

Summary of TMB runs

Lena Morrill

24/05/2021

```
## Warning in .recacheSubclasses(def@class$className, def, env): undefined subclass
## "numericVector" of class "Mnumeric"; definition not updated

## Loading required package: viridisLite

## Loading required package: ggplot2

## Loading required package: tikzDevice

## Loading required package: coda

## Loading required package: MASS

## Warning in .recacheSubclasses(def@class$className, def, env): undefined subclass
## "numericVector" of class "Mnumeric"; definition not updated

## ##
## ## Markov Chain Monte Carlo Package (MCMCpack)

## ## Copyright (C) 2003-2022 Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park

## ##
## ## Support provided by the U.S. National Science Foundation

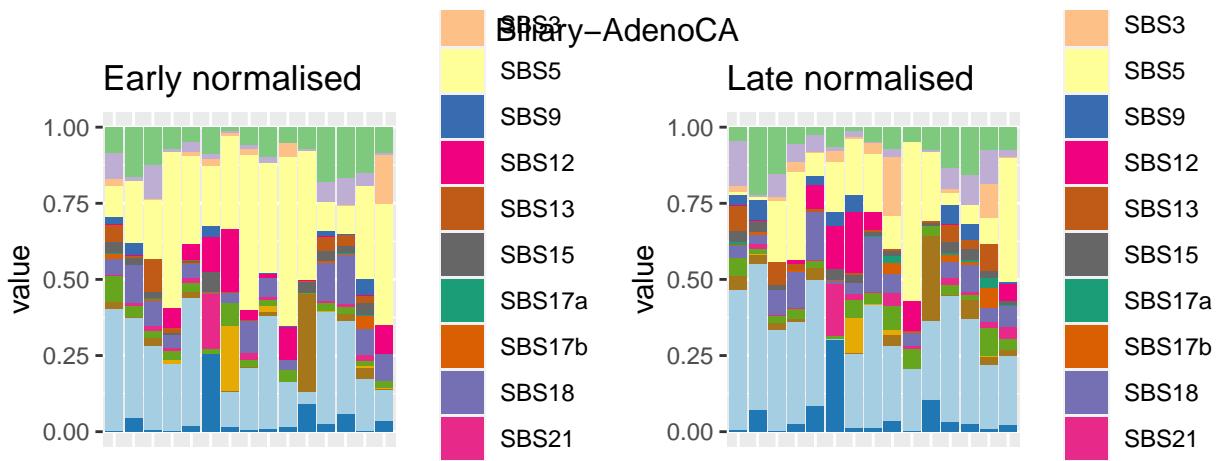
## ## (Grants SES-0350646 and SES-0350613)

## ##

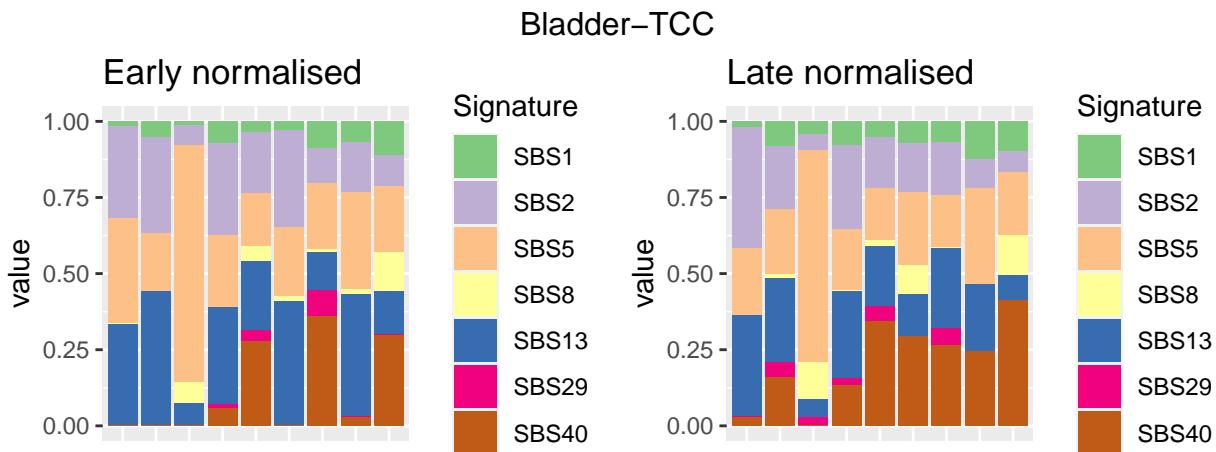
## 
## Attaching package: 'ggthemr'

## The following objects are masked from 'package:ggpubr':
## 
##   rotate_x_text, rotate_y_text

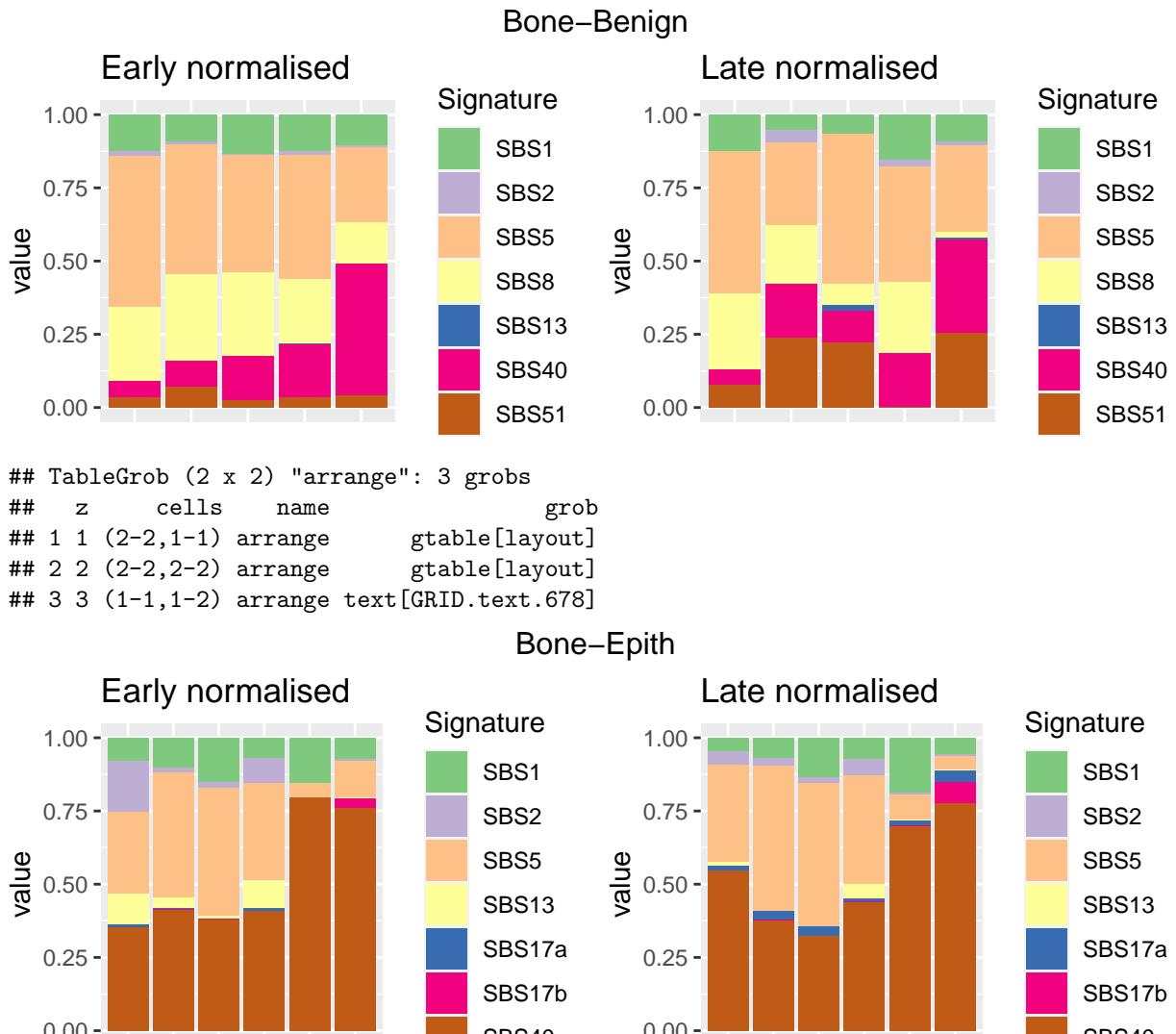
pcawg_palette <- pcawg.colour.palette(gsub("\\.*", "", enough_samples), scheme = "tumour.subtype")
names(pcawg_palette) <- enough_samples
pcawg_palette[names(pcawg_palette) == 'Lung-SCC'] <- '#ffff29' # #a8a800
```



```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name      grob
## 1 1 (2-2,1-1) arrange    gtable[layout]
## 2 2 (2-2,2-2) arrange    gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.328]
```

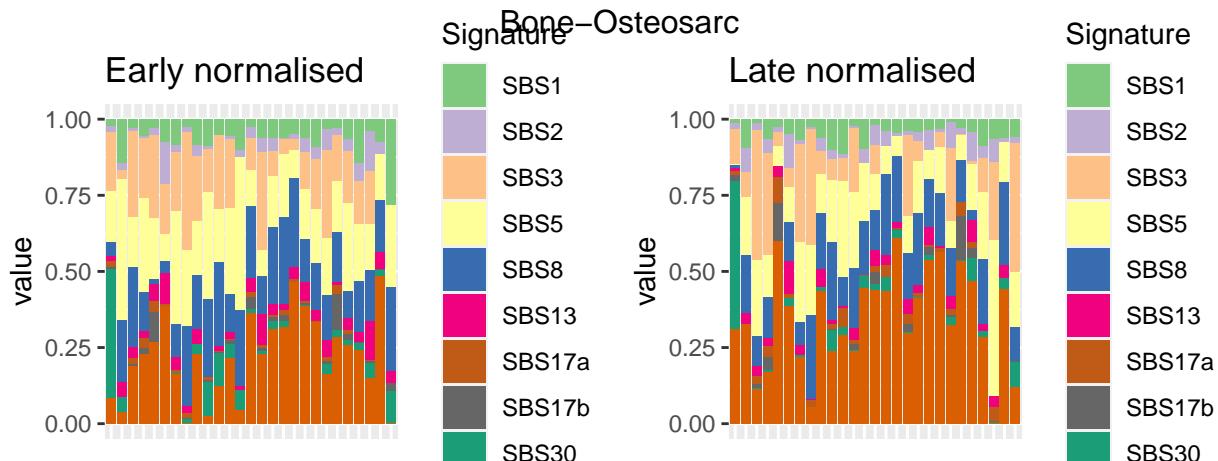


```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name      grob
## 1 1 (2-2,1-1) arrange    gtable[layout]
## 2 2 (2-2,2-2) arrange    gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.503]
```

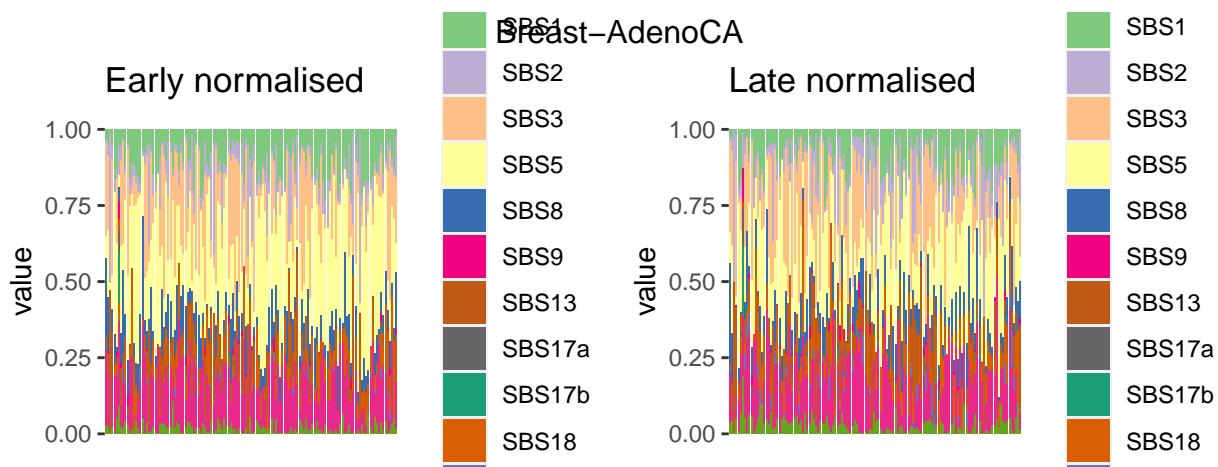


```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name      grob
## 1 1 (2-2,1-1) arrange    gtable[layout]
## 2 2 (2-2,2-2) arrange    gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.678]
```

```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name      grob
## 1 1 (2-2,1-1) arrange    gtable[layout]
## 2 2 (2-2,2-2) arrange    gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.853]
```

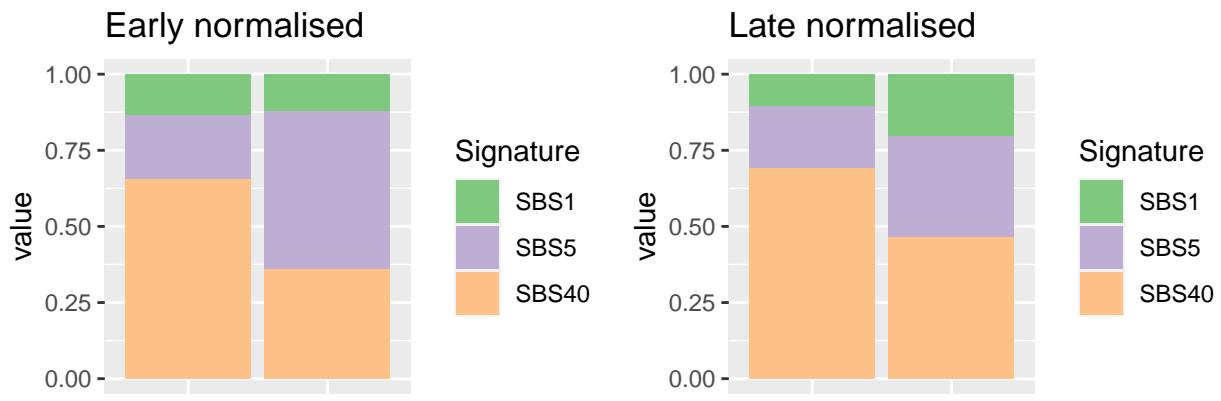


```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells  name          grob
## 1 1 (2-2,1-1) arrange    gtable[layout]
## 2 2 (2-2,2-2) arrange    gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1070]
```



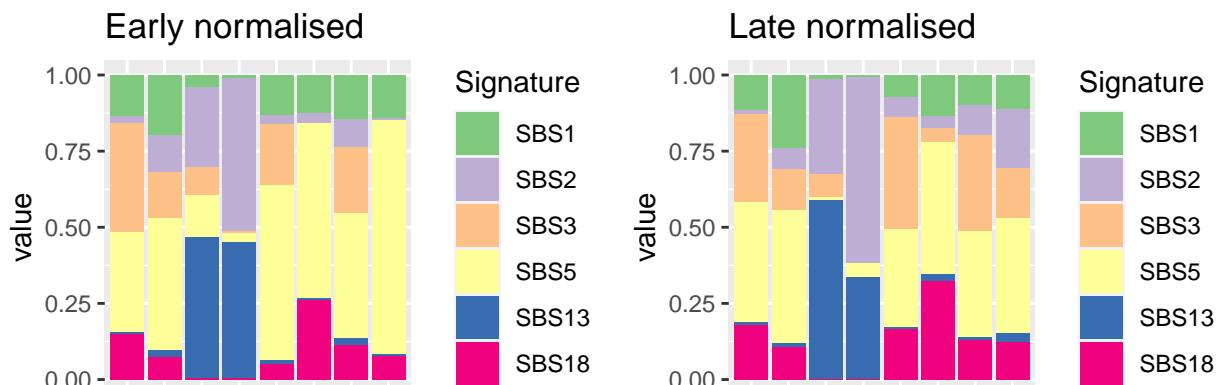
```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells  name          grob
## 1 1 (2-2,1-1) arrange    gtable[layout]
## 2 2 (2-2,2-2) arrange    gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1329]
```

Breast-DCIS



```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells  name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1448]
```

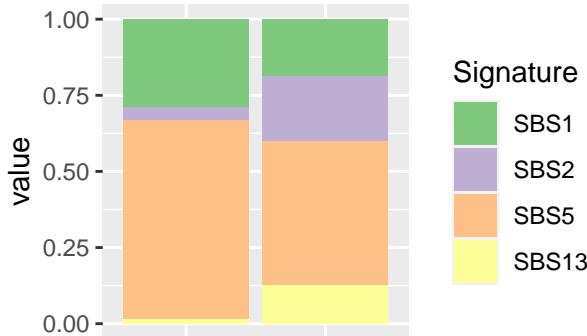
Breast-LobularCA



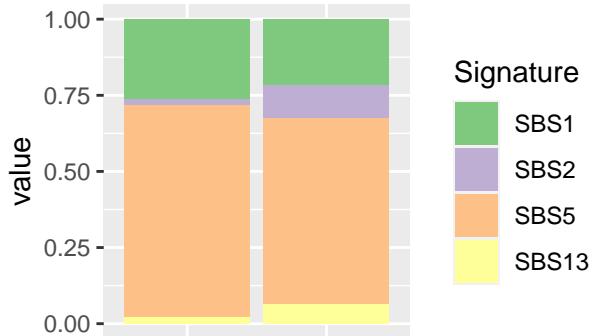
```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells  name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1609]
```

Cervix–AdenoCA

Early normalised



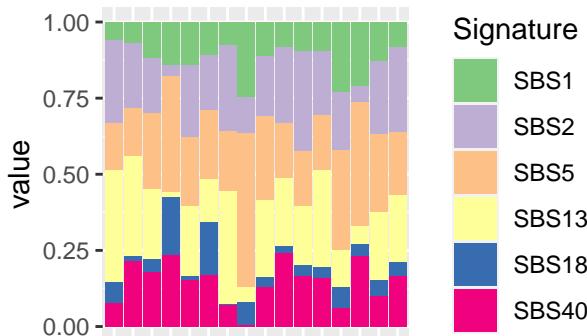
Late normalised



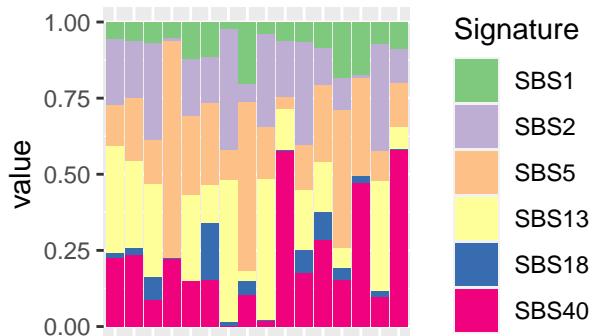
```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells  name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1742]
```

Cervix–SCC

Early normalised

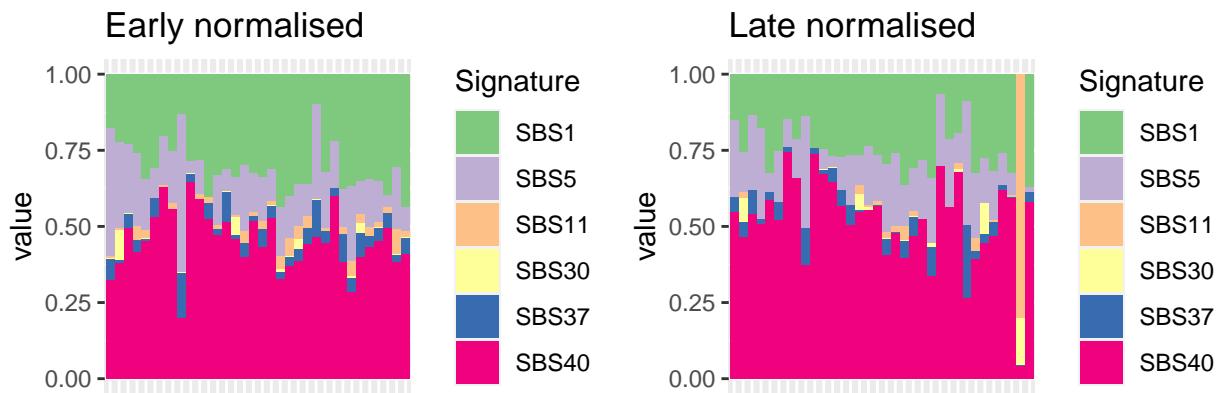


Late normalised



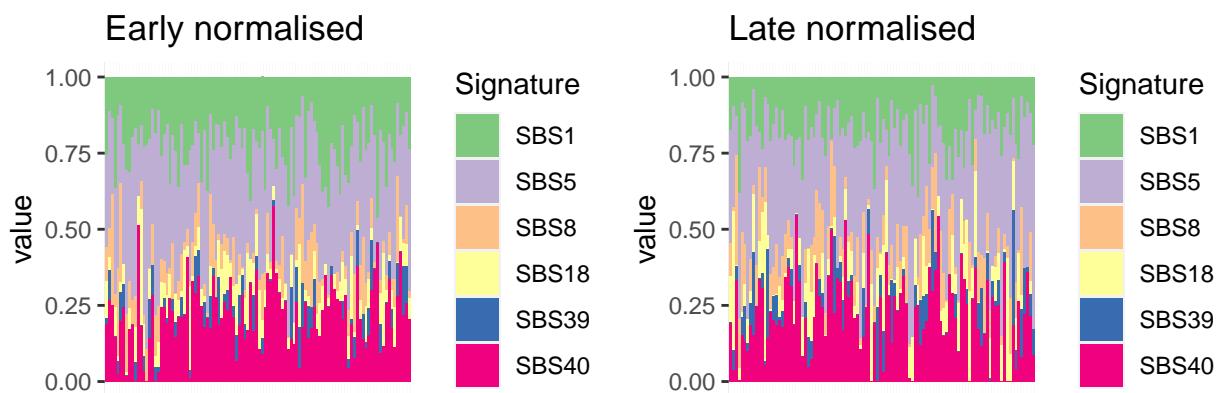
```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells  name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1903]
```

CNS–GBM

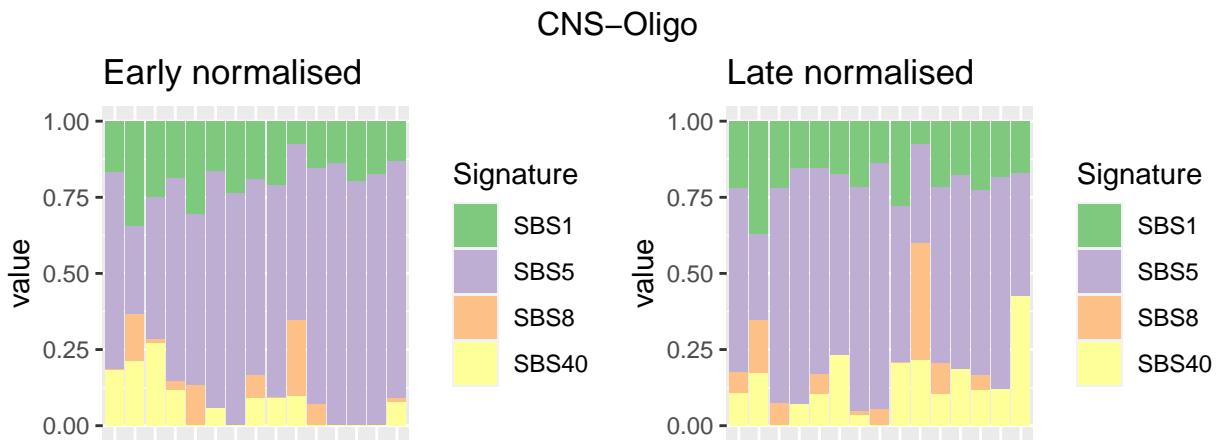


```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells  name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2064]
```

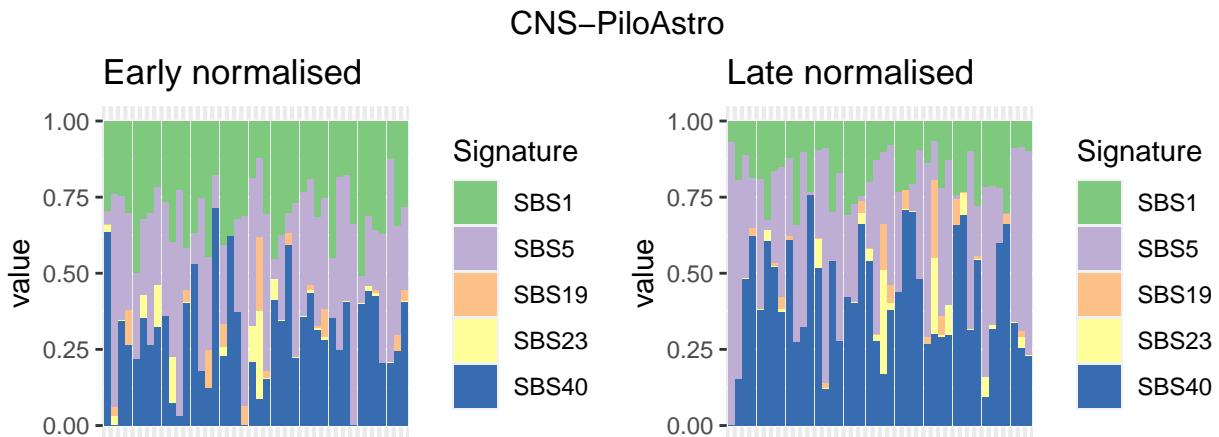
CNS–Medullo



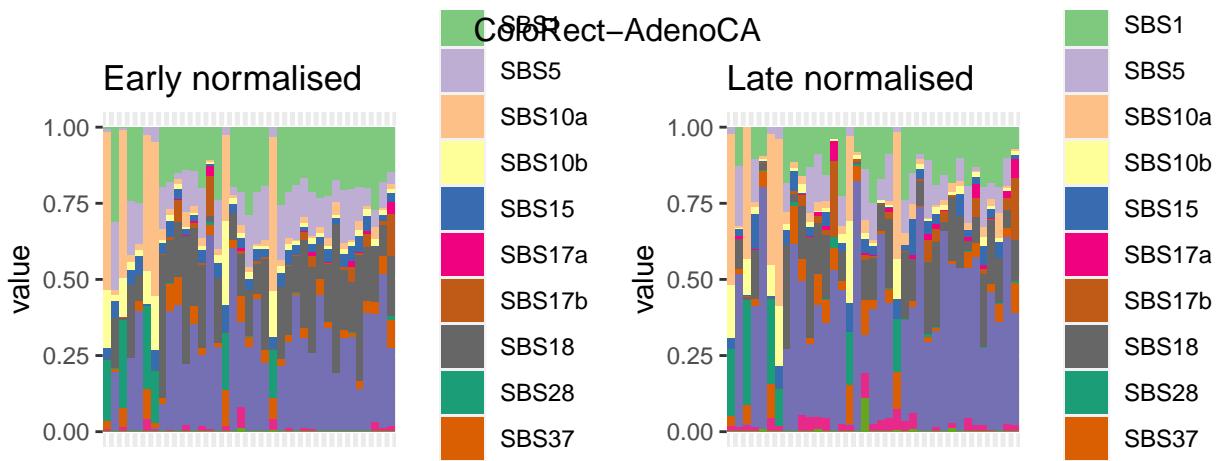
```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells  name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2225]
```



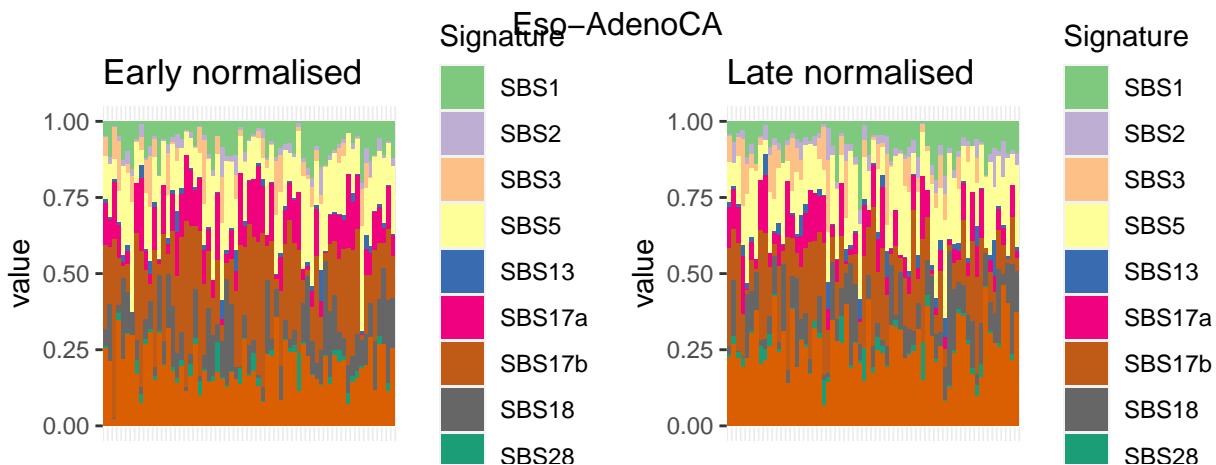
```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z    cells    name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2358]
```



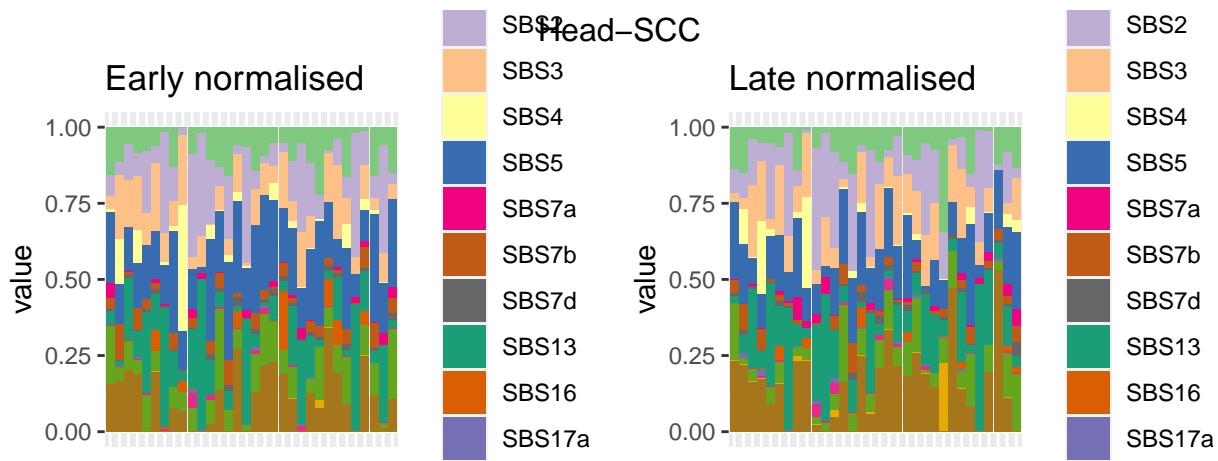
```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z    cells    name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2505]
```



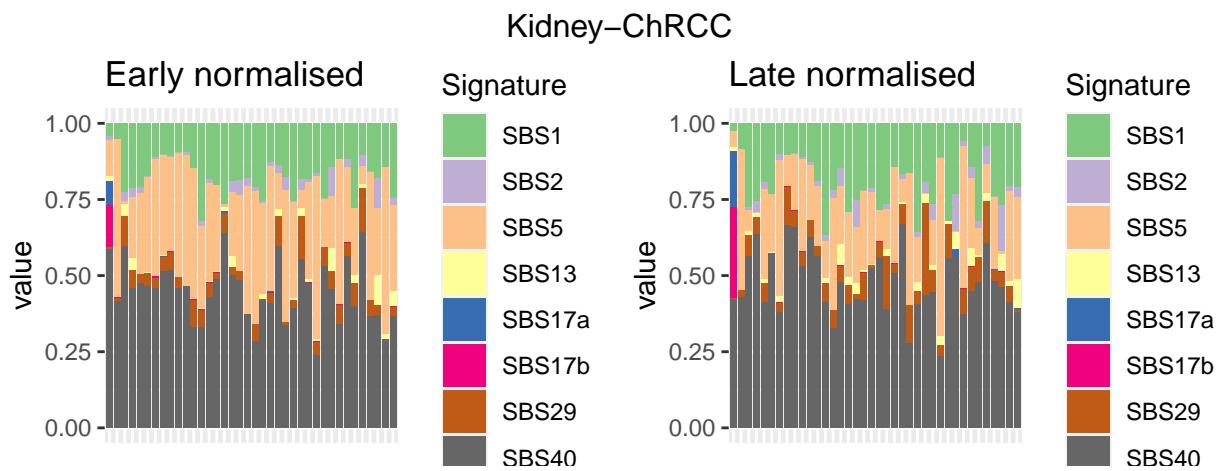
```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2764]
```



```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2981]
```

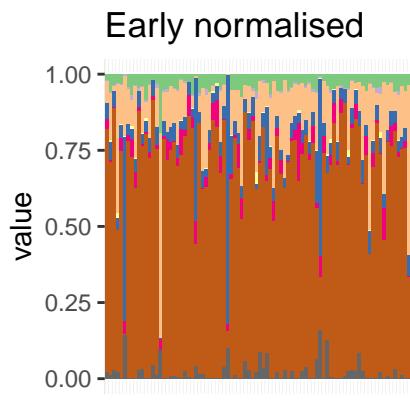


```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.3268]
```



```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.3457]
```

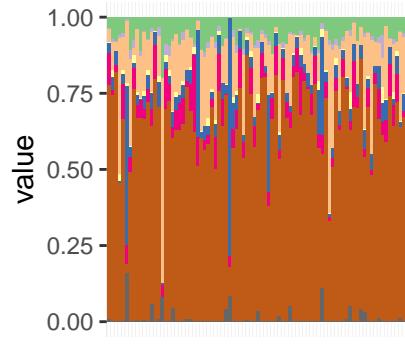
Kidney–RCC.clearcell



Signature

- SBS1
- SBS2
- SBS5
- SBS13
- SBS22
- SBS29
- SBS40
- SBS41

Late normalised

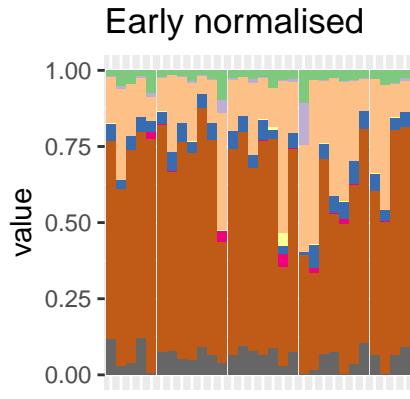


Signature

- SBS1
- SBS2
- SBS5
- SBS13
- SBS22
- SBS29
- SBS40
- SBS41

```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.3646]
```

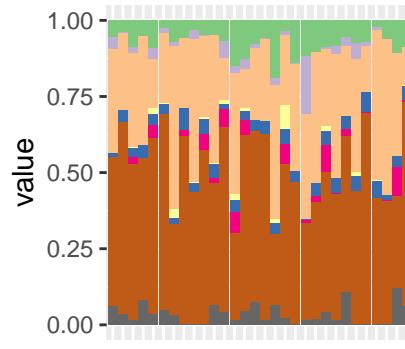
Kidney–RCC.papillary



Signature

- SBS1
- SBS2
- SBS5
- SBS13
- SBS22
- SBS29
- SBS40
- SBS41

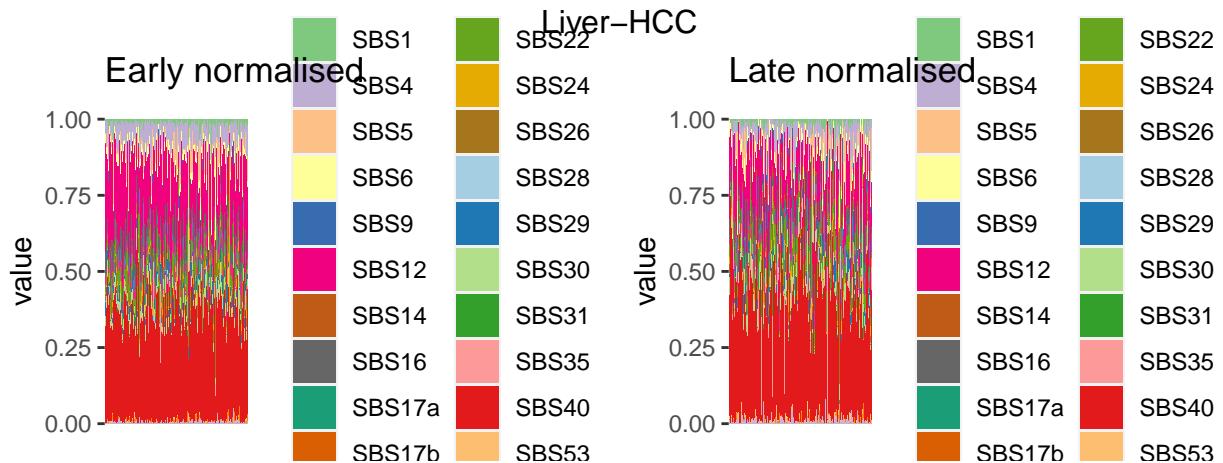
Late normalised



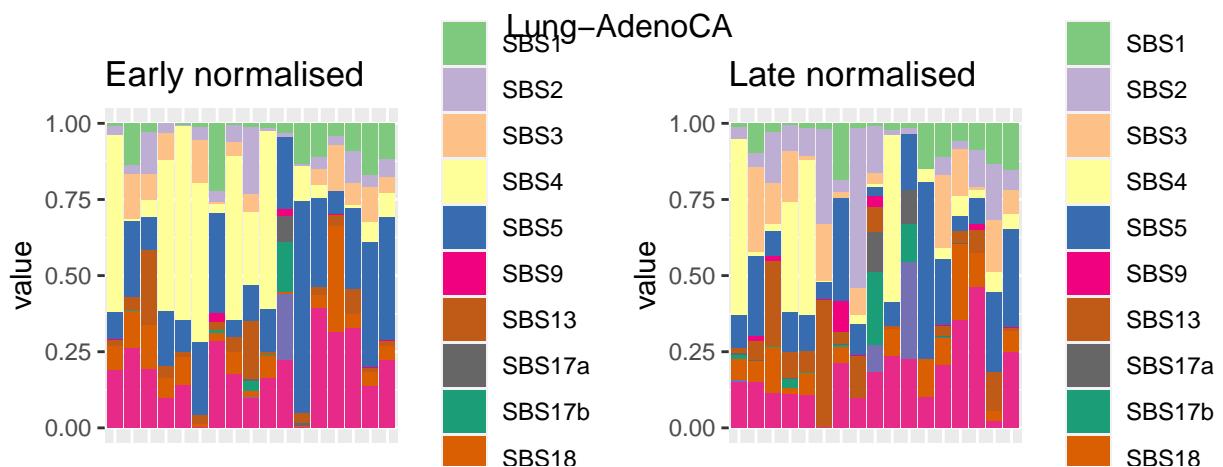
Signature

- SBS1
- SBS2
- SBS5
- SBS13
- SBS22
- SBS29
- SBS40
- SBS41

```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.3835]
```



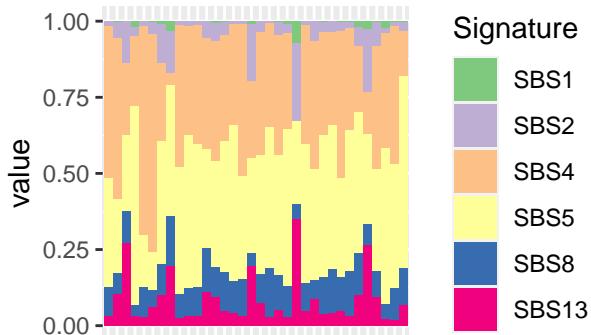
```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.4248]
```



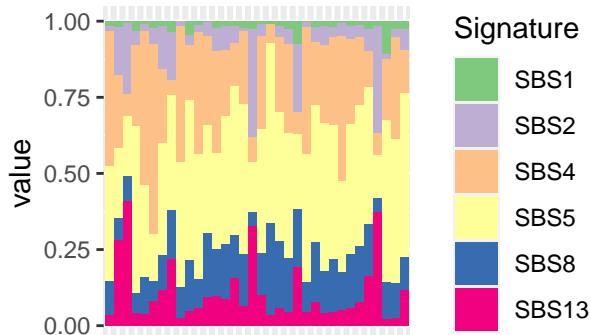
```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.4493]
```

Lung-SCC

Early normalised

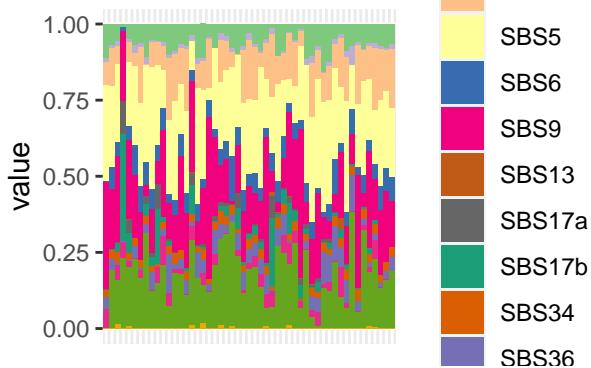


Late normalised

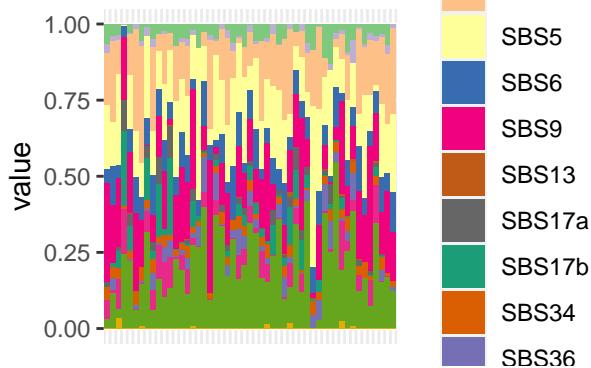


```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.4654]
```

Early normalised



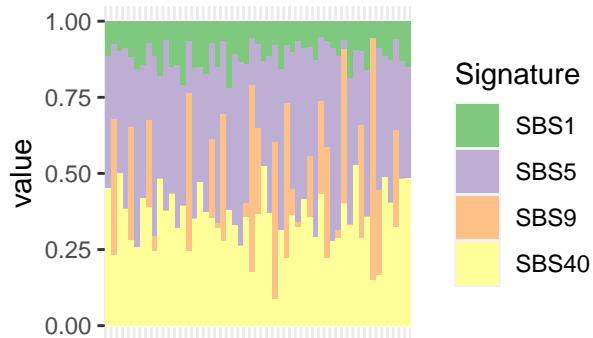
Late normalised



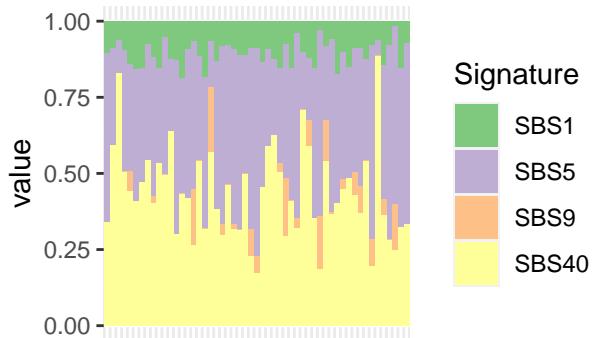
```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.4927]
```

Lymph-CLL

Early normalised



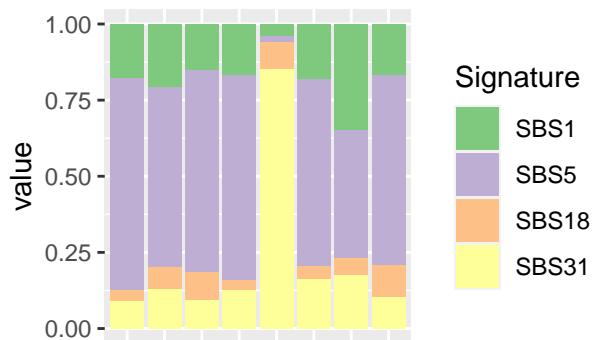
Late normalised



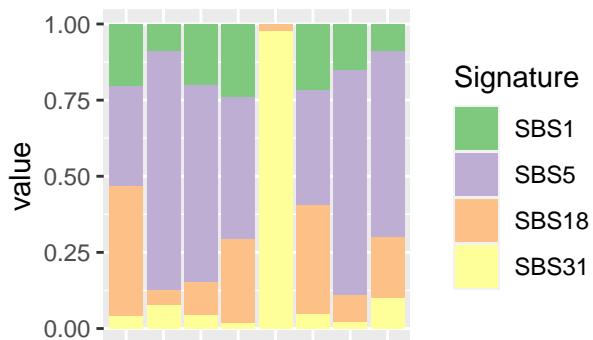
```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name      grob
## 1 1 (2-2,1-1) arrange    gtable[layout]
## 2 2 (2-2,2-2) arrange    gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.5060]
```

Myeloid-AML

Early normalised

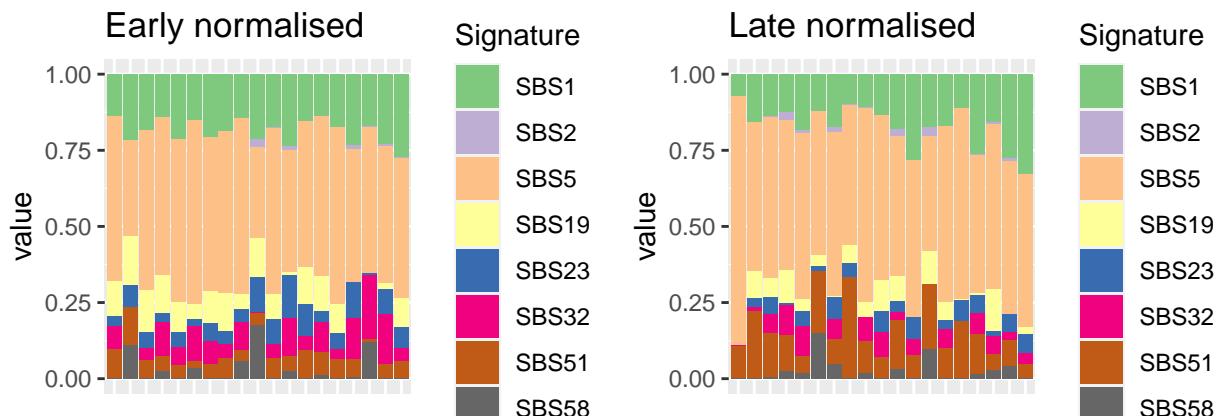


Late normalised

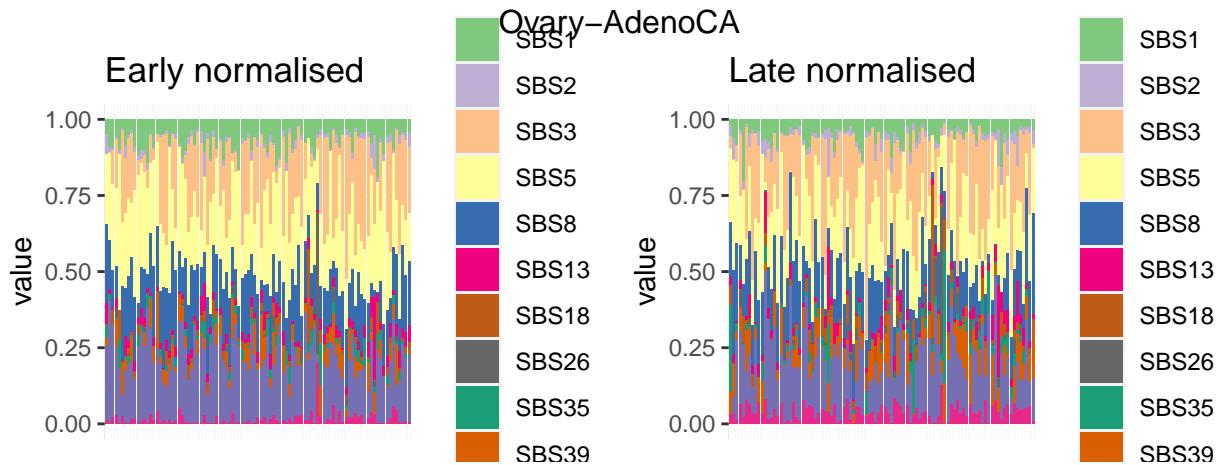


```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name      grob
## 1 1 (2-2,1-1) arrange    gtable[layout]
## 2 2 (2-2,2-2) arrange    gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.5193]
```

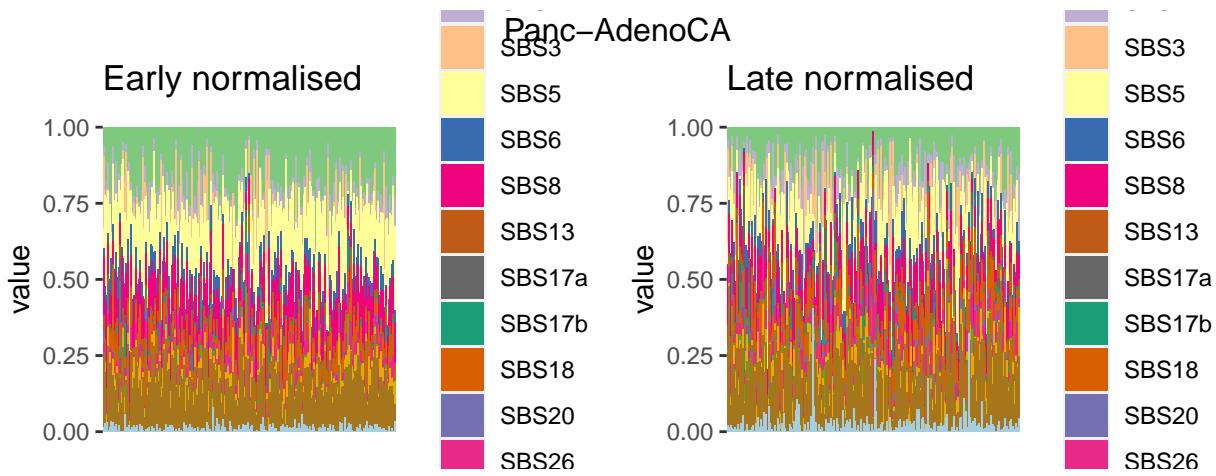
Myeloid–MPN



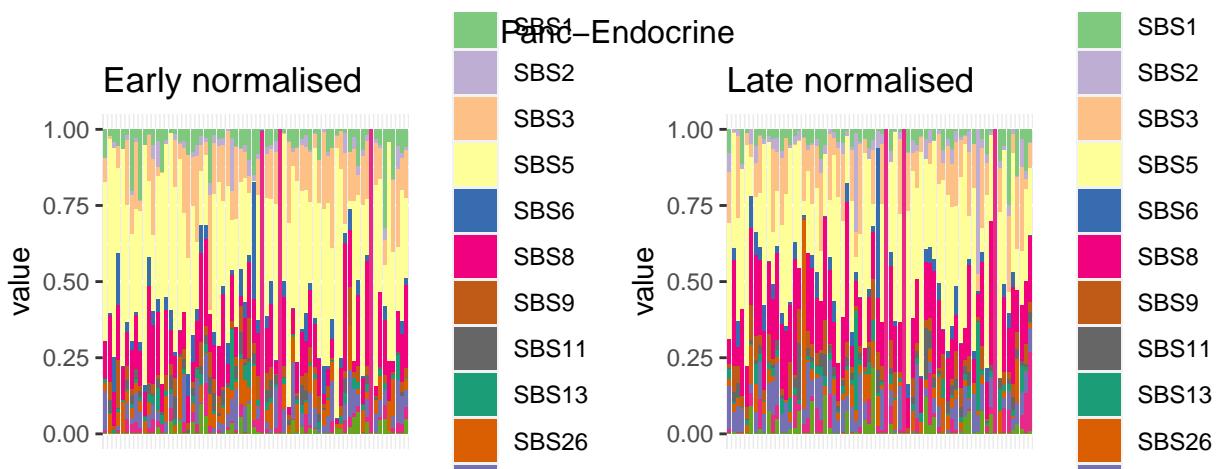
```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells  name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.5382]
```



