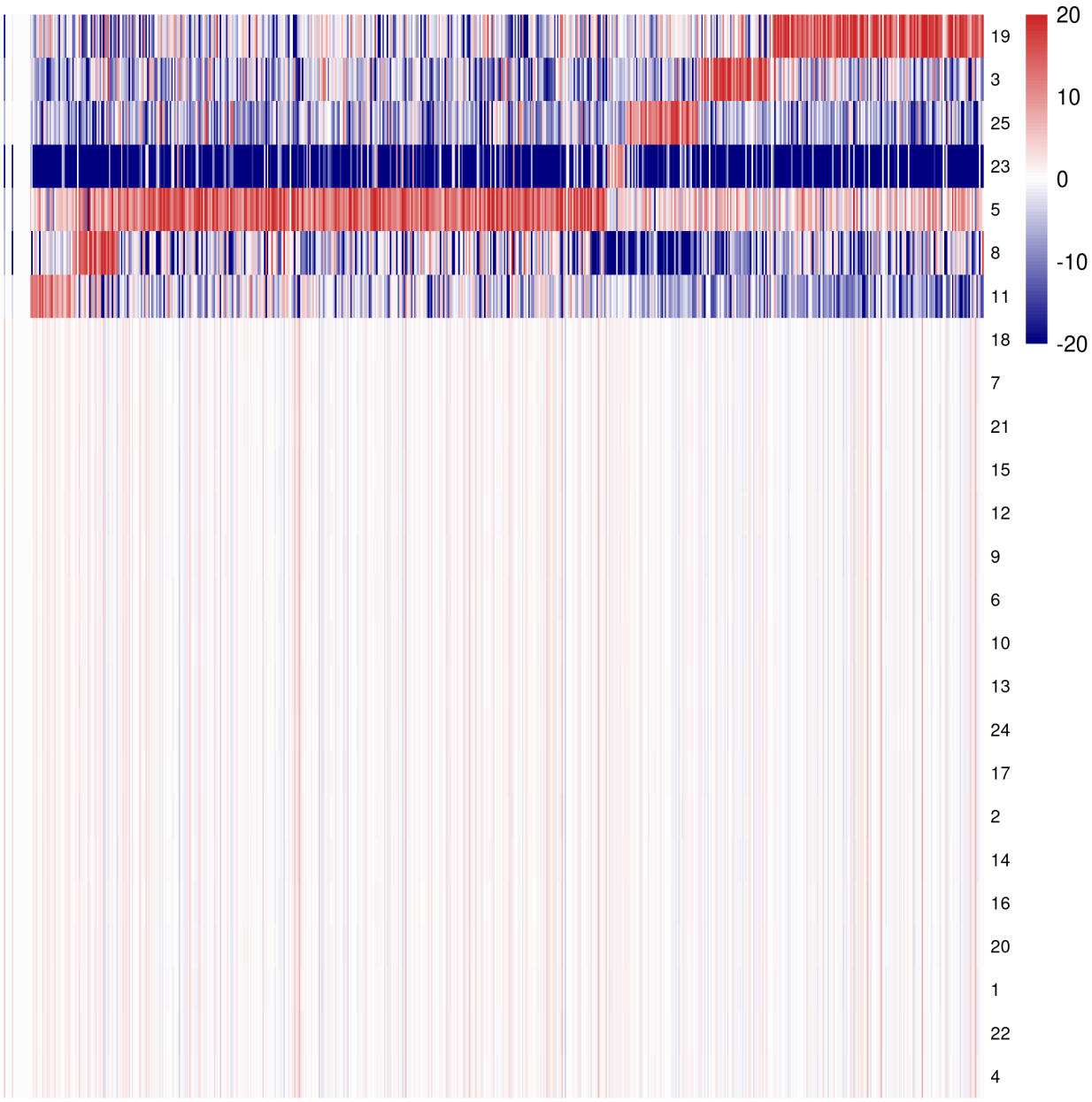

Figure 14: Beta - all ligand genes