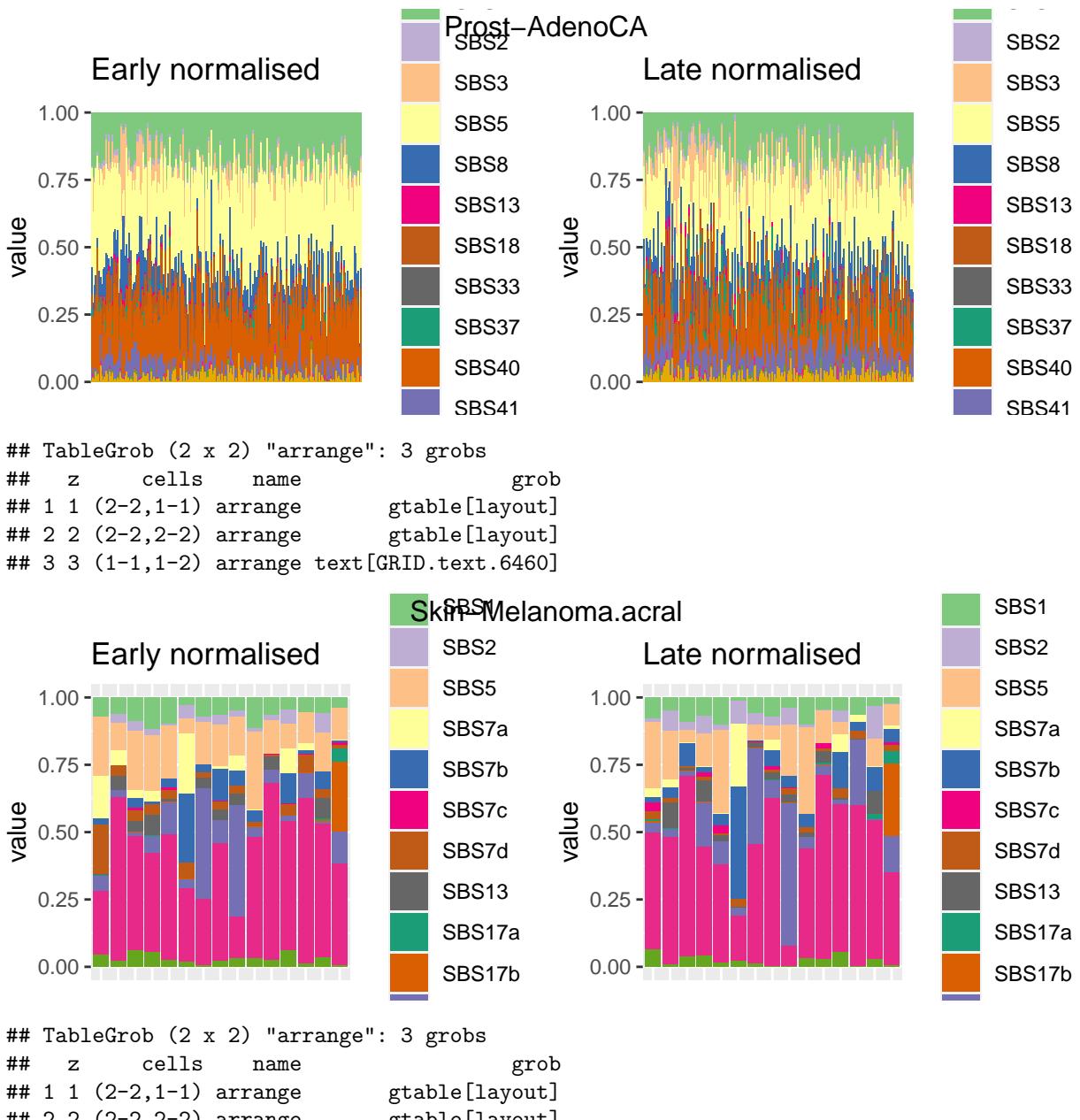
```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells  name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.5627]
```

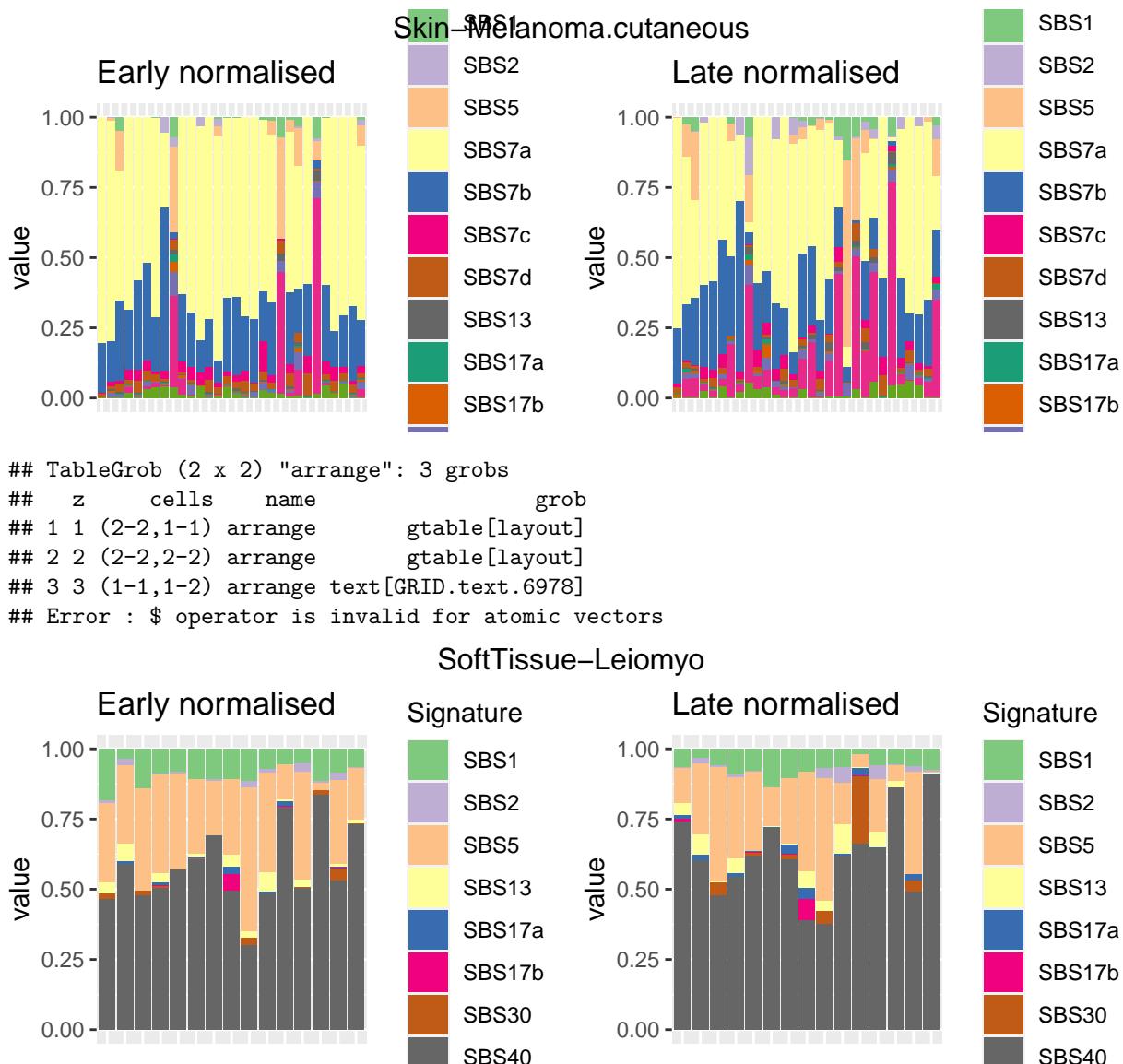


```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name      grob
## 1 1 (2-2,1-1) arrange    gtable[layout]
## 2 2 (2-2,2-2) arrange    gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.5928]
```

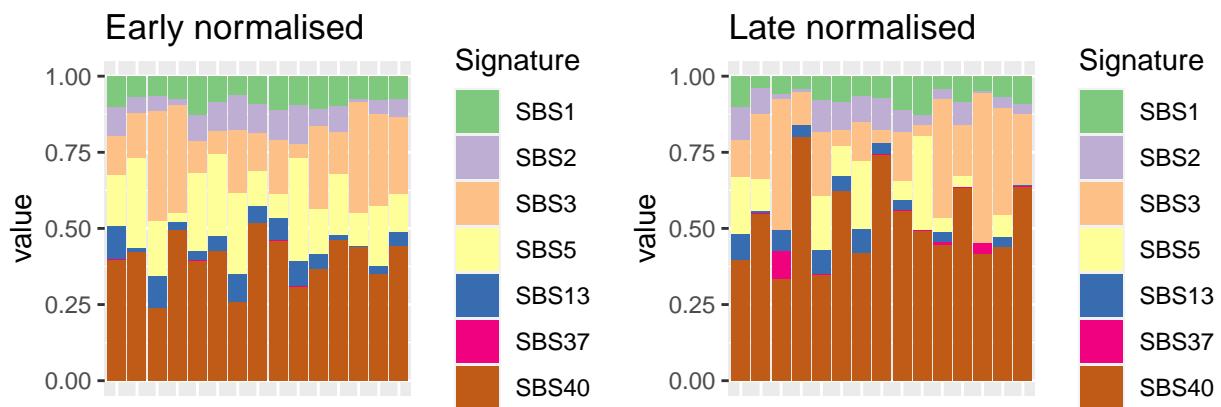


```
## # TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name      grob
## 1 1 (2-2,1-1) arrange    gtable[layout]
## 2 2 (2-2,2-2) arrange    gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.6187]
```

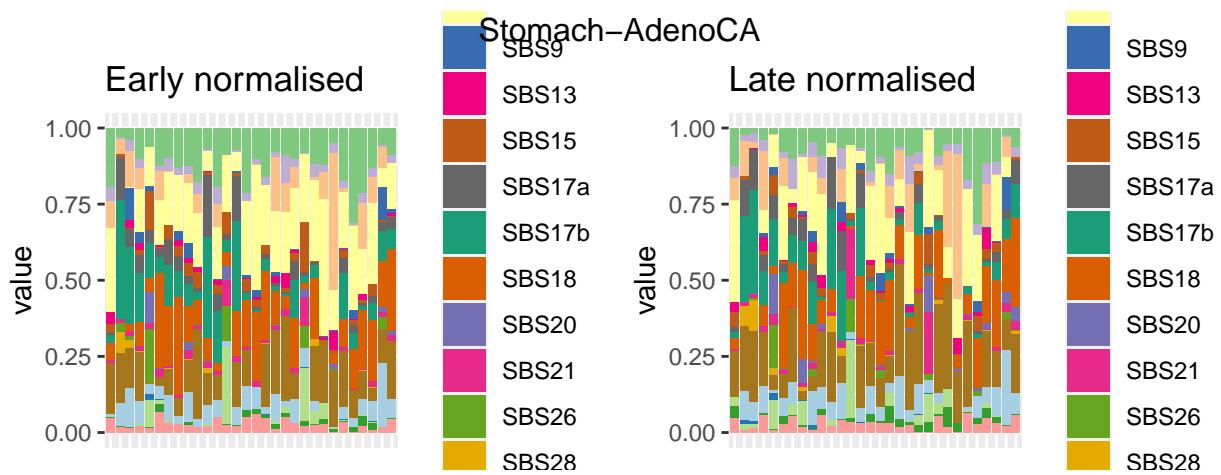




SoftTissue–Liposarc



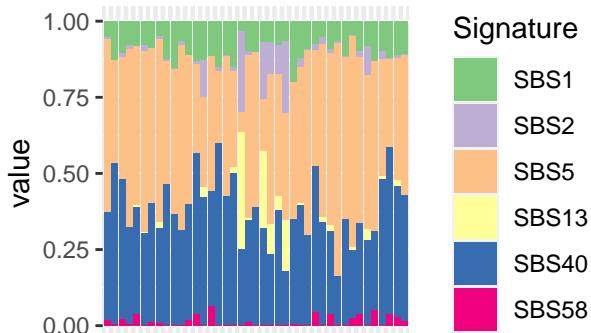
```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.7342]
```



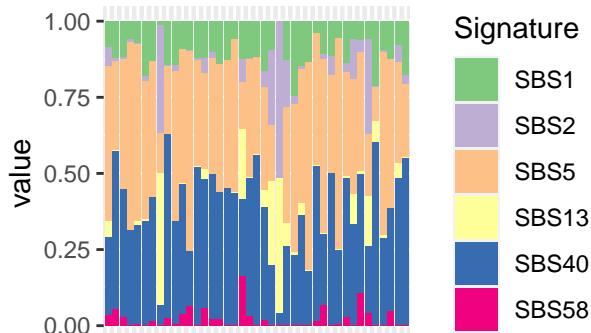
```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.7699]
```

Thy–AdenoCA

Early normalised



Late normalised



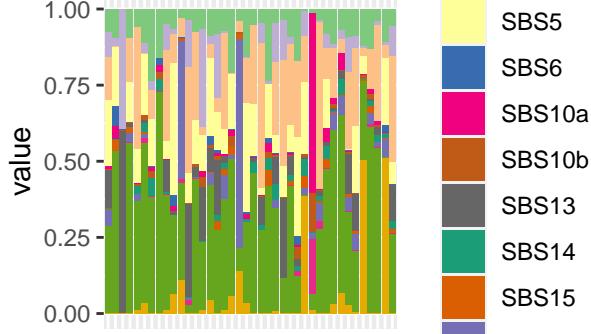
```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.7860]
```

Early normalised

Uterus–AdenoCA

- SBS2
- SBS3
- SBS5
- SBS6
- SBS10a
- SBS10b
- SBS13
- SBS14
- SBS15
- SBS26

Late normalised



```
## TableGrob (2 x 2) "arrange": 3 grobs
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.8133]
```

```
## $`Biliary-AdenoCA`
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.328]
##
## $`Bladder-TCC`
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.503]
```

```

## 
## $`Bone-Benign` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.678]
##
## $`Bone-Epith` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.853]
##
## $`Bone-Osteosarc` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1070]
##
## $`Breast-AdenoCA` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1329]
##
## $`Breast-DCIS` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1448]
##
## $`Breast-LobularCA` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1609]
##
## $`Cervix-AdenoCA` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1742]
##
## $`Cervix-SCC` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.1903]
##
## $`CNS-GBM` 
##   z   cells   name           grob

```

```

## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2064]
##
## $`CNS-Medullo`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2225]
##
## $`CNS-Oligo`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2358]
##
## $`CNS-PiloAstro`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2505]
##
## $`ColoRect-AdenoCA`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2764]
##
## $`Eso-AdenoCA`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.2981]
##
## $`Head-SCC`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.3268]
##
## $`Kidney-ChRCC`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.3457]
##
## $`Kidney-RCC.clearcell`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.3646]

```

```

## 
## $`Kidney-RCC.papillary` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.3835]
##
## $`Liver-HCC` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.4248]
##
## $`Lung-AdenoCA` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.4493]
##
## $`Lung-SCC` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.4654]
##
## $`Lymph-BNHL` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.4927]
##
## $`Lymph-CLL` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.5060]
##
## $`Myeloid-AML` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.5193]
##
## $`Myeloid-MPN` 
##   z   cells   name           grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.5382]
##
## $`Ovary-AdenoCA` 
##   z   cells   name           grob

```

```

## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.5627]
##
## $`Panc-AdenoCA`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.5928]
##
## $`Panc-Endocrine`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.6187]
##
## $`Prost-AdenoCA`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.6460]
##
## $`Skin-Melanoma.acral`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.6719]
##
## $`Skin-Melanoma.cutaneous`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.6978]
##
## $`Skin-Melanoma.mucosal`
## [1] "Error : $ operator is invalid for atomic vectors\n"
## attr(,"class")
## [1] "try-error"
## attr(,"condition")
## <simpleError: $ operator is invalid for atomic vectors>
##
## $`SoftTissue-Leiomyo`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (1-1,1-2) arrange text[GRID.text.7167]
##
## $`SoftTissue-Liposarc`
##   z   cells   name          grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]

```



```

pvals_fullRE_M_nonexo_SP <- sapply(fullRE_M_nonexo_SP, function(i) try(wald_TMB_wrapper(i)))

## Check data - slope appears to be of length one (binomial)
pvals_fullRE_DMSL_nonexo_SP <- sapply(fullRE_DMSL_nonexo_SP, function(i) try(wald_TMB_wrapper(i)))

## Check data - slope appears to be of length one (binomial)
pvals_fullREDMnoscaling_SP_nonexo_subsets_and_amalgamations <- sapply(fullREDMnoscaling_SP_nonexo_subsets_and_amalgamations, function(i) try(wald_TMB_wrapper(i)))

## Check data - slope appears to be of length one (binomial)
pvals_fullREDMonefixedlambdanonexo_SP <- sapply(fullREDMonefixedlambdanonexo_SP, function(i) try(wald_TMB_wrapper(i)))

## Check data - slope appears to be of length one (binomial)
pvals_fullREDMonefixedlambdanonexo_SPSaA <- sapply(fullREDMonefixedlambdanonexo_SPSaA, function(i) try(wald_TMB_wrapper(i)))

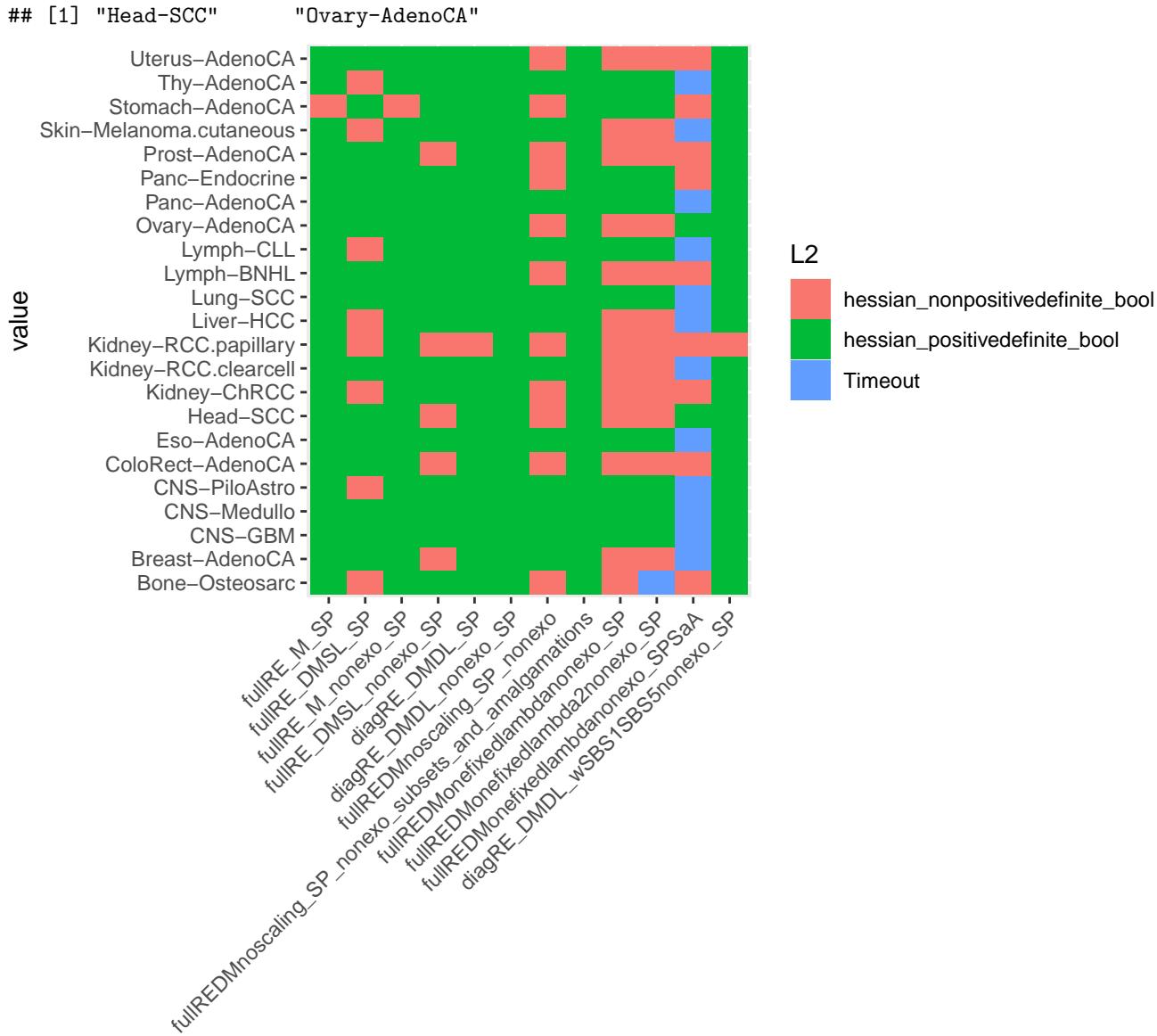
names(fullREDMonefixedlambdanonexo_SPSaA) <- names(pvals_fullREDMonefixedlambdanonexo_SP) <- names(pvals_fullRE_DMSL_nonexo_SP)
names(pvals_fullRE_DMSL_nonexo_SP) <- enough_samples

pvals_diagRE_DMDL_SP

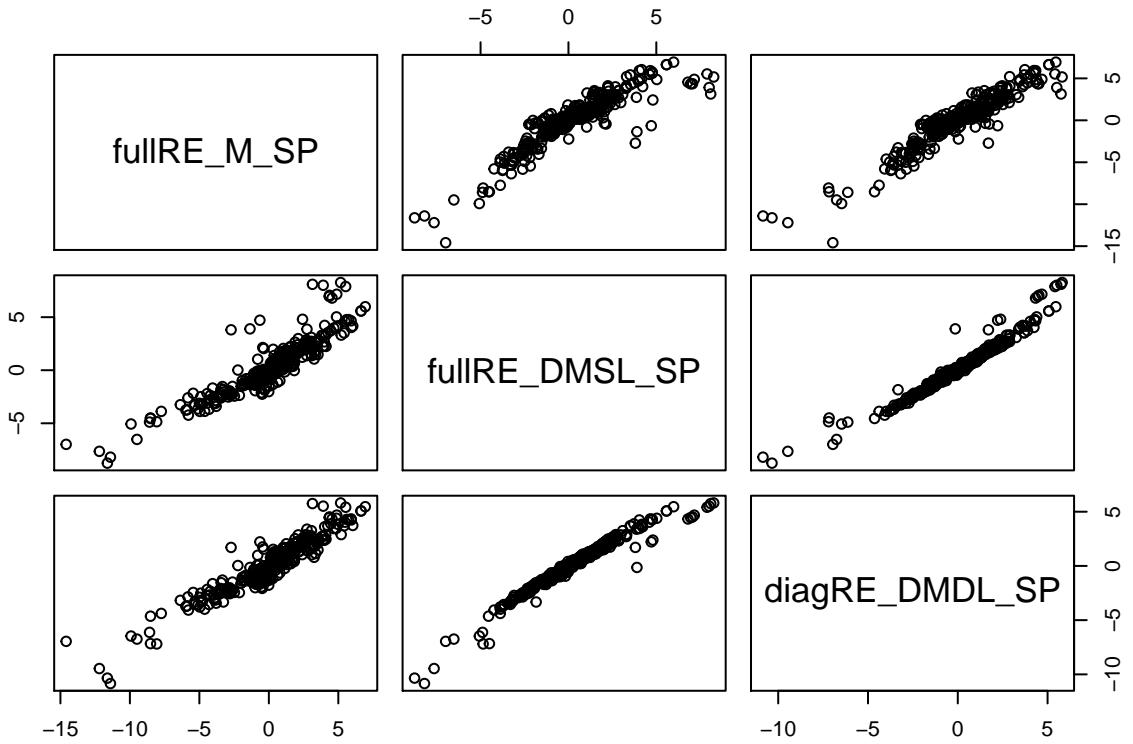
##          Bone-Osteosarc      Breast-AdenoCA          CNS-GBM
##          1.080828e-04      2.239756e-28      3.390137e-03
##          CNS-Medullo       CNS-PiloAstro       ColoRect-AdenoCA
##          8.431463e-03      5.615238e-04      6.356131e-26
##          Eso-AdenoCA        Head-SCC           Kidney-ChRCC
##          5.329093e-21      4.975610e-05      1.562125e-09
##          Kidney-RCC.clearcell Kidney-RCC.papillary Liver-HCC
##          4.027485e-18          NA                 4.747822e-107
##          Lung-SCC            Lymph-BNHL         Lymph-CLL
##          7.747310e-22      3.908637e-19      6.611927e-20
##          Ovary-AdenoCA       Panc-AdenoCA       Panc-Endocrine
##          8.965185e-38      4.096402e-119     3.987099e-10
##          Prost-AdenoCA      Skin-Melanoma.cutaneous Stomach-AdenoCA
##          6.474116e-99      9.272113e-25      1.715150e-06
##          Thy-AdenoCA        Uterus-AdenoCA
##          8.821583e-06      4.819867e-10

## $Timeout
## [1] "Breast-AdenoCA"          "CNS-GBM"
## [3] "CNS-Medullo"             "CNS-PiloAstro"
## [5] "Eso-AdenoCA"              "Kidney-RCC.clearcell"
## [7] "Liver-HCC"                "Lung-SCC"
## [9] "Lymph-CLL"                "Panc-AdenoCA"
## [11] "Skin-Melanoma.cutaneous" "Thy-AdenoCA"
##
## $hessian_nonpositivedefinite_bool
## [1] "Bone-Osteosarc"          "ColoRect-AdenoCA"      "Kidney-ChRCC"
## [4] "Kidney-RCC.papillary"    "Lymph-BNHL"           "Panc-Endocrine"
## [7] "Prost-AdenoCA"           "Stomach-AdenoCA"       "Uterus-AdenoCA"
##
## $hessian_positivedefinite_bool

```



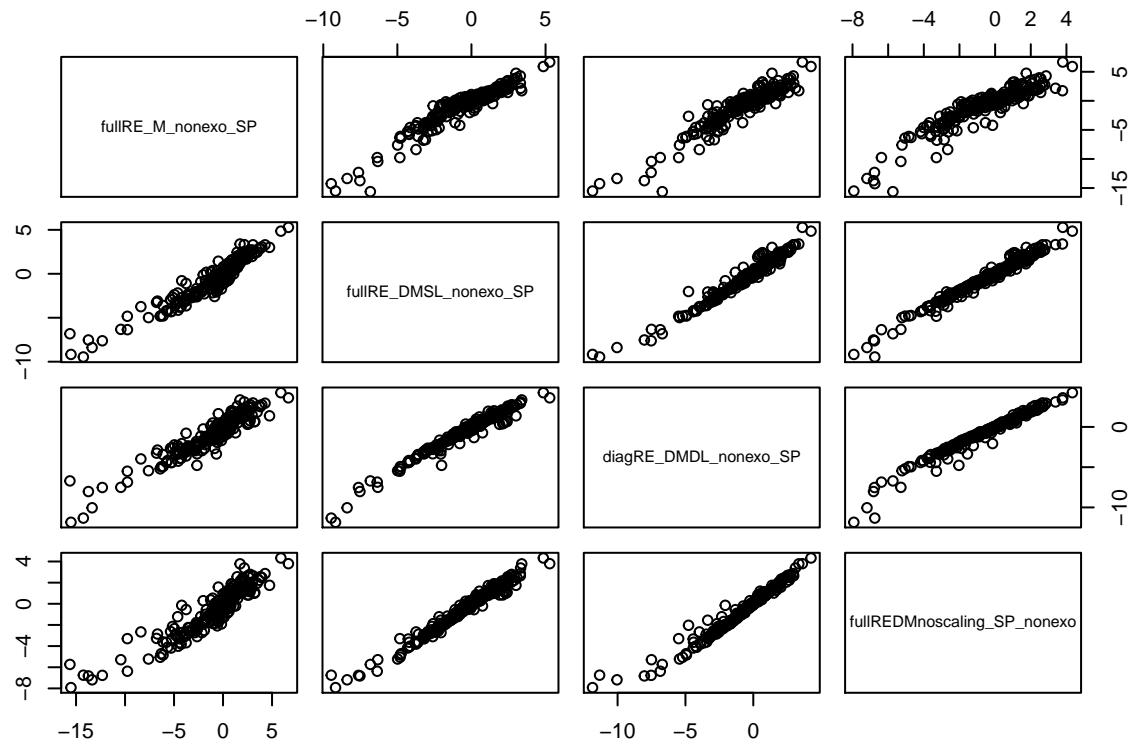
```
## comparison of betas
# give_betas(get(list_models_SP)[[1]][[1]])
all_betas_SP <- do.call('cbind', lapply(c( 'fullRE_M_SP', 'fullRE_DMSL_SP',
                                             'diagRE_DMDL_SP'), function(j) do.call('c', sapply(get(j), function(i) as.vector(give_b
colnames(all_betas_SP) <- c( 'fullRE_M_SP', 'fullRE_DMSL_SP',
                             'diagRE_DMDL_SP')
pairs(all_betas_SP)
```



```

all_betas_SP_nonexo <- do.call('cbind', lapply(c('fullRE_M_nonexo_SP','fullRE_DMSL_nonexo_SP',
                                                 'diagRE_DMDL_nonexo_SP', 'fullREDMnoscaling_SP_nonexo'), function(j) do.call('c', sapply
colnames(all_betas_SP_nonexo) <- c('fullRE_M_nonexo_SP','fullRE_DMSL_nonexo_SP',
                                         'diagRE_DMDL_nonexo_SP', 'fullREDMnoscaling_SP_nonexo')
pairs(all_betas_SP_nonexo)

```

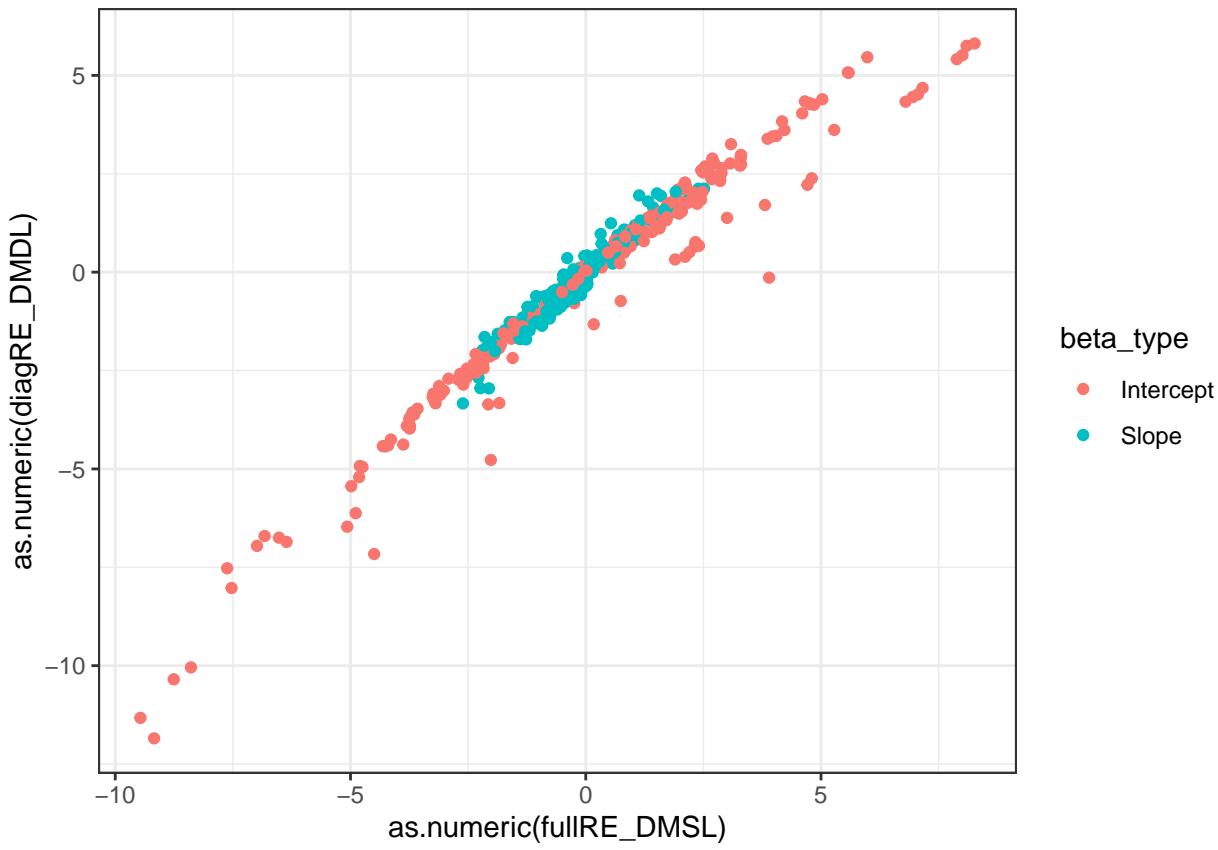


```

##   fullRE_DMSL diagRE_DMDL   fullRE_M beta_type      ct  ct2    sigs
## 1 -0.31531015 -0.1086854 -0.2237241 Slope Bone-Osteosarc B0 nonexo
## 2 -0.71770886 -0.6897242 -0.6712599 Slope Bone-Osteosarc B0 nonexo
## 3 -0.86019179 -0.7960993 -0.6637865 Slope Bone-Osteosarc B0 nonexo
## 4 -0.69897610 -0.4851852 -0.4948419 Slope Bone-Osteosarc B0 nonexo
## 5 -0.09619079  0.1160637  0.1756435 Slope Bone-Osteosarc B0 nonexo
## 6 -0.64276669 -0.4495100 -0.1553669 Slope Bone-Osteosarc B0 nonexo

```

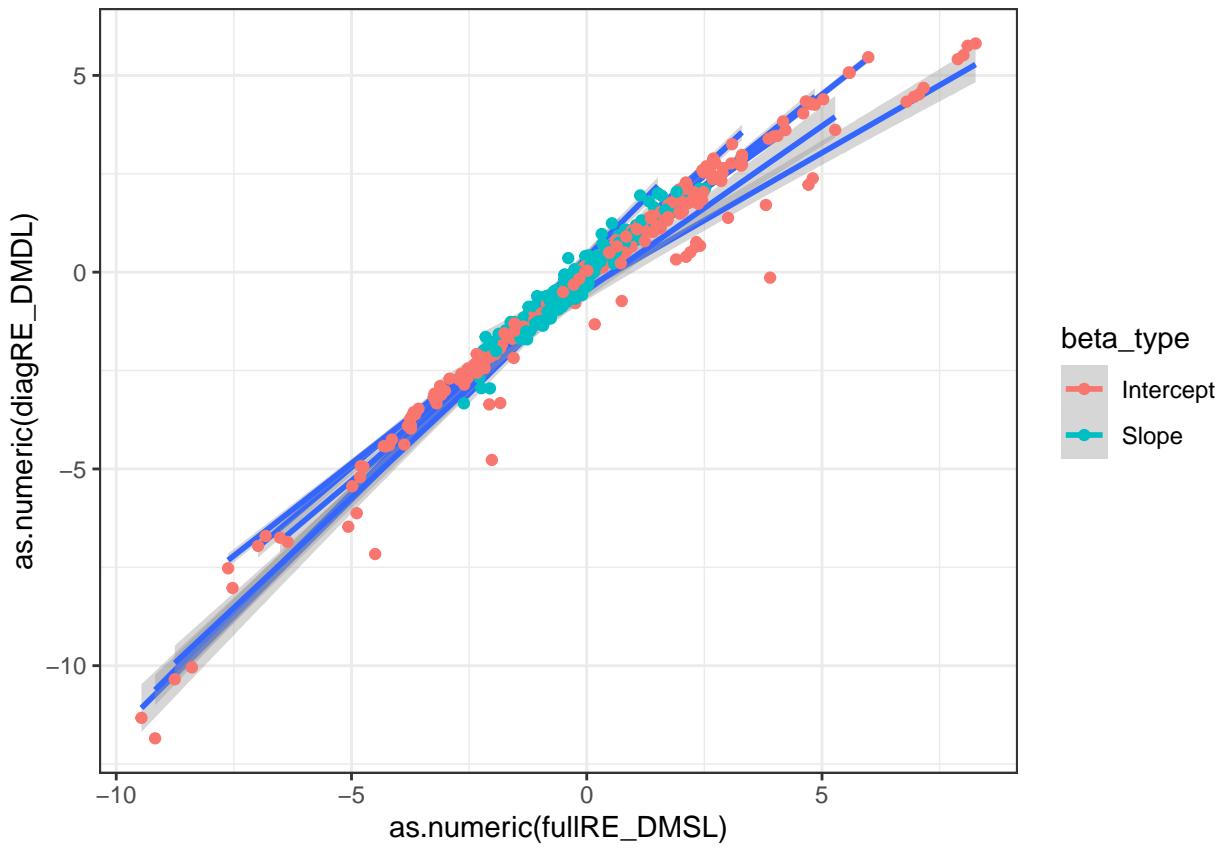
Warning: Removed 218 rows containing missing values (geom_point).



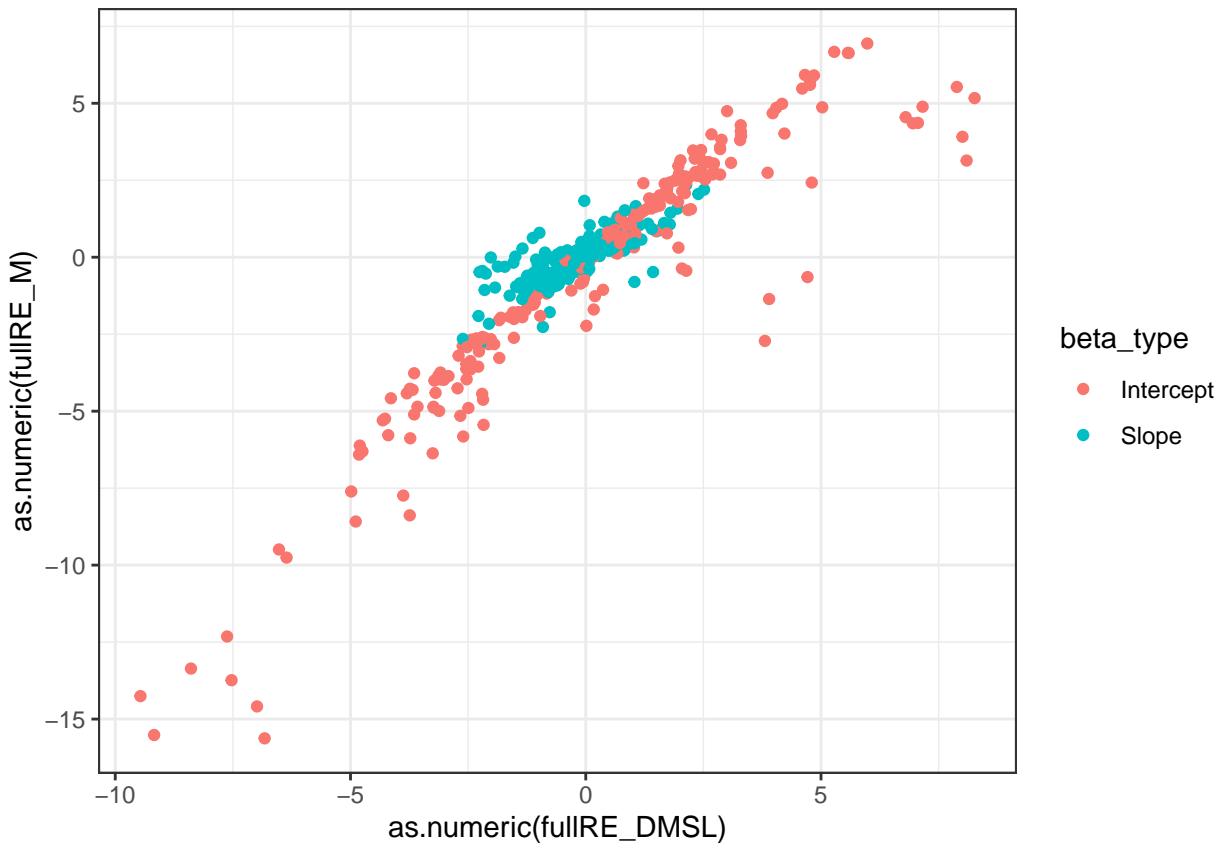
```

## `geom_smooth()` using formula 'y ~ x'
## Warning: Removed 218 rows containing non-finite values (stat_smooth).
## Warning in qt((1 - level)/2, df): NaNs produced
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
## Warning: Removed 218 rows containing missing values (geom_point).

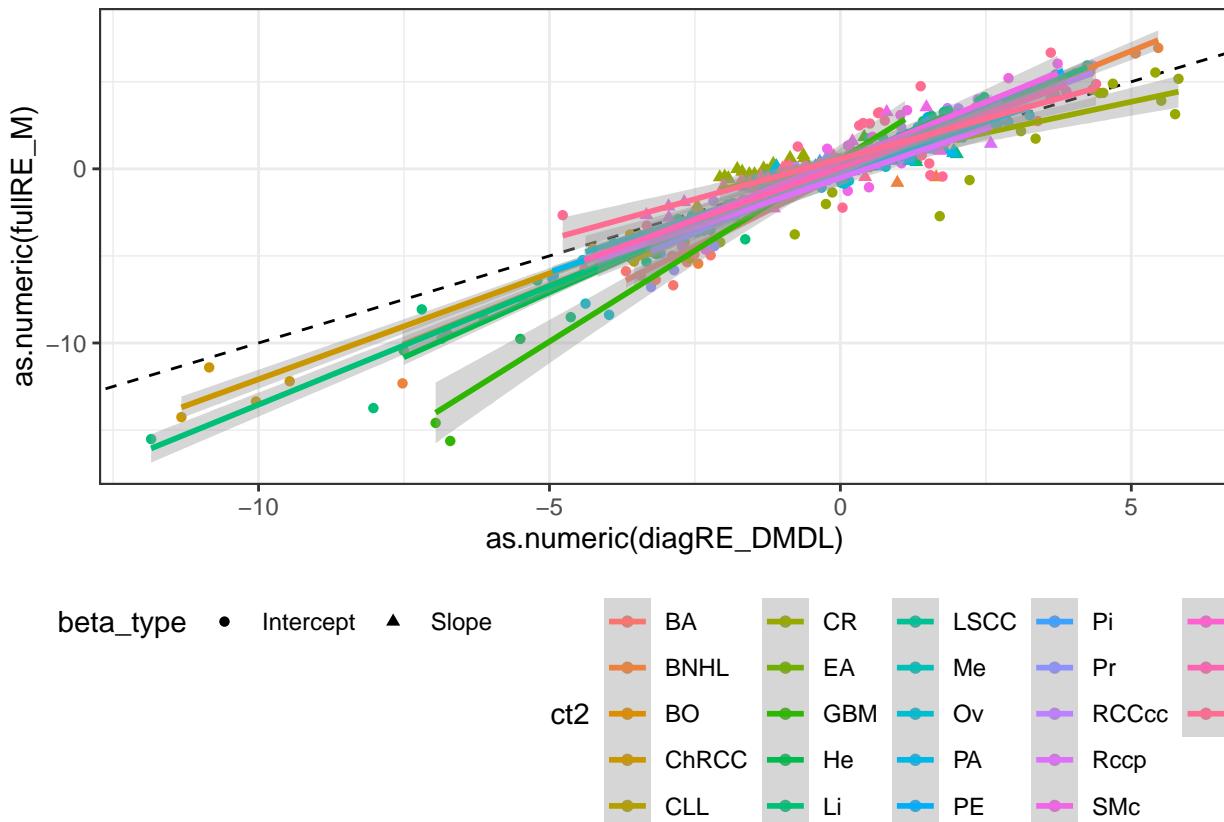
```



```
## Warning: Removed 284 rows containing missing values (geom_point).
```



```
## `geom_smooth()` using formula 'y ~ x'  
## Warning: Removed 80 rows containing non-finite values (stat_smooth).  
## Warning: Removed 80 rows containing missing values (geom_point).
```



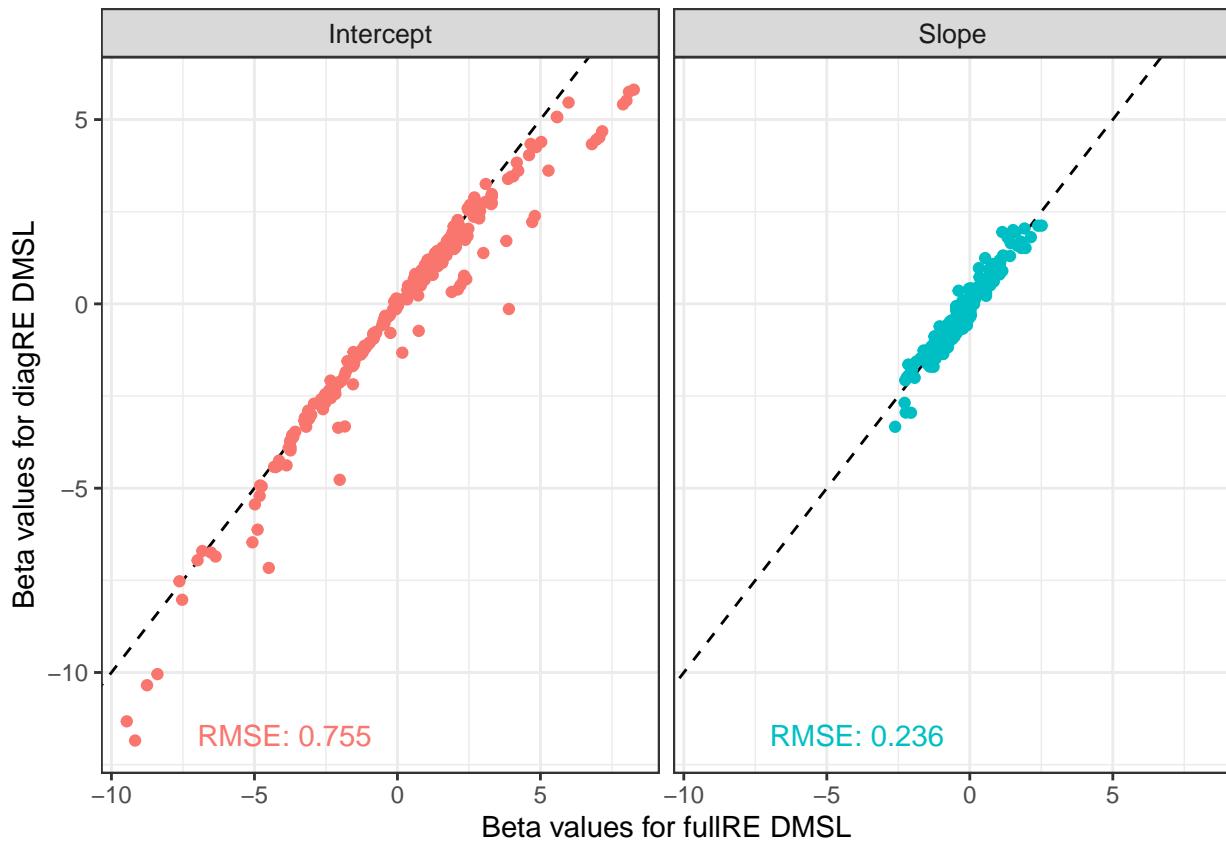
```

##   rmse_diag_full_DMSL rmse_fullDMSL_fullM beta_type
## 1      0.2362400          0.5467457     Slope
## 2      0.7551778          1.7127751 Intercept

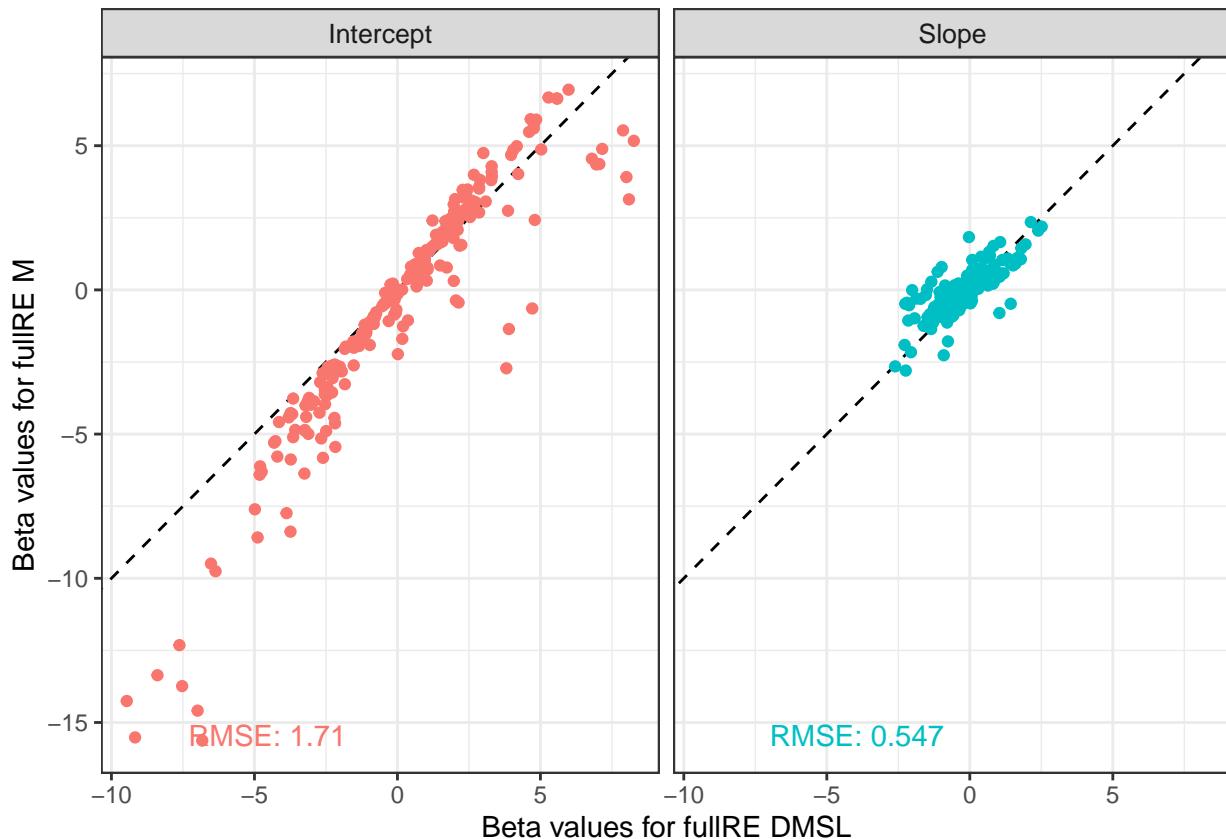
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Removed 218 rows containing missing values (geom_point).

```



```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
## Warning: Removed 284 rows containing missing values (geom_point).
```



```

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Removed 218 rows containing missing values (geom_point).

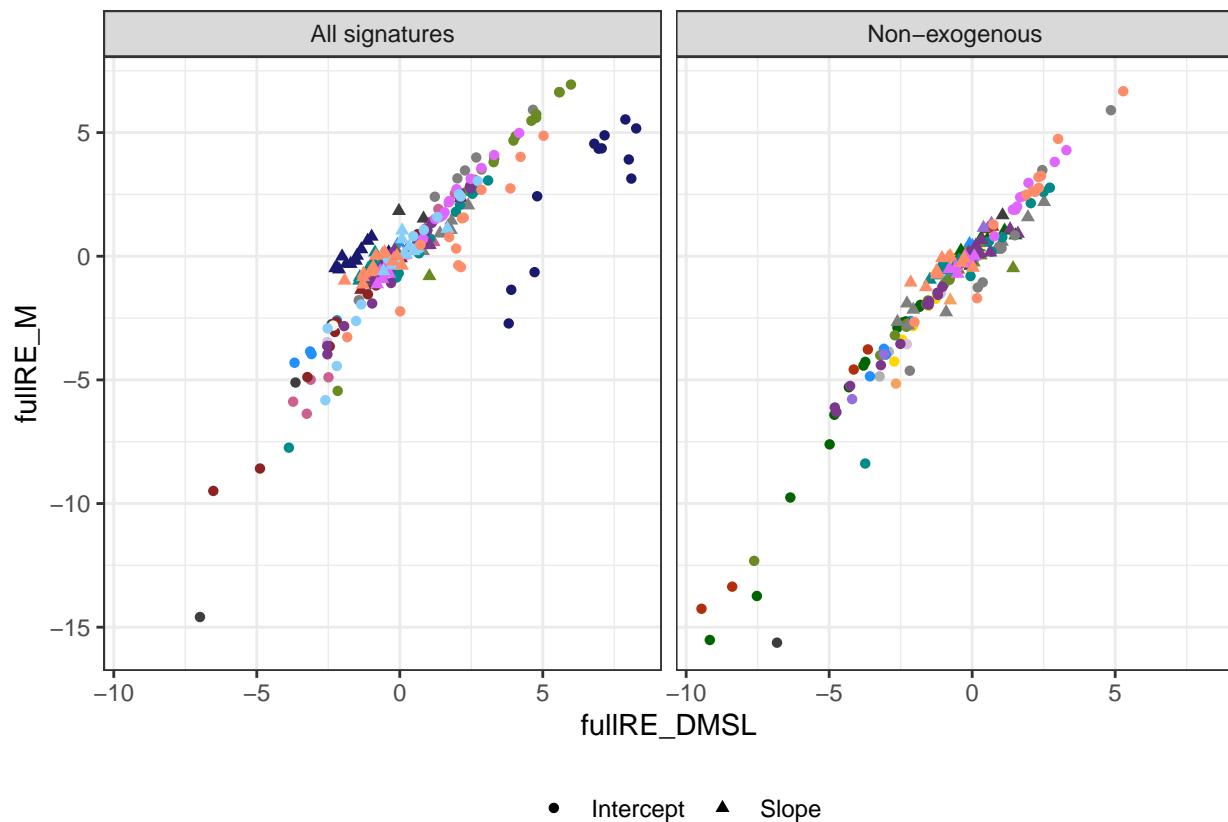
## Warning: Removed 284 rows containing missing values (geom_point).

## pdf
## 2

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

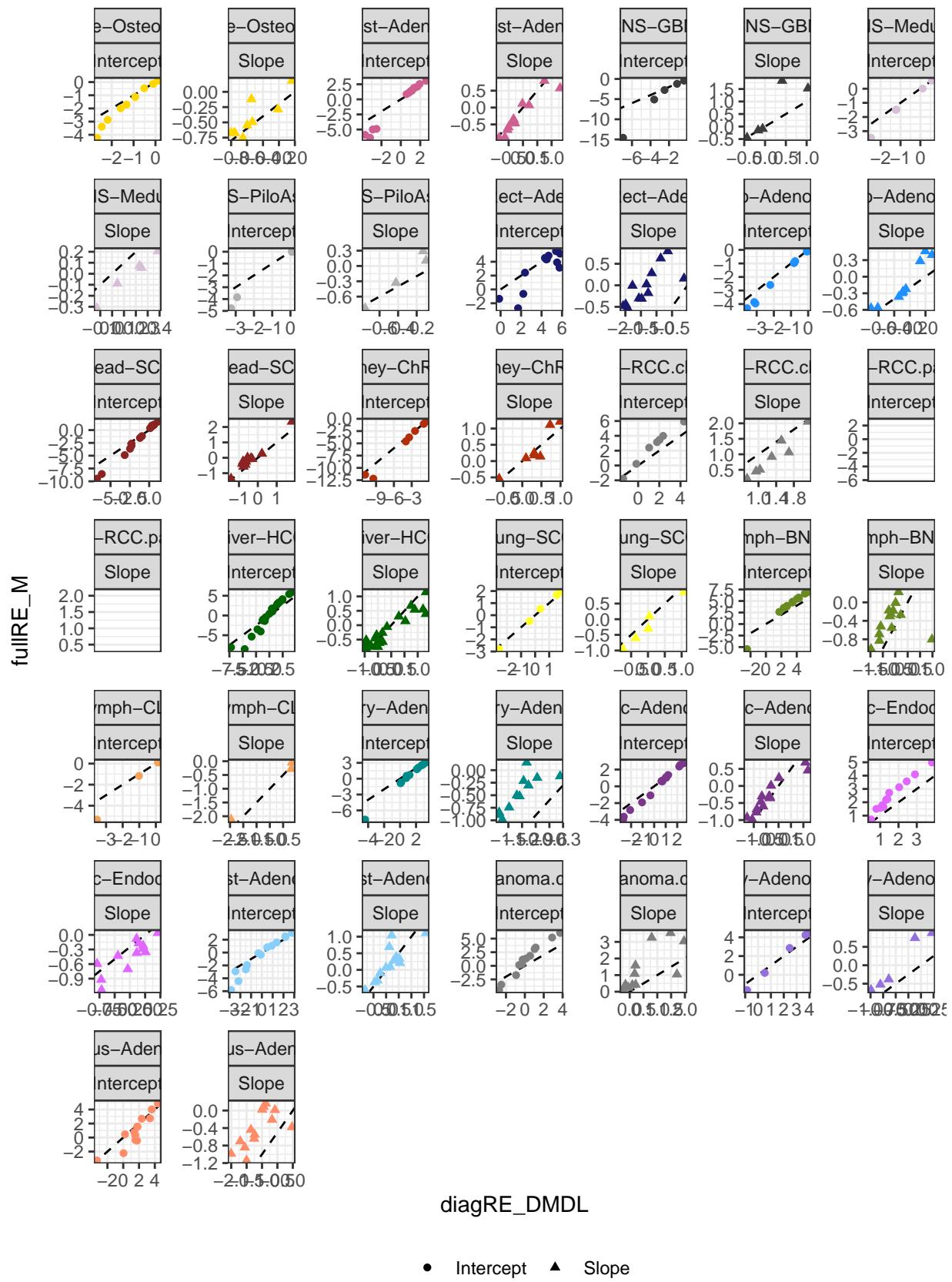
## Warning: Removed 218 rows containing missing values (geom_point).

```



Comparison of beta coefficients for fullRE M and DM

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
## Warning: Removed 14 rows containing missing values (geom_point).
```



● Intercept ▲ Slope

```

head(comparison_betas_models_rbind)

##   fullRE_DMSL diagRE_DMDL   fullRE_M beta_type          ct  ct2      sigs
## 1 -0.31531015 -0.1086854 -0.2237241 Slope Bone-Osteosarc BO Non-exogenous
## 2 -0.71770886 -0.6897242 -0.6712599 Slope Bone-Osteosarc BO Non-exogenous
## 3 -0.86019179 -0.7960993 -0.6637865 Slope Bone-Osteosarc BO Non-exogenous
## 4 -0.69897610 -0.4851852 -0.4948419 Slope Bone-Osteosarc BO Non-exogenous
## 5 -0.09619079  0.1160637  0.1756435 Slope Bone-Osteosarc BO Non-exogenous
## 6 -0.64276669 -0.4495100 -0.1553669 Slope Bone-Osteosarc BO Non-exogenous

comparison_betas_models_rbind_stats_per_ct <- rbind.data.frame(
  cbind.data.frame(comparison_betas_models_rbind %>% filter(beta_type == 'Slope') %>%
    group_by(ct) %>%
    summarise(rmse_diag_full_DMDL=sqrt(mean( (diagRE_DMDL-fullRE_DMSL)^2, na.rm = T )), 
              slope_diag_full_DMDL=as.numeric(try(coefficients(lm(y~x, data = cbind.data
                rmse_fullDMSL_fullM=sqrt(mean( (fullRE_M-fullRE_DMSL)^2, na.rm = T )), 
                slope_fullDMSL_fullM=as.numeric(try(coefficients(lm(y~x, data = cbind.data
      beta_type='Slope'),
  cbind.data.frame(comparison_betas_models_rbind %>% filter(beta_type == 'Intercept') %>%
    group_by(ct) %>%
    summarise(rmse_diag_full_DMDL=sqrt(mean( (diagRE_DMDL-fullRE_DMSL)^2, na.rm = T )), 
              slope_diag_full_DMDL=as.numeric(try(coefficients(lm(y~x, data = cbind.data
                rmse_fullDMSL_fullM=sqrt(mean( (fullRE_M-fullRE_DMSL)^2, na.rm = T )), 
                slope_fullDMSL_fullM=as.numeric(try(coefficients(lm(y~x, data = cbind.data
      beta_type='Intercept'))))

## Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
##   0 (non-NA) cases

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
##   0 (non-NA) cases

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
##   0 (non-NA) cases

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
##   0 (non-NA) cases

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
##   0 (non-NA) cases

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
##   0 (non-NA) cases

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
##   0 (non-NA) cases

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...):
##   0 (non-NA) cases

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

comparison_betas_models_rbind_stats_per_ct$ct2=renaming_pcawg[,2][match(comparison_betas_models_rbind_st

```

```

comparison_betas_models_rbind_stats_per_ct_pool_beta <- cbind.data.frame(comparison_betas_models_rbind%>%
  group_by(ct) %>%
  summarise(rmse_diag_full_DMDL=sqrt(mean( (diagRE_DMDL-fullRE_DMSL)^2, na.rm = T )), 
            slope_diag_full_DMDL=as.numeric(try(coefficients(lm(y~x, data = cbind.data,
            rmse_fullDMSL_fullM=sqrt(mean( (fullRE_M-fullRE_DMSL)^2, na.rm = T )), 
            slope_fullDMSL_fullM=as.numeric(try(coefficients(lm(y~x, data = cbind.data

## Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##   0 (non-NA) cases

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##   0 (non-NA) cases

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##   0 (non-NA) cases

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

comparison_betas_models_rbind_stats_per_ct_pool_beta$ct2=renaming_pcawg[,2] [match(comparison_betas_models

pcawg_palette <- pcawg.colour.palette(gsub("\\.*", "", comparison_betas_models_rbind_stats_per_ct$ct),
                                         scheme = "tumour_subtype")
pcawg_palette[names(pcawg_palette) == 'Lung-SCC'] <- '#ffff29'
names(pcawg_palette) <- comparison_betas_models_rbind_stats_per_ct$ct

ggplot(comparison_betas_models_rbind, aes(x=as.numeric(fullRE_DMSL), y=as.numeric(fullRE_M), col=ct))+geom_abline(slope = 1, intercept = 0, lty='dashed')+theme_bw()+
  geom_point()+theme_bw()+
  facet_wrap(~beta_type)+
  labs(x='Beta values for fullRE DMSL', y='Beta values for fullRE M')+
  geom_smooth(aes(group=ct2), method = "lm")+
  theme(legend.title=element_blank(),
        strip.text.x = element_text(size = 10),
        legend.text=element_text(size=10), legend.position = "bottom")+guides(col=FALSE)+
  scale_color_manual(values = pcawg_palette)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 284 rows containing non-finite values (stat_smooth).

## Warning in qt((1 - level)/2, df): NaNs produced

## Warning in qt((1 - level)/2, df): NaNs produced

## Warning in qt((1 - level)/2, df): NaNs produced

## Warning in qt((1 - level)/2, df): NaNs produced

```

```

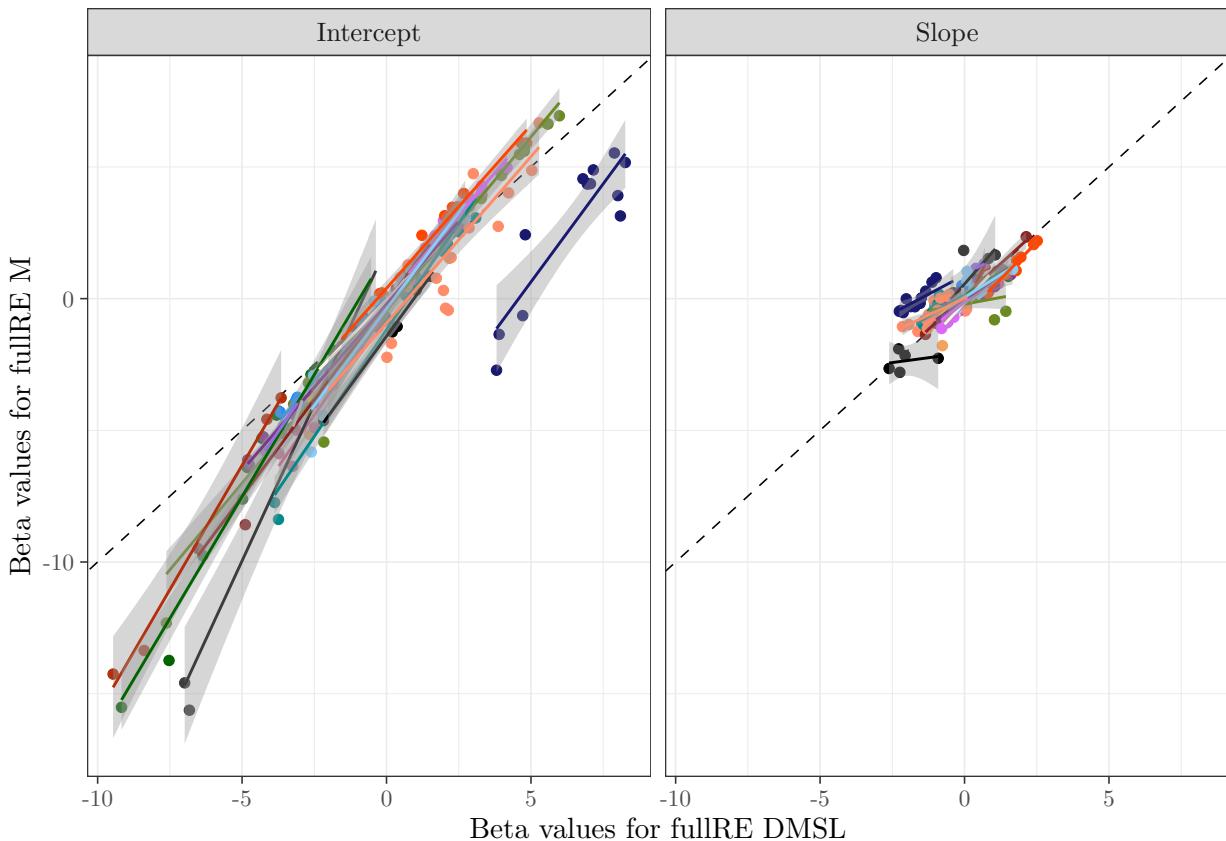
## Warning: Removed 284 rows containing missing values (geom_point).
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

```



```

# ggplot(comparison_betas_models_rbind_stats_per_ct_pool_beta, aes(x=rmse_fullDMSL_fullM,
# #                                     y=slope_fullDMSL_fullM, col=ct, label=ct2))+ 
#   geom_point()+geom_label_repel()+theme_bw()+labs(x='RMSE of betas between\nfullDMSL and fullM',
# #                                     y = 'Slope of linear model of betas\n(fullDMSL ~ fullM)')

# ggplot(comparison_betas_models_rbind_stats_per_ct, aes(x=rmse_fullDMSL_fullM,
# #                                     y=slope_fullDMSL_fullM, col=ct, label=ct))+ 
#   geom_point()+geom_label_repel()+theme_bw()+labs(x='RMSE of betas between\nfullRE DMSL and fullRE M',
# #                                     y = 'Slope of linear model of betas\n(fullRE DMSL ~ fullRE M)')
#   facet_wrap(~beta_type)+theme(strip.text.x = element_text(size = 10))

```

```

# head(comparison_betas_models_rbind_stats_per_ct)

ggplot(comparison_betas_models_rbind_stats_per_ct, aes(x=rmse_fullDMSL_fullM,
                                                       y=slope_fullDMSL_fullM, col=ct, label=ct))+
  geom_point() + geom_label_repel() + theme_bw() + labs(x='RMSE of betas between\nfullRE DMSL and fullRE M',
                                                       y = 'Slope of linear model of betas\n(fullRE DMSL ~ fullRE M)') +
  facet_wrap(~beta_type) + theme(strip.text.x = element_text(size = 10)) +
  scale_color_manual(values = pcawg_palette)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

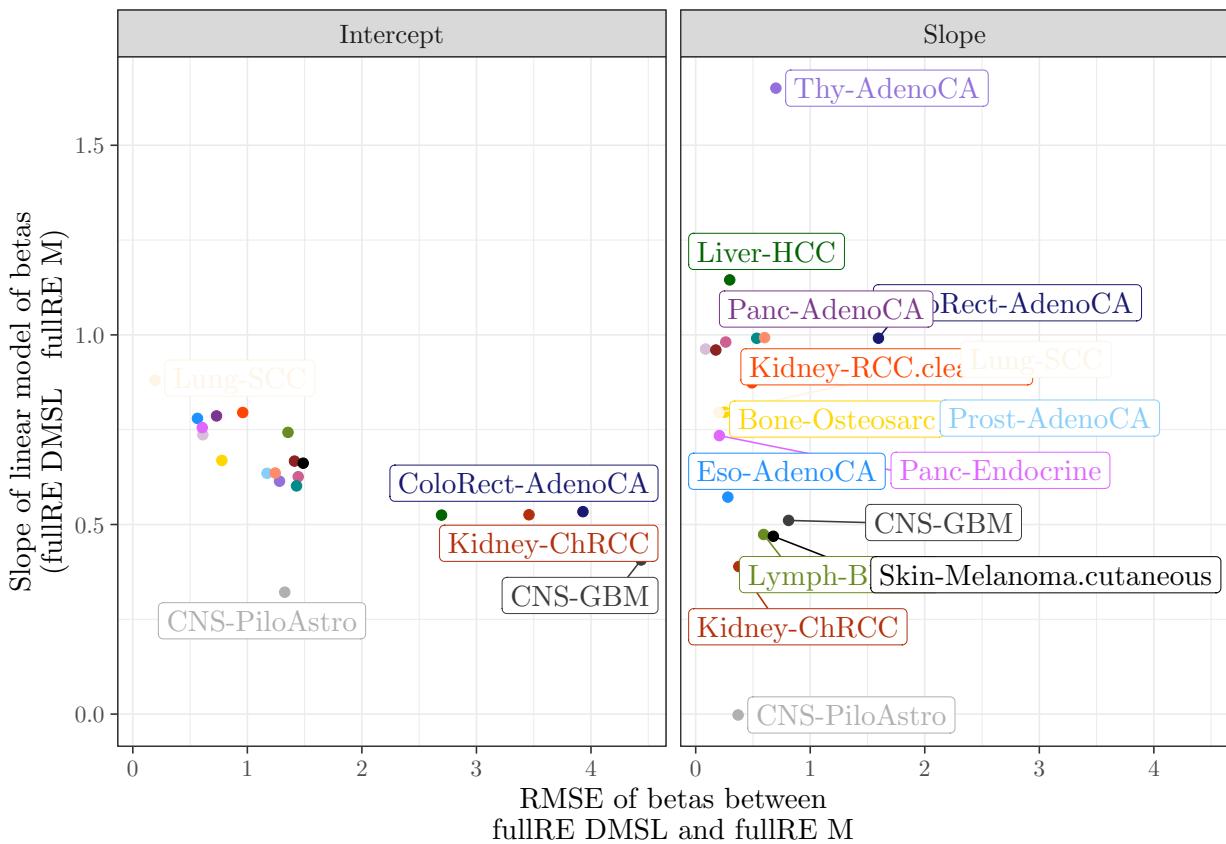
## Warning: Removed 6 rows containing missing values (geom_point).

## Warning: Removed 6 rows containing missing values (geom_label_repel).

## Warning: ggrepel: 15 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

## Warning: ggrepel: 5 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```



```

tikz( 'summary_TMB_PCAWG_SP_files/figure-latex/beta_coef_differences_in_models_plots_tikz2.tex',
      height = 2.5, width=5.5)
cowplot::plot_grid(ggplot(comparison_betas_models_rbind_cut, aes(x=fullRE_DMSL, y=fullRE_M, col=ct, shape=ct),
                           geom_abline(slope = 1, intercept = 0, lty='dashed'))+

```

```

geom_point() + theme_bw() + facet_wrap(~sigs) +
scale_color_manual(values = pcawg_palette) + guides(col=F, shape=F) +
labs(shape='', x='fullRE DMSL', y='fullRE M'), #+theme(legend.position = "bottom"),
ggplot(comparison_betas_models_rbind_cut, aes(x=fullRE_DMSL, y=diagRE_DMDL, col=ct, shape=beta_type)) +
  geom_abline(slope = 1, intercept = 0, lty='dashed') +
geom_point() + theme_bw() + facet_wrap(~sigs) +
scale_color_manual(values = pcawg_palette) + guides(col=F, shape=F) +
labs(shape='', x='fullRE DMSL', y='diagRE DMDL'), rel_widths = c(1, 1), ncol = 2)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Removed 218 rows containing missing values (geom_point).

## Warning: Removed 218 rows containing missing values (geom_point).

dev.off()

## tikz output
##          2

tikz('summary_TMB_PCAWG_SP_files/figure-latex/beta_coef_differences_in_models_plots_tikz3.tex',
      height = 2.5, width=5.5)
ggplot(comparison_betas_models_rbind_stats_per_ct, aes(x=rmse_fullDMSL_fullM,
                                                       y=slope_fullDMSL_fullM, col=ct, label=ct)) +
  geom_point() + geom_label_repel() + theme_bw() + labs(x='RMSE of betas between\nfullRE DMSL and fullRE M',
                                                       y = 'Slope of linear model of betas\n(fullRE DMSL ~ ful
  facet_wrap(~beta_type) + theme(strip.text.x = element_text(size = 10)) +
  scale_color_manual(values = pcawg_palette) #+theme(legend.position = "bottom")

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Removed 6 rows containing missing values (geom_point).

## Warning: Removed 6 rows containing missing values (geom_label_repel).

## Warning: ggrepel: 18 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

## Warning: ggrepel: 17 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

dev.off()

## tikz output
##          2

head(comparison_betas_models_rbind_stats_per_ct)

##           ct rmse_diag_full_DMDL slope_diag_full_DMDL rmse_fullDMSL_fullM
## 1   Bone-Osteosarc      0.18479169      0.85364222      0.25351523
## 2   Breast-AdenoCA     0.10332852      0.91205139      0.26250850
## 3       CNS-GBM        0.28542654      1.01099718      0.81084551

```

```

## 4      CNS-Medullo      0.09067728      0.94821655      0.08629327
## 5      CNS-PiloAstro    0.32907695     -0.00807609      0.37025033
## 6 ColoRect-AdenoCA    0.24657998      0.95097590      1.59564225
## slope_fullDMSL_fullM beta_type ct2
## 1      0.796157804    Slope B0
## 2      0.981125224    Slope BA
## 3      0.510734188    Slope GBM
## 4      0.962647186    Slope Me
## 5      -0.002269125   Slope Pi
## 6      0.991218105   Slope CR

tikz( 'summary_TMB_PCAWG_SP_files/figure-latex/beta_coef_differences_in_models_plots_tikz3b.tex',
      height = 2.5, width=5.5)
ggplot(comparison_betas_models_rbind_stats_per_ct, aes(x=rmse_fullDMSL_fullM,
                                                       y=rmse_diag_full_DMDL, col=ct, label=ct))+
  geom_point() + geom_label_repel(size=3) + theme_bw() +
  # labs(x='RMSE of betas between\nfullRE DMSL and fullRE M',
  #       y = 'Slope of linear model of betas\n(fullRE DMSL ~ j'
  guides(col=F) +
  labs(x='rmse fullDMSL fullM', y='rmse diag full DMDL') +
  facet_wrap(~beta_type, scales="free") + theme(strip.text.x = element_text(size = 10)) +
  scale_color_manual(values = pcawg_palette) #+theme(legend.position = "bottom")

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Removed 4 rows containing missing values (geom_point).

## Warning: Removed 4 rows containing missing values (geom_label_repel).

## Warning: ggrepel: 16 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

## Warning: ggrepel: 13 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

dev.off()

## tikz output
##      2

## `summarise()` has grouped output by 'ct'. You can override using the `groups` argument.
## `summarise()` has grouped output by 'ct'. You can override using the `groups` argument.

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## pdf
##      2

```

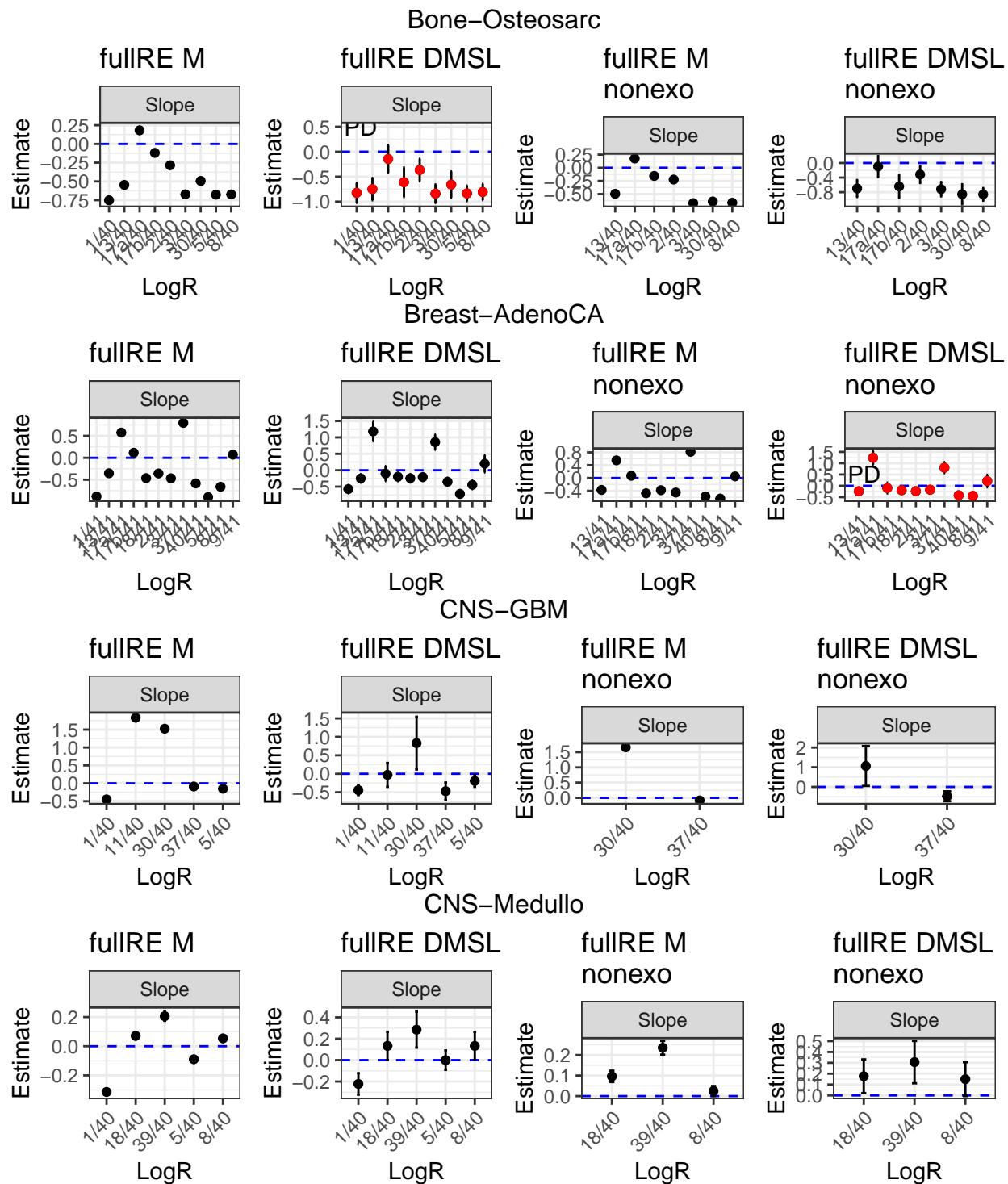
differences between signatures from sigprofiler from the paper, and the ones I get

See Rmd signatures_from_PCAWG_paper.Rmd for the figures

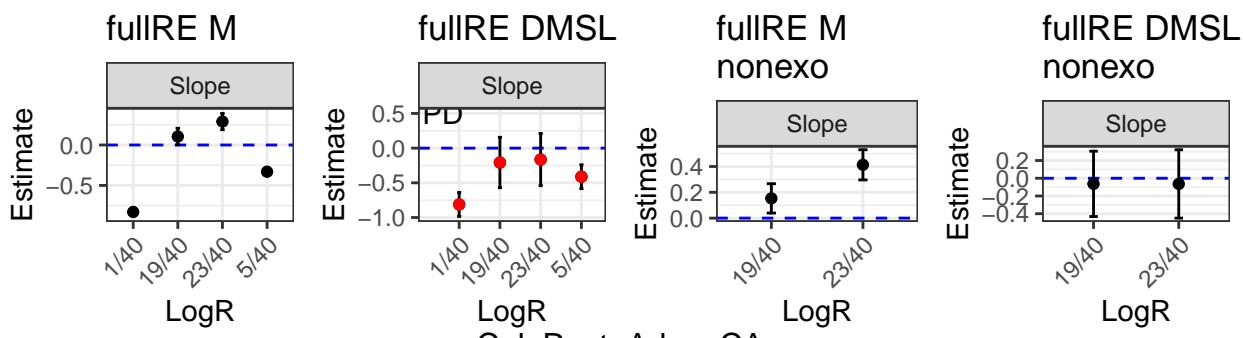
- biliary adenoca similar, though not exact
- bladder tcc: very similar
- bone benign: similar

- bone epith: extremely similar
- bone osteosarc: I have a lot of SBS8, which they don't. Other than that, similar
- breast adenicaL I have a lot of SBS9, which they don't. Other than that, similar
- breast DCIS: I have more SBS40 than they do
- breast lobularca: very similar
- cervix adenoca: very similar
- cervix SCC: very similar, although I have more SBS40
- CNS GBM: very similar, mine seem to be more homogeneous
- CNS medullo: very similar, mine seem to be more homogeneous
- CNS oligo: very similar, mine seem to be more homogeneous
- CNS piloastro: very similar, mine seem to be more homogeneous
- Colorect adenoca: quite similar
- eso adenoca: very similar
- head scc: very similar, mine seem to be more homogeneous
- kidney chrc: very similar
- kidney rcc.clearcell: very similar, although I have more SBS29
- kidney papillary: only I have it
- liver hcc: different. I have a lot of SBS40 and SBS12, they have mostly SBS5 ***
- lung adenoca: very similar
- lung SCC: very similar, though I have more SBS8
- lymph BNHL: very similar
- Lymph CLL: very similar, althpugh theirs are much more sparse
- myeloid AML: very similar, although I don't have any SBS60 and they seem to have
- myeloid MDS: I don't have it
- myeloid MPN: similar, although mine is much more sparse
- ovary adenoca: different. I have a lot of SBS40, which in their case is rare, and they have much more of SBS3 than I do ***
- panc-adenoca: different. I have a lot of SBS8 that they don't have. ***
- panc-endocrine: sort of similar. I have more SBS8 than they and they have more SBS5
- Prost-adenoca: sort of similar, I have more SBS8
- skin-melanoma.acral: they don't have this category. They have "skin-melanoma", which might be both together? (!!!) Similar exposures...
- softtissue-leiomyo: very similar exposures
- softtissue-liposarc: very similar exposures
- stomach adenoca: very similar, mine seem to be more homogeneous
- thy-adenoca: very similar
- uterus-adenoca: very similar

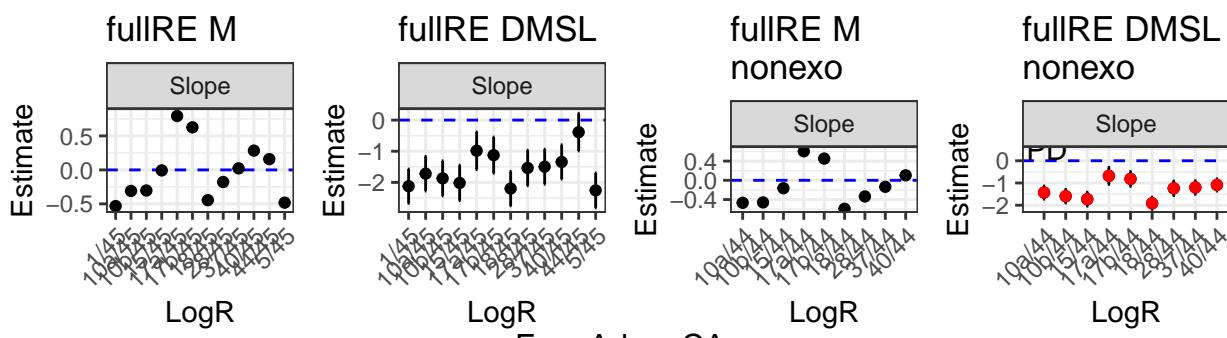
Betas from PCAWG subset of signatures



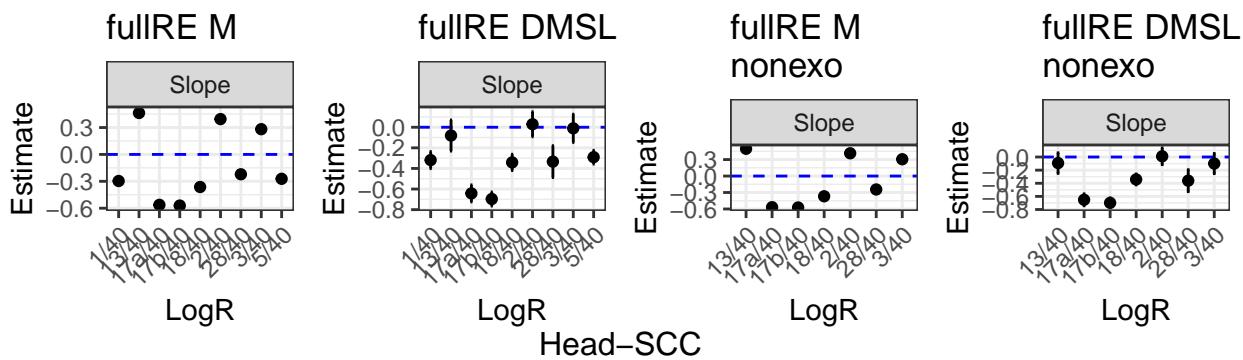
CNS–PiloAstro



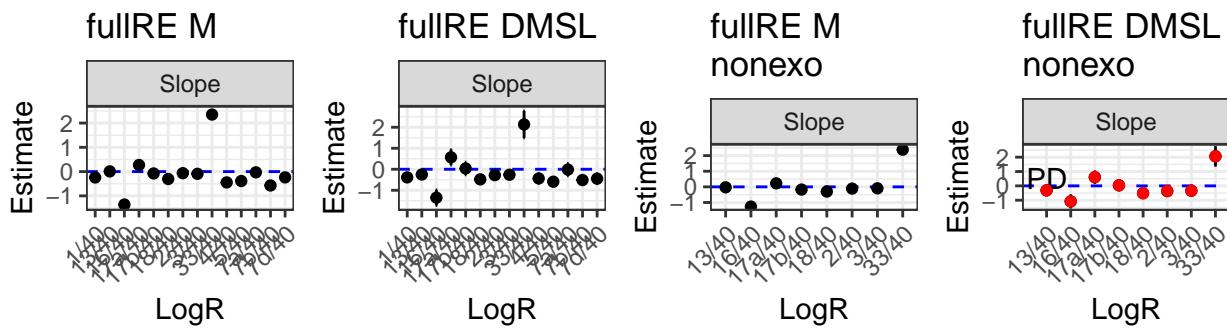
ColoRect–AdenoCA

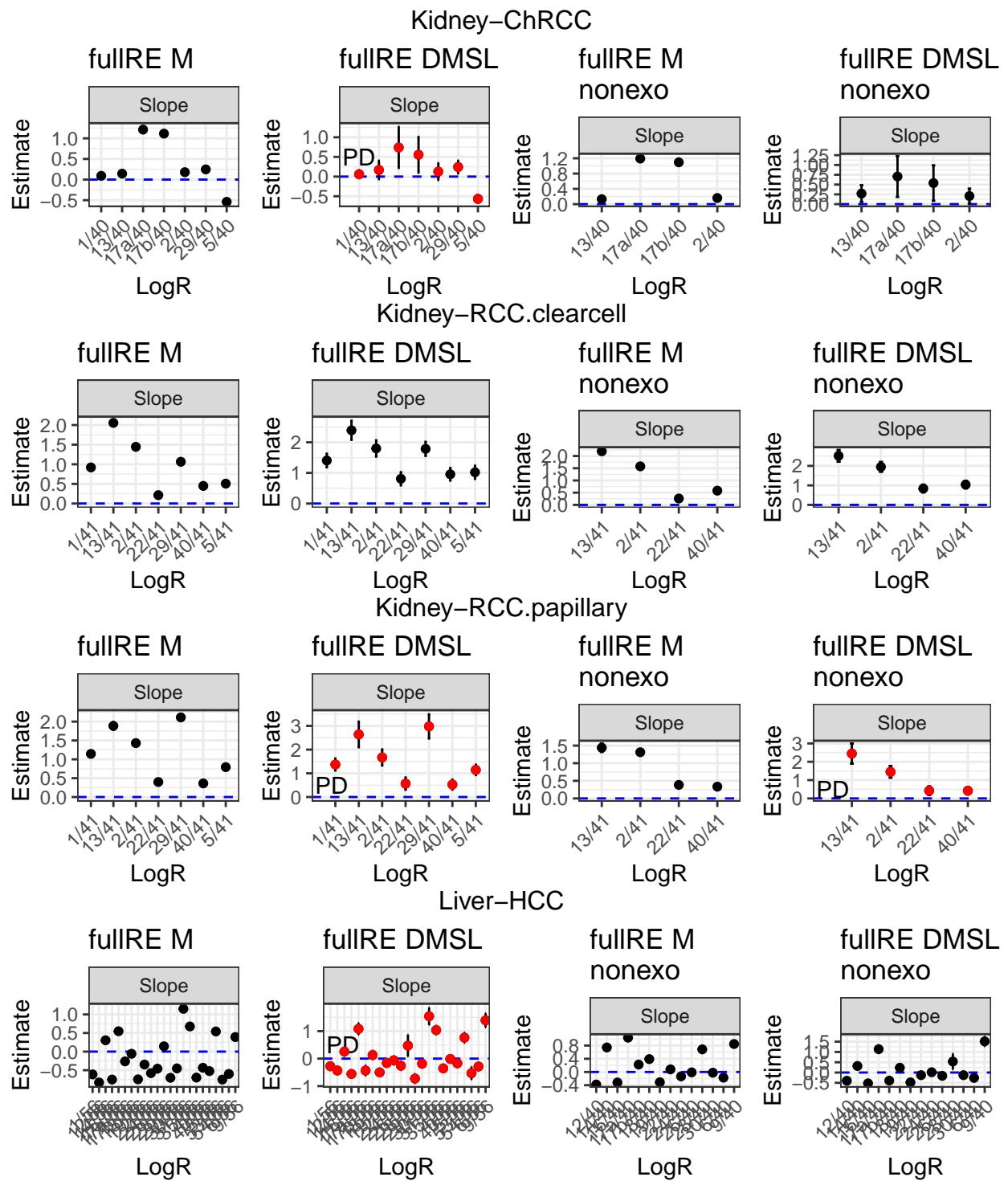


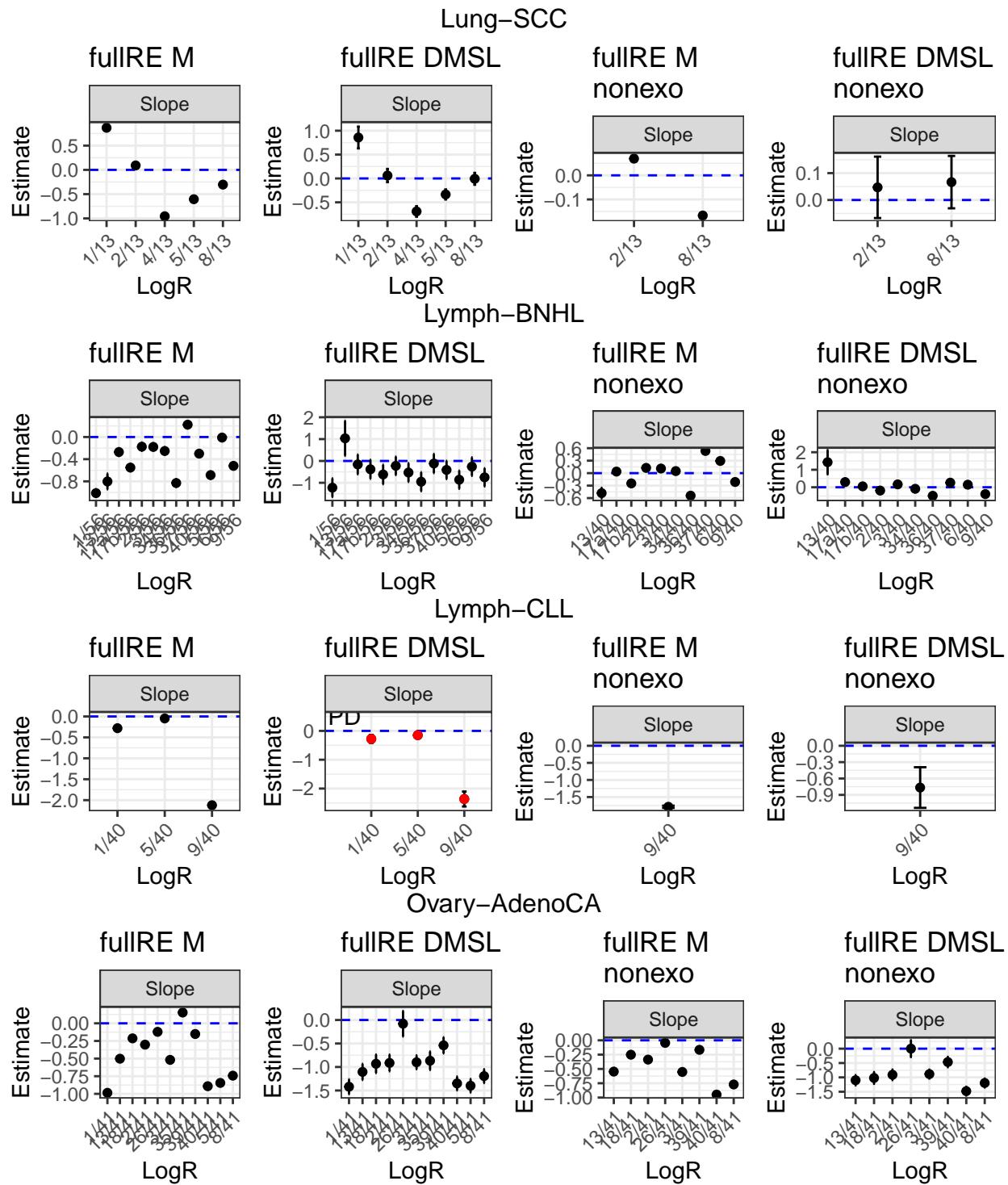
Eso–AdenoCA

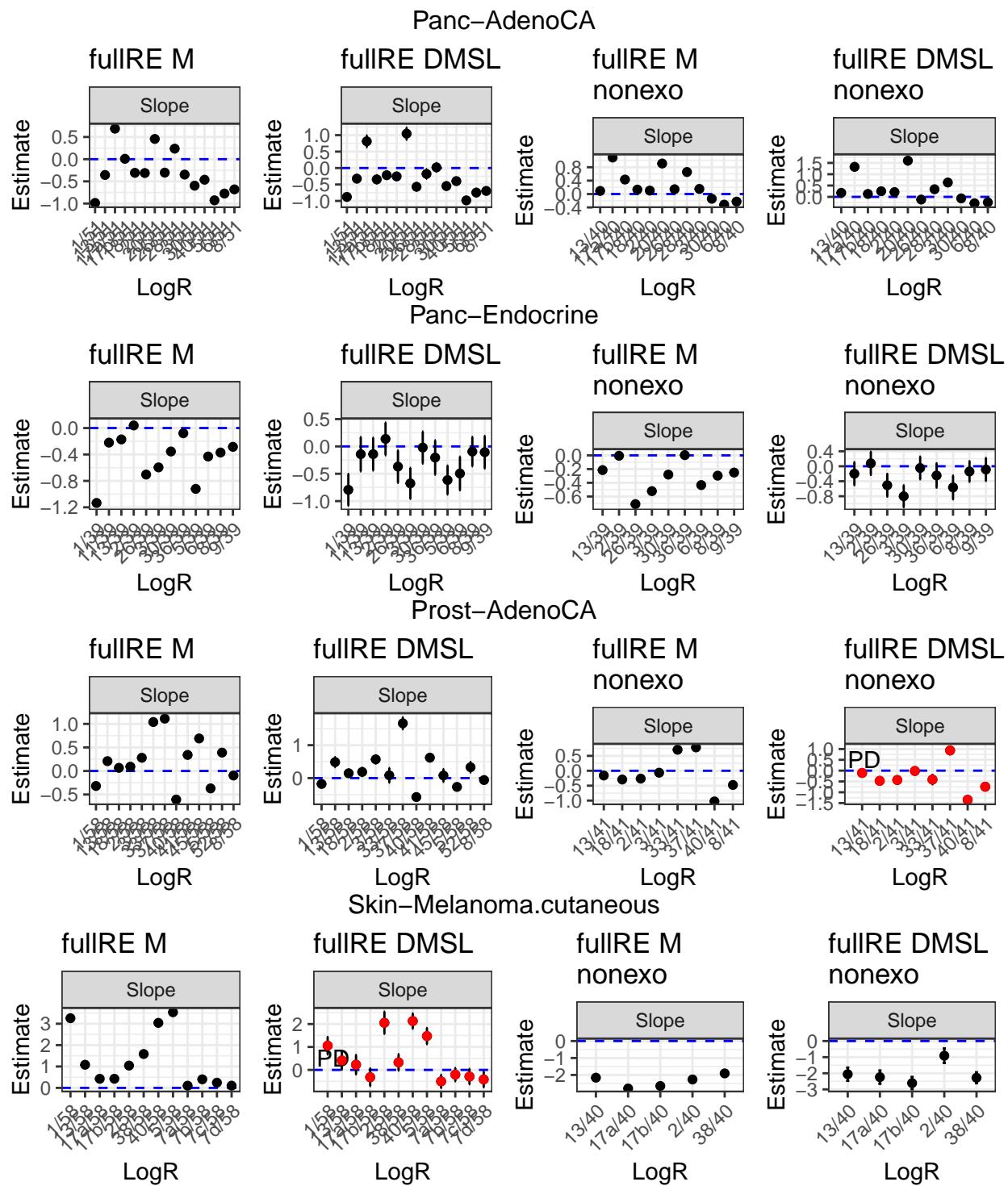


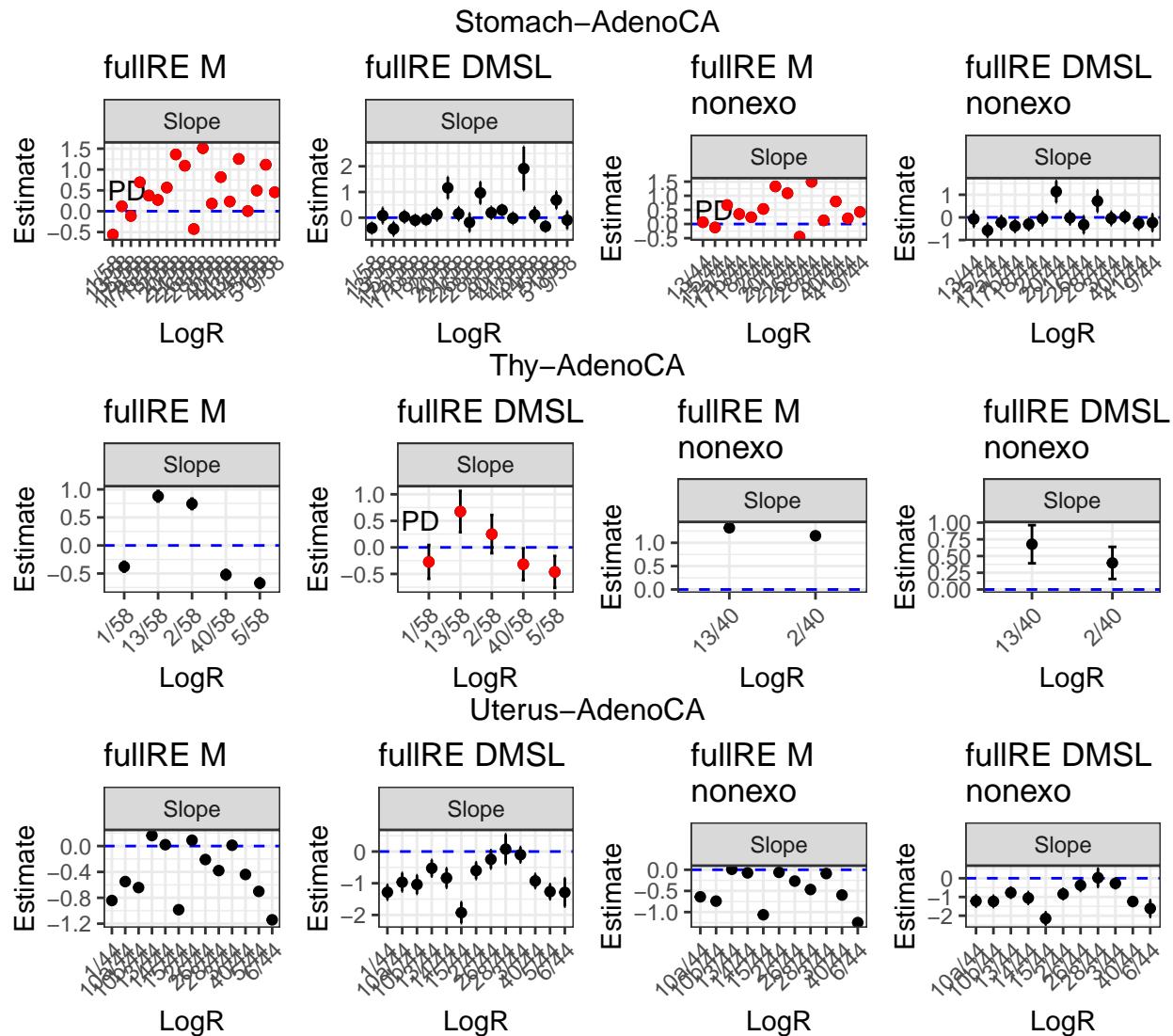
Head–SCC











Overdispersion parameters in double-lambda DM

```

ovrdisp <- do.call('rbind.data.frame', lapply(1:length(diagRE_DMDL_nonexo_SP), try(function(idx){
  if(diagRE_DMDL_nonexo_SP[[idx]]$pdHess){
    cbind.data.frame( plot_lambdas(diagRE_DMDL_nonexo_SP[[idx]], return_df=T, plot=F), ct=names(diagRE_DM
} else{
  c(NA, NA)
}
})))
ovrdisp[ovrdisp$name == 'Lambda 1','name'] = 'Early'
ovrdisp[ovrdisp$name == 'Lambda 2','name'] = 'Late'

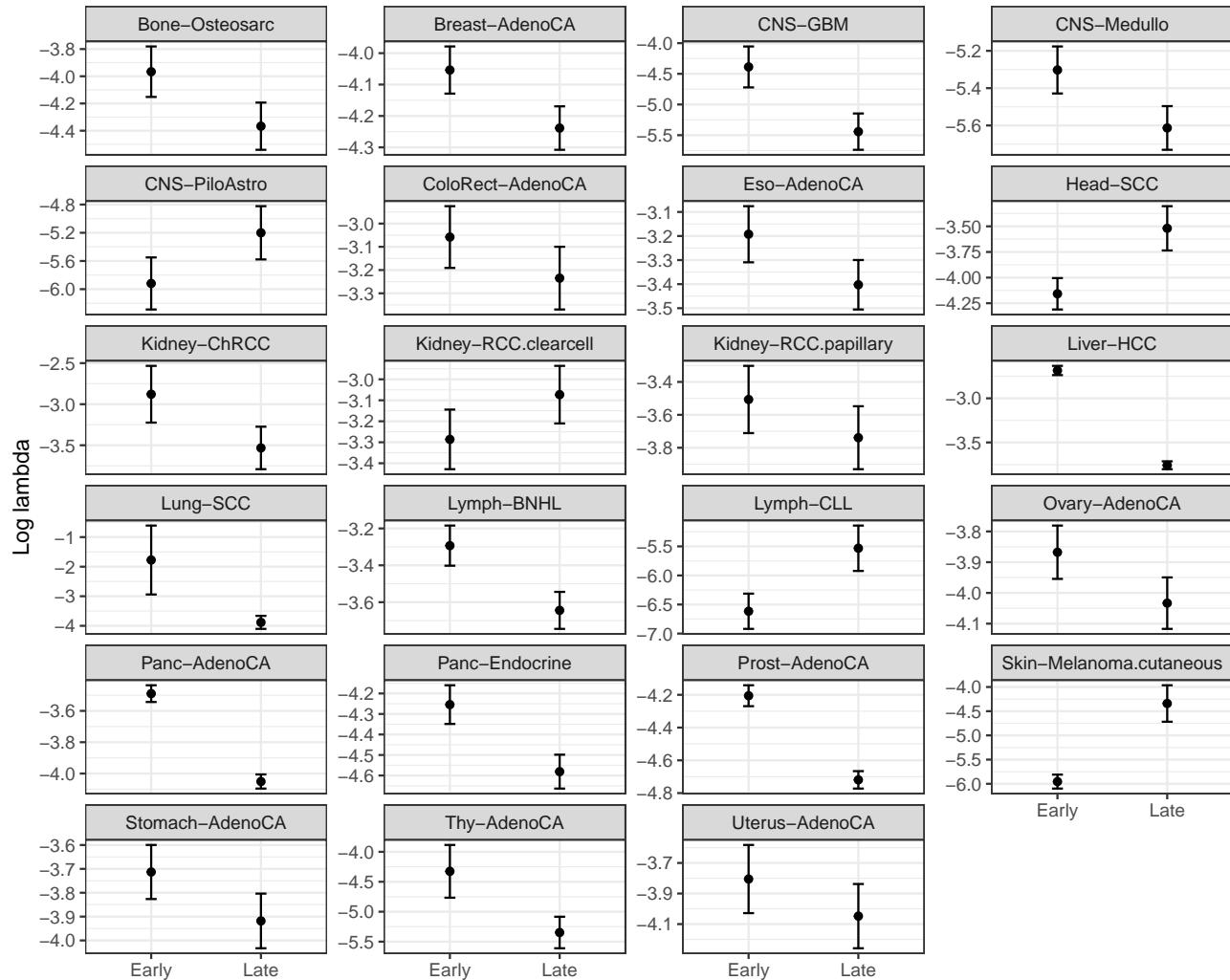
ggplot(ovrdisp, aes(x=name, y=Estimate))+
  geom_point()+
  geom_errorbar(aes(ymin=Estimate-Std..Error,