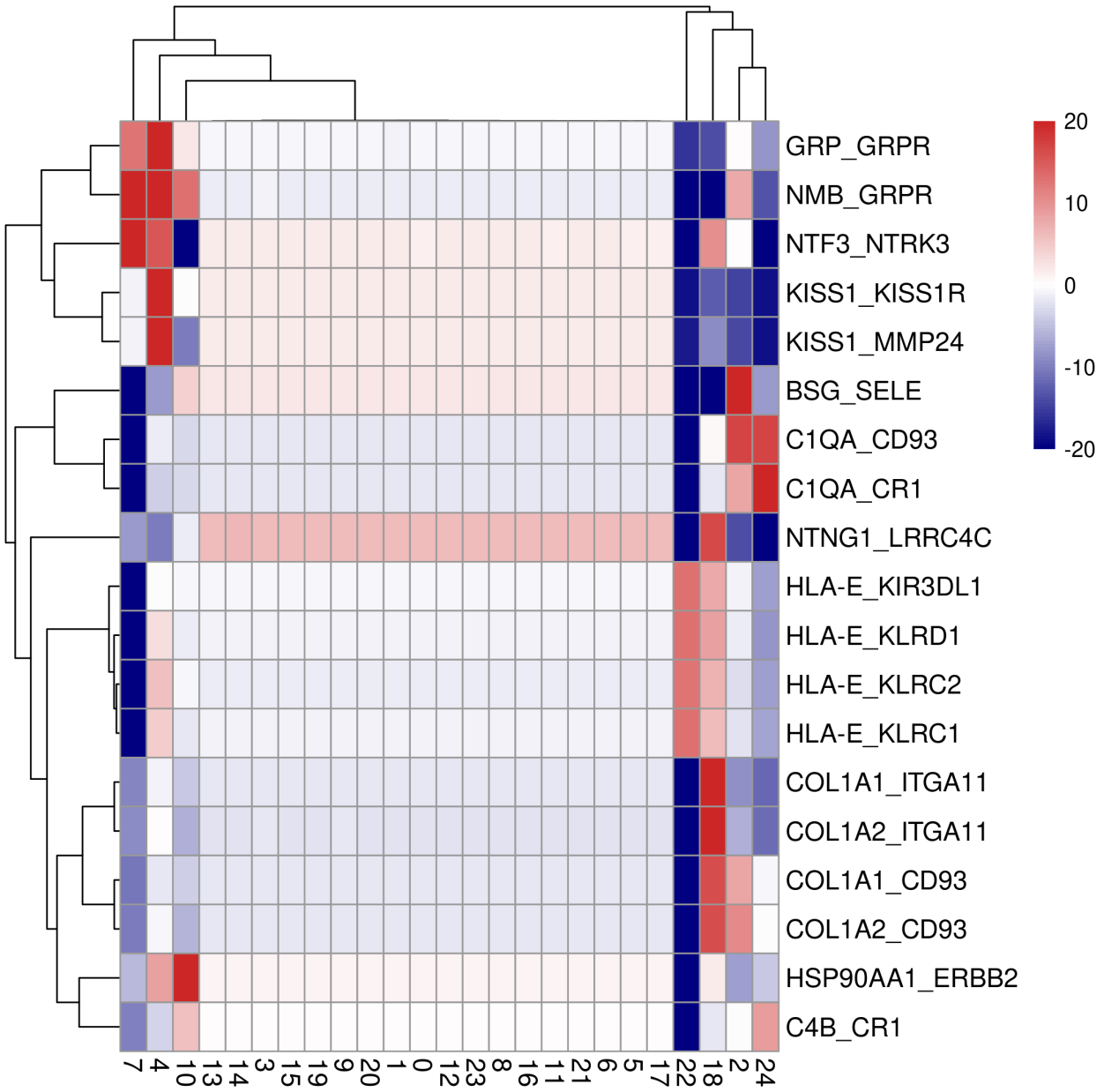

Figure 15: Beta - top lr pair genes