```

```

  ymax=`Estimate`+`Std..Error`), width=.1)+
theme_bw()+
facet_wrap(~ct, scales = "free_y", nrow=6)+
labs(x=' ', y='Log lambda')

```



Test for differential precision (1/overdispersion) parameter

```

differential_precision <- p.adjust(sapply(diagRE_DMDL_nonexo_SP, wald_TMB_wrapper_overdisp), method = 'fdr')
names(differential_precision) <- names(diagRE_DMDL_nonexo_SP)
sort(differential_precision)

```

##	Liver-HCC	Panc-AdenoCA	Skin-Melanoma.cutaneous
##	5.160159e-58	7.096293e-18	9.923539e-09
##	Prost-AdenoCA	Lung-SCC	Head-SCC
##	1.219769e-07	1.139987e-05	8.968677e-02
##	Panc-Endocrine	CNS-GBM	Lymph-CLL
##	8.968677e-02	1.094428e-01	1.257732e-01
##	Thy-AdenoCA	Lymph-BNHL	CNS-Medullo
##	1.332616e-01	1.707526e-01	2.670091e-01

```

##          Breast-AdenoCA      Kidney-ChRCC      CNS-PiloAstro
##          3.645074e-01      3.883654e-01      4.524310e-01
##          Stomach-AdenoCA     Uterus-AdenoCA     Bone-Osteosarc
##          4.524310e-01      4.524310e-01      4.745719e-01
##          Eso-AdenoCA        Ovary-AdenoCA      Kidney-RCC.clearcell
##          4.815932e-01      4.979627e-01      5.431631e-01
##          ColoRect-AdenoCA   Kidney-RCC.papillary
##          6.589716e-01      6.806308e-01



|    | Liver-HCC     | Lung-SCC                | Panc-AdenoCA |
|----|---------------|-------------------------|--------------|
| ## | 5.160159e-58  | 1.139987e-05            | 7.096293e-18 |
| ## | Prost-AdenoCA | Skin-Melanoma.cutaneous |              |
| ## | 1.219769e-07  | 9.923539e-09            |              |



```

ovrdisp$differentially_abundant = ifelse(ovrdisp$ct %in% names(differential_precision[(differential_precision <= 0.05)]), 1, 0)
ovrdisp$differentially_abundant
```

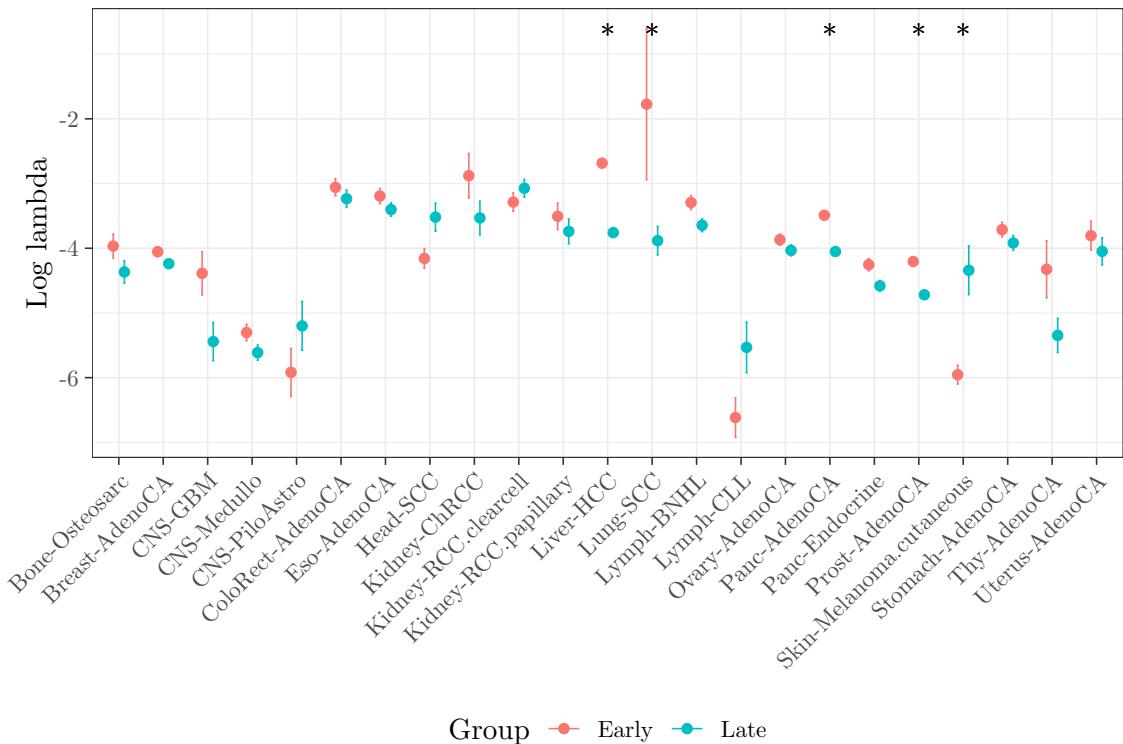


```

[1] "" "" "" "" "" "" "" "" "" "" "" "" "" "" "" "" ""
[20] "" "" "*" "*" "*" "*" "" "" "" "" "" "*" "*" "" "" "*"
[39] "*" "*" "" "" "" "" "" ""

ggplot(ovrdisp, aes(x=ct, y=Estimate, group=name, col=name))+
 geom_point(position=position_dodge(width=0.5))+
 geom_errorbar(aes(ymin=Estimate-Std..Error,
 ymax=Estimate+Std..Error), width=.1, position=position_dodge(width=0.5))+
 theme_bw()+
 geom_text(aes(y=Inf, label=differentially_abundant, vjust=1.8), col='black')+
 theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+
 labs(x=' ', y='Log lambda', col='Group')+theme(legend.position = "bottom")+
 theme(
 legend.margin=margin(0,0,0,0),
 legend.box.margin=margin(-10,-10,-10,-10),
 plot.margin = unit(c(1,1,1,1), "cm"))
```


```



Group ● Early ● Late

Test for differential precision (1/overdispersion) parameter

```
differential_precision_2 <- p.adjust(sapply(diagRE_DMDL_nonexo_SP, ttest_TMB_wrapper_overdisp), method = names(differential_precision_2) <- names(diagRE_DMDL_nonexo_SP)
sort(differential_precision_2)
```

##	Liver-HCC	Panc-AdenoCA	Prost-AdenoCA
##	8.145123e-24	2.860615e-07	1.126237e-04
##	Skin-Melanoma.cutaneous	CNS-GBM	Head-SCC
##	1.723631e-02	2.834265e-01	2.834265e-01
##	Lymph-BNHL	Panc-Endocrine	Lung-SCC
##	2.834265e-01	2.834265e-01	3.053481e-01
##	Lymph-CLL	Thy-AdenoCA	Breast-AdenoCA
##	3.053481e-01	3.147768e-01	3.578140e-01
##	CNS-Medullo	Bone-Osteosarc	Kidney-ChRCC
##	3.578140e-01	4.331744e-01	4.331744e-01
##	CNS-PiloAstro	Eso-AdenoCA	Ovary-AdenoCA
##	4.352908e-01	4.352908e-01	4.352908e-01
##	Stomach-AdenoCA	Kidney-RCC.clearcell	ColoRect-AdenoCA
##	4.502790e-01	5.128995e-01	5.601231e-01
##	Kidney-RCC.papillary	Uterus-AdenoCA	
##	5.770837e-01	5.770837e-01	

```
table(differential_precision_2 <= 0.05)
```

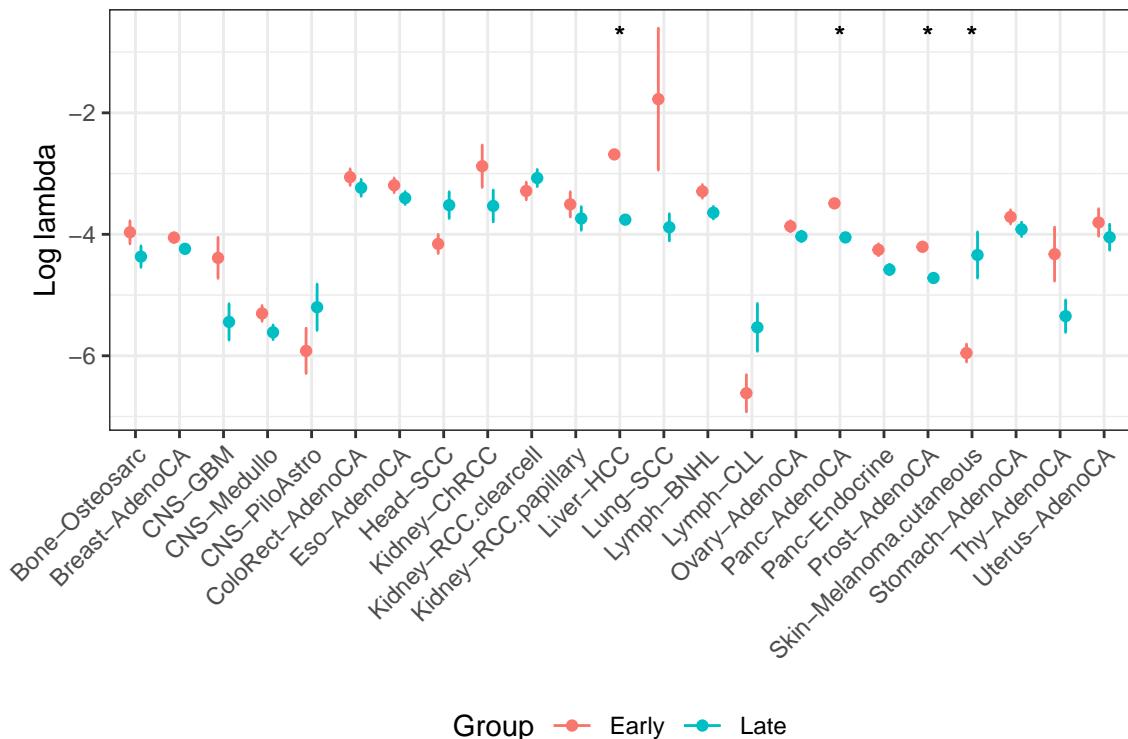
```
##  
## FALSE TRUE  
## 19 4
```

```
differential_precision_2[(differential_precision_2 <= 0.05)]
```

```
##          Liver-HCC      Panc-AdenoCA      Prost-AdenoCA
## 8.145123e-24 2.860615e-07 1.126237e-04
## Skin-Melanoma.cutaneous
## 1.723631e-02

ovrdisp$differential_precision_2 = ifelse(ovrdisp$ct %in% names(differential_precision_2)[(differential_precision_2 <= 0.05)], differential_precision_2, 0)

ggplot(ovrdisp, aes(x=ct, y=Estimate, group=name, col=name))+
  geom_point(position=position_dodge(width=0.5))+
  geom_errorbar(aes(ymin=Estimate-Std..Error,
                     ymax=Estimate+Std..Error), width=.1, position=position_dodge(width=0.5))+
  theme_bw()+
  geom_text(aes(y=Inf, label=differential_precision_2, vjust=1.8), col='black')+
  theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+
  labs(x='', y='Log lambda', col='Group')+theme(legend.position = "bottom")+
  theme(
    legend.margin=margin(0,0,0,0),
    legend.box.margin=margin(-10,-10,-10,-10),
    plot.margin = unit(c(1,1,1,1), "cm"))
```



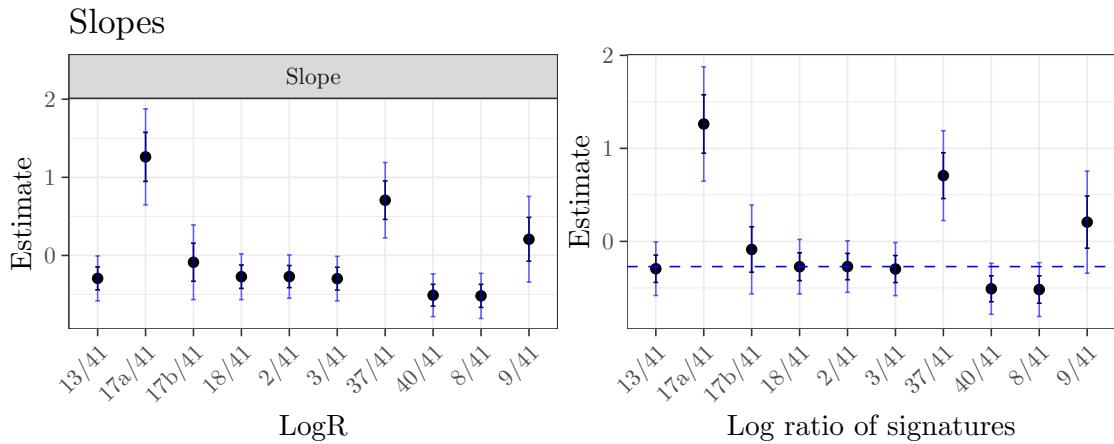
Minimal perturbation

Note: this is using the original, non SP, signatures (hence not the best version).

```
minimalpert_L2 <- function(i){
  sum(i)/sum(i^2)
}
```

```

#
# betas_breast <- data.frame(plot_betas(diagRE_DMSL_nonexo[["Breast-AdenoCA"]], names_cats= logR_nonexo_r
#                               return_df=T, plot=F))
#
# .slopes_minpert <- betas_breast %>% dplyr::filter(type_beta == "Slope") %>% dplyr::select(Estimate) %>%
# #
# # minimalpert_L2(softmax(c(.slopes_minpert, 0)))
# median(c(.slopes_minpert, 0))
#
# aa <- plot_betas(diagRE_DMSL_nonexo[["Breast-AdenoCA"]], names_cats= logR_nonexo_notsorted[["Breast-AdenoCA"]]
#                   return_df=F, plot=F, only_slope = T, line_zero=F)
# # aa <- geom_hline(yintercept = 0)+geom_vline(xintercept = 1)+geom_hline(yintercept = median(c(.slopes_minpert, 0)))
# aa + geom_hline(yintercept = median(c(.slopes_minpert, 0)), lty='dashed', col='blue')+geom_hline(yintercept = mean(c(.slopes_minpert, 0)), lty='dashed', col='red')
# 
```



For the thesis:

```

pdf("../results/results_TMB/pcawg/minimalperturbation_SP_all.pdf", width = 4, height = 3)
for(ct in names(diagRE_DMDL_nonexo_SP)){
  .betas_ct_it <- data.frame(plot_betas(TMB_obj = diagRE_DMDL_nonexo_SP[[ct]],
                                         names_cats= logR_nonexo_notsorted_SP[[ct]],
                                         return_df=T, plot=F))
  .slopes_minpert <- .betas_ct_it %>% dplyr::filter(type_beta == "Slope") %>% dplyr::select(Estimate) %>%
  print(aaa <- plot_betas(TMB_obj = diagRE_DMDL_nonexo_SP[[ct]], names_cats= logR_nonexo_notsorted_SP[[ct]],
                           return_df=F, plot=F, only_slope = T, line_zero=F, add_confint = T, return_ggplot=TRUE)
  +
  geom_hline(yintercept = median(c(.slopes_minpert, 0)), lty='dashed', col='blue')+ggtitle(ct))
}
dev.off()

## pdf
## 2

\subsection{Minimal perturbation in diagRE_DMDL_nonexo_S}

perturbed_betas_diagRE_DMDL_nonexo_SP <- lapply(names(diagRE_DMDL_nonexo_SP), try(give_min_pert))

perturbed_betas_diagRE_DMDL_nonexo_SP_vec <- do.call('c', sapply(perturbed_betas_diagRE_DMDL_nonexo_SP, f

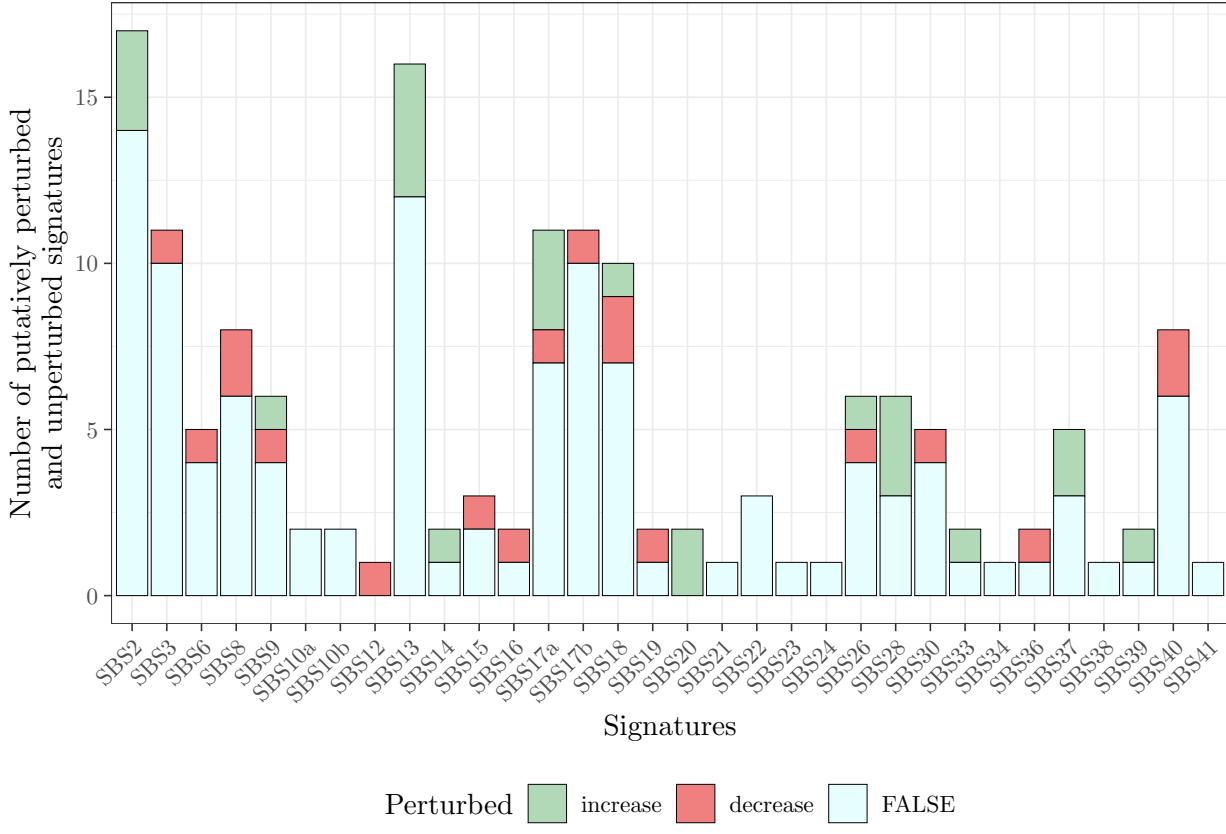
perturbed_betas_diagRE_DMDL_nonexo_SP_df <- cbind.data.frame(sig=gsub("betas_perturbed.", "", names(perturbed_betas_diagRE_DMDL_nonexo_SP_vec)))

```

```

perturbed=perturbed_betas_diagRE_DMDL_nonexo_SP_df
ggplot(perturbed_betas_diagRE_DMDL_nonexo_SP_df, aes(x=factor(sig, levels=gtools::mixedsort(unique(sig))), fill=factor(perturbed, levels=c('increase', 'decrease'))))
  geom_bar(col='black', size=0.001)+theme_bw()+
  scale_fill_manual(values=c( '#b1d8b7', '#f08080', '#e7feff'))+
  theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1), legend.position = "bottom")+
  labs(x='Signatures', y='Number of putatively perturbed\nand unperturbed signatures', fill='Perturbed')

```



Minimal perturbation per signature

```

names(perturbed_betas_diagRE_DMDL_nonexo_SP) <- names(diagRE_DMDL_nonexo_SP)
ggplot(reshape2::melt(lapply(perturbed_betas_diagRE_DMDL_nonexo_SP, `[, 'betas_perturbed'])),
  aes(x=L1, fill=factor(value, levels=c('increase', 'decrease', 'FALSE')))+geom_bar(col='black', size=0.001)+scale_fill_manual(values=c('#b1d8b7', '#f08080', '#e7feff'))+theme(legend.position = "bottom")+
  theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+labs(x='Cancer type',
  y='Number of putatively perturbed and unperturbed signatures', fill='Perturbed')

```

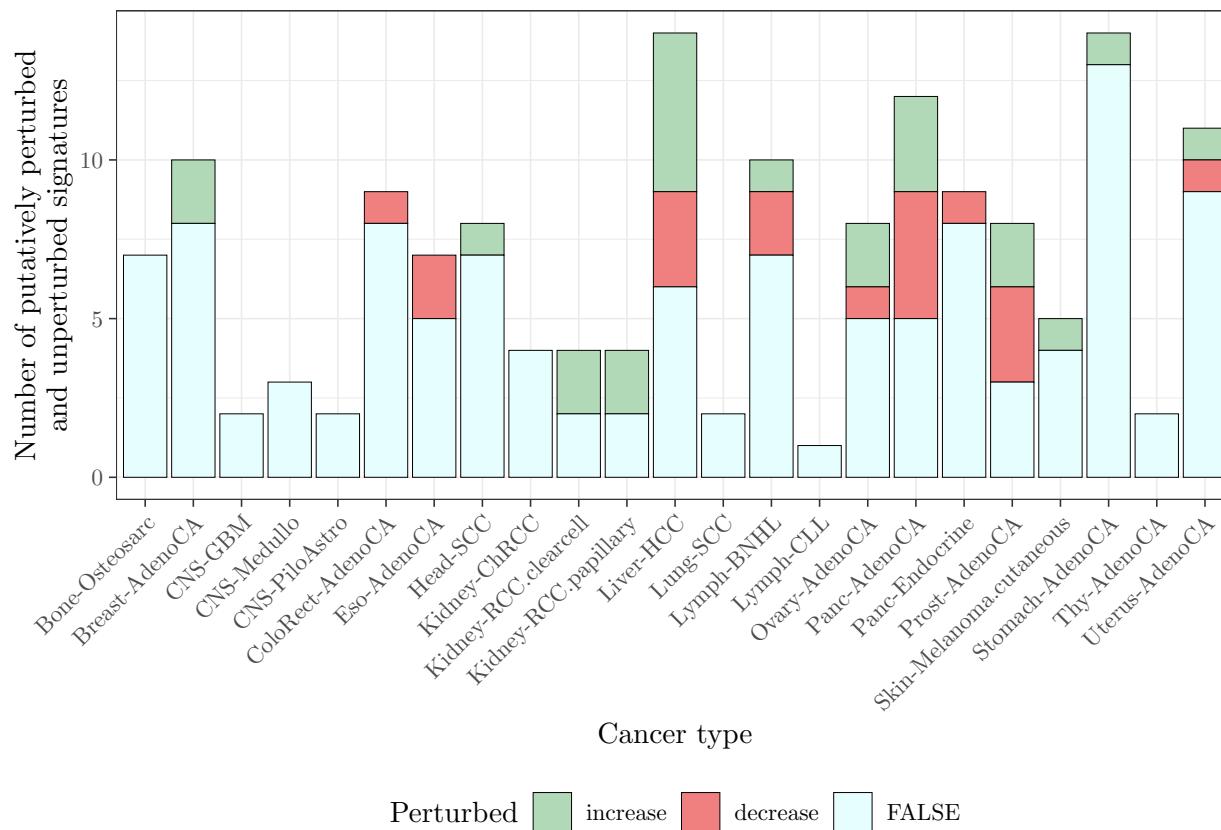


Table of perturbed signatures

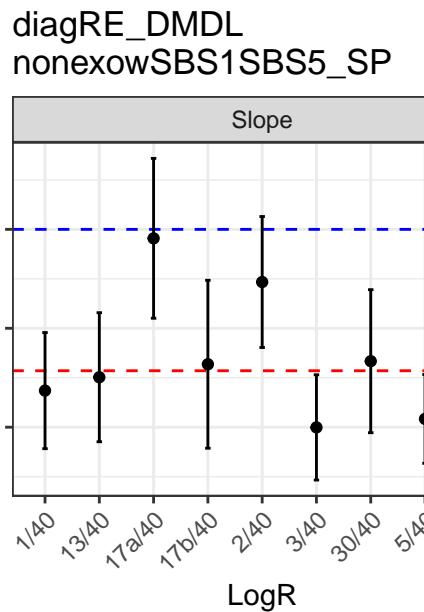
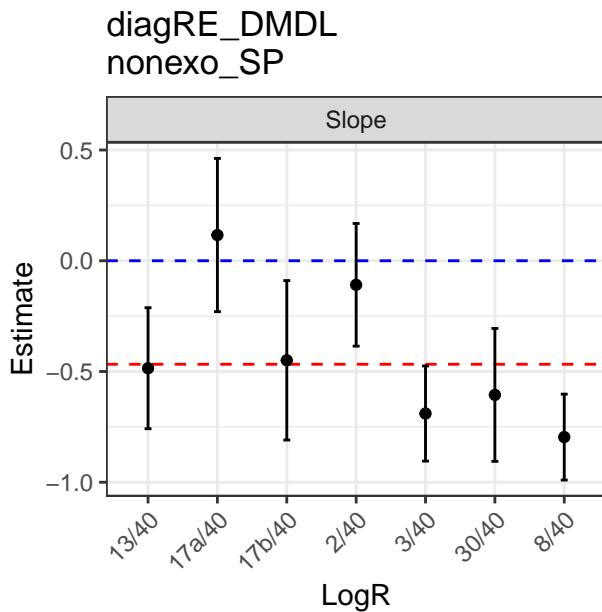
```

relevel_perturbation <- cbind(c('increase', 'decrease'), c('+', '-'))

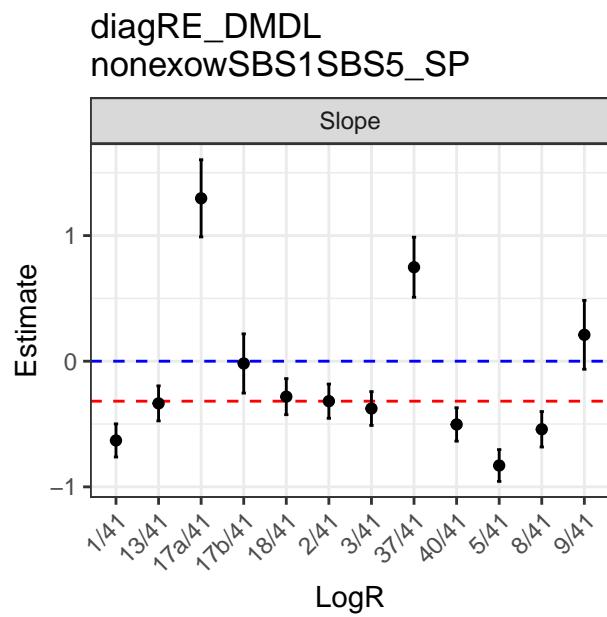
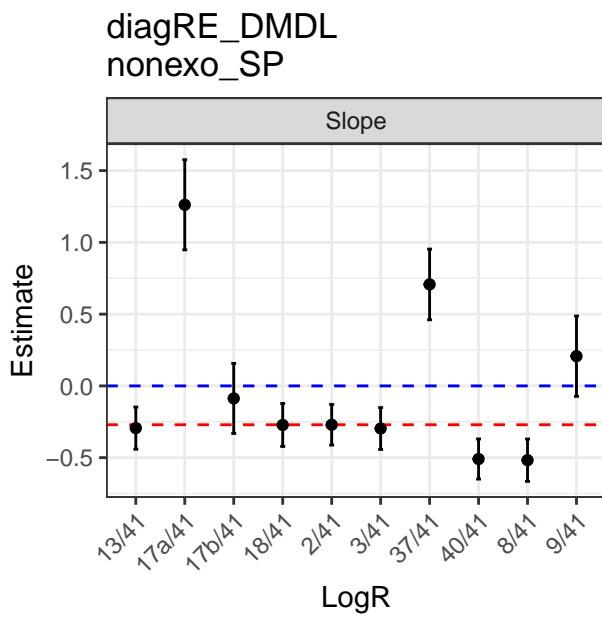
write("", ".../results/results_TMB/pcawg/minimal_perturbation_sigs.txt", append = F)
# for(i in names(perturbed_betas_diagRE_DMDL_nonexo_SP)){
#   write(paste0(i, '&', paste0(names(perturbed_betas_diagRE_DMDL_nonexo_SP[[i]]$betas_perturbed)[perturb
# })
for(i in names(perturbed_betas_diagRE_DMDL_nonexo_SP)){
  # write(x = paste0(i, '&', paste0(sapply(which(perturbed_betas_diagRE_DMDL_nonexo_SP[[i]]$betas_perturb
  #   paste0(names(perturbed_betas_diagRE_DMDL_nonexo_SP[[i]]$betas_perturbed)[j], '(', relevel_pertur
  #   }), collapse=' ', sep=''), '\n\n'), file=".../results/results_TMB/pcawg/minimal_perturbation_s
  write(x = paste0(paste0(i, '&', paste0(sapply(which(perturbed_betas_diagRE_DMDL_nonexo_SP[[i]]$betas_
    paste0(names(perturbed_betas_diagRE_DMDL_nonexo_SP[[i]]$betas_perturbed)[j], '(', relevel_perturb
  }), collapse=' ', sep=''), '\n\n')), file=".../results/results_TMB/pcawg/minimal_perturbation_s
}

```

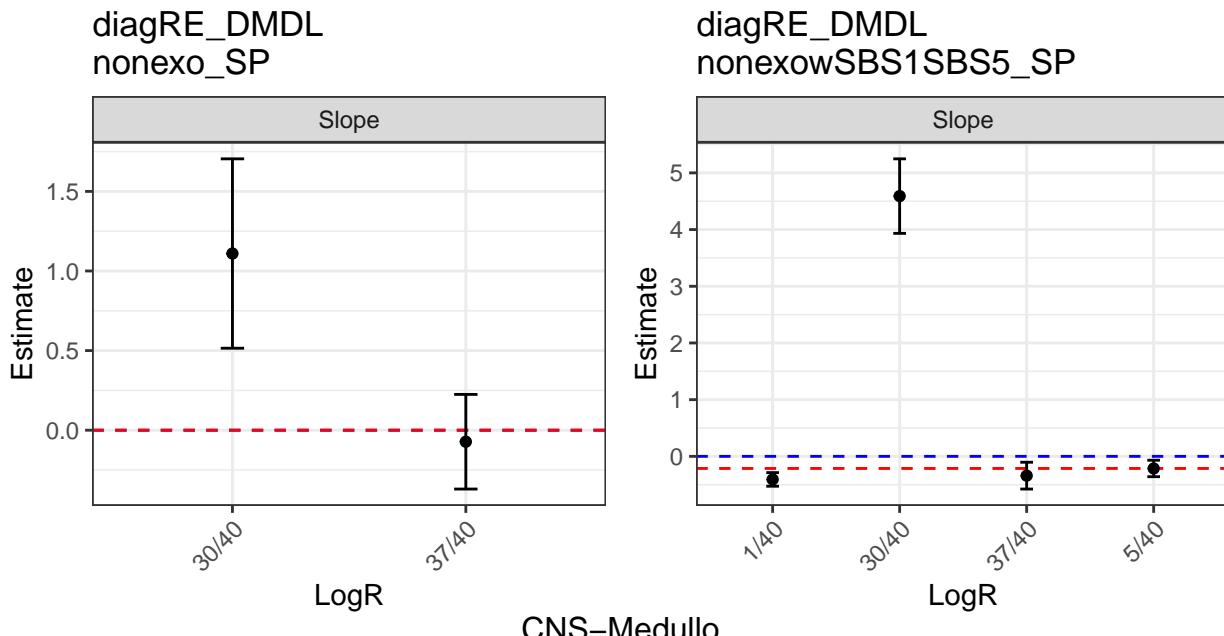
Bone–Osteosarc



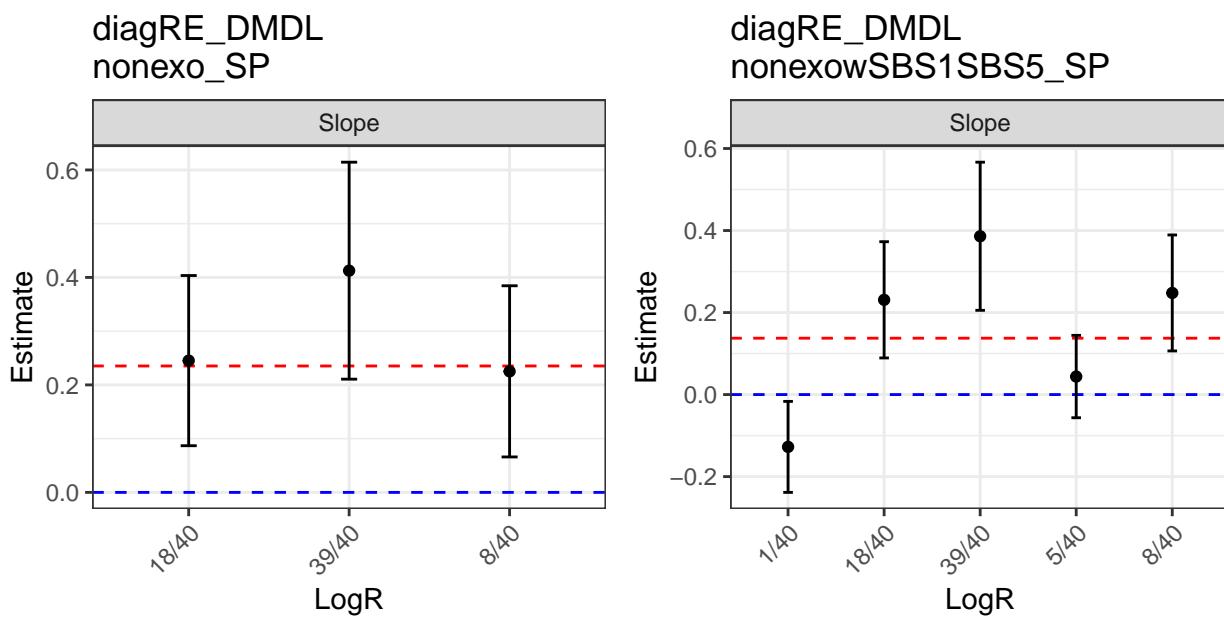
Breast–AdenoCA



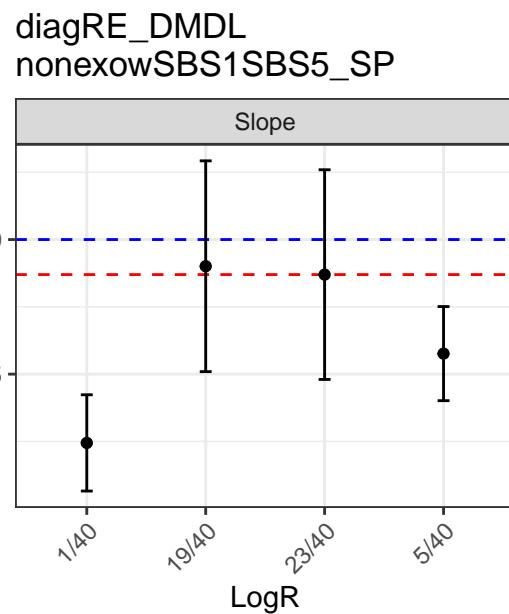
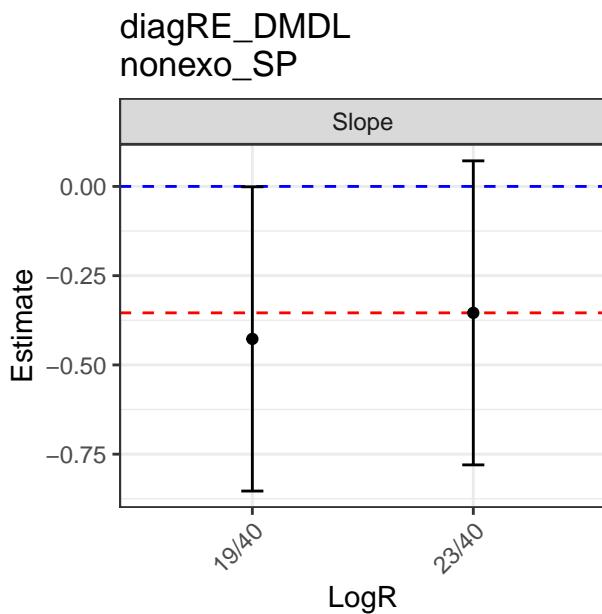
CNS–GBM



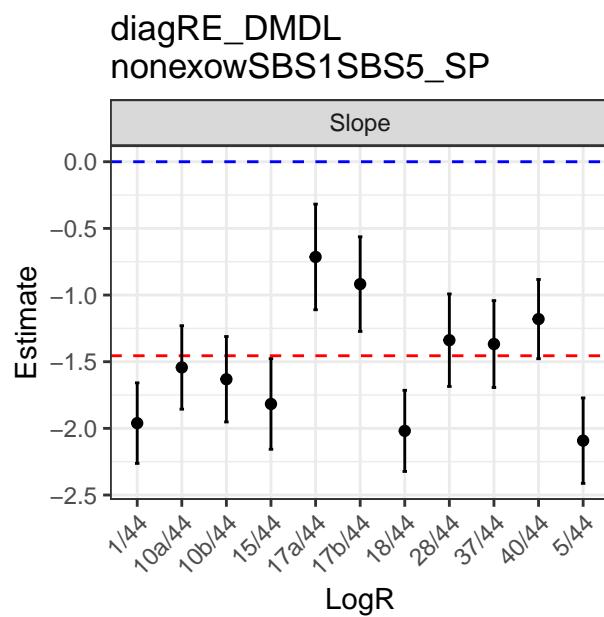
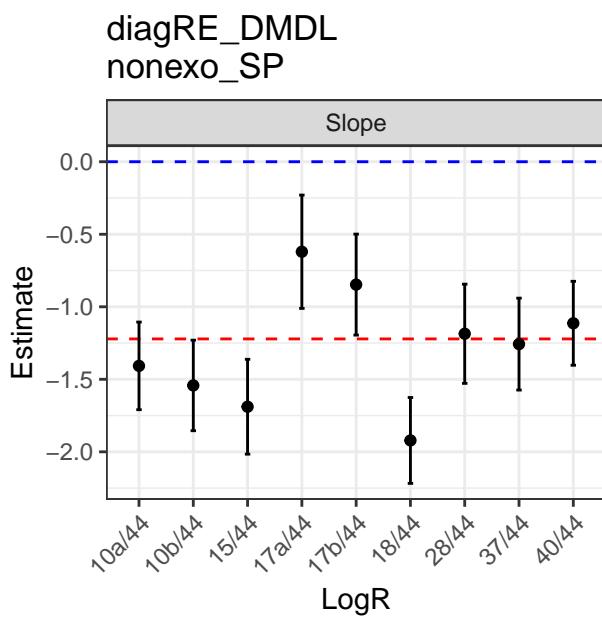
CNS–Medullo



CNS–PiloAstro

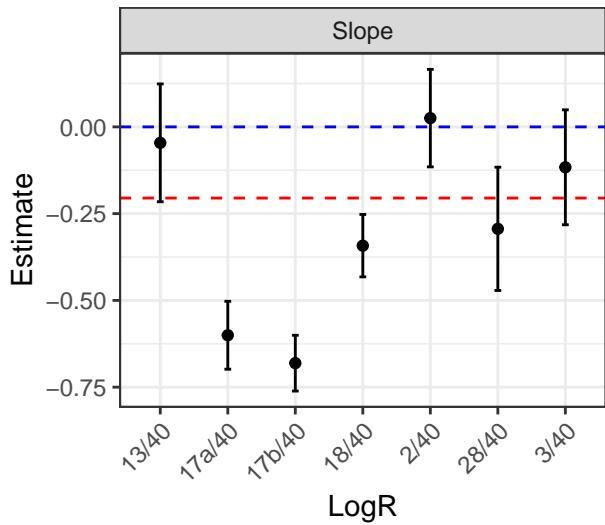


ColoRect–AdenoCA

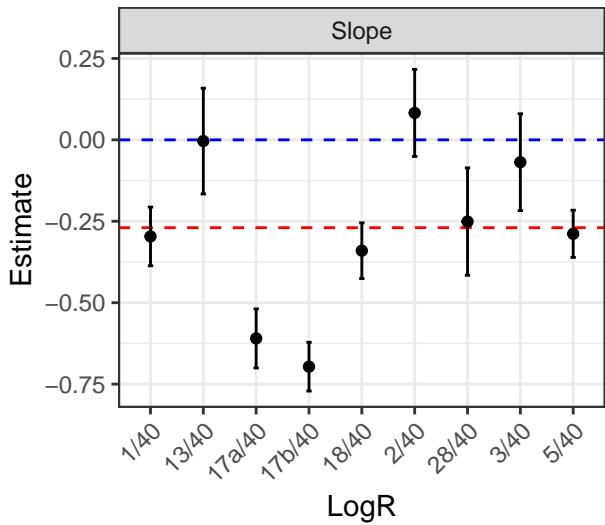


Eso-AdenoCA

diagRE_DMDL
nonexo_SP

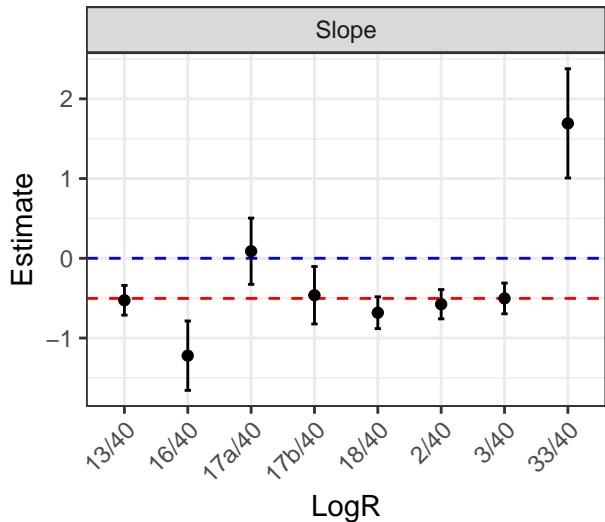


diagRE_DMDL
nonexowSBS1SBS5_SP

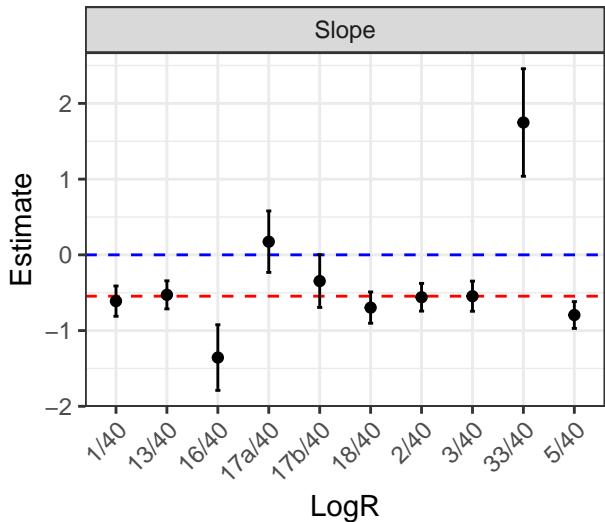


Head-SCC

diagRE_DMDL
nonexo_SP

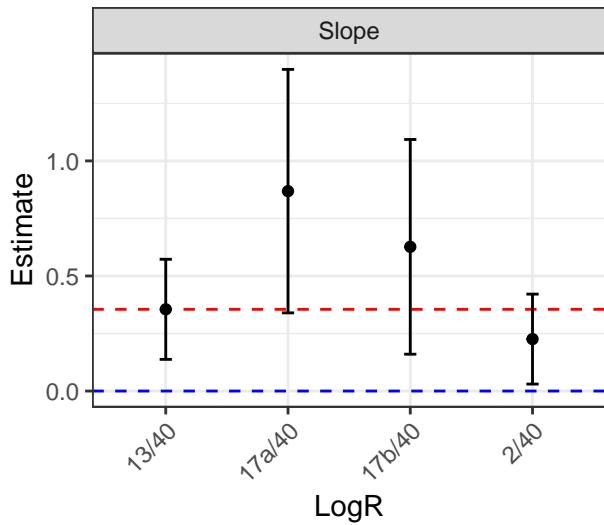


diagRE_DMDL
nonexowSBS1SBS5_SP

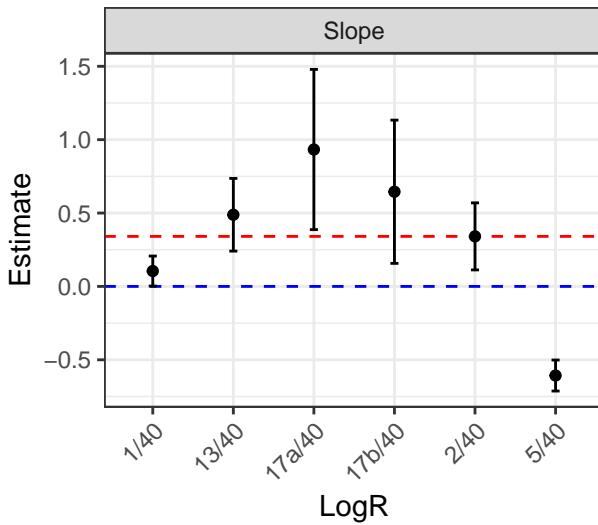


Kidney–ChRCC

diagRE_DMDL
nonexo_SP

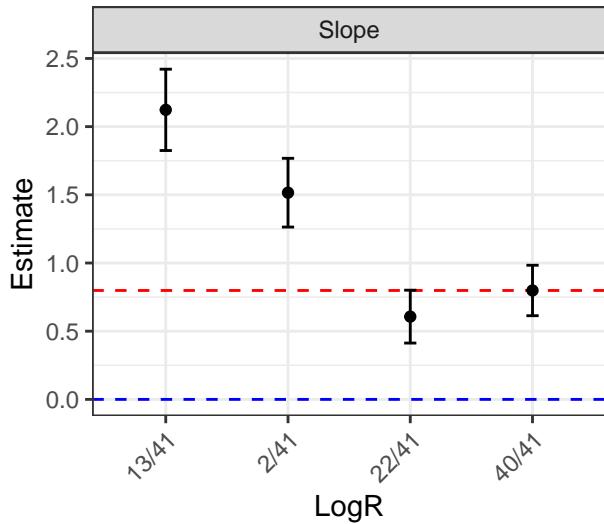


diagRE_DMDL
nonexowSBS1SBS5_SP

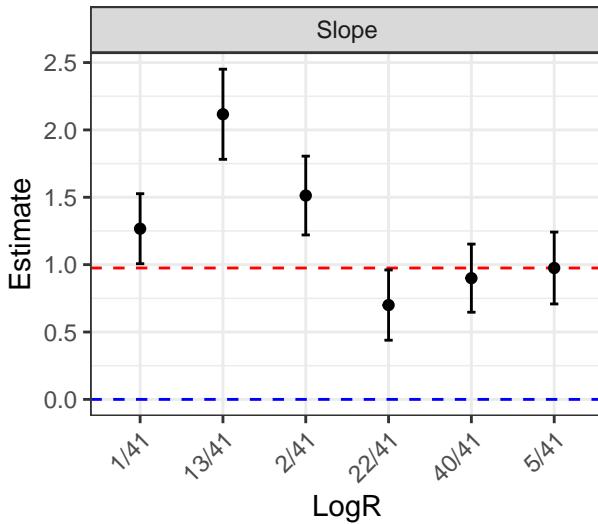


Kidney–RCC.clearcell

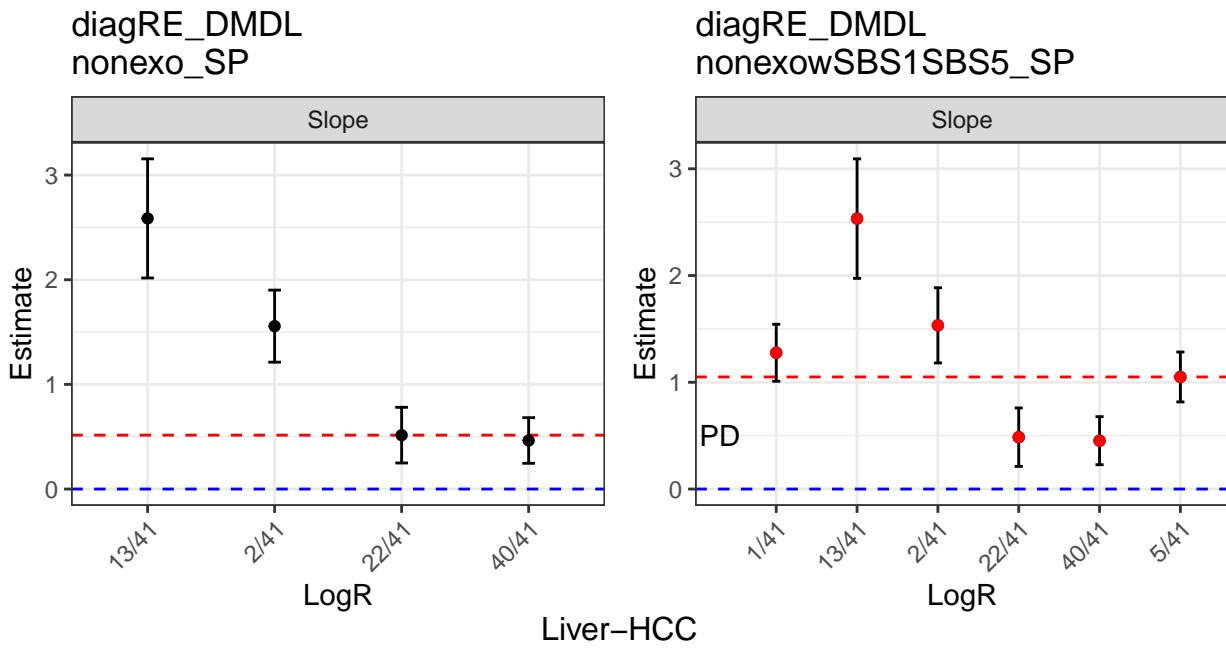
diagRE_DMDL
nonexo_SP



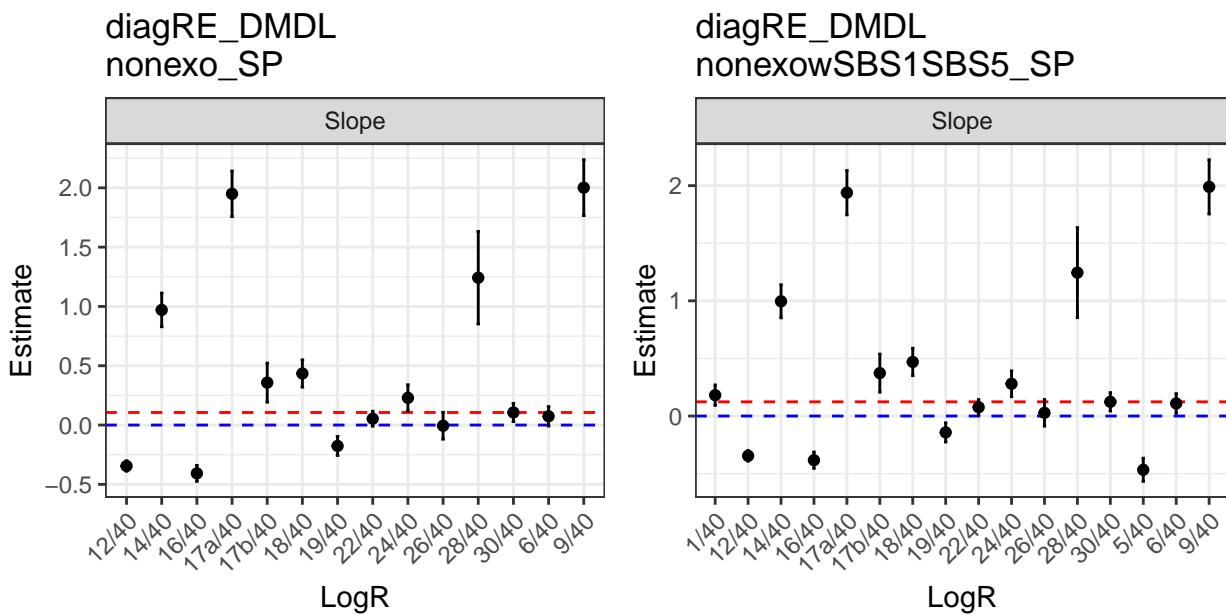
diagRE_DMDL
nonexowSBS1SBS5_SP



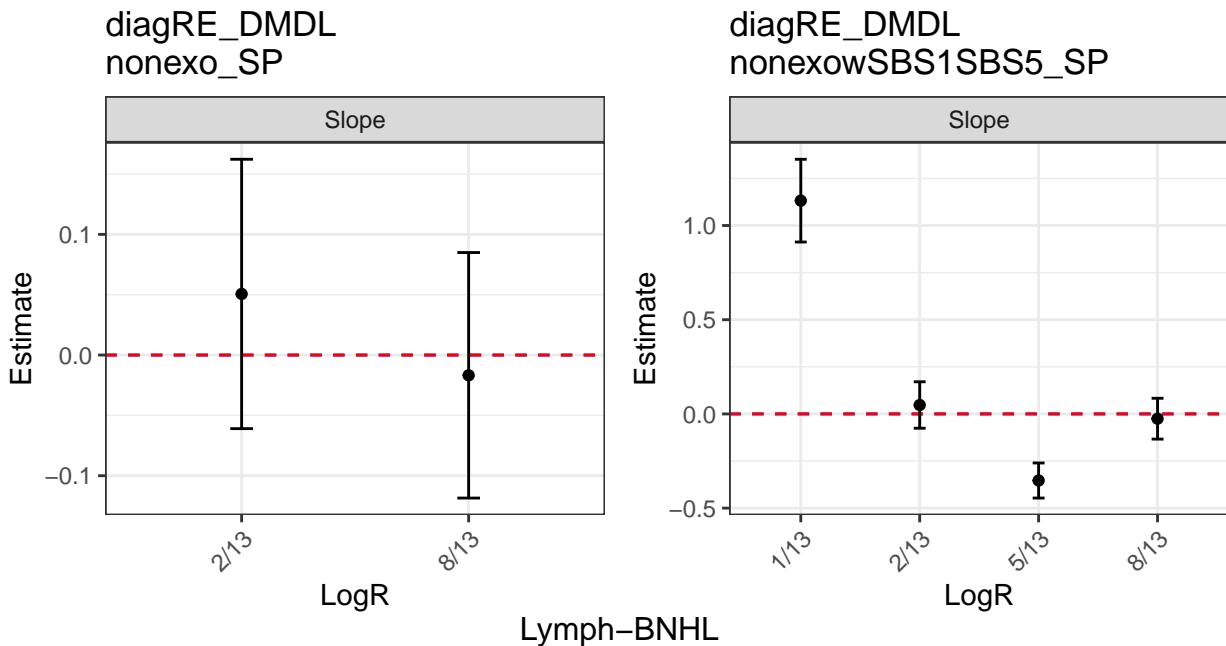
Kidney–RCC.papillary



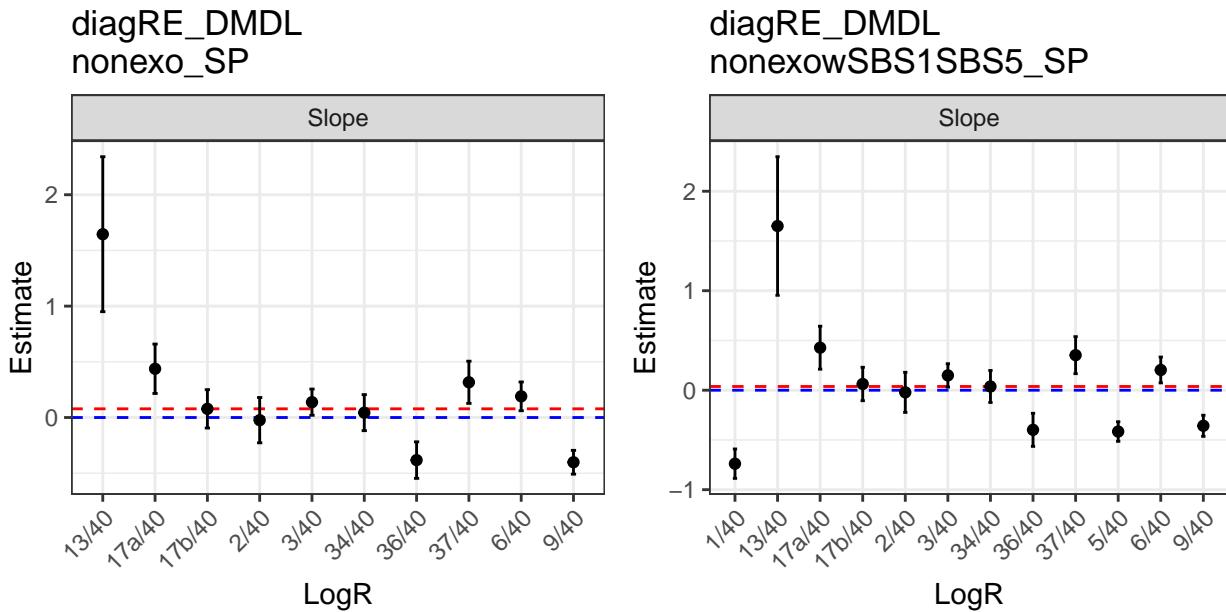
Liver–HCC



Lung-SCC

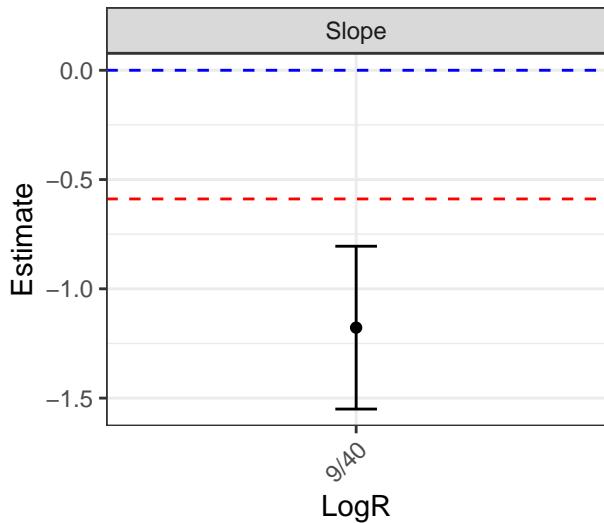


Lymph-BNHL

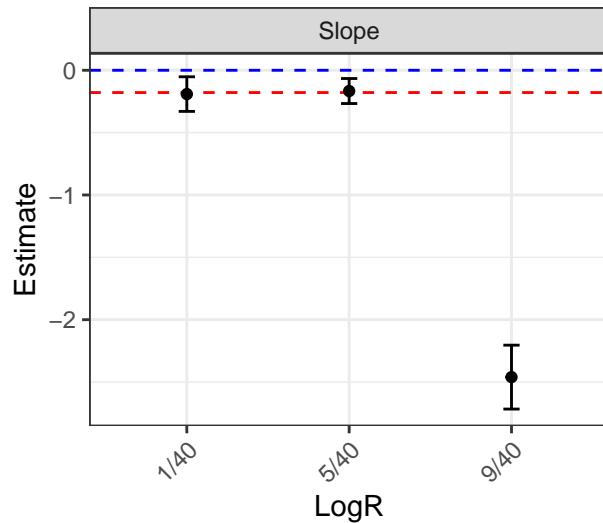


Lymph-CLL

diagRE_DMDL
nonexo_SP

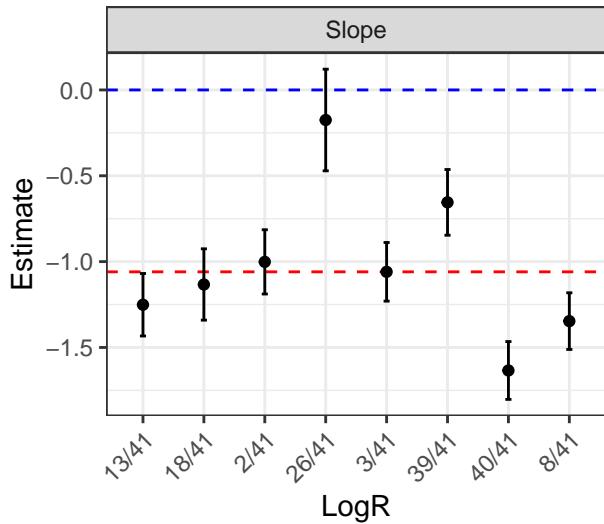


diagRE_DMDL
nonexowSBS1SBS5_SP

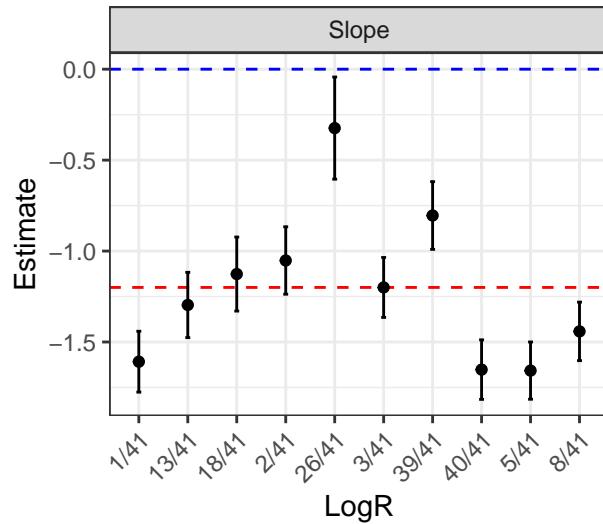


Ovary-AdenoCA

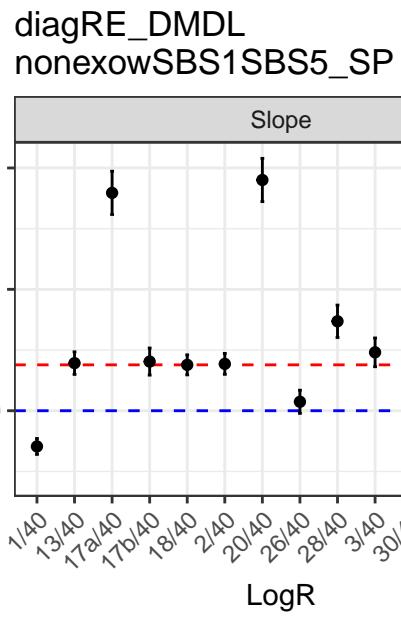
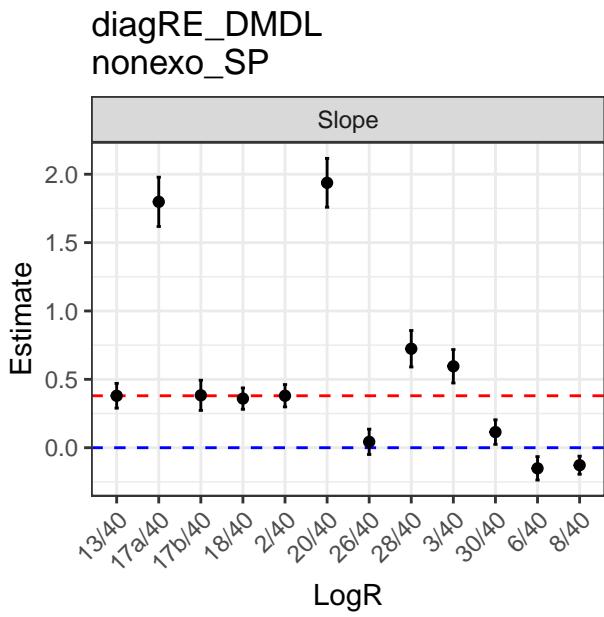
diagRE_DMDL
nonexo_SP



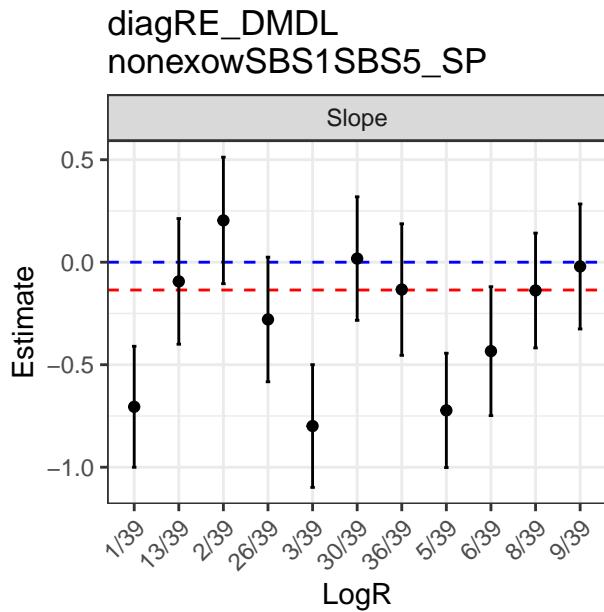
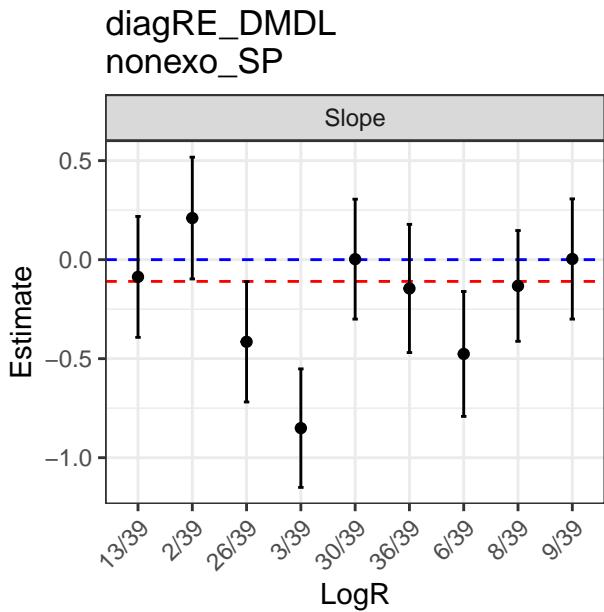
diagRE_DMDL
nonexowSBS1SBS5_SP



Panc–AdenoCA

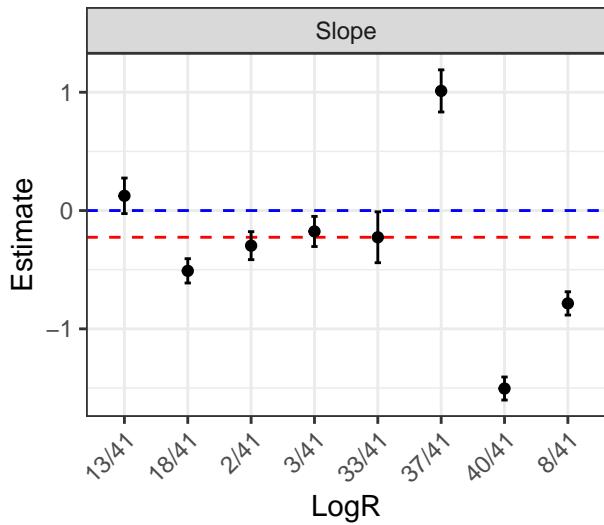


Panc–Endocrine

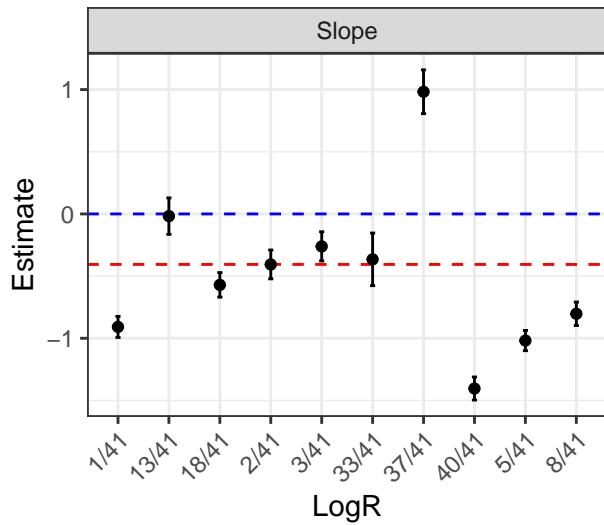


Prost–AdenoCA

diagRE_DMDL
nonexo_SP

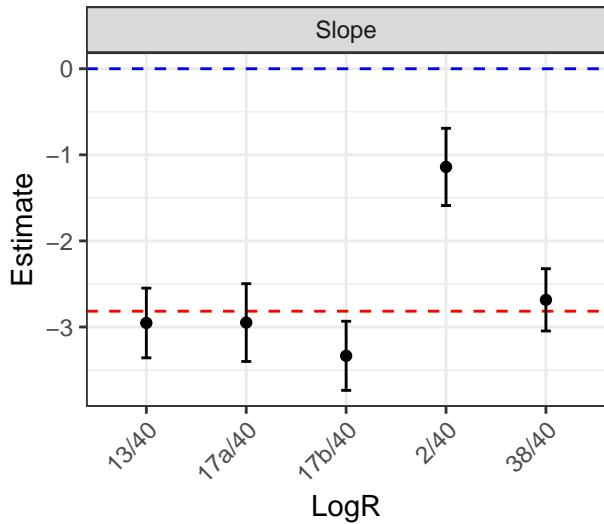


diagRE_DMDL
nonexowSBS1SBS5_SP

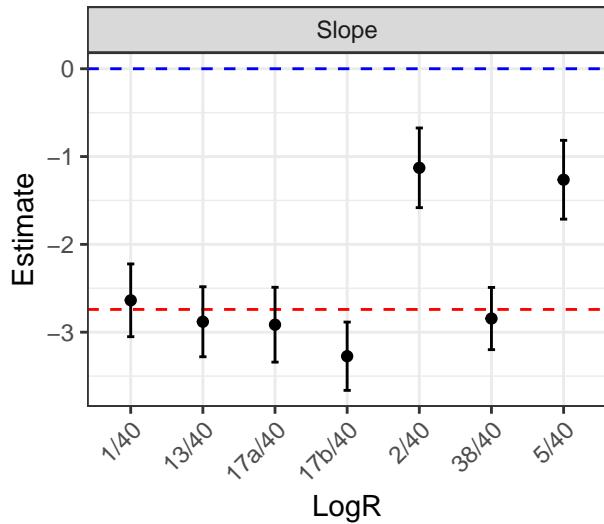


Skin–Melanoma.cutaneous

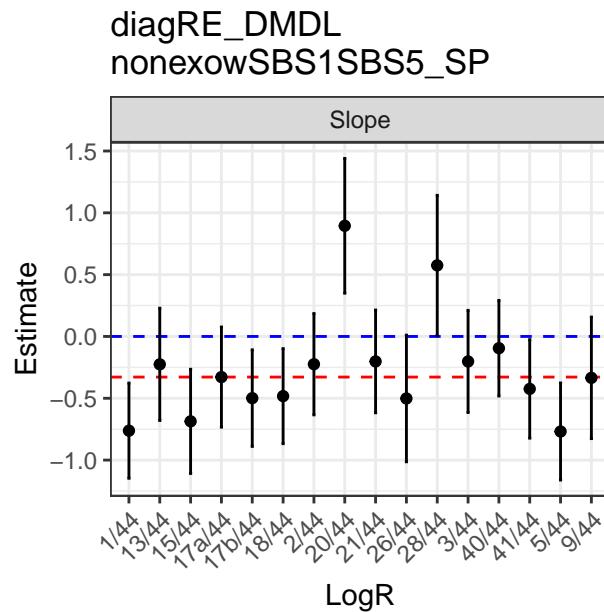
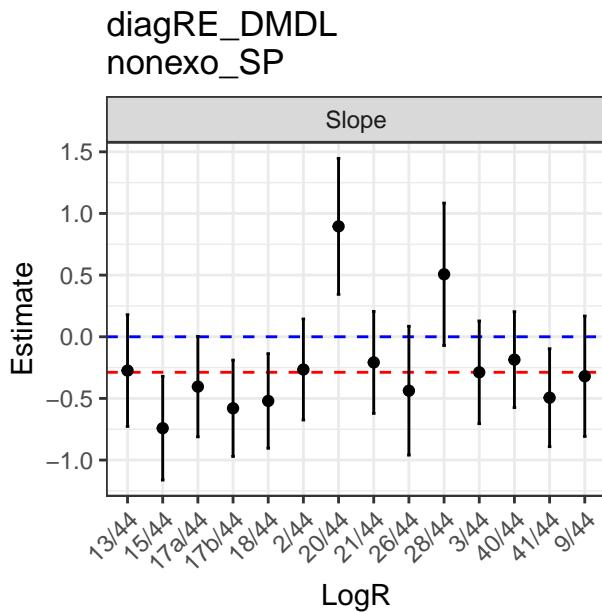
diagRE_DMDL
nonexo_SP



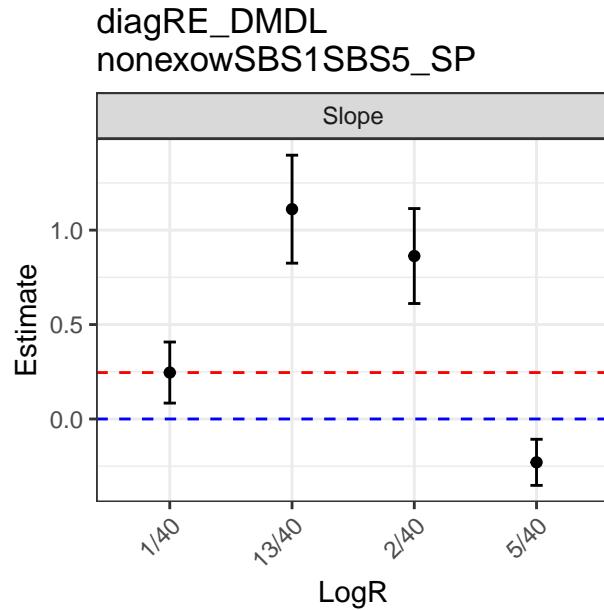
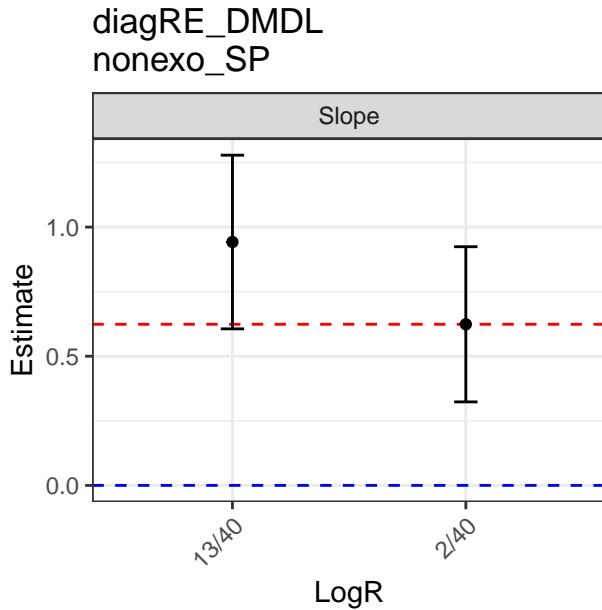
diagRE_DMDL
nonexowSBS1SBS5_SP



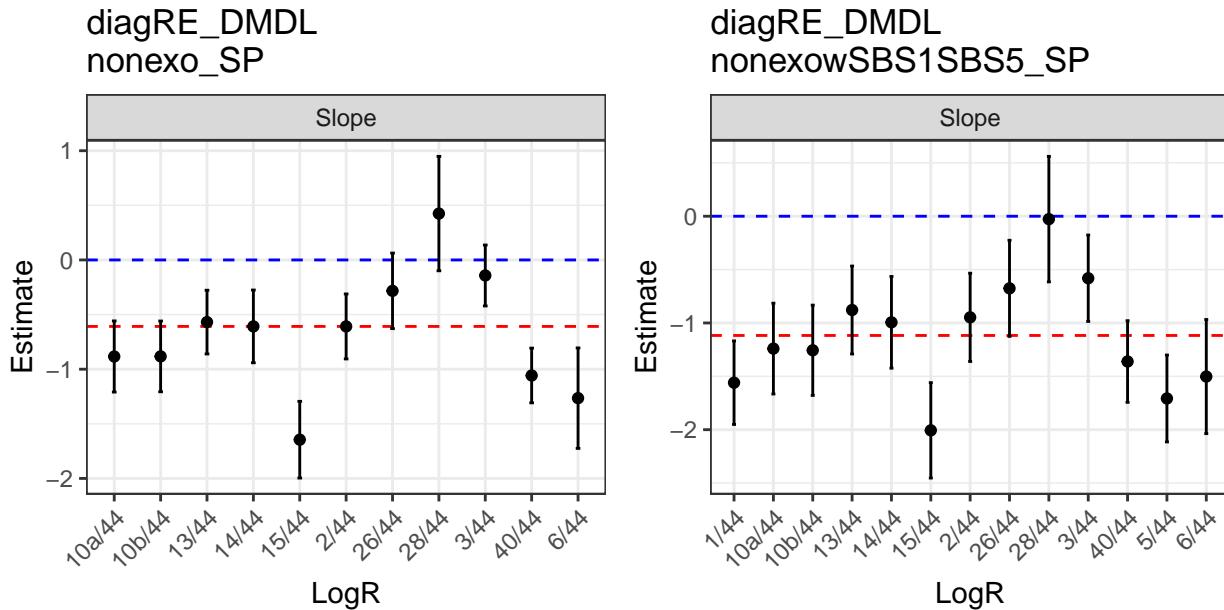
Stomach–AdenoCA



Thy–AdenoCA



Uterus–AdenoCA



Comparing the s1/s5 baseline to the minimal perturbation results

```

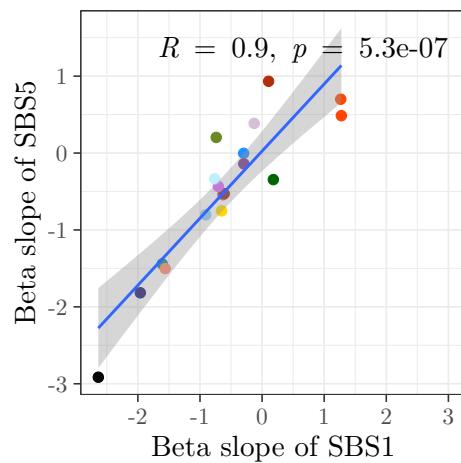
##      [,1]     [,2]     [,3]     [,4]     [,5]     [,6]     [,7]
## SBS1 "FALSE" "decrease" "FALSE" "decrease" "decrease" "FALSE" "FALSE"
## SBS5 "FALSE" "decrease" "FALSE" "FALSE"   "FALSE"   "decrease" "FALSE"
##      [,8]     [,9]     [,10]    [,11]    [,12]    [,13]    [,14]
## SBS1 "FALSE" "decrease" "FALSE" "FALSE"   "FALSE"   "increase" "decrease"
## SBS5 "FALSE" "decrease" "FALSE" "FALSE"   "decrease" "decrease" "decrease"
##      [,15]    [,16]    [,17]    [,18]    [,19]    [,20]    [,21]
## SBS1 "FALSE" "decrease" "decrease" "FALSE"   "decrease" "FALSE" "FALSE"
## SBS5 "FALSE" "decrease" "decrease" "decrease" "decrease" "increase" "FALSE"
##      [,22]    [,23]
## SBS1 "FALSE"   "FALSE"
## SBS5 "decrease" "FALSE"

## [1] 2 23
## [1] "Ct for which SBS1 and SBS5 are not DA"
## [1] 9
## [1] "Ct for which SBS1 and SBS5 both decrease"
## [1] 6

```

Plotting the betas of SBS1 and SBS5, and their correlation

```
## `geom_smooth()` using formula 'y ~ x'
```



Correlation between the two APOBEC signatures, SBS2 and SBS13:

[diagRE_DMDL_wSBS1SBS5nonexo_SP_betas_SBS1SBS5_0](#)

```
##      LogR Bone-Osteosarc Breast-AdenoCA    CNS-GBM CNS-Medullo CNS-PiloAstro
## 1      1   -0.65205448   -0.63056226  -0.4068313  -0.12757480  -0.75547436
## 2      2   -0.21291191   -0.31810871  -0.2135513   0.04392387  -0.42374188
## 3      3   -0.80086778   -0.37655392   4.5904007   0.24778355  -0.09911198
## 4      4   -0.76647219   -0.83005858  -0.3402184   0.23102511  -0.13019202
## 5      5   -0.75193369   -0.54174300        NA   0.38604360        NA
## 6      6   -0.59793592   0.21011709        NA        NA        NA
## 7      7   -0.03626699   -0.33566311        NA        NA        NA
## 8      8   -0.54542495   1.29663200        NA        NA        NA
## 9      9   -0.53311606   -0.01798581        NA        NA        NA
## 10    10        NA   -0.28202010        NA        NA        NA
## 11    11        NA   0.74818117        NA        NA        NA
## 12    12        NA   -0.50339545        NA        NA        NA
## 13    13        NA        NA        NA        NA        NA
## 14    14        NA        NA        NA        NA        NA
## 15    15        NA        NA        NA        NA        NA
## 16    16        NA        NA        NA        NA        NA
##      ColoRect-AdenoCA Eso-AdenoCA Head-SCC Kidney-ChRCC Kidney-RCC.clearcell
## 1      -1.9608217  -0.296603923  -0.6109625   0.1048328   1.2666589
## 2      -2.0921571   0.082743769  -0.5598662   0.3410101   1.5127991
## 3      -1.5429485  -0.068557764  -0.5458705  -0.6070669   0.9750606
## 4      -1.6315098  -0.288525263  -0.7945682   0.4884064   2.1161471
## 5      -1.8169416  -0.003715624  -0.5281156   0.9328886   0.6999134
## 6      -0.7141478  -0.609730394  -1.3562882   0.6452427   0.8996817
## 7      -0.9178427  -0.696390350  0.1740217        NA        NA
## 8      -2.0189069  -0.340353080  -0.3459108        NA        NA
## 9      -1.3386046  -0.250926253  -0.6965822        NA        NA
## 10     -1.3676077        NA   1.7476965        NA        NA
## 11     -1.1803477        NA        NA        NA        NA
## 12        NA        NA        NA        NA        NA
## 13        NA        NA        NA        NA        NA
## 14        NA        NA        NA        NA        NA
## 15        NA        NA        NA        NA        NA
```

	## 16	NA	NA	NA	NA	NA	NA
## 1	Kidney-RCC.papillary	1.2765283	0.18121701	1.13205500	-0.73853075	-0.1909831	
## 2		1.5336255	-0.46670474	0.04730124	-0.02155235	-0.1664975	
## 3		1.0498660	0.10929054	-0.35347146	0.14873148	-2.4606035	
## 4		2.5333762	1.98952435	-0.02528508	-0.41539701		NA
## 5		0.4858476	-0.34536130		NA	0.20340232	NA
## 6		0.4530121	0.99614133		NA	-0.35782076	NA
## 7			NA	-0.38332262	NA	1.65033823	NA
## 8			NA	1.93883092	NA	0.42738903	NA
## 9			NA	0.37266116	NA	0.06248822	NA
## 10			NA	0.46956392	NA	0.03747916	NA
## 11			NA	-0.14208627	NA	-0.39838546	NA
## 12			NA	0.07649181	NA	0.35234432	NA
## 13			NA	0.28004806	NA		NA
## 14			NA	0.02771816	NA		NA
## 15			NA	1.24493307	NA		NA
## 16			NA	0.12339364	NA		NA
## 1	Ovary-AdenoCA	Panc-AdenoCA	Panc-Endocrine	Prost-AdenoCA			
## 2	-1.6082797	-0.29422601	-0.70515367	-0.90886623			
## 3	-1.0518667	0.38598957	0.20384815	-0.40619540			
## 4	-1.1999800	0.48104136	-0.79909092	-0.26111647			
## 5	-1.6572683	-0.50265940	-0.72276391	-1.01842704			
## 6	-1.4416134	-0.14008512	-0.43372724	-0.80323807			
## 7	-1.2964332	-0.10780668	-0.13794423	-0.01776658			
## 8	-1.1264324	0.39219508	-0.02073439	-0.57061143			
## 9	-0.3239953	1.79488708	-0.09355158	-0.36523367			
## 10	-0.8045413	0.40504435	-0.27920660	0.98264553			
## 11	-1.6518010	0.37786323	0.01775660	-1.40425704			
## 12		NA	1.90077896	-0.13335340		NA	
## 13		NA	0.07381949		NA		NA
## 14		NA	0.73689208		NA		NA
## 15		NA	0.13742745		NA		NA
## 16		NA	NA		NA		NA
## 1	Skin-Melanoma.cutaneous	Stomach-AdenoCA	Thy-AdenoCA	Uterus-AdenoCA			
## 2	-2.637102	-0.76210875	0.2458919	-1.55971527			
## 3	-1.128145	-0.22457673	0.8629161	-0.94737555			
## 4	-1.263913	-0.20226748	-0.2294617	-0.58104103			
## 5	-2.881089	-0.76879634	1.1110284	-1.70751291			
## 6	-2.914800	-0.33486170		NA	-1.50212706		
## 7	-3.273806	-0.22552842		NA	-1.24033653		
## 8	-2.844686	-0.68667840		NA	-1.25607734		
## 9		NA	-0.32863748		NA	-0.87878860	
## 10		NA	-0.49898640		NA	-0.99418452	
## 11		NA	-0.48274065		NA	-2.00610779	
## 12		NA	0.89496038		NA	-0.67554877	
## 13		NA	-0.20196073		NA	-0.02725474	
## 14		NA	-0.50159888		NA	-1.36125477	
## 15		NA	0.57491183		NA		NA
		NA	-0.09507625		NA		NA

```

## 16                               NA      -0.42431574      NA      NA
## [1] 27   8
## [1] 27   8
## [1] 136  11
## [1] 136  11
## [1] 34   3
## [1] 34   3
## [1] 106  4
## [1] 106  4
## [1] 41   3
## [1] 41   3
## [1] 37   10
## [1] 37   10
## [1] 65   8
## [1] 65   8
## [1] 32   9
## [1] 32   9
## [1] 38   5
## [1] 38   5
## [1] 86   5
## [1] 86   5
## [1] 30   5
## [1] 30   5
## [1] 207  15
## [1] 207  15
## [1] 34   3
## [1] 34   3
## [1] 51   11
## [1] 51   11
## [1] 53   2
## [1] 53   2
## [1] 97   9
## [1] 97   9
## [1] 193  13
## [1] 193  13
## [1] 70   10
## [1] 70   10
## [1] 208  9
## [1] 208  9
## [1] 29   6
## [1] 29   6
## [1] 30   15
## [1] 30   15
## [1] 41   3
## [1] 41   3
## [1] 40   12
## [1] 40   12

##          Bone-Osteosarc      Breast-AdenoCA      CNS-GBM
##                      8                  11                  3
##          CNS-Medullo      CNS-PiloAstro      ColoRect-AdenoCA

```

```

##          4          3          10
##      Eso-AdenoCA     Head-SCC Kidney-ChRCC
##          8          9          5
## Kidney-RCC.clearcell Kidney-RCC.papillary Liver-HCC
##          5          5          15
##      Lung-SCC       Lymph-BNHL Lymph-CLL
##          3          11          2
## Ovary-AdenoCA      Panc-AdenoCA Panc-Endocrine
##          9          13          10
## Prost-AdenoCA Skin-Melanoma.cutaneous Stomach-AdenoCA
##          9          6          15
## Thy-AdenoCA        Uterus-AdenoCA
##          3          12

## Error in apply(pert, 1, function(i) sqrt(sum((i - 1/(ncol(exposures_cancertype_obj$Y))))^2))) :
##   dim(X) must have a positive length
## Error in apply(pert, 1, function(i) sqrt(sum((i - 1/(ncol(exposures_cancertype_obj$Y))))^2))) :
##   dim(X) must have a positive length

## Warning: NAs introduced by coercion

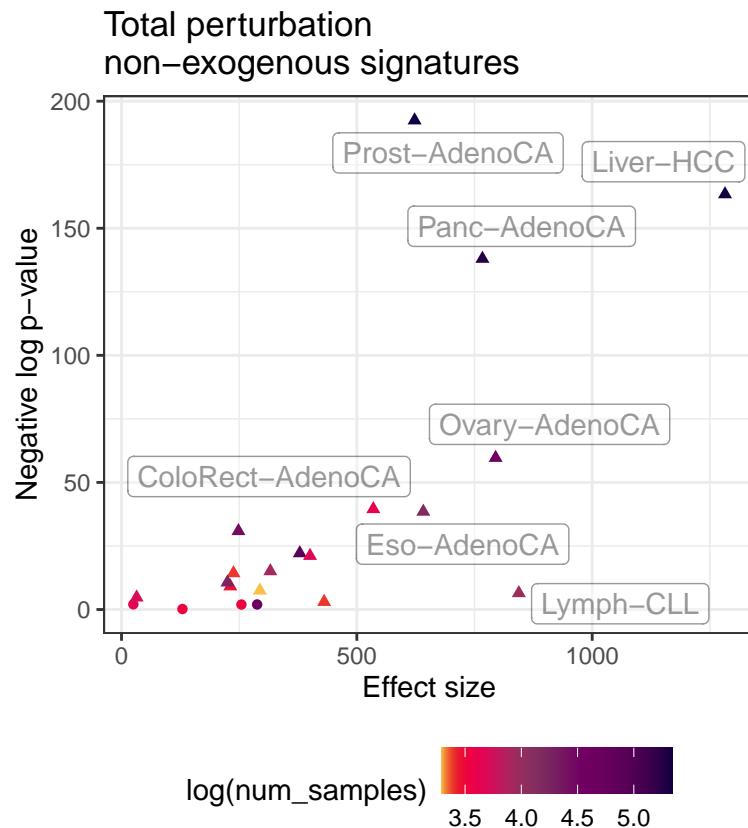
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Removed 2 rows containing missing values (geom_point).

## Warning: Removed 2 rows containing missing values (geom_label_repel).

## Warning: ggrepel: 14 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```



##	Bone-Osteosarc	Breast-AdenoCA	CNS-GBM
##	320.02704	469.29125	524.08517
##	CNS-Medullo	CNS-PiloAstro	ColoRect-AdenoCA
##	217.31161	27.02067	1031.46320
##	Eso-AdenoCA	Head-SCC	Kidney-ChRCC
##	522.66804	221.26052	103.16727
##	Kidney-RCC.clearcell	Kidney-RCC.papillary	Liver-HCC
##	247.53963	156.30958	1332.22216
##	Lung-SCC	Lymph-BNHL	Lymph-CLL
##	86.07143	345.88535	442.78303
##	Ovary-AdenoCA	Panc-AdenoCA	Panc-Endocrine
##	755.21919	829.19711	220.55650
##	Prost-AdenoCA	Skin-Melanoma.cutaneous	Stomach-AdenoCA
##	459.55271	1017.96421	540.50633
##	Thy-AdenoCA	Uterus-AdenoCA	
##	82.57481	458.24059	
##	Bone-Osteosarc	Breast-AdenoCA	CNS-GBM
##	1.080828e-04	2.239756e-28	3.390137e-03
##	CNS-Medullo	CNS-PiloAstro	ColoRect-AdenoCA
##	8.431463e-03	5.615238e-04	6.356131e-26
##	Eso-AdenoCA	Head-SCC	Kidney-ChRCC
##	5.329093e-21	4.975610e-05	1.562125e-09
##	Kidney-RCC.clearcell	Kidney-RCC.papillary	Liver-HCC
##	4.027485e-18	NA	4.747822e-107
##	Lung-SCC	Lymph-BNHL	Lymph-CLL

```

##          7.747310e-22      3.908637e-19      6.611927e-20
## Ovary-AdenoCA      Panc-AdenoCA      Panc-Endocrine
##          8.965185e-38      4.096402e-119     3.987099e-10
## Prost-AdenoCA Skin-Melanoma.cutaneous Stomach-AdenoCA
##          6.474116e-99      9.272113e-25     1.715150e-06
## Thy-AdenoCA       Uterus-AdenoCA
##          8.821583e-06      4.819867e-10

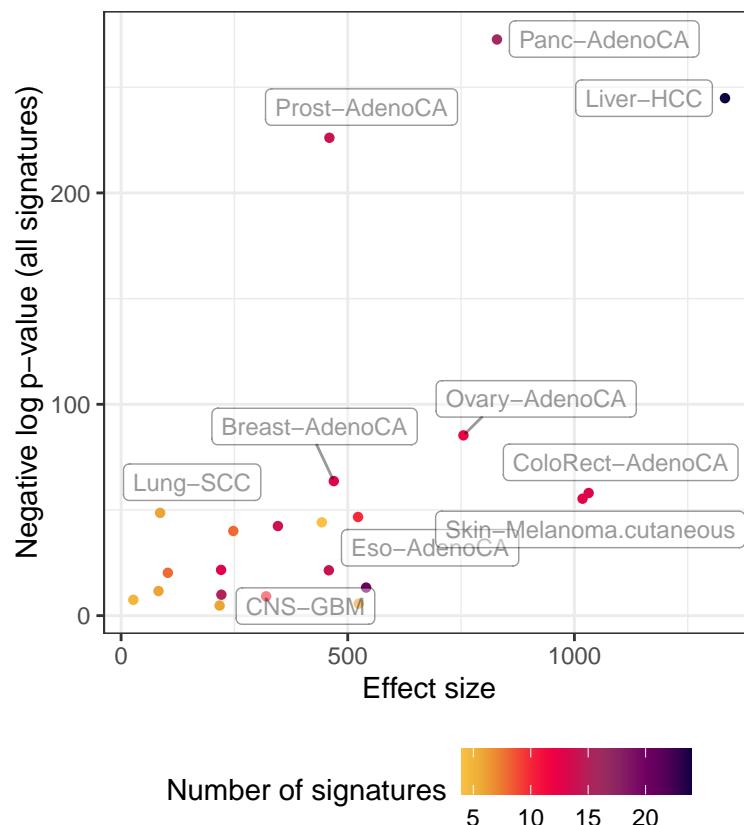
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_label_repel).

## Warning: ggrepel: 12 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```



```

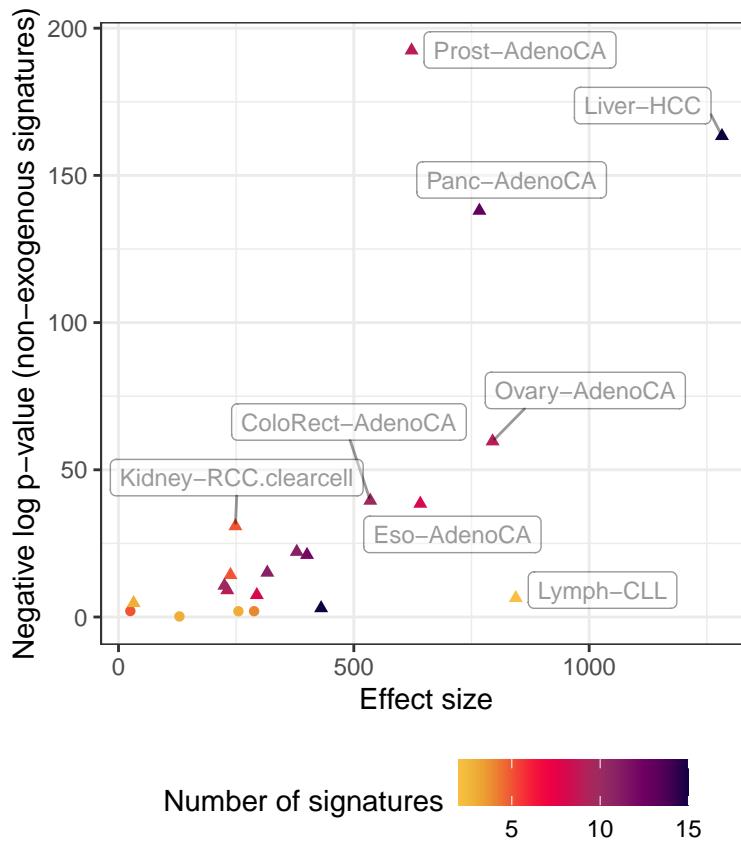
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Removed 2 rows containing missing values (geom_point).

## Warning: Removed 2 rows containing missing values (geom_label_repel).

## Warning: ggrepel: 13 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```



HMP

Simple HMP tests to see what is differentially abundant

```

## Error in signatures_PCAWG_it["x"][[1]][, 2] :
##   incorrect number of dimensions

## Error in signature_roo_it[[2]] : subscript out of bounds

## Warning in loglikDM(data, gamma.MoM): full precision may not have been achieved
## in 'lgamma'

## Error in rowSums(sig_obj$Y) :
##   'x' must be an array of at least two dimensions
## Error in rowSums(sig_obj$Y) :
##   'x' must be an array of at least two dimensions
## Error : $ operator is invalid for atomic vectors

## [1] "Biliary-AdenoCA.p value"           "Bladder-TCC.p value"
## [3] "Bone-Benign.p value"                "Bone-Epith.p value"
## [5] "Bone-Osteosarc.p value"             "Breast-AdenoCA.p value"
## [7] "Breast-DCIS.p value"                "Breast-LobularCA.p value"
## [9] "Cervix-AdenoCA.p value"              "Cervix-SCC.p value"
## [11] "CNS-GBM.p value"                   "CNS-Medullo.p value"
## [13] "CNS-Oligo.p value"                 "CNS-PiloAstro.p value"
## [15] "ColoRect-AdenoCA.p value"           "Eso-AdenoCA.p value"

```

```

## [17] "Head-SCC.p value"           "Kidney-ChRCC.p value"
## [19] "Kidney-RCC.clearcell.p value" "Kidney-RCC.papillary.p value"
## [21] "Liver-HCC.p value"          "Lung-AdenoCA.p value"
## [23] "Lung-SCC.p value"          "Lymph-BNHL.p value"
## [25] "Lymph-CLL.p value"          "Myeloid-AML.p value"
## [27] "Myeloid-MPN.p value"        "Ovary-AdenoCA.p value"
## [29] "Panc-AdenoCA.p value"       "Panc-Endocrine.p value"
## [31] "Prost-AdenoCA.p value"      "Skin-Melanoma.acral.p value"
## [33] "Skin-Melanoma.cutaneous.p value" "Skin-Melanoma.mucosal"
## [35] "SoftTissue-Leiomyo.p value"   "SoftTissue-Liposarc.p value"
## [37] "Stomach-AdenoCA.p value"     "Thy-AdenoCA.p value"
## [39] "Uterus-AdenoCA.p value"

## [1] "Bone-Osteosarc"            "Breast-AdenoCA"
## [3] "CNS-GBM"                  "CNS-Medullo"
## [5] "CNS-PiloAstro"            "ColoRect-AdenoCA"
## [7] "Eso-AdenoCA"              "Head-SCC"
## [9] "Kidney-ChRCC"             "Kidney-RCC.clearcell"
## [11] "Kidney-RCC.papillary"     "Liver-HCC"
## [13] "Lung-SCC"                 "Lymph-BNHL"
## [15] "Lymph-CLL"                "Ovary-AdenoCA"
## [17] "Panc-AdenoCA"             "Panc-Endocrine"
## [19] "Prost-AdenoCA"            "Skin-Melanoma.cutaneous"
## [21] "Stomach-AdenoCA"          "Thy-AdenoCA"
## [23] "Uterus-AdenoCA"

##      Bone-Osteosarc.p value      Breast-AdenoCA.p value
##                      6.918553e-01    1.000000e+00
##      CNS-GBM.p value           CNS-Medullo.p value
##                      1.000000e+00    3.245284e-01
##      CNS-PiloAstro.p value     ColoRect-AdenoCA.p value
##                      1.000000e+00    1.000000e+00
##      Eso-AdenoCA.p value       Head-SCC.p value
##                      1.000000e+00    1.000000e+00
##      Kidney-ChRCC.p value     Kidney-RCC.clearcell.p value
##                      1.000000e+00    1.000000e+00
##      Kidney-RCC.papillary.p value Liver-HCC.p value
##                      1.000000e+00    1.000000e+00
##      Lung-SCC.p value          Lymph-BNHL.p value
##                      1.000000e+00    1.000000e+00
##      Lymph-CLL.p value          Ovary-AdenoCA.p value
##                      1.000000e+00    1.000000e+00
##      Panc-AdenoCA.p value      Panc-Endocrine.p value
##                      1.000000e+00    1.000000e+00
##      Prost-AdenoCA.p value     Skin-Melanoma.cutaneous.p value
##                      1.000000e+00    1.000000e+00
##      Stomach-AdenoCA.p value    Thy-AdenoCA.p value
##                      1.000000e+00    1.000000e+00
##      Uterus-AdenoCA.p value
##                      7.559522e-10

##      Bone-Osteosarc.p value      Breast-AdenoCA.p value
##                      0.8538039    0.8538039

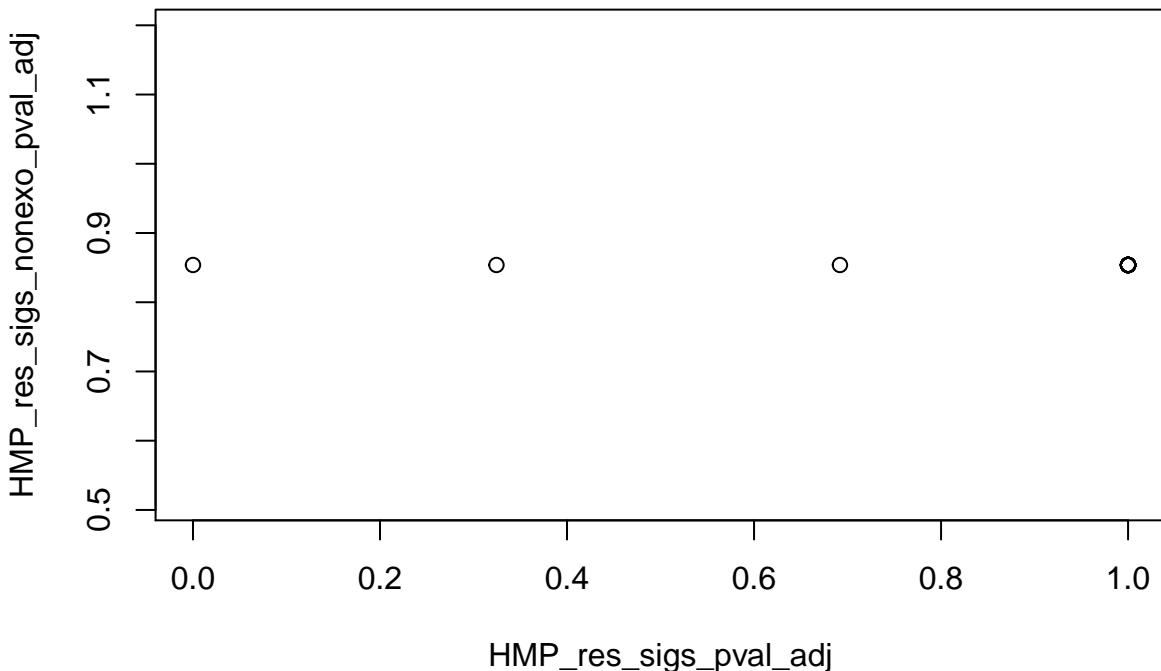
```

```

##          CNS-GBM.p value          CNS-Medullo.p value
##          0.8538039          0.8538039
##          CNS-PiloAstro.p value    ColoRect-AdenoCA.p value
##          0.8538039          0.8538039
##          Eso-AdenoCA.p value     Head-SCC.p value
##          0.8538039          0.8538039
##          Kidney-ChRCC.p value   Kidney-RCC.clearcell.p value
##          0.8538039          0.8538039
##          Kidney-RCC.papillary.p value Liver-HCC.p value
##          0.8538039          0.8538039
##          Lung-SCC.p value       Lymph-BNHL.p value
##          0.8538039          0.8538039
##          Lymph-CLL.p value      Ovary-AdenoCA.p value
##          0.8538039          0.8538039
##          Panc-AdenoCA.p value   Panc-Endocrine.p value
##          0.8538039          0.8538039
##          Prost-AdenoCA.p value  Skin-Melanoma.cutaneous.p value
##          0.8538039          0.8538039
##          Stomach-AdenoCA.p value Thy-AdenoCA.p value
##          0.8538039          0.8538039
##          Uterus-AdenoCA.p value
##          0.8538039

## Uterus-AdenoCA.p value
## 7.559522e-10

```



```

pvals_diagRE_DMDL_SP_adj <- p.adjust(pvals_diagRE_DMDL_SP, 'fdr')
pvals_diagRE_DMDL_nonexo_SP_adj <- p.adjust(pvals_diagRE_DMDL_nonexo_SP, 'fdr')
pvals_fullREDMnoscaling_SP_nonexo_subsets_and_amalgamations_adj <- p.adjust(pvals_fullREDMnoscaling_SP_no
pvals_fullRE_M_nonexo_SP_adj <- p.adjust(pvals_fullRE_M_nonexo_SP, 'fdr')
pvals_fullRE_DMSL_nonexo_SP_adj <- p.adjust(pvals_fullRE_DMSL_nonexo_SP, 'fdr')

```

```

names(pvals_diagRE_DMDL_SP_adj) <- names(pvals_diagRE_DMDL_nonexo_SP_adj) <- names(pvals_fullRE_M_nonexo_)

table(diagRE_DMDL_DA=pvals_diagRE_DMDL_nonexo_SP_adj <= 0.05,
      fullRE_DMSL_DA=pvals_fullRE_DMSL_nonexo_SP_adj <= 0.05)

##          fullRE_DMSL_DA
## diagRE_DMDL_DA FALSE TRUE
##      FALSE      5     1
##      TRUE       1    11

names(pvals_diagRE_DMDL_nonexo_SP_adj)[which((pvals_diagRE_DMDL_nonexo_SP_adj <= 0.05) & (fullRE_DMSL_DA=0.05))

## [1] "Lymph-CLL"

names(pvals_diagRE_DMDL_nonexo_SP_adj)[which((pvals_diagRE_DMDL_nonexo_SP_adj > 0.05) & (fullRE_DMSL_DA=0.05))

## [1] "Stomach-AdenoCA"

table(diagRE_DMDL_DA=pvals_diagRE_DMDL_nonexo_SP_adj <= 0.05,
      fullRE_M_DA=pvals_fullRE_M_nonexo_SP_adj <= 0.05)

##          fullRE_M_DA
## diagRE_DMDL_DA TRUE
##      FALSE      5
##      TRUE      17

# table(diagRE_DMDL_DA=pvals_diagRE_DMDL_nonexo_SP_adj <= 0.05,
#       HMP_DA=HMP_res_sigs_pval_adj <= 0.05)

# table(diagRE_DMDL_DA=pvals_diagRE_DMDL_nonexo_SP_adj < 0.05, HMP_res_sigs_DA=HMP_res_sigs < 0.05)

p.adjust(pvals_diagRE_DMDL_SP, method = "BH")

##          Bone-Osteosarc      Breast-AdenoCA          CNS-GBM
##      1.251484e-04      9.854928e-28      3.551572e-03
##          CNS-Medullo      CNS-PiloAstro      ColoRect-AdenoCA
##      8.431463e-03      6.176762e-04      2.330581e-25
##          Eso-AdenoCA      Head-SCC          Kidney-ChRCC
##      1.302667e-20      6.081301e-05      2.291117e-09
##      Kidney-RCC.clearcell      Kidney-RCC.papillary      Liver-HCC
##      7.383723e-18                  NA      5.222605e-106
##          Lung-SCC      Lymph-BNHL          Lymph-CLL
##      2.130510e-21      7.817274e-19      1.454624e-19
##          Ovary-AdenoCA      Panc-AdenoCA      Panc-Endocrine
##      4.930852e-37      9.012085e-118      6.747398e-10
##          Prost-AdenoCA      Skin-Melanoma.cutaneous      Stomach-AdenoCA
##      4.747685e-98      2.914093e-24      2.358331e-06
##          Thy-AdenoCA      Uterus-AdenoCA
##      1.141617e-05      7.574077e-10

p.adjust(pvals_diagRE_DMDL_nonexo_SP, method = "BH")

##          Bone-Osteosarc      Breast-AdenoCA          CNS-GBM
##      8.712593e-04      6.123721e-10      1.557247e-01

```

```

##          CNS-Medullo      CNS-PiloAstro    ColoRect-AdenoCA
## 1.557247e-01 5.270878e-01 3.141298e-17
##          Eso-AdenoCA      Head-SCC       Kidney-ChRCC
## 7.328650e-17 1.788732e-04 1.557247e-01
## Kidney-RCC.clearcell Kidney-RCC.papillary Liver-HCC
## 1.137607e-13 1.190455e-06 1.263498e-70
##          Lung-SCC        Lymph-BNHL     Lymph-CLL
## 8.301558e-01 5.881884e-07 2.256505e-03
##          Ovary-AdenoCA    Panc-AdenoCA   Panc-Endocrine
## 6.973340e-26 8.893577e-60 4.169125e-05
##          Prost-AdenoCA   Skin-Melanoma.cutaneous Stomach-AdenoCA
## 6.044387e-83 9.919659e-17 6.320633e-02
##          Thy-AdenoCA      Uterus-AdenoCA
## 1.170750e-02 1.568961e-09

df_pvals_DMDL_SP <- cbind.data.frame(pvals_DM=pvals_diagRE_DMDL_SP_adj,
                                         pvals_DMnonexo=pvals_diagRE_DMDL_nonexo_SP_adj,
                                         num_samples=as.numeric(num_samples_all_SP),
                                         num_sigs_nonexo=as.numeric(num_sigs_nonexo_SP),
                                         ct=enough_samples,
                                         pvals_DM_censored=sapply(-log(pvals_diagRE_DMDL_SP_adj),
                                         function(i) min(i, 25)),
                                         pvals_DMnonexo_censored=sapply(-log(pvals_diagRE_DMDL_nonexo_SP_adj),
                                         function(i) min(i, 25)),
                                         bool_censored=(( -log(pvals_diagRE_DMDL_nonexo_SP_adj) > 25 ) | ( -log(pvals_diagRE_))

ggplot(df_pvals_DMDL_SP,
       aes(x=pvals_DM_censored, y=pvals_DMnonexo_censored,
           # size=num_samples,
           label=ct, size=bool_censored))+geom_point(aes (col=ct))+geom_hline(yintercept = -log(0.05), lty='dashed')+geom_vline(xintercept = -log(0.05), lty='dashed')+geom_label_repel(size=3.2, alpha=0.6, max.overlaps = 30)+ theme_bw()+
theme(legend.position = "bottom", legend.text=element_text(size=8))+labs(x=' - Log p-value all signatures', y=' - Log p-value nonexogenous signatures')+guides(size=FALSE, col=FALSE)+ #, col=guide_legend(ncol=4),
scale_color_manual(values = pcawg_palette)+theme(legend.position = "bottom")+
lims(x=c(0, 30), y=c(0,30))

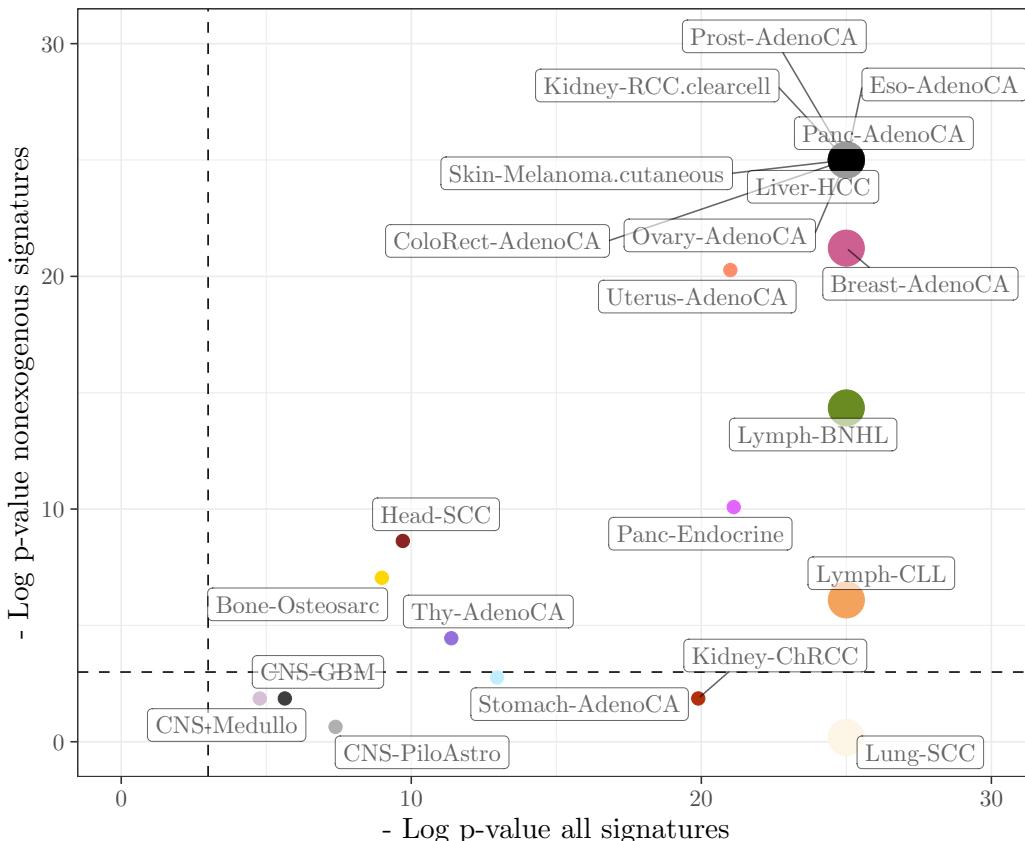
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Using size for a discrete variable is not advised.

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_label_repel).

```



```
t.test(df_pvals_DMDL_SP[df_pvals_DMDL_SP$pvals_DMnonexo <= 0.05, 'num_samples'],
       df_pvals_DMDL_SP[df_pvals_DMDL_SP$pvals_DMnonexo > 0.05, 'num_samples'])
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: df_pvals_DMDL_SP[df_pvals_DMDL_SP$pvals_DMnonexo <= 0.05, "num_samples"] and df_pvals_DMDL_SP[df_pvals_DMDL_SP$pvals_DMnonexo > 0.05, "num_samples"]
## t = 1.8016, df = 19.213, p-value = 0.08733
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -5.663104 76.055261
```

```
## sample estimates:
```

```
## mean of x mean of y
```

```
## 82.52941 47.33333
```

```
t.test(df_pvals_DMDL_SP[df_pvals_DMDL_SP$pvals_DMnonexo <= 0.05, 'num_sigs_nonexo'],
       df_pvals_DMDL_SP[df_pvals_DMDL_SP$pvals_DMnonexo > 0.05, 'num_sigs_nonexo'])
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: df_pvals_DMDL_SP[df_pvals_DMDL_SP$pvals_DMnonexo <= 0.05, "num_sigs_nonexo"] and df_pvals_DMDL_SP[df_pvals_DMDL_SP$pvals_DMnonexo > 0.05, "num_sigs_nonexo"]
## t = 1.466, df = 7.0453, p-value = 0.1858
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -1.886557 8.063027
```

```

## sample estimates:
## mean of x mean of y
## 8.588235 5.500000
mean(df_pvals_DMDL_SP[df_pvals_DMDL_SP$pvals_DMnonexo <= 0.05, 'num_sigs_nonexo'])

## [1] 8.588235
mean(df_pvals_DMDL_SP[df_pvals_DMDL_SP$pvals_DMnonexo > 0.05, 'num_sigs_nonexo'])

## [1] 5.5

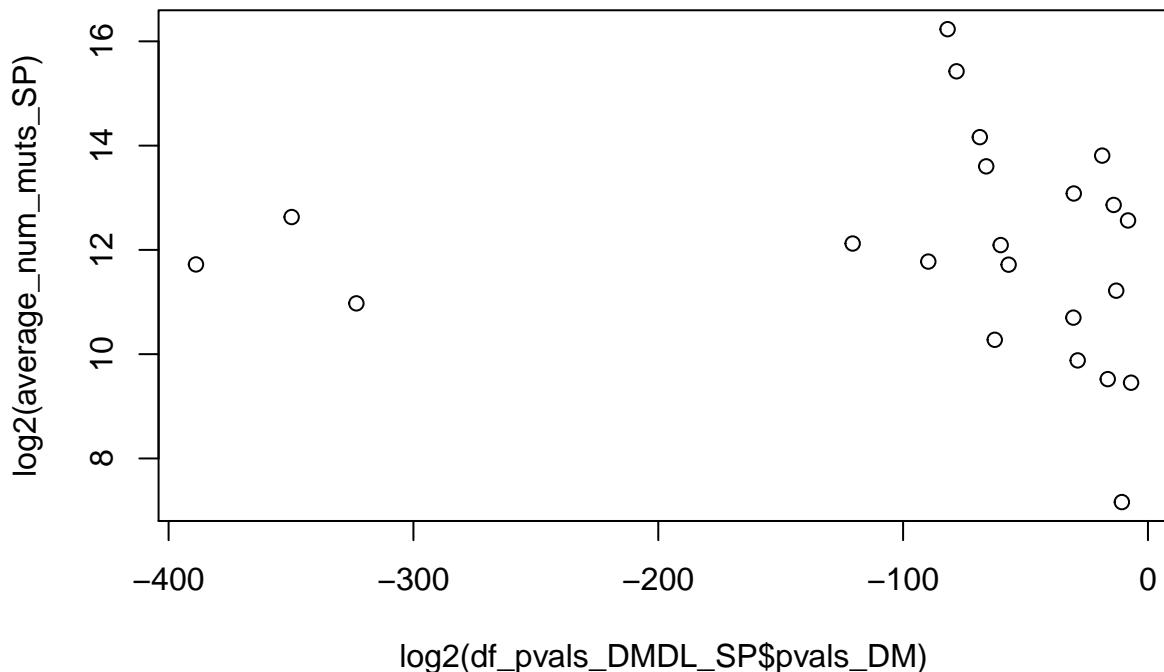
Does DA or not scale with the average number of mutations in the observed exposures (i.e. per patient and group) of the relevant ct?

average_num_muts_SP <- sapply(enough_samples, function(ct){
  .xx <- all_objects_SP[[ct]]
  try(mean(rowSums(.xx$Y)))
})

average_num_muts_SP

##          Bone-Osteosarc      Breast-AdenoCA      CNS-GBM
##                2382.1852        3507.5588       6057.6618
##          CNS-Medullo      CNS-PiloAstro  ColoRect-AdenoCA
##                  700.8915        143.3452      77035.7838
##          Eso-AdenoCA      Head-SCC      Kidney-ChRCC
##                 12439.3692        7445.4219       942.5263
## Kidney-RCC.clearcell  Kidney-RCC.papillary      Liver-HCC
##                   3365.9477        2740.9833      6340.5121
##          Lung-SCC      Lymph-BNHL      Lymph-CLL
##                 18340.1471        4367.0392      1239.9057
##          Ovary-AdenoCA      Panc-AdenoCA  Panc-Endocrine
##                   4458.4691        3376.0415      1664.5857
## Prost-AdenoCA Skin-Melanoma.cutaneous      Stomach-AdenoCA
##                  2011.8966        43998.4833     14345.9667
##          Thy-AdenoCA      Uterus-AdenoCA
##                   734.8902        8670.1125
plot(log2(df_pvals_DMDL_SP$pvals_DM), log2(average_num_muts_SP))

```



`log2(average_num_mutations_SP)`

```
t.test(log2(average_num_mutations_SP)[df_pvals_DMDL_SP$pvals_DMnonexo <= 0.05],
       log2(average_num_mutations_SP)[df_pvals_DMDL_SP$pvals_DMnonexo > 0.05])
```

##

Welch Two Sample t-test

##

data: log2(average_num_mutations_SP)[df_pvals_DMDL_SP\$pvals_DMnonexo <= 0.05] and log2(average_num_mutations_SP)[df_pvals_DMDL_SP\$pvals_DMnonexo > 0.05]

t = 0.85052, df = 6.3966, p-value = 0.4257

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1.88287 3.93586

sample estimates:

mean of x mean of y

12.19856 11.17207

```
pcawg_palette <- pcawg.colour.palette(gsub("\\.*", "", enough_samples), scheme = "tumour.subtype")
names(pcawg_palette) <- enough_samples
```

`pvals_diagRE_DMDL_nonexo_SP_adj`

##	Bone-Osteosarc	Breast-AdenoCA	CNS-GBM
##	8.712593e-04	6.123721e-10	1.557247e-01
##	CNS-Medullo	CNS-PiloAstro	ColoRect-AdenoCA
##	1.557247e-01	5.270878e-01	3.141298e-17
##	Eso-AdenoCA	Head-SCC	Kidney-ChRCC
##	7.328650e-17	1.788732e-04	1.557247e-01
##	Kidney-RCC.clearcell	Kidney-RCC.papillary	Liver-HCC
##	1.137607e-13	1.190455e-06	1.263498e-70
##	Lung-SCC	Lymph-BNHL	Lymph-CLL
##	8.301558e-01	5.881884e-07	2.256505e-03
##	Ovary-AdenoCA	Panc-AdenoCA	Panc-Endocrine

```

##          6.973340e-26          8.893577e-60          4.169125e-05
## Prost-AdenoCA Skin-Melanoma.cutaneous      Stomach-AdenoCA
##          6.044387e-83          9.919659e-17          6.320633e-02
## Thy-AdenoCA           Uterus-AdenoCA
##          1.170750e-02          1.568961e-09

pvals_fullREDMnoscaling_SP_nonexo_subsets_and_amalgamations_adj

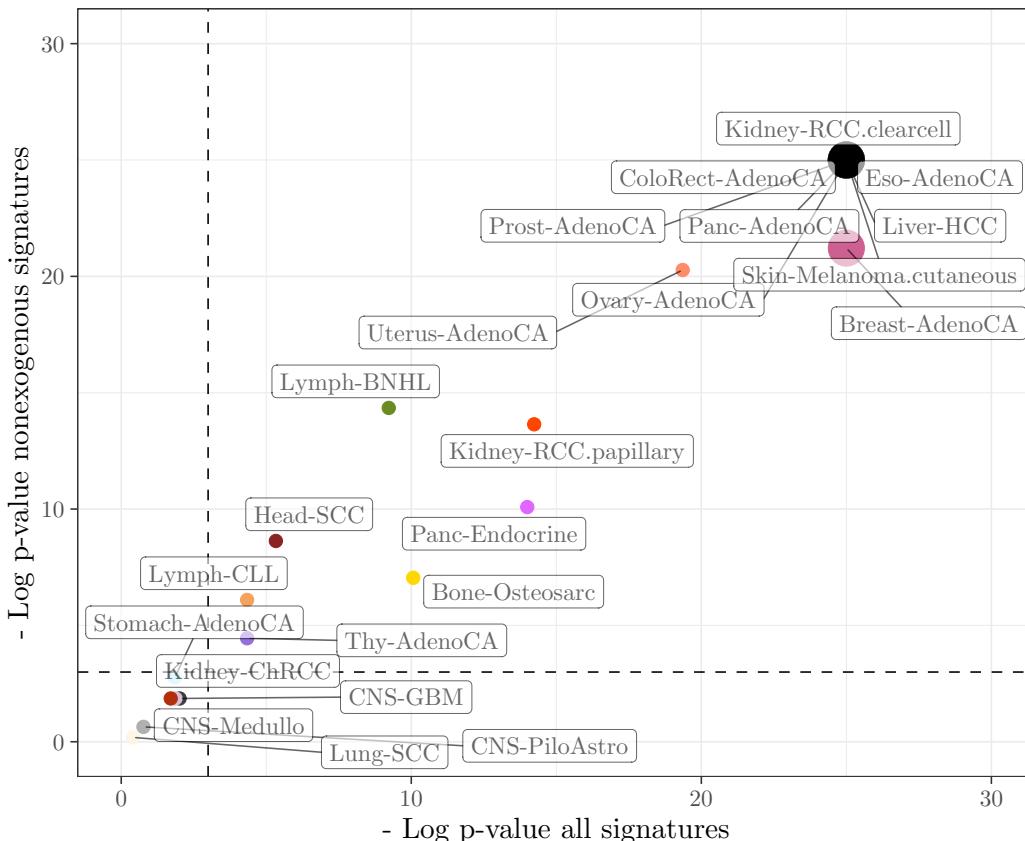
##          Bone-Osteosarc      Breast-AdenoCA          CNS-GBM
##          4.248457e-05          7.759783e-12          1.336685e-01
##          CNS-Medullo        CNS-PiloAstro       ColoRect-AdenoCA
##          1.538381e-01          4.648237e-01          1.023484e-15
##          Eso-AdenoCA          Head-SCC            Kidney-ChRCC
##          3.759670e-18          4.834276e-03          1.814391e-01
## Kidney-RCC.clearcell    Kidney-RCC.papillary      Liver-HCC
##          3.652385e-14          6.557344e-07          3.704923e-63
##          Lung-SCC             Lymph-BNHL            Lymph-CLL
##          6.693019e-01          9.868396e-05          1.303426e-02
##          Ovary-AdenoCA        Panc-AdenoCA       Panc-Endocrine
##          1.231778e-22          7.710539e-59          8.327076e-07
##          Prost-AdenoCA Skin-Melanoma.cutaneous      Stomach-AdenoCA
##          6.572758e-67          3.732926e-14          1.576047e-01
##          Thy-AdenoCA          Uterus-AdenoCA
##          1.303426e-02          3.911013e-09

df_pvals_fullREDMnoscaling_SP_nonexo_subsets_and_amalgamations <- cbind.data.frame(pvals_DMnonexo_nonscal-
                                         pvals_DMnonexo=pvals_diagRE_DMDL_nonexo_SP_adj,
                                         num_samples=as.numeric(num_samples_all_SP),
                                         num_sigs_nonexo=as.numeric(num_sigs_nonexo_SP),
                                         ct=enough_samples,
                                         pvals_DM_nonscaling_censored=sapply(-log(pvals_fullREDMnoscaling_SP_nonexo_subset
                                             function(i) min(i, 25)),
                                         pvals_DMnonexo_censored=sapply(-log(pvals_diagRE_DMDL_nonexo_SP_adj),
                                             function(i) min(i, 25)),
                                         bool_censored=(( -log(pvals_diagRE_DMDL_nonexo_SP_adj) > 25 ) | ( -log(pvals_fullRE
ggplot(df_pvals_fullREDMnoscaling_SP_nonexo_subsets_and_amalgamations,
  aes(x=pvals_DM_nonscaling_censored, y=pvals_DMnonexo_censored,
      # size=num_samples,
      label=ct, size=bool_censored))+geom_point(aes (col=ct))+
  geom_hline(yintercept = -log(0.05), lty='dashed')+geom_vline(xintercept = -log(0.05), lty='dashed')+#
  geom_label_repel(size=3.2, alpha=0.6, max.overlaps = 30)+ theme_bw()+
  theme(legend.position = "bottom", legend.text=element_text(size=8))+#
  labs(x=' - Log p-value all signatures', y=' - Log p-value nonexogenous signatures')+#
  guides(size=FALSE, col=FALSE)+ #, col=guide_legend(ncol=4),
  scale_color_manual(values = pcawg_palette)+theme(legend.position = "bottom")+
  lims(x=c(0, 30), y=c(0,30))

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Using size for a discrete variable is not advised.

```



See script PCAWG_HMP_and_alternative_methods.R for the analyses of PCAWG data using alternative models.

Tracksig

```

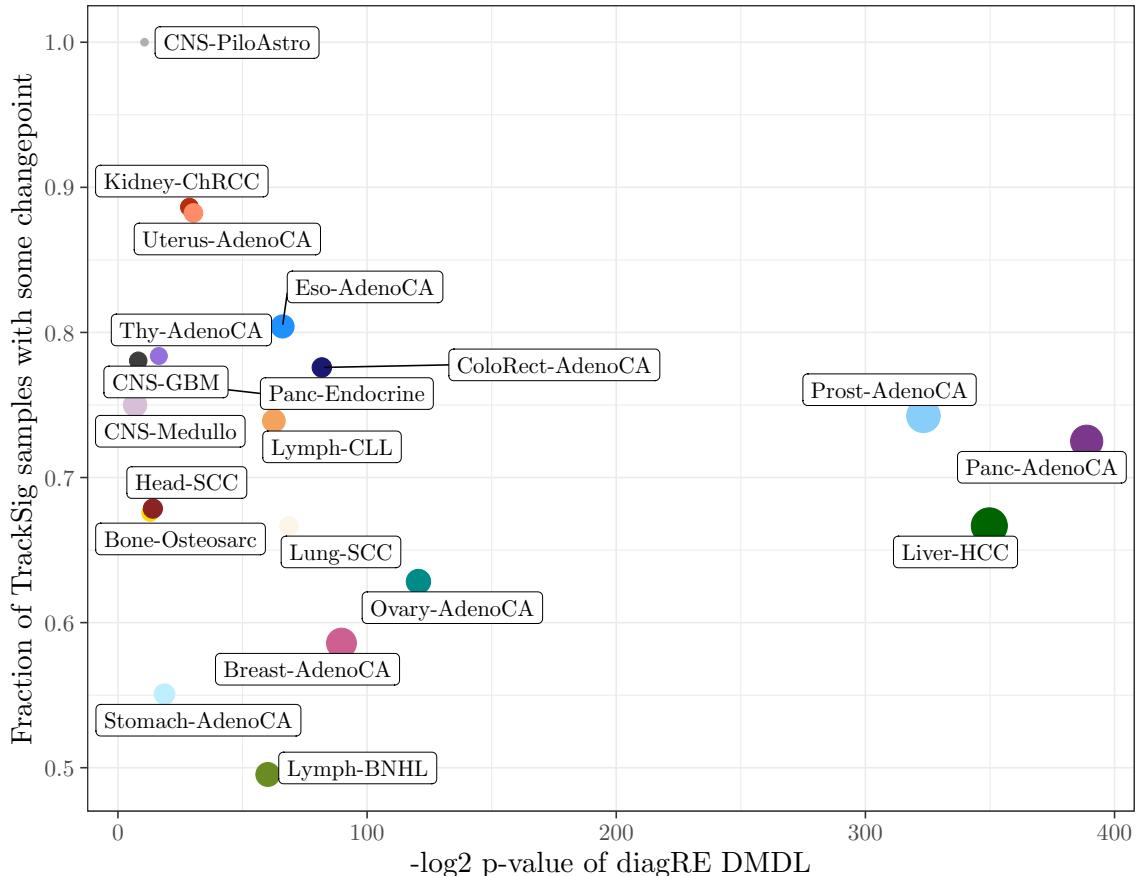
tracksig = read.csv("../data/restricted/tracksig/changepoints_stats_tracksig.csv", stringsAsFactors =
tracksig = tracksig %>% group_by(type) %>%
  dplyr::summarize(count = n(), bool_changepoints=sum(n_changepoints > 0)) %>%
  mutate(tracksig_frac= bool_changepoints/count)
tracksig = cbind.data.frame(pvals_diagRE_DMDL_SP_adj,
                           tracksig[match(names(pvals_diagRE_DMDL_SP_adj), tracksig$type),],
                           effect_size3_SP=effect_size3_SP[match(names(pvals_diagRE_DMDL_SP_adj), names(effect_size3_SP))])
tracksig$ct = rownames(tracksig)
tracksig$minpvals = -log2(tracksig$pvals_diagRE_DM)

pcawg_palette <- pcawg.colour.palette(gsub("\\\\..*", "", tracksig$ct), scheme = "tumour.subtype")
names(pcawg_palette) <- tracksig$ct

ggplot(tracksig, aes(x=-log2(pvals_diagRE_DMDL_SP_adj), y=tracksig_frac, label=ct, size=count))+geom_point()
  labs(x='-\log2 p-value of diagRE DMDL', y='Fraction of TrackSig samples with some changepoint')+ 
  scale_color_manual(values = pcawg_palette)+theme(legend.position = "bottom")+
  # scale_x_continuous(trans = "log2")+
  theme_bw() + theme(legend.position = "bottom")

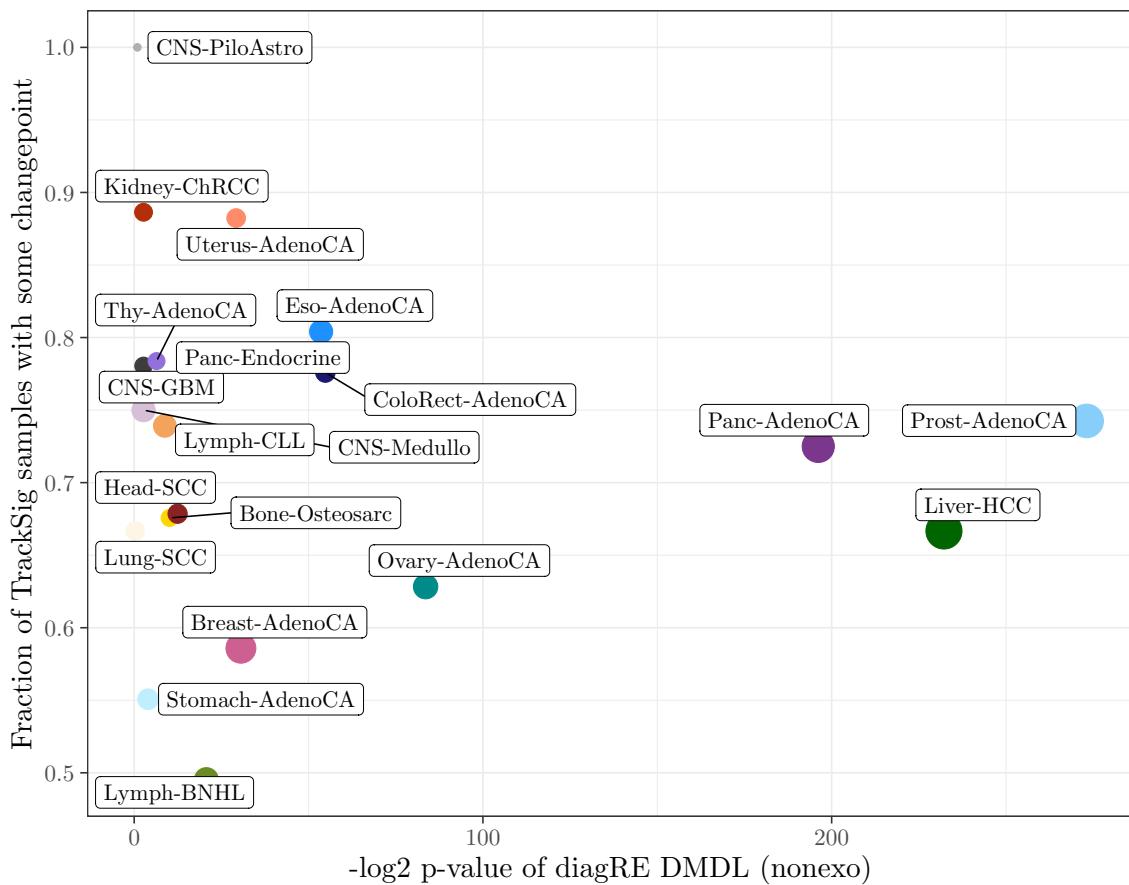
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
## Warning: Removed 3 rows containing missing values (geom_label_repel).
```



```
ggplot(tracksig, aes(x=-log2(pvals_diagRE_DMDL_nonexo_SP_adj), y=tracksig_frac, label=ct, size=count))+
  geom_point()+
  labs(x='-\log2 p-value of diagRE DMDL (nonexo)', y='Fraction of TrackSig samples with some changepoint')+
  scale_color_manual(values = pcawg_palette)+theme(legend.position = "bottom")+
  theme_bw()+theme(legend.position = "bottom")
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
## Warning: Removed 3 rows containing missing values (geom_label_repel).
```



```

  300      ● Bone-Osteosarc   ● ColoRect-AdenoCA   ● Kidney-RCC.papillary   ● Ovary-Aden
ct      ● Breast-AdenoCA   ● Eso-AdenoCA       ● Liver-HCC                 ● Panc-Adeno
      ● CNS-GBM           ● Head-SCC          ● Lung-SCC                  ● Panc-Endoc
      ● CNS-Medullo        ● Kidney-ChRCC       ● Lymph-BNHL                ● Prost-Adenc
      ● CNS-PiloAstro      ● Kidney-RCC.clearcell ● Lymph-CLL                 ● Skin-Melan

```

```

ggplot(tracksig, aes(x=effect_size3_SP, y=tracksig_frac, label=ct, col=ct))+  

  geom_point(aes(size=minpvals))+geom_label_repel(max.overlaps = 5)+  

  labs(x='Effect size', y='Fraction of TrackSig samples with some changepoint', col="")  

  + theme_bw()  

  + scale_color_manual(values = pcawg_palette)+theme(legend.position = "bottom")  

  + guides(col=guide_legend(ncol=4), size=FALSE)

```

```

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  

## "none")` instead.  

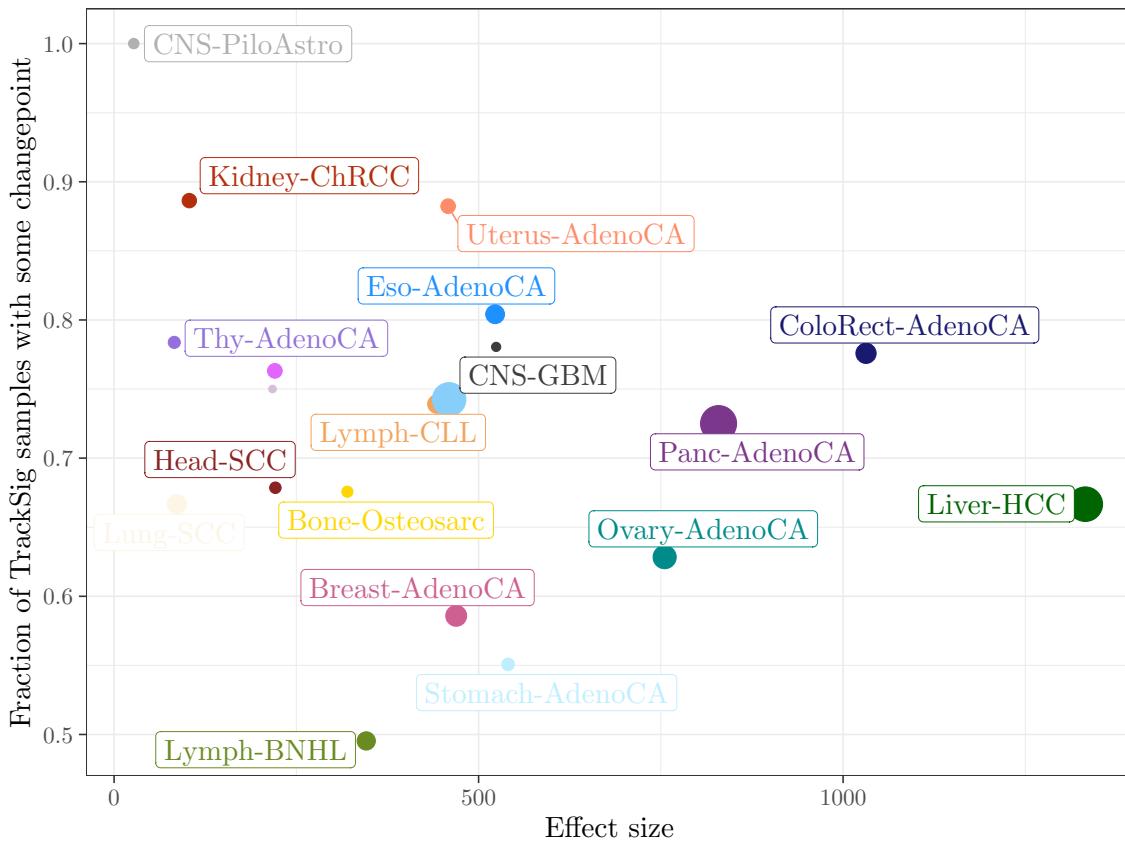
## Warning: Removed 3 rows containing missing values (geom_point).  

## Warning: Removed 3 rows containing missing values (geom_label_repel).  

## Warning: ggrepel: 3 unlabeled data points (too many overlaps). Consider  

## increasing max.overlaps

```



a	Bone-Osteosarc	a	Eso-AdenoCA	a	Lung-SCC	a	Prost-AdenoCA
a	Breast-AdenoCA	a	Head-SCC	a	Lymph-BNHL	a	Skin-Melanoma.cutaneo
a	CNS-GBM	a	Kidney-ChRCC	a	Lymph-CLL	a	Stomach-AdenoCA
a	CNS-Medullo	a	Kidney-RCC.clearcell	a	Ovary-AdenoCA	a	Thy-AdenoCA
a	CNS-PiloAstro	a	Kidney-RCC.papillary	a	Panc-AdenoCA	a	Uterus-AdenoCA
a	ColoRect-AdenoCA	a	Liver-HCC	a	Panc-Endocrine		

Same plots, but smaller, for images

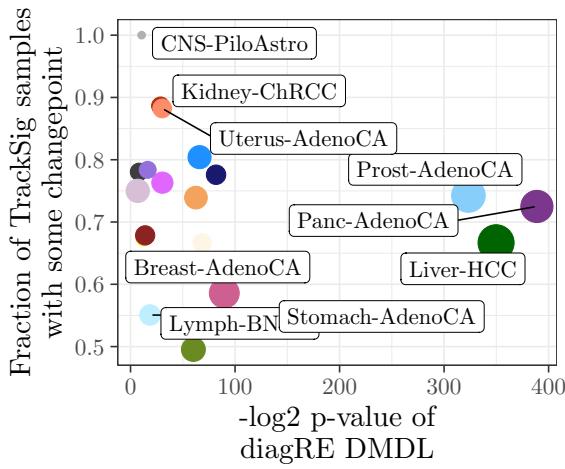
```
ggplot(tracksig, aes(x=-log2(pvals_diagRE_DMDL_SP_adj), y=tracksig_frac, label=ct, size=count))+geom_point()
  labs(x='-log2 p-value of\n diagRE DMDL', y='Fraction of TrackSig samples\n with some changepoint')+ 
  scale_color_manual(values = pcawg_palette)+theme(legend.position = "bottom")+
  # scale_x_continuous(trans = "log2")+
  theme_bw()+theme(legend.position = "bottom")+guides(col=FALSE)+labs(size='N. obs')

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Removed 3 rows containing missing values (geom_point).

## Warning: Removed 3 rows containing missing values (geom_label_repel).

## Warning: ggrepel: 11 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



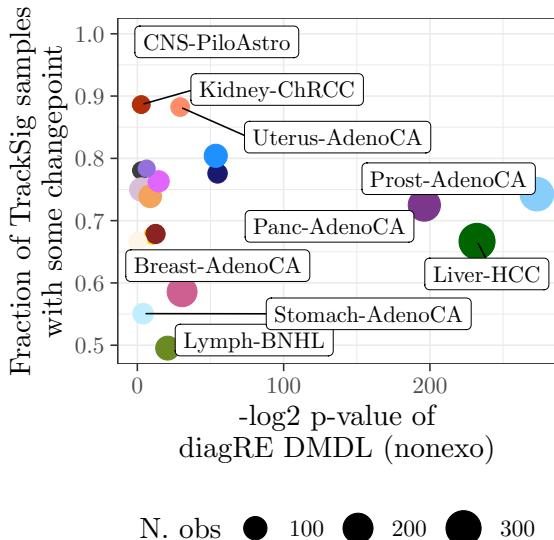
```
ggplot(tracksig, aes(x=-log2(pvals_diagRE_DMDL_nonexo_SP_adj), y=tracksig_frac, label=ct, size=count))+
  geom_label_repel(size=3, col='black', max.overlaps = 4)+
  labs(x='-\log2 p-value of\n diagRE DMDL (nonexo)', y='Fraction of TrackSig samples\n with some changepoint')+
  scale_color_manual(values = pcawg_palette)+theme(legend.position = "bottom")+
  theme_bw()+theme(legend.position = "bottom")+guides(col=FALSE)+labs(size='N. obs')
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Removed 3 rows containing missing values (geom_point).

## Warning: Removed 3 rows containing missing values (geom_label_repel).

## Warning: ggrepel: 11 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
plot_for_ct_legend <- ggplot(tracksig, aes(x=-log2(pvals_diagRE_DMDL_nonexo_SP_adj), y=tracksig_frac, label=ct, size=count))+
  geom_label_repel(size=3, col='black', max.overlaps = 2)+
```

```

  labs(x=-log2 p-value of\n diagRE DMDL (nonexo)', y='Fraction of TrackSig samples\n with some changepoi
    scale_color_manual(values = pcawg_palette)+theme(legend.position = "bottom")+
  theme_bw()+theme(legend.position = "bottom")+labs(size='N. observations')+guides(size=FALSE)+labs(col='')

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

legend_ct <- (cowplot::get_legend(plot_for_ct_legend))

## Warning: Removed 3 rows containing missing values (geom_point).

## Warning: Removed 3 rows containing missing values (geom_label_repel).

# pdf("../results/results_TMB/pcawg/legend_cts.pdf", height = 1, width = 6.5)
grid.newpage()
grid.draw(legend_ct)

  ● Bone-Osteosarc      ● ColoRect-AdenoCA      ● Kidney-RCC.papillary      ● Ovary-AdenoCA      ● Stomach-AdenoCA
  ● Breast-AdenoCA     ● Eso-AdenoCA        ● Liver-HCC            ● Panc-AdenoCA       ● Thy-AdenoCA
  ● CNS-GBM             ● Head-SCC          ● Lung-SCC           ● Panc-Endocrine     ● Uterus-AdenoCA
  ● CNS-Medullo         ● Kidney-ChRCC        ● Lymph-BNHL          ● Prost-AdenoCA
  ● CNS-PiloAstro       ● Kidney-RCC.clearcell ● Lymph-CLL           ● Skin-Melanoma.cutaneous

# dev.off()

tracksig

##                                     pvals_diagRE_DMDL_SP_adj      type  count
## Bone-Osteosarc                  1.251484e-04 Bone-Osteosarc 37
## Breast-AdenoCA                 9.854928e-28 Breast-AdenoCA 198
## CNS-GBM                         3.551572e-03 CNS-GBM        41
## CNS-Medullo                     8.431463e-03 CNS-Medullo   100
## CNS-PiloAstro                   6.176762e-04 CNS-PiloAstro  3
## ColoRect-AdenoCA                2.330581e-25 ColoRect-AdenoCA 58
## Eso-AdenoCA                     1.302667e-20 Eso-AdenoCA  97
## Head-SCC                        6.081301e-05 Head-SCC        56
## Kidney-ChRCC                    2.291117e-09 Kidney-ChRCC 44
## Kidney-RCC.clearcell           7.383723e-18 <NA>          NA
## Kidney-RCC.papillary            NA              <NA>          NA
## Liver-HCC                       5.222605e-106 Liver-HCC      321
## Lung-SCC                        2.130510e-21 Lung-SCC        48
## Lymph-BNHL                      7.817274e-19 Lymph-BNHL     107
## Lymph-CLL                        1.454624e-19 Lymph-CLL        92
## Ovary-AdenoCA                   4.930852e-37 Ovary-AdenoCA 113
## Panc-AdenoCA                    9.012085e-118 Panc-AdenoCA 240
## Panc-Endocrine                  6.747398e-10 Panc-Endocrine 76
## Prost-AdenoCA                   4.747685e-98 Prost-AdenoCA 268
## Skin-Melanoma.cutaneous         2.914093e-24 <NA>          NA
## Stomach-AdenoCA                 2.358331e-06 Stomach-AdenoCA 69
## Thy-AdenoCA                     1.141617e-05 Thy-AdenoCA  37
## Uterus-AdenoCA                  7.574077e-10 Uterus-AdenoCA 51
##                                     bool_changepoints tracksig_frac effect_size3_SP
## Bone-Osteosarc                  25          0.6756757      320.02704

```

```

## Breast-AdenoCA           116    0.5858586   469.29125
## CNS-GBM                  32     0.7804878   524.08517
## CNS-Medullo              75     0.7500000   217.31161
## CNS-PiloAstro             3      1.0000000   27.02067
## ColoRect-AdenoCA          45     0.7758621   1031.46320
## Eso-AdenoCA               78     0.8041237   522.66804
## Head-SCC                  38     0.6785714   221.26052
## Kidney-ChRCC              39     0.8863636   103.16727
## Kidney-RCC.clearcell       NA     NA           247.53963
## Kidney-RCC.papillary       NA     NA           156.30958
## Liver-HCC                 214    0.6666667   1332.22216
## Lung-SCC                  32     0.6666667   86.07143
## Lymph-BNHL                53     0.4953271   345.88535
## Lymph-CLL                  68     0.7391304   442.78303
## Ovary-AdenoCA              71     0.6283186   755.21919
## Panc-AdenoCA               174    0.7250000   829.19711
## Panc-Endocrine              58     0.7631579   220.55650
## Prost-AdenoCA              199    0.7425373   459.55271
## Skin-Melanoma.cutaneous    NA     NA           1017.96421
## Stomach-AdenoCA             38     0.5507246   540.50633
## Thy-AdenoCA                29     0.7837838   82.57481
## Uterus-AdenoCA              45     0.8823529   458.24059
##                                         ct    minpvals
## Bone-Osteosarc              Bone-Osteosarc 12.964072
## Breast-AdenoCA              Breast-AdenoCA 89.713141
## CNS-GBM                      CNS-GBM        8.137327
## CNS-Medullo                 CNS-Medullo   6.890001
## CNS-PiloAstro                CNS-PiloAstro 10.660862
## ColoRect-AdenoCA             ColoRect-AdenoCA 81.827512
## Eso-AdenoCA                  Eso-AdenoCA   66.057093
## Head-SCC                      Head-SCC      14.005260
## Kidney-ChRCC                 Kidney-ChRCC  28.701302
## Kidney-RCC.clearcell         Kidney-RCC.clearcell 56.910357
## Kidney-RCC.papillary         Kidney-RCC.papillary NA
## Liver-HCC                      Liver-HCC    349.739609
## Lung-SCC                      Lung-SCC     68.669291
## Lymph-BNHL                     Lymph-BNHL   60.149968
## Lymph-CLL                      Lymph-CLL    62.575988
## Ovary-AdenoCA                 Ovary-AdenoCA 120.609503
## Panc-AdenoCA                  Panc-AdenoCA 388.815654
## Panc-Endocrine                 Panc-Endocrine 30.464950
## Prost-AdenoCA                  Prost-AdenoCA 323.301729
## Skin-Melanoma.cutaneous       Skin-Melanoma.cutaneous 78.183228
## Stomach-AdenoCA                Stomach-AdenoCA 18.693802
## Thy-AdenoCA                   Thy-AdenoCA   16.418562
## Uterus-AdenoCA                 Uterus-AdenoCA 30.298211
gerstung_changing_sigs_early_late <- readxl::read_excel("/Users/morril01/Documents/PhD/GlobalDA/data/rest

## New names:
## * `` -> ...

```

```

gerstung_changing_sigs_clonalsubclonal <- readxl::read_excel("/Users/morril01/Documents/PhD/GlobalDA/data/gerstung_changing_sigs_clonalsubclonal.xlsx")
## New names:
## * `` -> ...
# gerstung_changing_sigs <- readxl::read_excel("/Users/morril01/Documents/PhD/GlobalDA/data/restricted/gerstung_changing_sigs.xlsx")
# gerstung_changing_sigs <- readxl::read_excel("/Users/morril01/Documents/PhD/GlobalDA/data/restricted/gerstung_changing_sigs.xlsx")
# gerstung_constant_sigs <- readxl::read_excel("/Users/morril01/Documents/PhD/GlobalDA/data/restricted/gerstung_constant_sigs.xlsx")
# gerstung_changing_sigs
# gerstung_constant_sigs

gerstung_changing_sigs_earlylate <- gerstung_changing_sigs_early_late #gerstung_changing_sigs[gerstung_changing_sigs=="Early"]
# gerstung_changing_sigs_clonalsubclonal <- gerstung_changing_sigs_clonal_subclonal #gerstung_changing_sigs[gerstung_changing_sigs=="Clonal"]

gerstung_changing_sigs_earlylate$signature[(gerstung_changing_sigs_earlylate$signature == "SBS6.14.15.20") | (gerstung_changing_sigs_earlylate$signature == "SBS6.14.15.21")]
gerstung_changing_sigs_clonalsubclonal$signature[(gerstung_changing_sigs_clonalsubclonal$signature == "SBS6.14.15.20") | (gerstung_changing_sigs_clonalsubclonal$signature == "SBS6.14.15.21")]
gerstung_changing_sigs_earlylate$signature <- gsub("_", ".", gerstung_changing_sigs_earlylate$signature)
gerstung_changing_sigs_clonalsubclonal$signature <- gsub("_", ".", gerstung_changing_sigs_clonalsubclonal$signature)

# df_changes_el <- gerstung_changing_sigs_earlylate %>% group_by(signature) %>% dplyr::summarise(median=mean)
# df_changes_cs <- gerstung_changing_sigs_clonalsubclonal %>% group_by(signature) %>% dplyr::summarise(median=mean)
df_changes_el <- gerstung_changing_sigs_earlylate %>% group_by(signature) %>% dplyr::summarise(median=mean)
df_changes_cs <- gerstung_changing_sigs_clonalsubclonal %>% group_by(signature) %>% dplyr::summarise(median=mean)

# grid.arrange(ggplot(gerstung_changing_sigs_earlylate, aes(x=factor(signature, levels=df_changes_el$signature), y=mean_change, group=signature, col=histologyAbbreviation))+geom_boxplot(), ggplot(gerstung_changing_sigs_clonalsubclonal, aes(x=factor(signature, levels=df_changes_cs$signature), y=mean_change, group=signature, col=histologyAbbreviation))+geom_boxplot(), nrow=2)
# theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+ggtitle('Early vs late'), theme_bw())
# guides(col=FALSE)+theme_bw()
# geom_hline(yintercept = 0, lty='dashed')+
# guides(col=FALSE)+theme_bw()
# theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+ggtitle('Clonal vs subclonal'), theme_bw())
# nrow=2)

## removing cancer types where there aren't many observations
select_self <- function(i) i[i]
gerstung_changing_sigs_earlylate <- gerstung_changing_sigs_earlylate %>% dplyr::filter(signature %in% names(select_self))
gerstung_changing_sigs_clonalsubclonal <- gerstung_changing_sigs_clonalsubclonal %>% dplyr::filter(signature %in% names(select_self))

pcawg_palette <- pcawg.colour.palette(x = gsub("-", ".", tolower(gsub("\\\\..*", "", sort(unique(gerstung_changing_sigs_earlylate$tumourType))))))

## Warning in pcawg.colour.palette(x = gsub("-", ".", tolower(gsub("\\\\..*", :, :
## Unrecognized input value for x. Default to fill.colour.
names(pcawg_palette) <- sort(unique(gerstung_changing_sigs_earlylate$tumour_type)))

# unique(gerstung_changing_sigs_clonal_subclonal$signature)
gerstung_changing_sigs_earlylate$signature <- factor(gerstung_changing_sigs_earlylate$signature,

```

```

levels=df_changes_el$signature[order(df_changes_el$signature)
gerstung_changing_sigs_clonalsubclonal$signature <- factor(gerstung_changing_sigs_clonalsubclonal$signature,
levels=df_changes_cs$signature[order(df_changes_cs$signature)

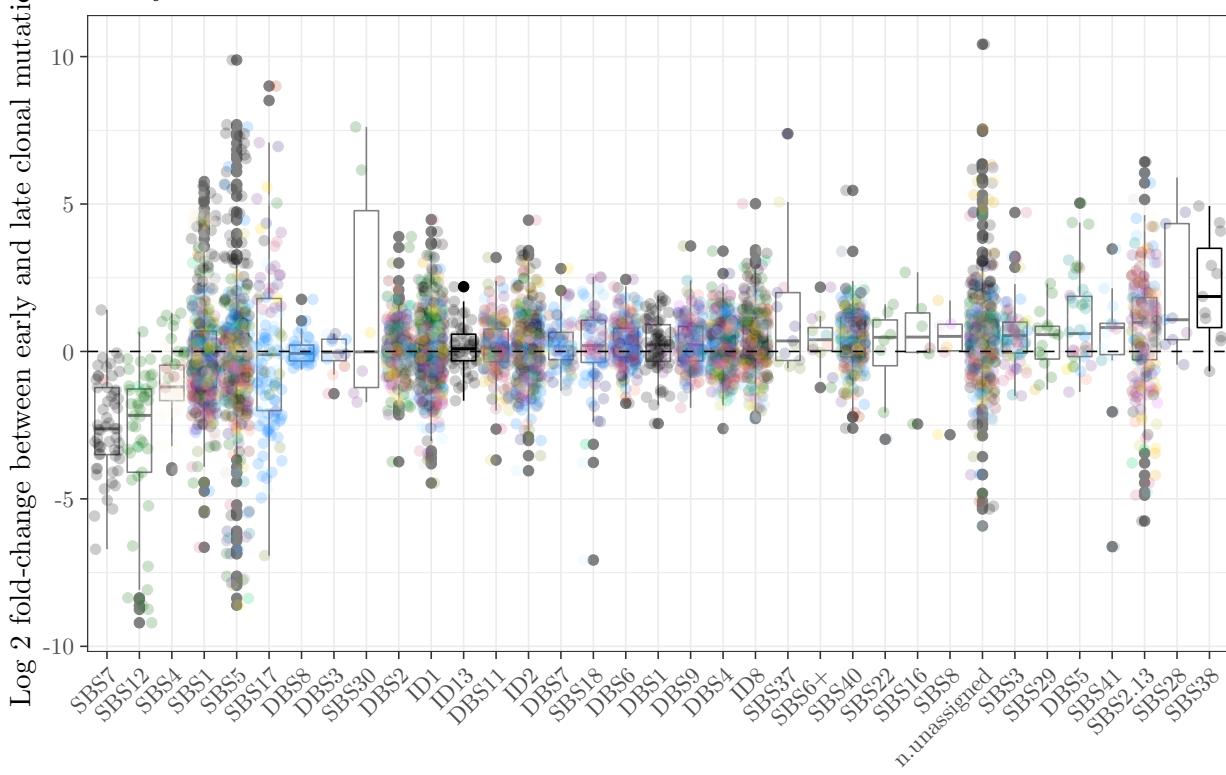
# grid.arrange(ggplot(gerstung_changing_sigs_earlylate, aes(x=signature,
#                                         y=log2fc_earlyLate, group=signature,col=tumour_
#                                         geom_hline(yintercept = 0, lty='dashed')+
#                                         guides(col=FALSE)+theme_bw()+
#                                         theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+ggtitle('Early vs late')+
#                                         scale_color_manual(values = pcawg_palette)+labs(x='', y='Log 2 fold-change between early
#                                         ggplot(gerstung_changing_sigs_clonalsubclonal, aes(x=signature,
#                                         y=log2fc_clonalSubclonal, group=signature)
#                                         geom_hline(yintercept = 0, lty='dashed')+
#                                         geom_boxplot()+ geom_jitter(alpha=0.2)+guides(col=FALSE)+theme_bw()+
#                                         theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+ggtitle('Clonal vs subclonal')+
#                                         scale_color_manual(values = pcawg_palette)+labs(x='', y='Log 2 fold-change between clonal
#                                         nrow=2)
ggplot(gerstung_changing_sigs_earlylate, aes(x=signature,
                                               y=log2fc_earlyLate, group=signature,col=tumour_
                                               geom_hline(yintercept = 0, lty='dashed')+
                                               guides(col=FALSE)+theme_bw()+
                                               theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+ggtitle('Early vs late')+
                                               scale_color_manual(values = pcawg_palette)+labs(x='', y='Log 2 fold-change between early and late')

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

```

Log 2 fold-change between early and late clonal mutations

Early vs late

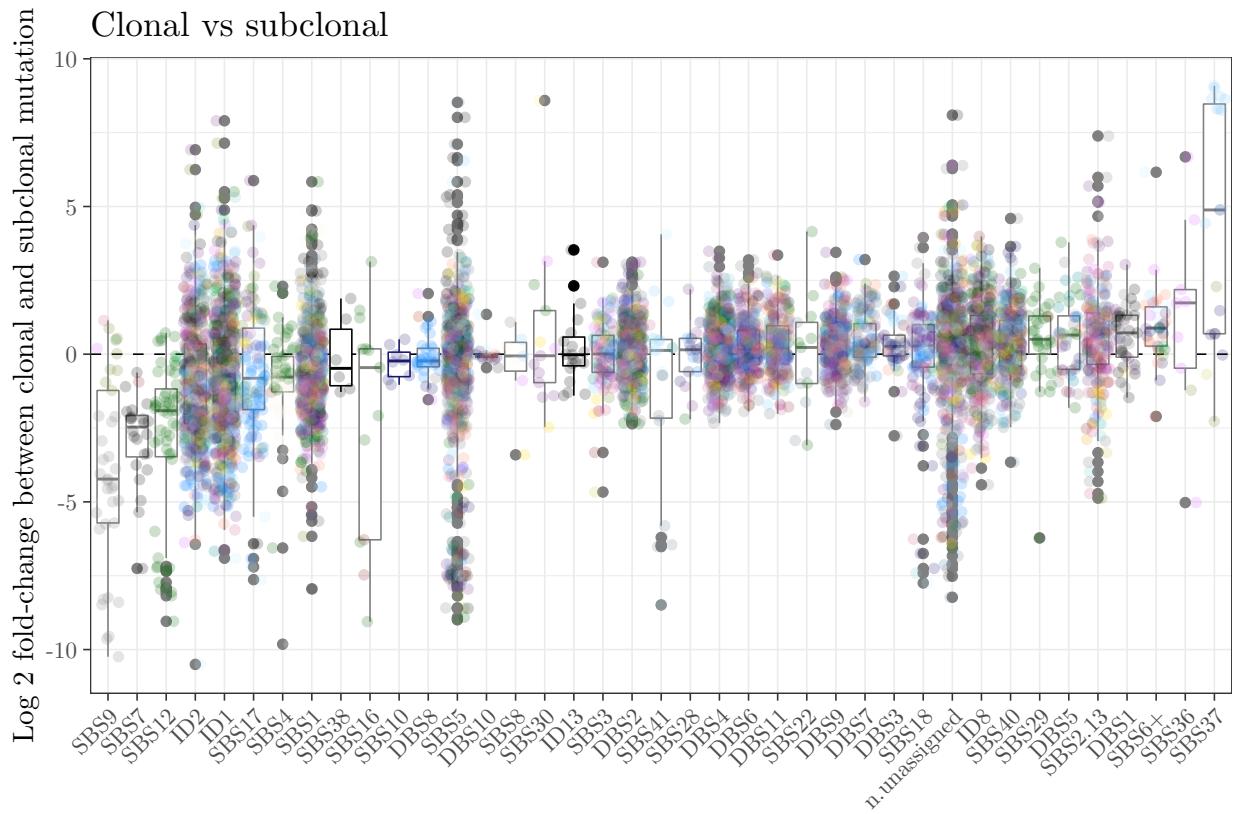


```
ggplot(gerstung_changing_sigs_clonalsubclonal, aes(x=signature,
                                                    y=log2fc_clonalSubclonal, group=signature)
       + geom_hline(yintercept = 0, lty='dashed')+
       geom_boxplot() + geom_jitter(alpha=0.2)+guides(col=FALSE)+theme_bw()+
       theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+ggtitle('Clonal vs subclonal')
       + scale_color_manual(values = pcawg_palette)+labs(x=' ', y='Log 2 fold-change between clonal
```

Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = "none")` instead.

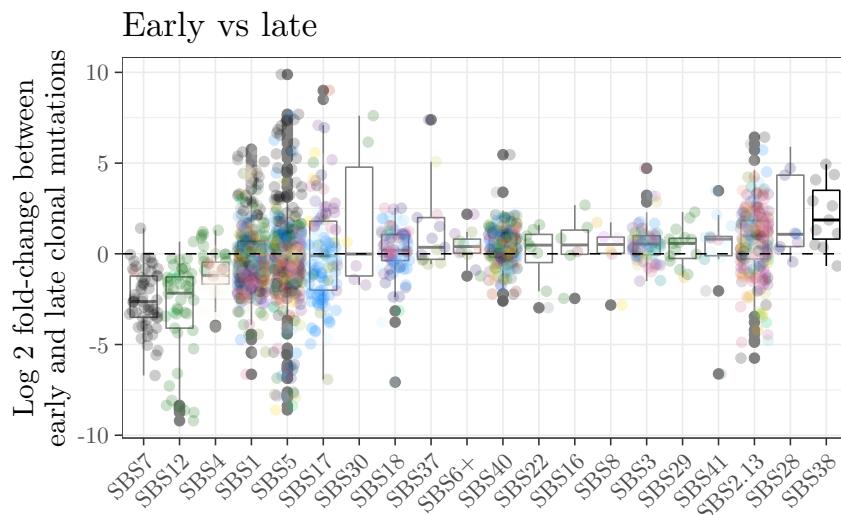
Warning: Removed 38 rows containing non-finite values (stat_boxplot).

Warning: Removed 38 rows containing missing values (geom_point).



```
ggplot(gerstung_changing_sigs_earlylate[grep('SBS', gerstung_changing_sigs_earlylate$signature),], aes(x=tumour, y=log2fc_earlyLate, group=signature, col=tumour))
  geom_hline(yintercept = 0, lty='dashed')+
  guides(col=FALSE)+theme_bw()+
  theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+ggtitle('Early vs late')+
  scale_color_manual(values = pcawg_palette)+labs(x='', y='Log 2 fold-change between \nearly and late mutations')
```

Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = "none")` instead.



```

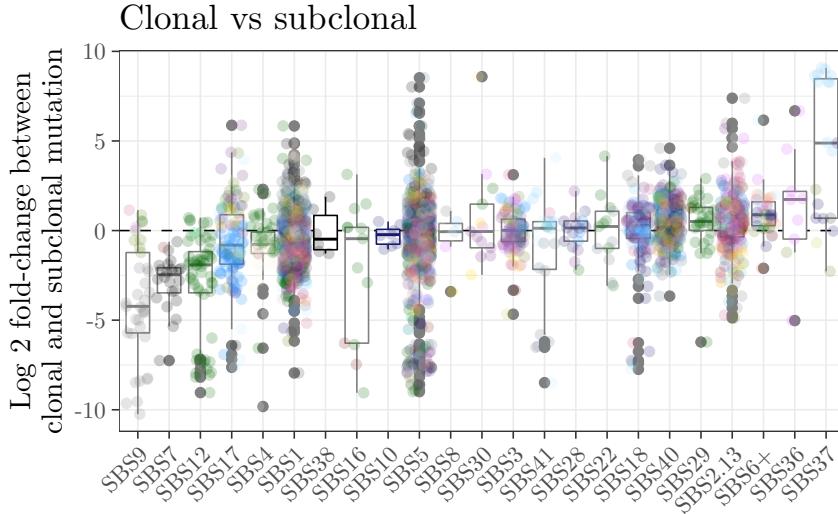
ggplot(gerstung_changing_sigs_clonalsubclonal[grepl('SBS', gerstung_changing_sigs_clonalsubclonal$signature),
                                               y=log2fc_clonalSubclonal, group=signature]
      + geom_hline(yintercept = 0, lty='dashed')+
      geom_boxplot() + geom_jitter(alpha=0.2)+guides(col=FALSE)+theme_bw()+
      theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+ggtitle('Clonal vs subclonal')
      + scale_color_manual(values = pcawg_palette)+labs(x='', y='Log 2 fold-change between \nclonal and subclonal mutation')

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: Removed 13 rows containing non-finite values (stat_boxplot).

## Warning: Removed 13 rows containing missing values (geom_point).

```



```

df_changes_el_persample <- gerstung_changing_sigs_earlylate %>% group_by(samplename) %>% dplyr:::summarise
df_changes_cs_persample <- gerstung_changing_sigs_clonalsubclonal %>% group_by(samplename) %>% dplyr:::summarise
gerstung_changing_sigs_earlylate$samplename <- factor(gerstung_changing_sigs_earlylate$samplename,
                                                       levels=df_changes_el_persample$samplename[order(df_changes_el_persample$n)])
gerstung_changing_sigs_clonalsubclonal$samplename <- factor(gerstung_changing_sigs_clonalsubclonal$samplename,
                                                               levels=df_changes_cs_persample$samplename[order(df_changes_cs_persample$n)])
table(is.na(gerstung_changing_sigs_earlylate$samplename))

```

```

## 
## FALSE
## 5347

```

```
table(is.na(gerstung_changing_sigs_clonalsubclonal$samplename))
```

```

## 
## FALSE
## 7216

```

```

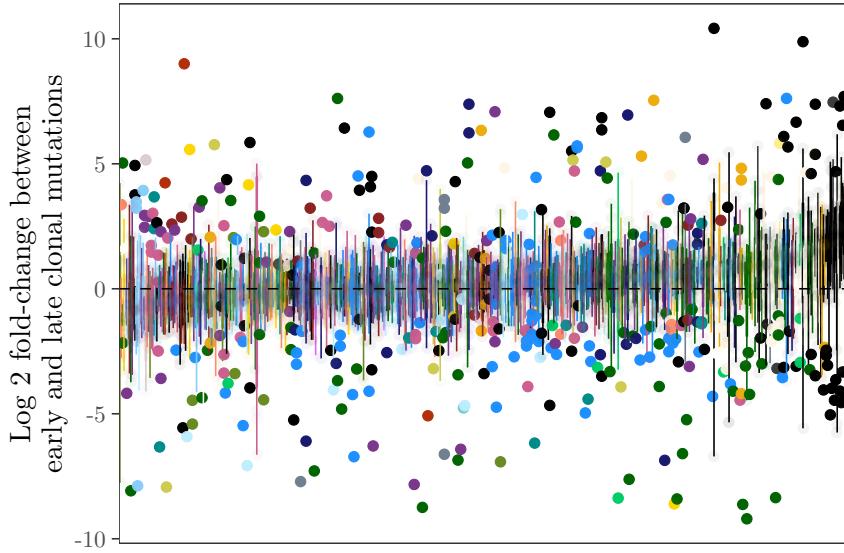
ggplot(gerstung_changing_sigs_earlylate, aes(x=samplename,
                                               y=log2fc_earlyLate, group=samplename,col=tumour_type))+geom_
      + geom_hline(yintercept = 0, lty='dashed')+
      guides(col=FALSE)+theme_bw()+
      theme(axis.title.x=element_blank(),
```

```

    axis.text.x=element_blank(),
    axis.ticks.x=element_blank())+
scale_color_manual(values = pcawg_palette)+labs(x='', y='Log 2 fold-change between\nnearly and late clonal mutations')
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())

```

Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = "none")` instead.



```

ggplot(gerstung_changing_sigs_clonalsubclonal, aes(x=samplename,
                                                    y=log2fc_clonalSubclonal, group=samplename, col=tumour_...
geom_hline(yintercept = 0, lty='dashed')+
geom_boxplot()+
geom_jitter(alpha=0.05)+guides(col=FALSE)+theme_bw()+
theme(axis.title.x=element_blank(),
      axis.text.x=element_blank(),
      axis.ticks.x=element_blank())+ggtitle('Clonal vs subclonal')+
scale_color_manual(values = pcawg_palette)+labs(x='', y='Log 2 fold-change between\nnclonal and subclonal mutations')
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())

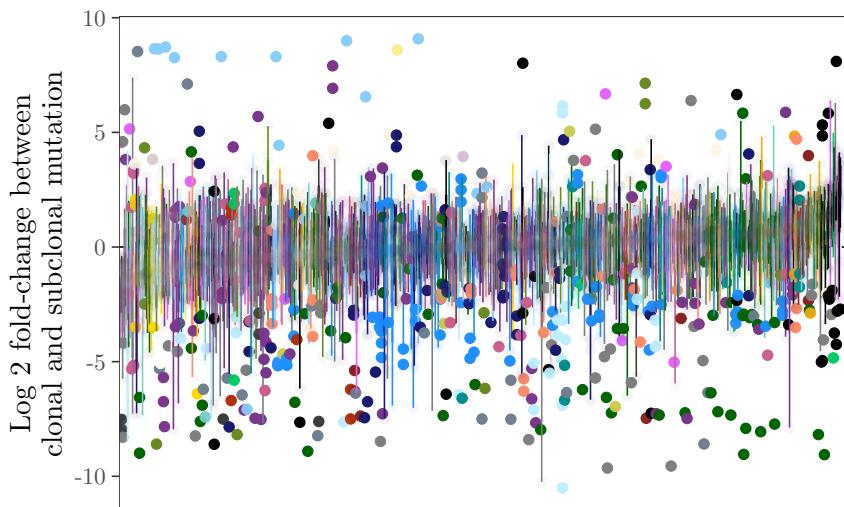
```

Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = "none")` instead.

Warning: Removed 38 rows containing non-finite values (stat_boxplot).

Warning: Removed 38 rows containing missing values (geom_point).

Clonal vs subclonal



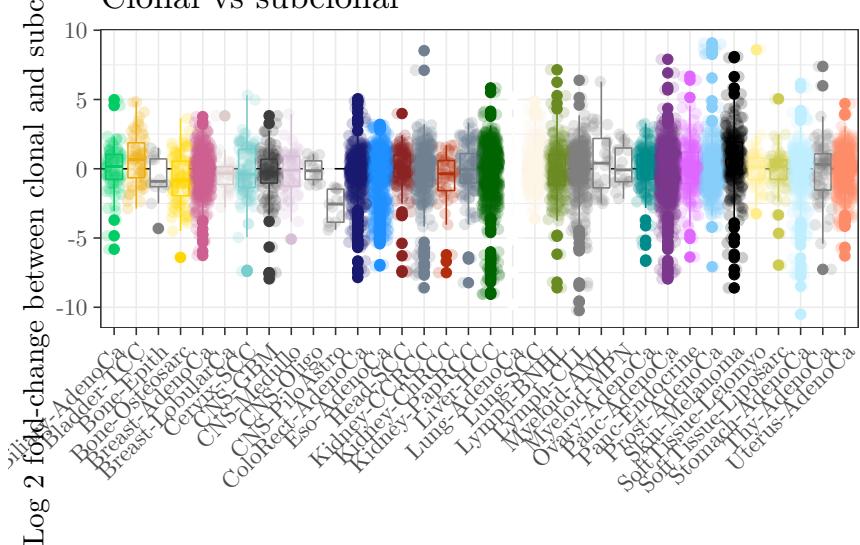
```
ggplot(gerstung_changing_sigs_clonalsubclonal, aes(x=tumour_type,
                                                    y=log2fc_clonalSubclonal, group=tumour_type,col=tumour_type))
  geom_hline(yintercept = 0, lty='dashed')+
  geom_boxplot() + geom_jitter(alpha=0.2)+guides(col=FALSE)+theme_bw()+
  theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+ggtitle('Early vs late')+
  ggtitle('Clonal vs subclonal')+
  scale_color_manual(values = pcawg_palette)+labs(x='', y='Log 2 fold-change between clonal and subclonal')

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = "none")` instead.

## Warning: Removed 38 rows containing non-finite values (stat_boxplot).

## Warning: Removed 38 rows containing missing values (geom_point).
```

Clonal vs subclonal

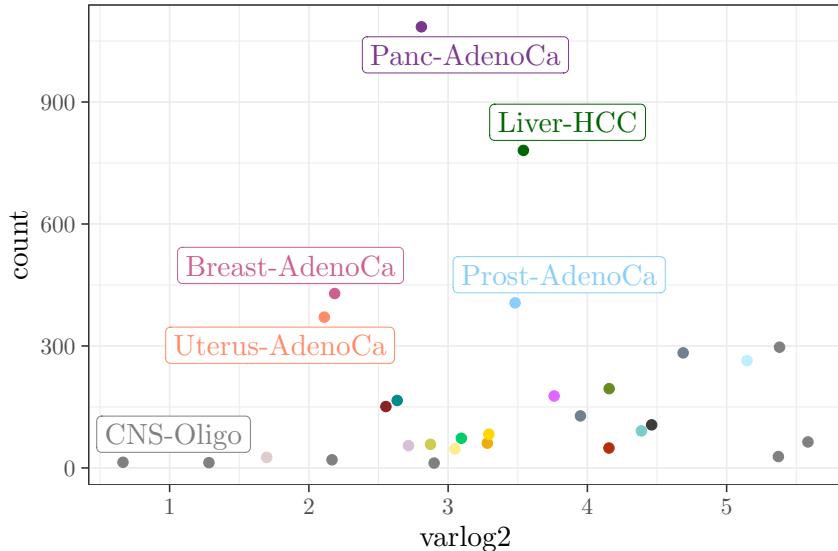


```
gerstung_changing_sigs_clonalsubclonal_var <- gerstung_changing_sigs_clonalsubclonal %>%
  dplyr::group_by(tumour_type) %>% summarise(varlog2=var(log2fc_clonalSubclonal), count=n())
```

```

ggplot(gerstung_changing_sigs_clonalsubclonal_var,
       aes(x=varlog2, y=count, label=tumour_type, col=tumour_type)) + geom_point() +
  scale_color_manual(values = pcawg_palette) + guides(col=FALSE) + theme_bw() + geom_label_repel(max.overlaps =
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
## Warning: Removed 4 rows containing missing values (geom_point).
## Warning: Removed 4 rows containing missing values (geom_label_repel).
## Warning: ggrepel: 24 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```



```

labs(x='Variance of log2 fold change in cancer type', y='Number of observations')

## $x
## [1] "Variance of log2 fold change in cancer type"
##
## $y
## [1] "Number of observations"
##
## attr(,"class")
## [1] "labels"

perturbed_betas_diagRE_DMDL_nonexo_SP_df_summary <- perturbed_betas_diagRE_DMDL_nonexo_SP_df %>% dplyr::group_by(perturbed) %>% summarise(meanperturbed=mean( !(perturbed == 'FALSE')))

perturbed_betas_diagRE_DMDL_nonexo_SP_df_summary

## # A tibble: 32 x 2
##   sig      meanperturbed
##   <chr>          <dbl>
## 1 SBS10a         0
## 2 SBS10b         0
## 3 SBS12          1

```

```

## 4 SBS13      0.25
## 5 SBS14      0.5
## 6 SBS15      0.333
## 7 SBS16      0.5
## 8 SBS17a     0.364
## 9 SBS17b     0.0909
## 10 SBS18     0.3
## # ... with 22 more rows

comparison_with_gerstung_earlylate <- cbind.data.frame(perturbed_betas_diagRE_DMDL_nonexo_SP_df_summary,
                                                       df_changes_el[match(perturbed_betas_diagRE_DM$signature, df_changes_el$signature),]

comparison_with_gerstung_earlylate

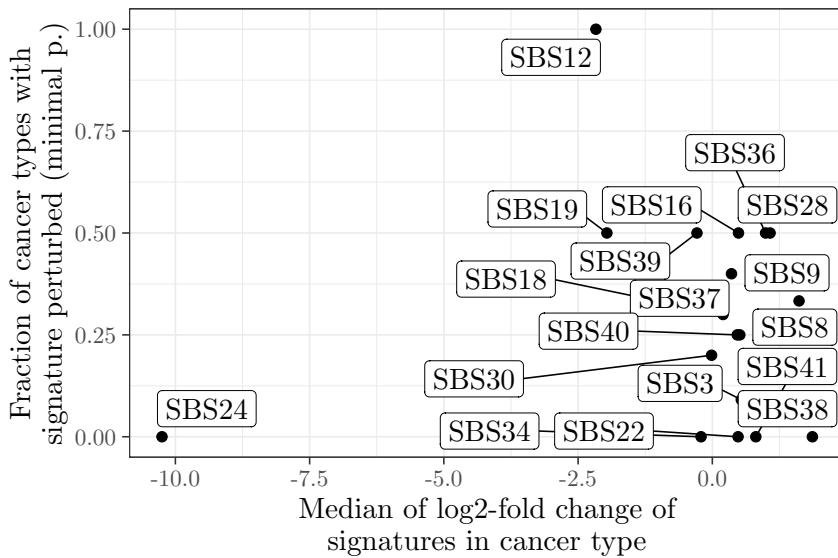
##      sig meanperturbed signature      median
## 1 SBS10a 0.000000000 <NA>        NA
## 2 SBS10b 0.000000000 <NA>        NA
## 3 SBS12  1.000000000 SBS12 -2.16839553
## 4 SBS13  0.250000000 <NA>        NA
## 5 SBS14  0.500000000 <NA>        NA
## 6 SBS15  0.333333333 <NA>        NA
## 7 SBS16  0.500000000 SBS16  0.48743401
## 8 SBS17a 0.36363636 <NA>        NA
## 9 SBS17b 0.09090909 <NA>        NA
## 10 SBS18 0.300000000 SBS18  0.19987446
## 11 SBS19 0.500000000 SBS19 -1.96353701
## 12 SBS2   0.17647059 <NA>        NA
## 13 SBS20 1.000000000 <NA>        NA
## 14 SBS21 0.000000000 <NA>        NA
## 15 SBS22 0.000000000 SBS22  0.47923554
## 16 SBS23 0.000000000 <NA>        NA
## 17 SBS24 0.000000000 SBS24 -10.24985022
## 18 SBS26 0.333333333 <NA>        NA
## 19 SBS28 0.500000000 SBS28  1.07561394
## 20 SBS3   0.09090909 SBS3   0.53915013
## 21 SBS30 0.200000000 SBS30 -0.01377077
## 22 SBS33 0.500000000 <NA>        NA
## 23 SBS34 0.000000000 SBS34 -0.20973256
## 24 SBS36 0.500000000 SBS36  0.99203837
## 25 SBS37 0.400000000 SBS37  0.35693398
## 26 SBS38 0.000000000 SBS38  1.86314214
## 27 SBS39 0.500000000 SBS39 -0.28583208
## 28 SBS40 0.250000000 SBS40  0.46436260
## 29 SBS41 0.000000000 SBS41  0.80987484
## 30 SBS6   0.200000000 <NA>        NA
## 31 SBS8   0.250000000 SBS8   0.51226887
## 32 SBS9   0.333333333 SBS9   1.61493581

comparison_with_gerstung_earlylate$medianlog2fcearlylate = comparison_with_gerstung_earlylate$median
ggplot(comparison_with_gerstung_earlylate, aes(x=medianlog2fcearlylate, y=meanperturbed, label=signature))
  labs(x='Median of log2-fold change of\nsignatures in cancer type', y='Fraction of cancer types with\nsignatures')

## Warning: Removed 14 rows containing missing values (geom_point).

```

```
## Warning: Removed 14 rows containing missing values (geom_label_repel).
```



Barplots of cancer types with and without differential abundance

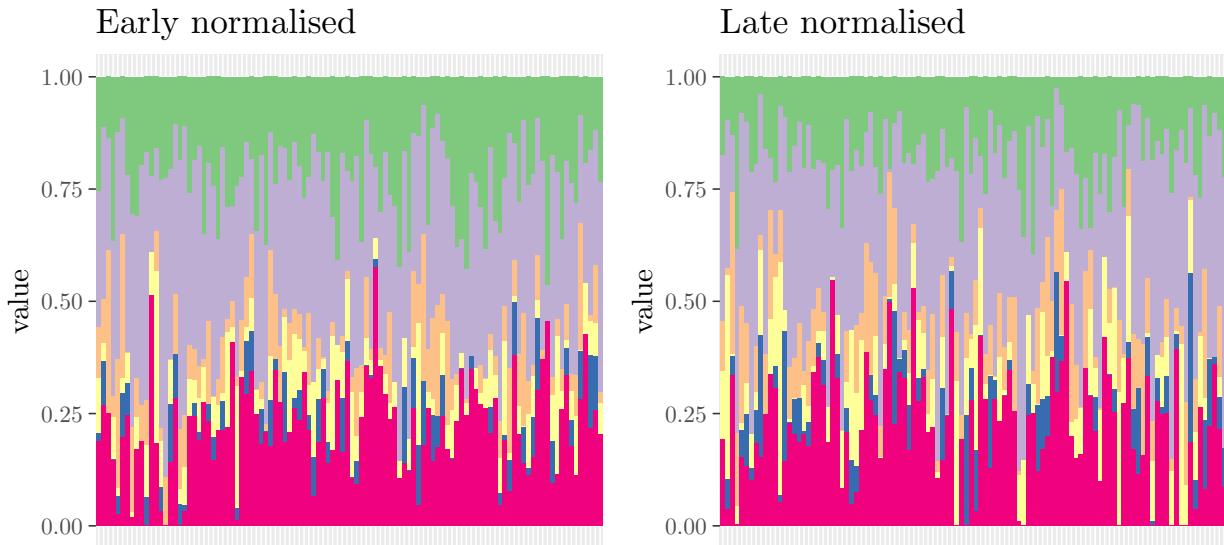
```
give_barplot_from_obj(obj = signatures_PCAWG[['CNS-Medullo']], legend_on = F,
                      nrow=1, verbose=F,
                      only_normalised=T)
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```



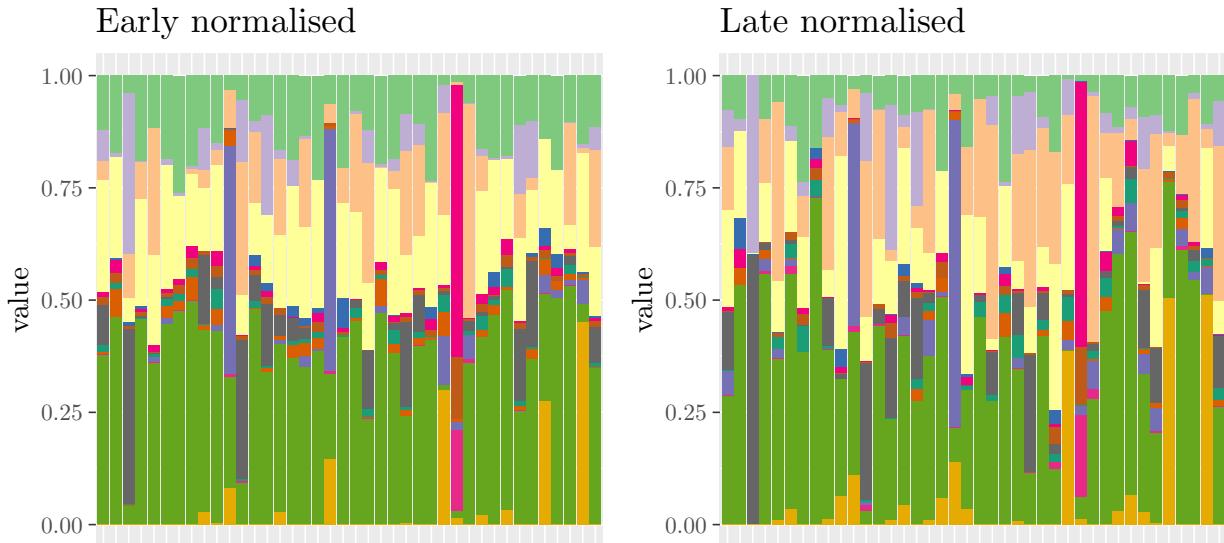
```
give_barplot_from_obj(obj = signatures_PCAWG[['Uterus-AdenoCA']], legend_on = F,
                      nrow=1, verbose=F,
                      only_normalised=T)
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```



```

give_barplot_from_obj(obj = signatures_PCAWG[['Ovary-AdenoCA']], legend_on = F,
                      nrow=1, verbose=F,
                      only_normalised=T)

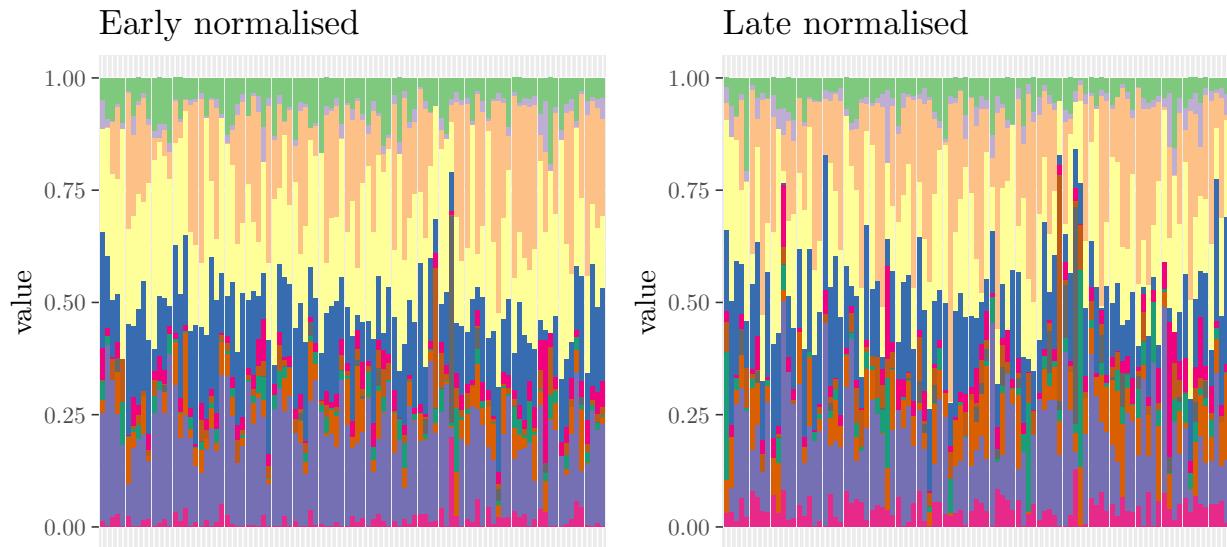
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

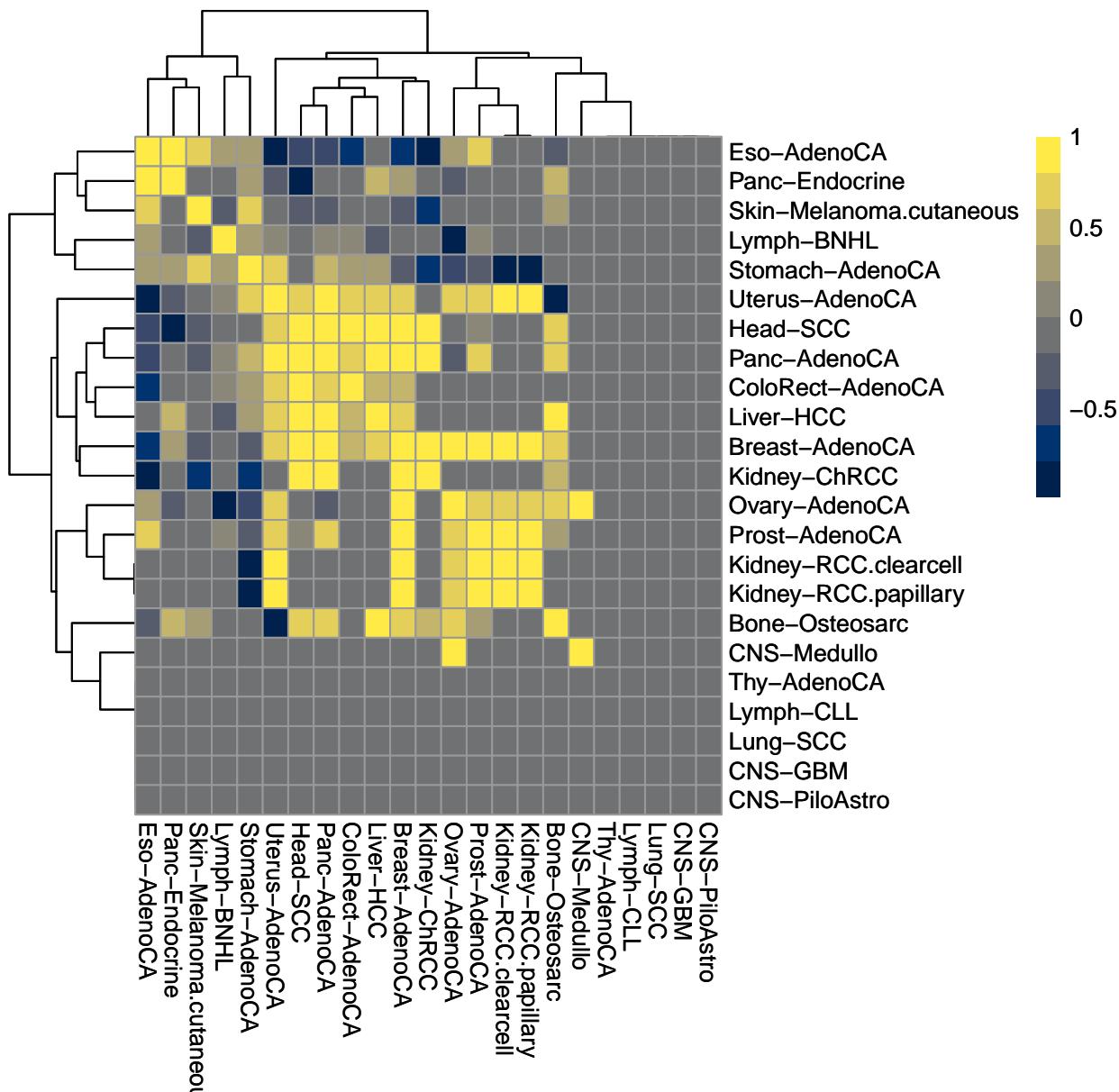
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.

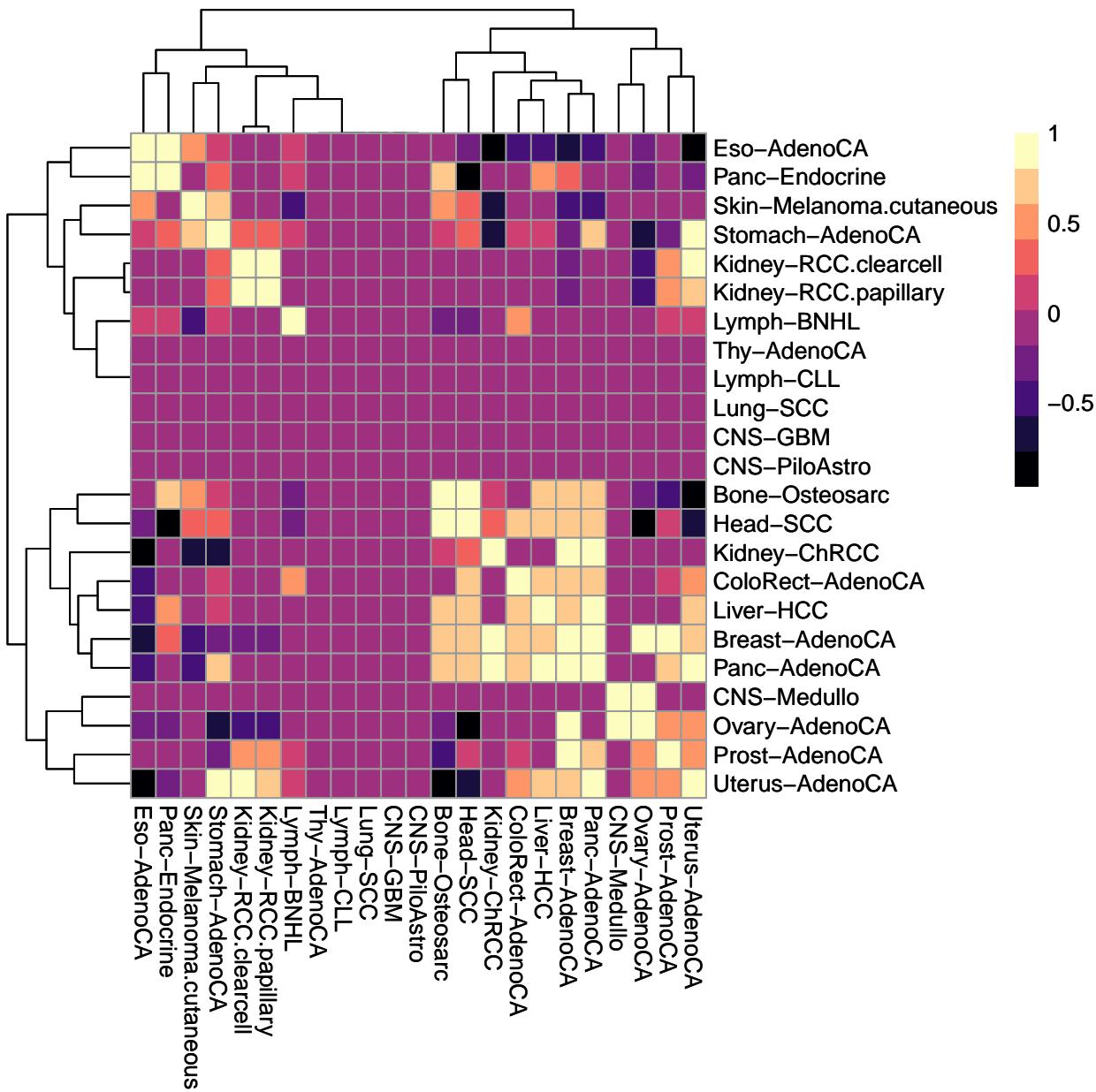
```



Correlations of cancer types and of signatures based on betas



```
## null device
##      1
## null device
##      1
```



```

## null device
##      1

## null device
##      1

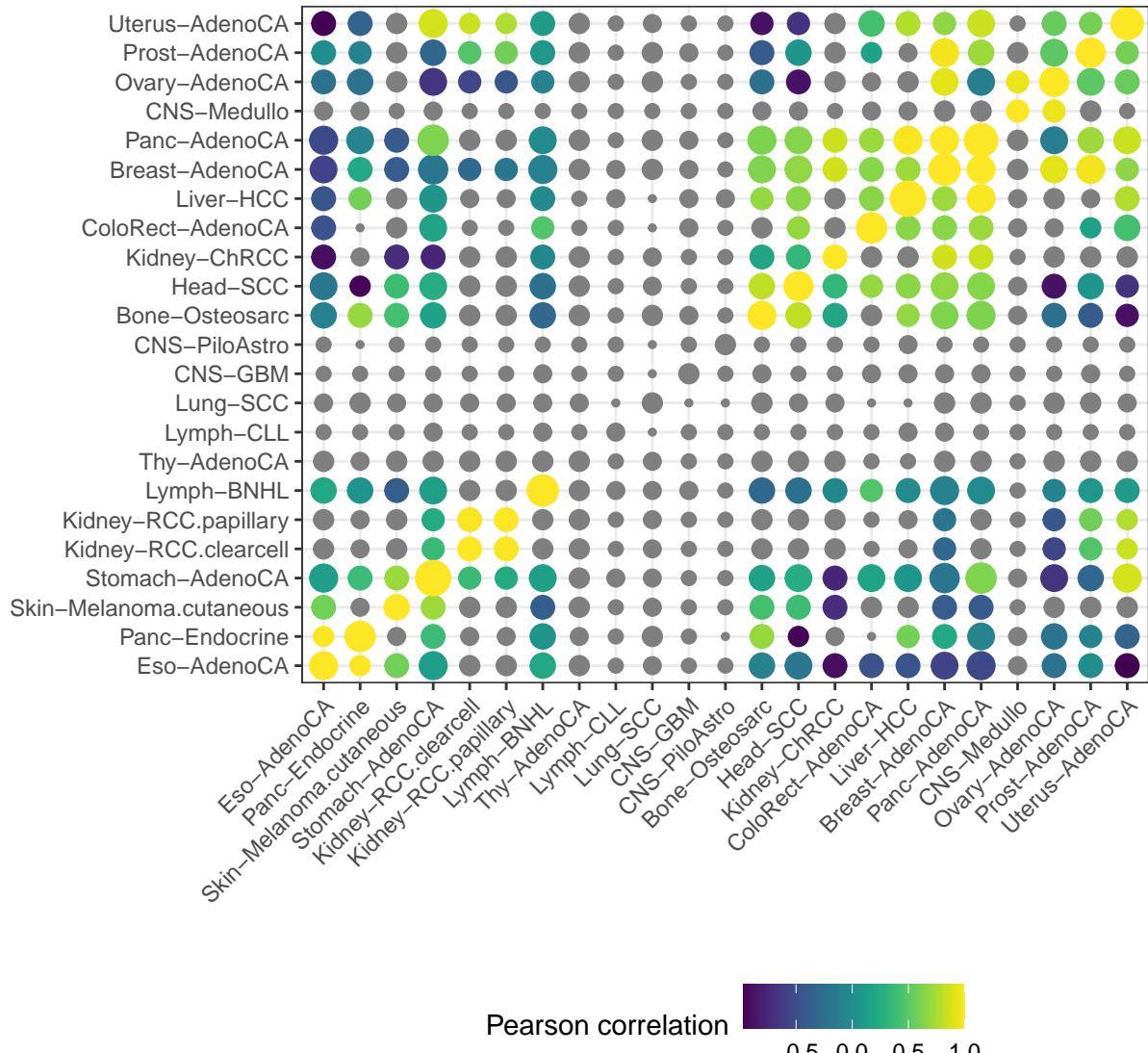
## cors_softmax.Var1 cors_softmax.Var2 cors_softmax.value num_common_sigs.Var1
## 1          1          1 1.0000000000000000 Bone-Osteosarc
## 2          2          1 0.6269273000000000 Breast-AdenoCA
## 3          3          1          NA CNS-GBM
## 4          4          1          NA CNS-Medullo
## 5          5          1          NA CNS-PiloAstro
## 6          6          1          NA ColoRect-AdenoCA

```

```

##   num_common_sigs.Var2 num_common_sigs.value
## 1      Bone-Osteosarc          8
## 2      Bone-Osteosarc          7
## 3      Bone-Osteosarc          2
## 4      Bone-Osteosarc          2
## 5      Bone-Osteosarc          1
## 6      Bone-Osteosarc          3

```

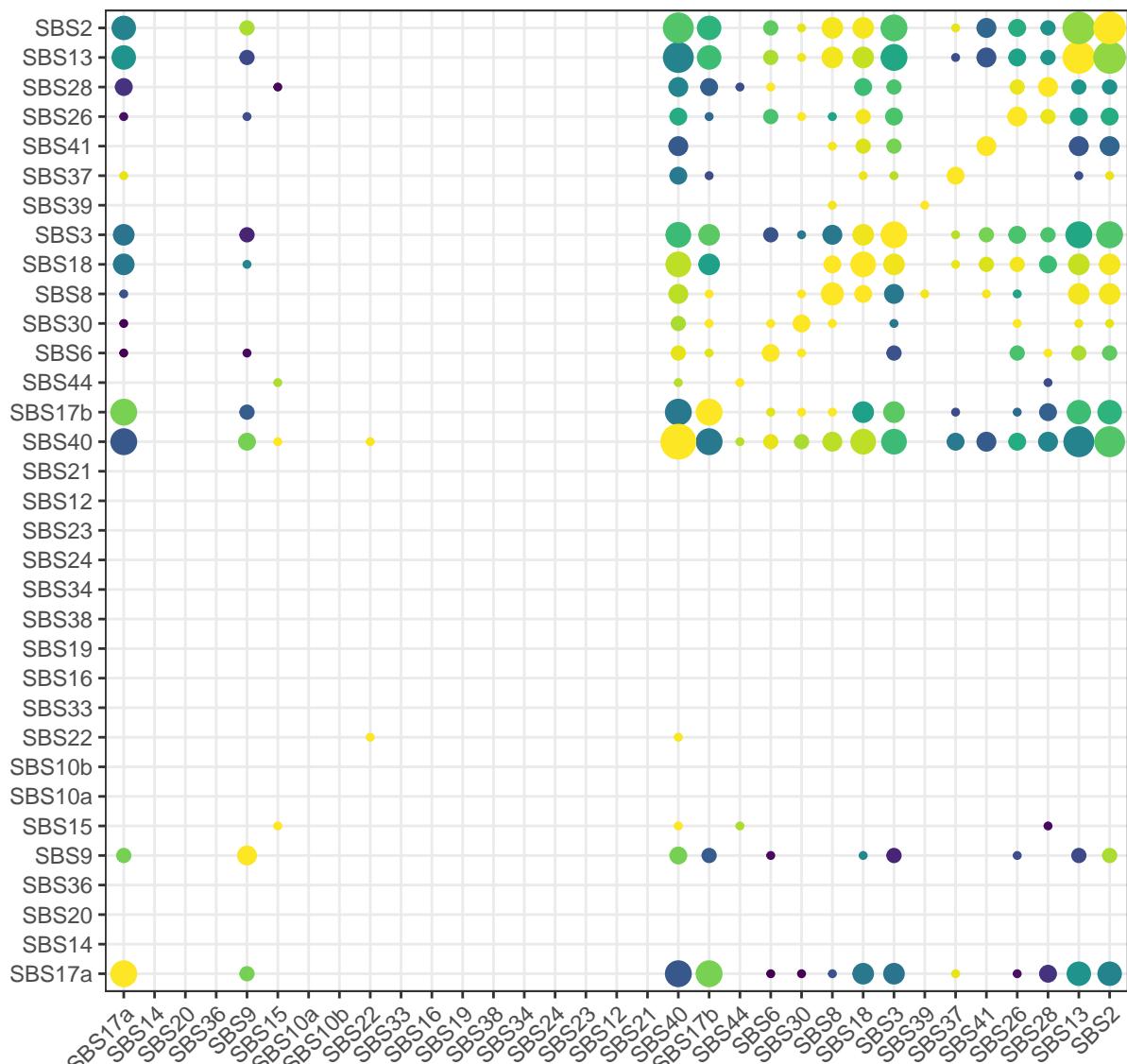


```

## Saving 6.5 x 6.5 in image
##   cors_softmax.Var1 cors_softmax.Var2 cors_softmax.value
## 1      SBS10a        SBS10a           1
## 2      SBS10b        SBS10a           1
## 3      SBS12         SBS10a          NA

```

```
## 4      SBS13      SBS10a      NA
## 5      SBS14      SBS10a      NA
## 6      SBS15      SBS10a      1
## num_common_samples.Var1 num_common_samples.Var2 num_common_samples.value
## 1      SBS10a      SBS10a      NA
## 2      SBS10b      SBS10a      NA
## 3      SBS12      SBS10a      NA
## 4      SBS13      SBS10a      NA
## 5      SBS14      SBS10a      NA
## 6      SBS15      SBS10a      NA
## Warning: Removed 890 rows containing missing values (geom_point).
```



Number of signatures in common ● 5 ● 10 ● 15 ● 20

```
# pcawg_palette <- pcawg.colour.palette(gsub("\\..*", "", all_pvals$ct),
#                                         scheme = "tumour.subtype")
#
# names(pcawg_palette) <- all_pvals$ct
#
# pcawg_palette
pcawg_palette[names(pcawg_palette) == 'Lung-SCC'] <- '#ffff29'
```

```
pcawg_palette
```

```
##      Biliary-AdenoCa      Bladder-TCC      Bone-Osteosarc      Breast-AdenoCa
## "#00CD66" "#EAD0E" "#FFD700" "#CD6090"
## Breast-LobularCa      Cervix-SCC      CNS-GBM      CNS-Medullo
## "#DDCDCD" "#79CDCD" "#3D3D3D" "#D8BFD8"
## ColoRect-AdenoCa      Eso-AdenoCa      Head-SCC      Kidney-CCRCC
## "#191970" "#1E90FF" "#8B2323" "slategrey"
## Kidney-ChRCC      Kidney-PapRCC      Liver-HCC      Lung-AdenoCa
## "#B32F0B" "slategrey" "#006400" "#FFFFFF"
## Lung-SCC      Lymph-BNHL      Ovary-AdenoCa      Panc-AdenoCa
## "#ffff29" "#698B22" "#008B8B" "#7A378B"
## Panc-Endocrine      Prost-AdenoCa      Skin-Melanoma  SoftTissue-Leiomyo
## "#E066FF" "#87CEFA" "#000000" "#FFEC8B"
## SoftTissue-Liposarc      Stomach-AdenoCa      Uterus-AdenoCa
## "#CDCB50" "#BFEFFF" "#FF8C69"
```

```
df_cor_pvals_n <- data.frame(num_samples=num_samples_all_SP,
                               pvalue=pvals_diagRE_DMDL_SP,
                               ct=gsub("CA$", "Ca", gsub("\\"..*", "", enough_samples)))
df_cor_pvals_n
```

	num_samples	pvalue	ct
## Bone-Osteosarc	27	1.080828e-04	Bone-Osteosarc
## Breast-AdenoCA	136	2.239756e-28	Breast-AdenoCa
## CNS-GBM	34	3.390137e-03	CNS-GBM
## CNS-Medullo	106	8.431463e-03	CNS-Medullo
## CNS-PiloAstro	42	5.615238e-04	CNS-PiloAstro
## ColoRect-AdenoCA	37	6.356131e-26	ColoRect-AdenoCa
## Eso-AdenoCA	65	5.329093e-21	Eso-AdenoCa
## Head-SCC	32	4.975610e-05	Head-SCC
## Kidney-ChRCC	38	1.562125e-09	Kidney-ChRCC
## Kidney-RCC.clearcell	86	4.027485e-18	Kidney-RCC
## Kidney-RCC.papillary	30	NA	Kidney-RCC
## Liver-HCC	207	4.747822e-107	Liver-HCC
## Lung-SCC	34	7.747310e-22	Lung-SCC
## Lymph-BNHL	51	3.908637e-19	Lymph-BNHL
## Lymph-CLL	53	6.611927e-20	Lymph-CLL
## Ovary-AdenoCA	97	8.965185e-38	Ovary-AdenoCa
## Panc-AdenoCA	193	4.096402e-119	Panc-AdenoCa
## Panc-Endocrine	70	3.987099e-10	Panc-Endocrine
## Prost-AdenoCA	208	6.474116e-99	Prost-AdenoCa
## Skin-Melanoma.cutaneous	30	9.272113e-25	Skin-Melanoma
## Stomach-AdenoCA	30	1.715150e-06	Stomach-AdenoCa
## Thy-AdenoCA	41	8.821583e-06	Thy-AdenoCa
## Uterus-AdenoCA	40	4.819867e-10	Uterus-AdenoCa

```
ggplot(df_cor_pvals_n, aes(x=num_samples, y=pvalue, label=ct, col=ct, group=1))+  
  geom_point() + scale_y_continuous(trans = "log2") +  
  geom_smooth(method = lm) + theme_bw() +  
  geom_label_repel() + geom_hline(yintercept = log2(0.05), lty='dashed') +  
  labs(x='Number of samples', y = 'p-value (log2)')
```

```

scale_color_manual(values = pcawg_palette)+  

guides(col=FALSE)

## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  

## "none")` instead.

## Warning in log(x, base): NaNs produced

## Warning: Transformation introduced infinite values in continuous y-axis

## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (stat_smooth).

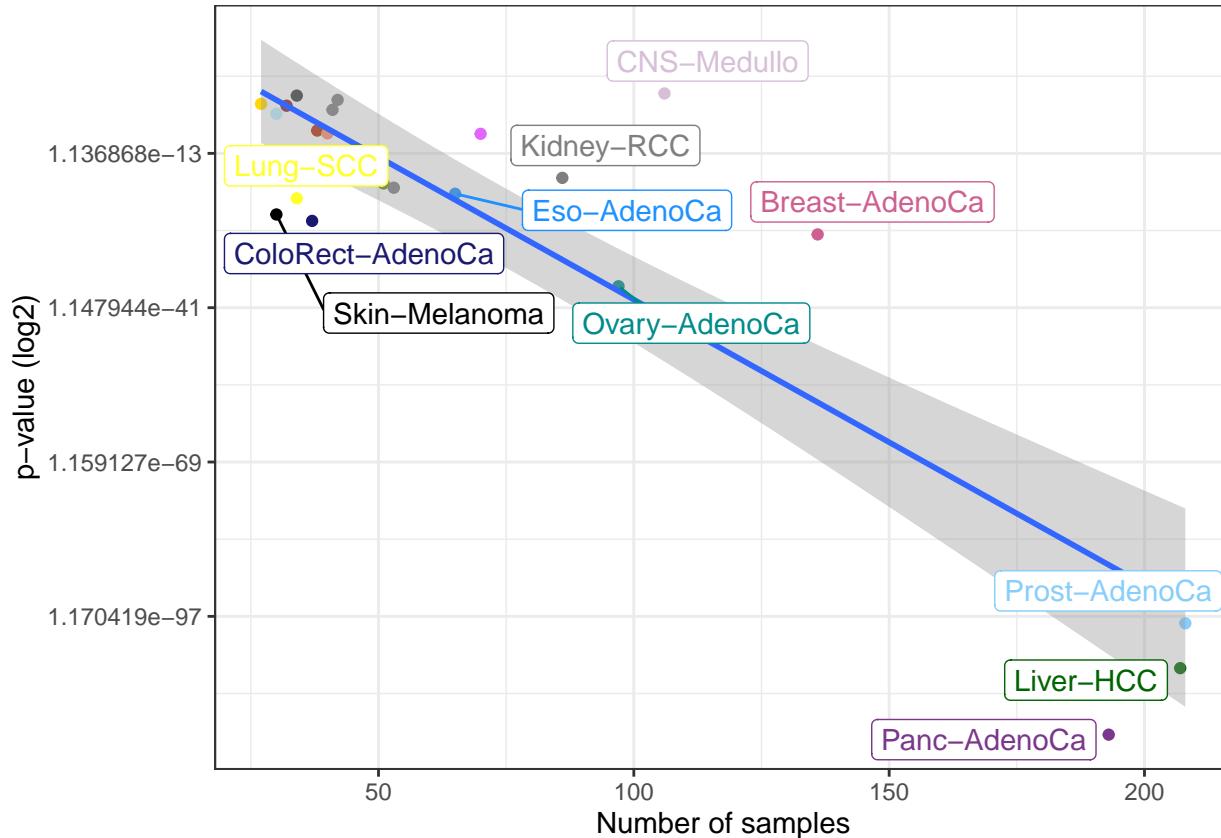
## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_label_repel).

## Warning: Removed 1 rows containing missing values (geom_hline).

## Warning: ggrepel: 11 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

```



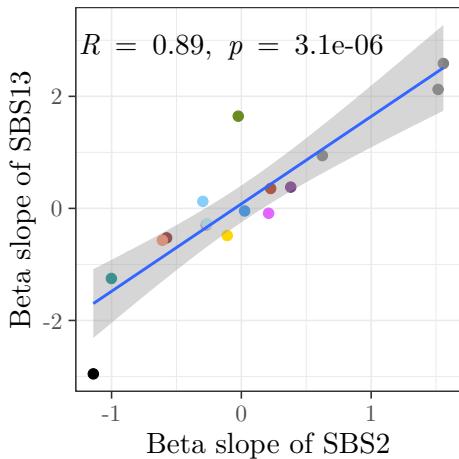
	Estimate numerator_LogR	ct
## beta.1	-0.10868535	2 Bone-Osteosarc
## beta.7	-0.48518518	13 Bone-Osteosarc
## beta.12	-0.27064004	2 Breast-AdenoCA
## beta.91	-0.29393557	13 Breast-AdenoCA
## beta.113	0.02518066	2 Eso-AdenoCA

```

## beta.54 -0.04596199 13 Eso-AdenoCA
## beta.115 -0.57582551 2 Head-SCC
## beta.55 -0.52635385 13 Head-SCC
## beta.117 0.22566046 2 Kidney-ChRCC
## beta.38 0.35508274 13 Kidney-ChRCC
## beta.118 1.51557028 2 Kidney-RCC.clearcell
## beta.39 2.12303828 13 Kidney-RCC.clearcell
## beta.119 1.55643670 2 Kidney-RCC.papillary
## beta.310 2.58590299 13 Kidney-RCC.papillary
## beta.121 0.05062637 2 Lung-SCC
## beta.122 -0.02400179 2 Lymph-BNHL
## beta.96 1.64560251 13 Lymph-BNHL
## beta.124 -1.00151692 2 Ovary-AdenoCA
## beta.710 -1.25120153 13 Ovary-AdenoCA
## beta.125 0.38030698 2 Panc-AdenoCA
## beta.98 0.37995820 13 Panc-AdenoCA
## beta.126 0.20999050 2 Panc-Endocrine
## beta.1114 -0.08708674 13 Panc-Endocrine
## beta.127 -0.29675703 2 Prost-AdenoCA
## beta.713 0.12467277 13 Prost-AdenoCA
## beta.128 -1.14063355 2 Skin-Melanoma.cutaneous
## beta.318 -2.95346899 13 Skin-Melanoma.cutaneous
## beta.129 -0.26552047 2 Stomach-AdenoCA
## beta.715 -0.27417496 13 Stomach-AdenoCA
## beta.130 0.62364260 2 Thy-AdenoCA
## beta.320 0.94221489 13 Thy-AdenoCA
## beta.140 -0.60872957 2 Uterus-AdenoCA
## beta.1117 -0.56888068 13 Uterus-AdenoCA

## `geom_smooth()` using formula 'y ~ x'

```

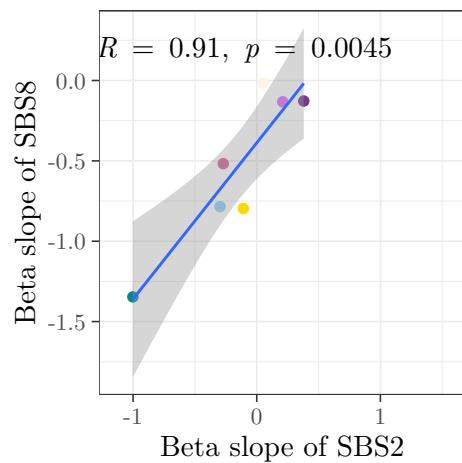


Correlations of beta slopes of SBS8 and APOBEC signatures

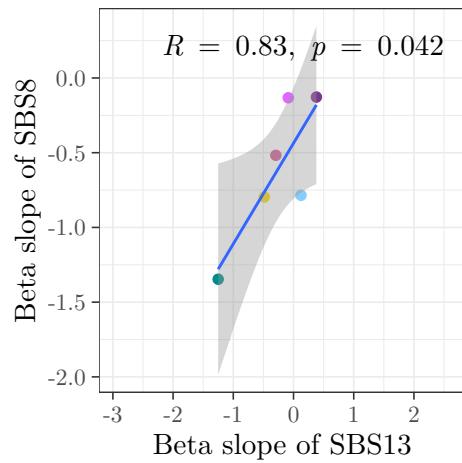
```

## `geom_smooth()` using formula 'y ~ x'

```

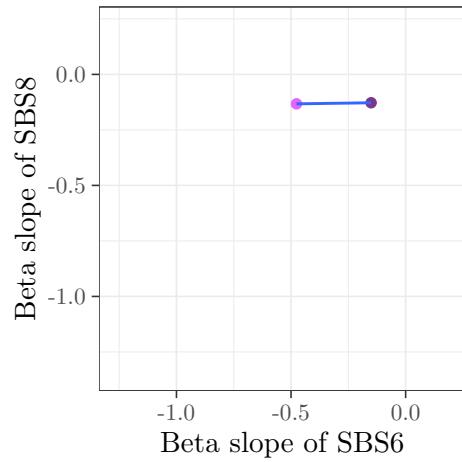


```
## `geom_smooth()` using formula 'y ~ x'
```

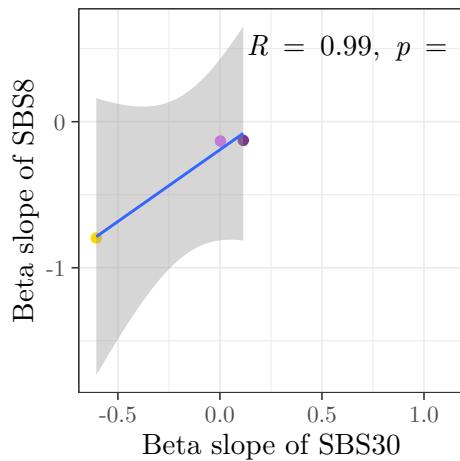


Correlations of beta slopes of SBS8 with SBS6, SBS18, SBS30

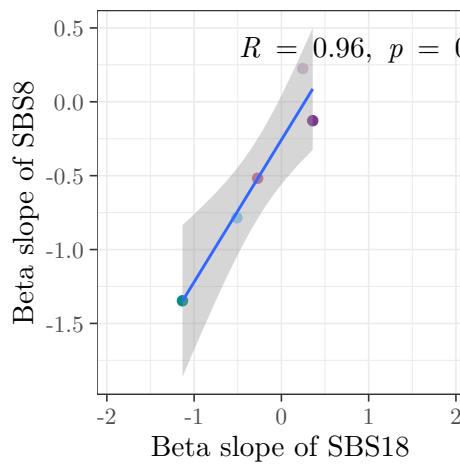
```
## `geom_smooth()` using formula 'y ~ x'
```



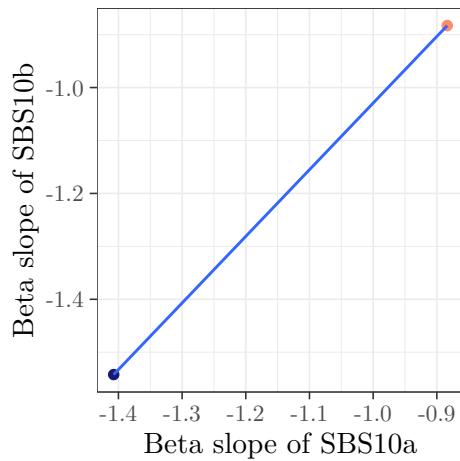
```
## `geom_smooth()` using formula 'y ~ x'
```



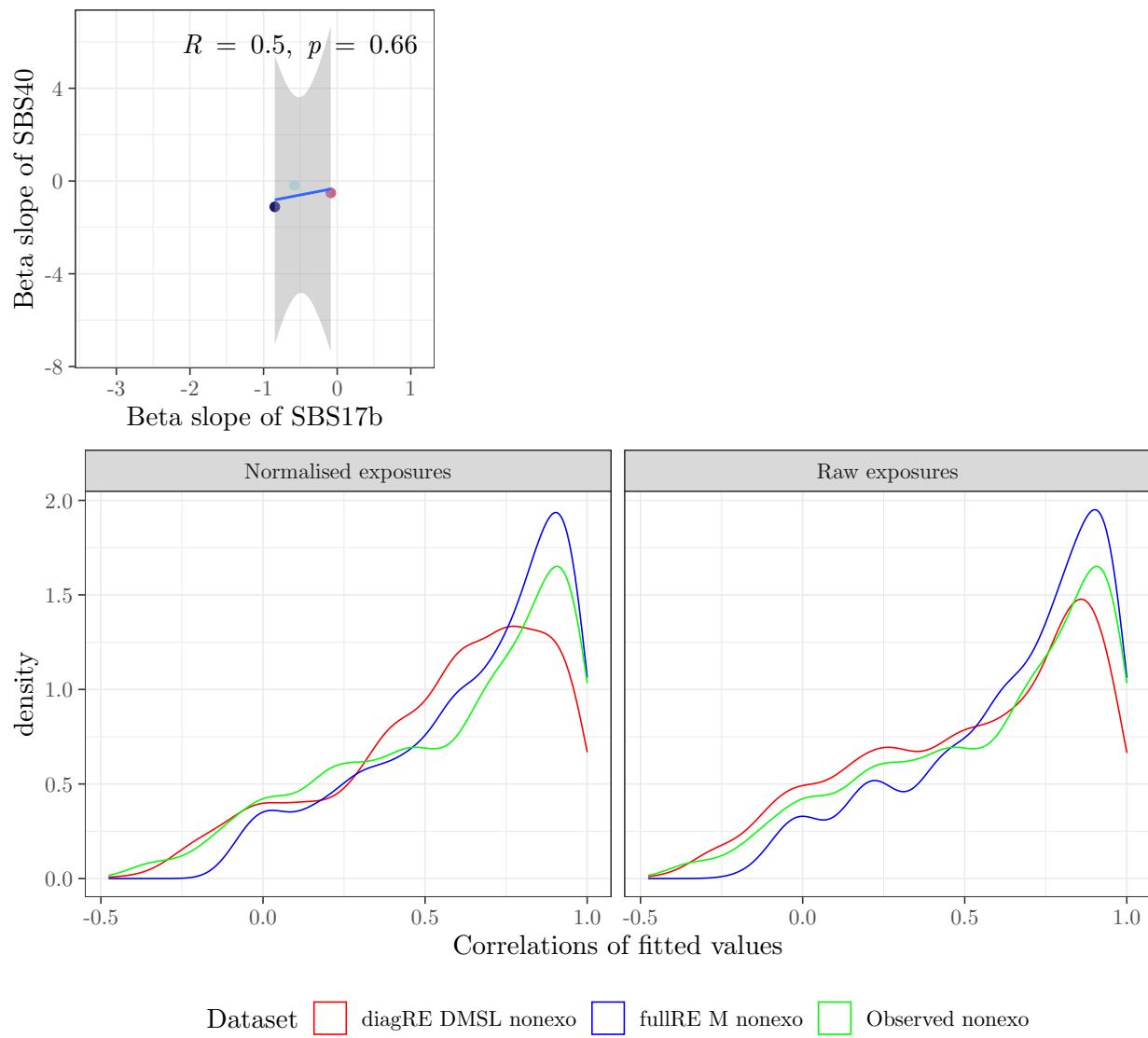
```
## `geom_smooth()` using formula 'y ~ x'
```

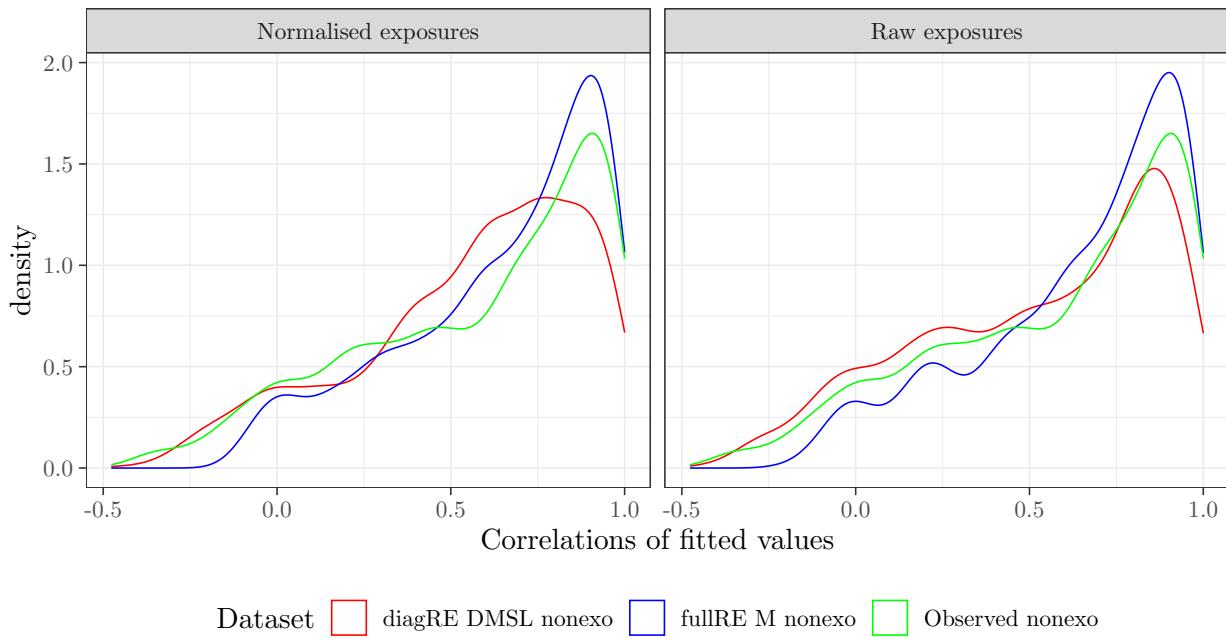


```
## `geom_smooth()` using formula 'y ~ x'
```



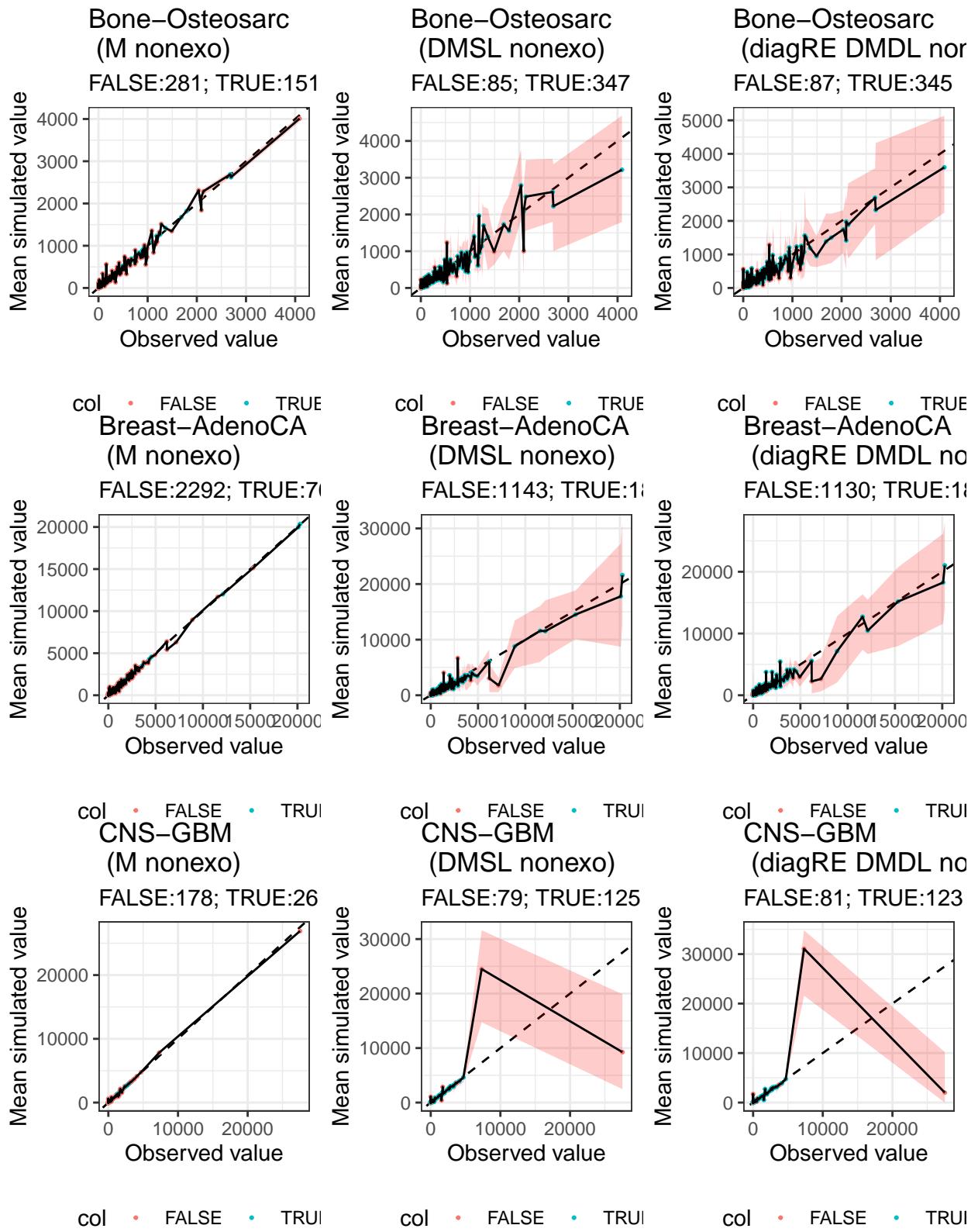
```
## `geom_smooth()` using formula 'y ~ x'
```

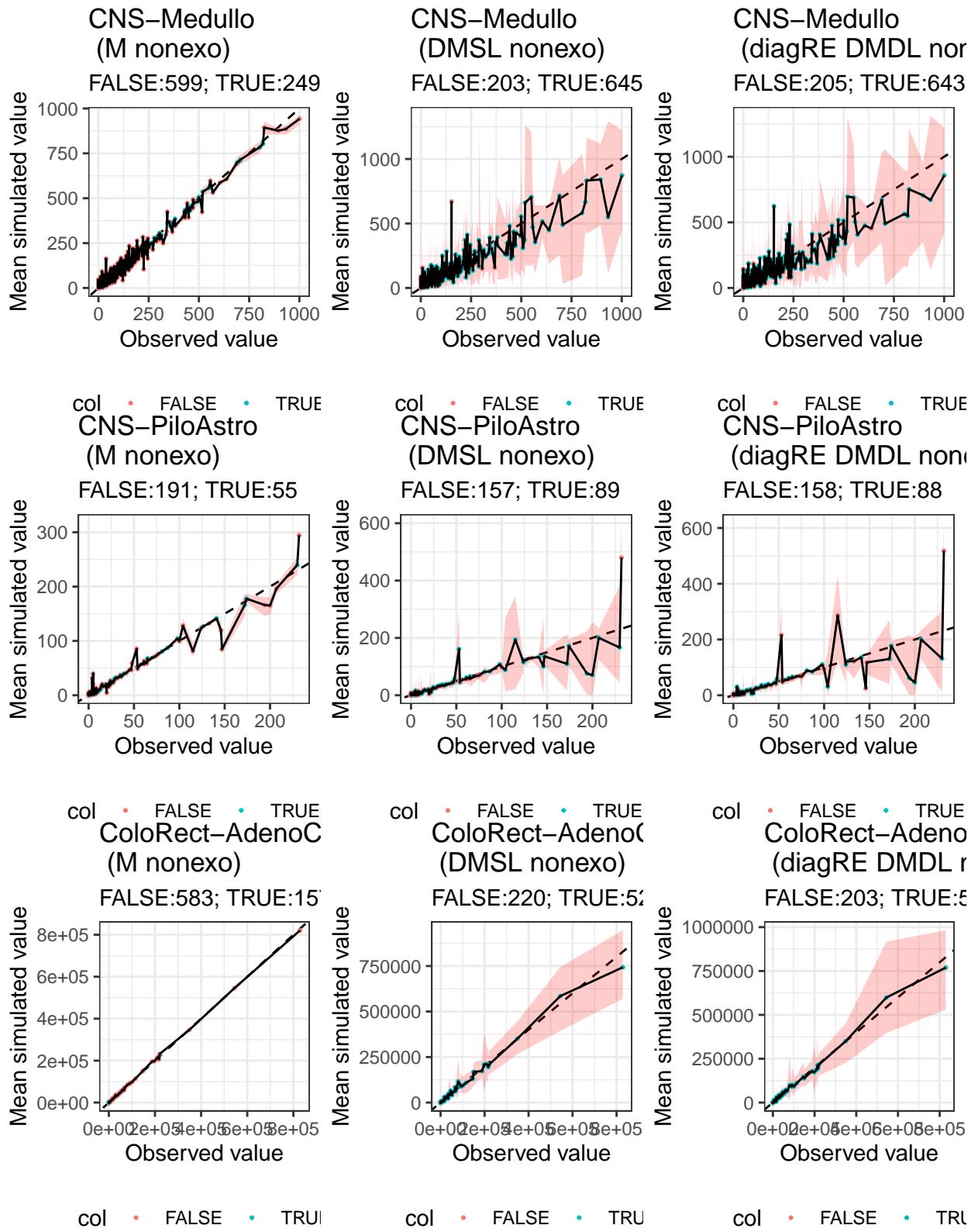


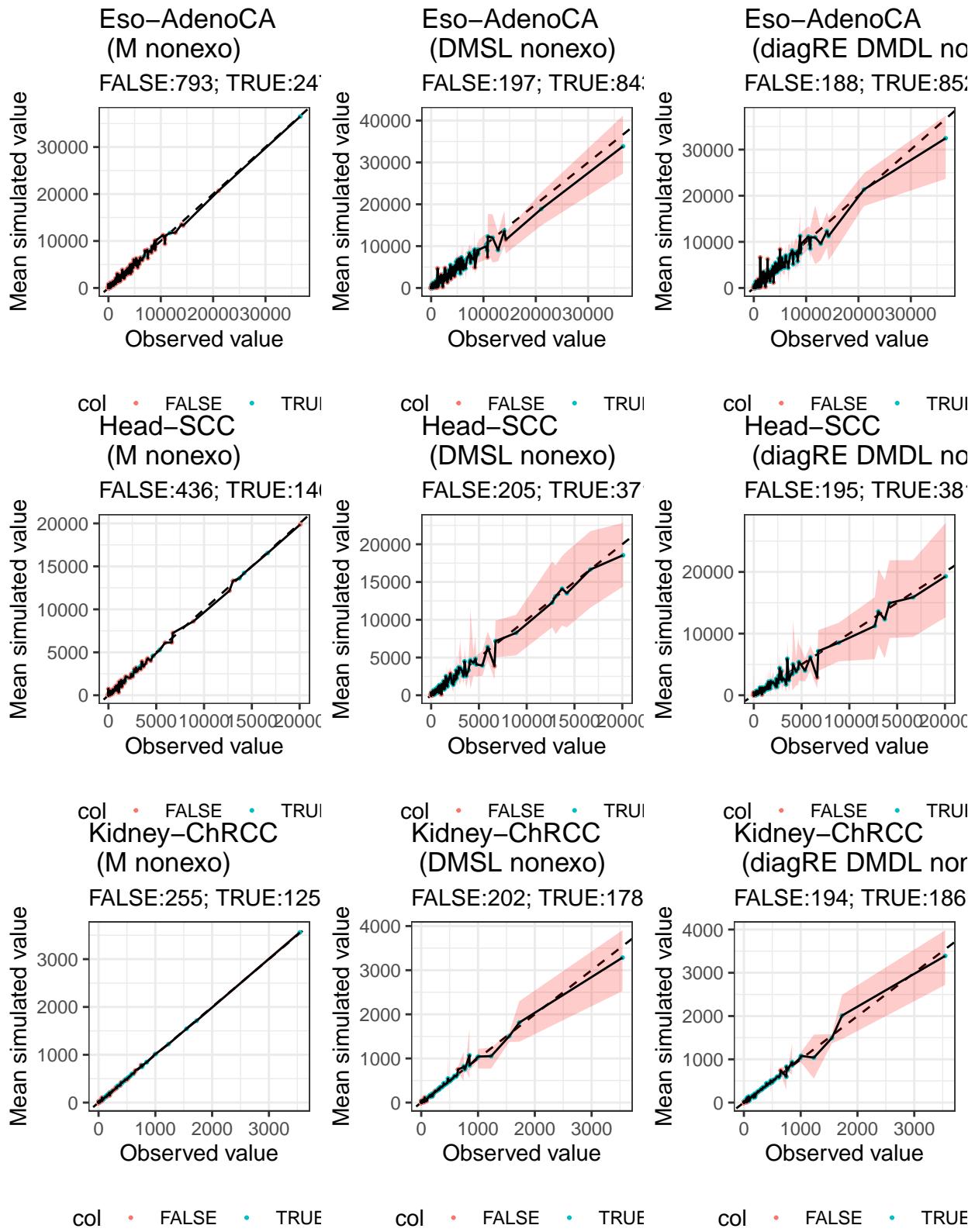


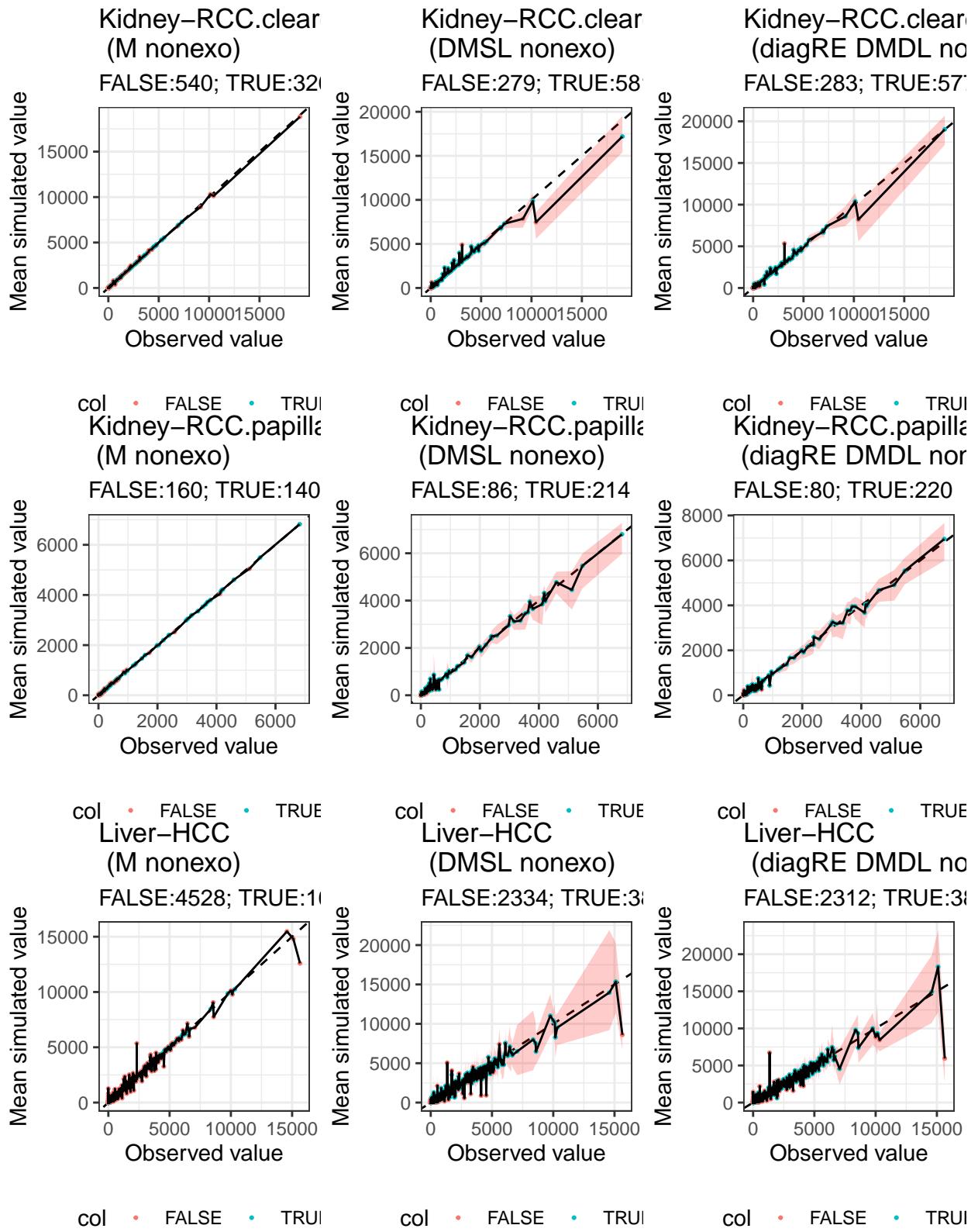
Ranked plots

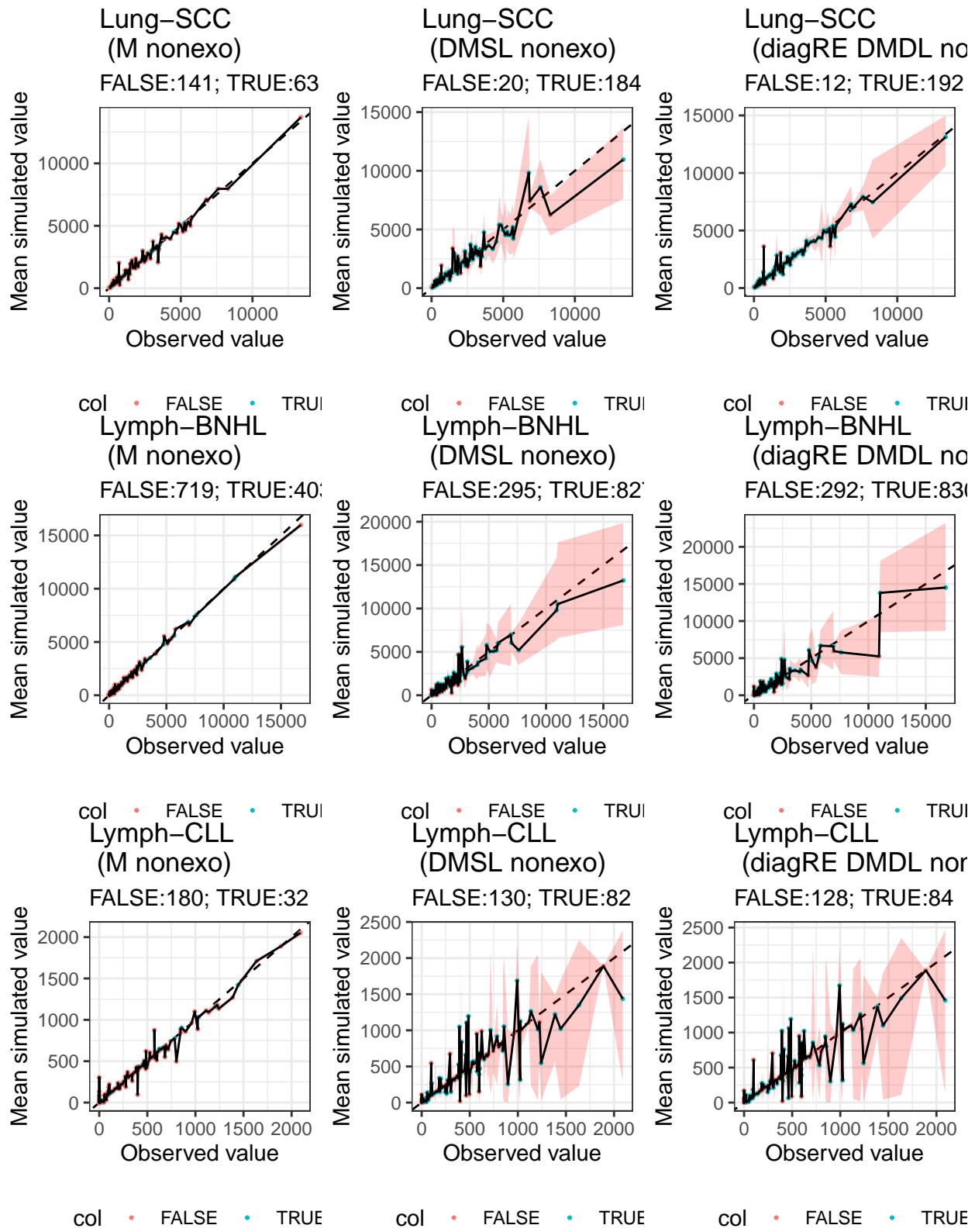
```
for(ct in enough_samples){
  integer_overdispersion_param_DMSL <- 1
  grid.arrange(give_interval_plots_2(df_rank = lapply(list(give_ranked_plot_simulation(tmb_fit_object =
    data_object = all_objects_nonexo_SP[[ct]],
    print_plot = F, nreps = 20, model = "M")),
    function(i){
      lapply(i, function(j) cbind.data.frame(sorted_value=as.vector(j),
                                             rank_number=1:length(j)))[[1]],
      data_object = all_objects_nonexo_SP[[ct]],
      loglog = F, title = paste0(ct, '\n (M nonexo)'), 
    give_interval_plots_2(df_rank = lapply(list(give_ranked_plot_simulation(tmb_fit_object = fullRE_DMSL_r,
      data_object = all_objects_nonexo_SP[[ct]],
      print_plot = F, nreps = 20, model = "DMSL", integer_overdispersion_param = integer_overdispersion_param_DMSL),
      function(i){
        lapply(i, function(j) cbind.data.frame(sorted_value=as.vector(j),
                                               rank_number=1:length(j)))[[1]],
        data_object = all_objects_nonexo_SP[[ct]],
        loglog = F, title = paste0(ct, '\n (DMSL nonexo)'), 
      give_interval_plots_2(df_rank = lapply(list(give_ranked_plot_simulation(tmb_fit_object = diagRE_DMDL_r,
        data_object = all_objects_nonexo_SP[[ct]],
        print_plot = F, nreps = 20, model = "DM", integer_overdispersion_param = 1000)),
        function(i){
          lapply(i, function(j) cbind.data.frame(sorted_value=as.vector(j),
                                                 rank_number=1:length(j)))[[1]],
          data_object = all_objects_nonexo_SP[[ct]],
          loglog = F, title = paste0(ct, '\n (diagRE DMDL nonexo)'), ncol=3
        }
      })
    })
  })
```

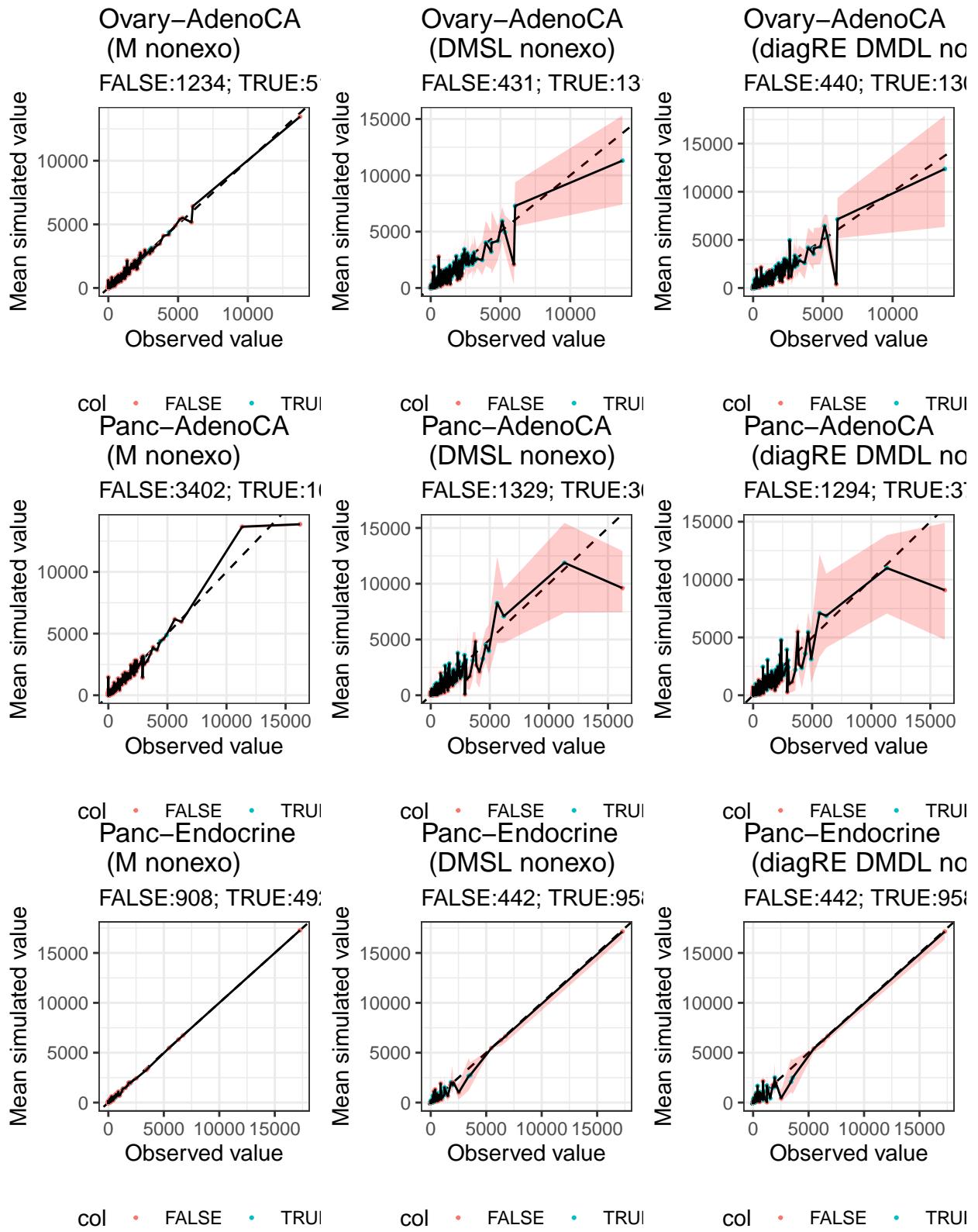


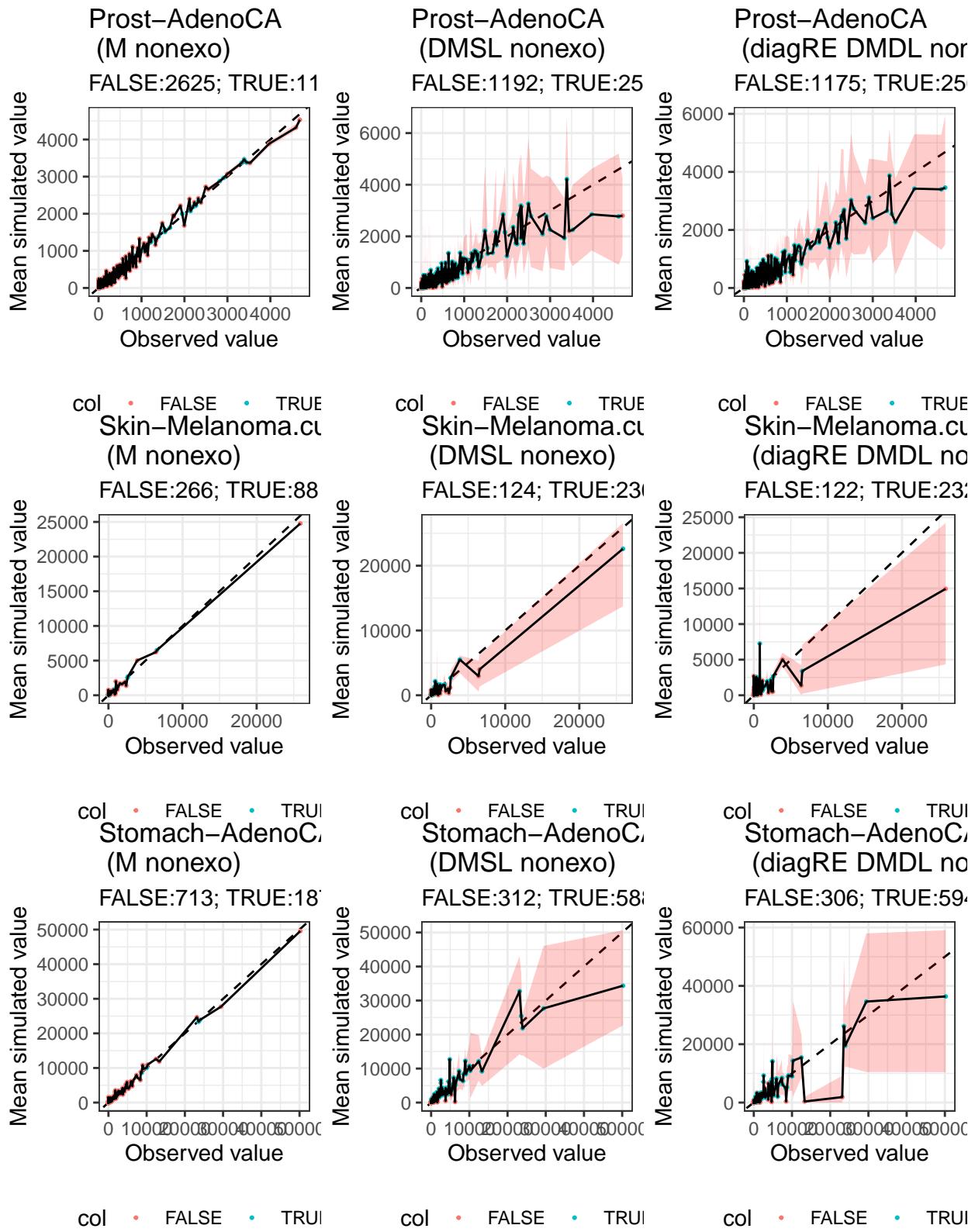


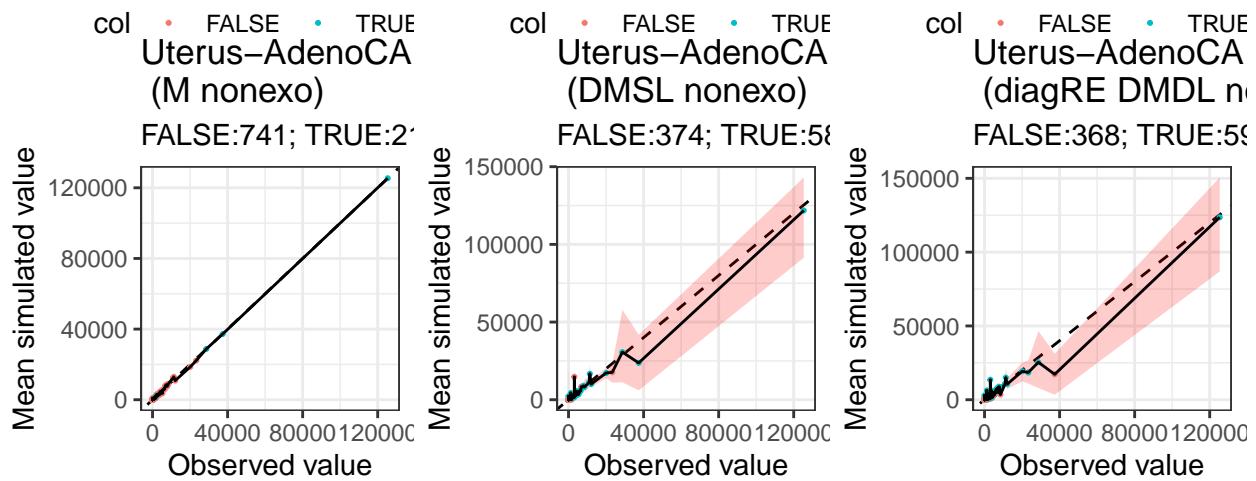
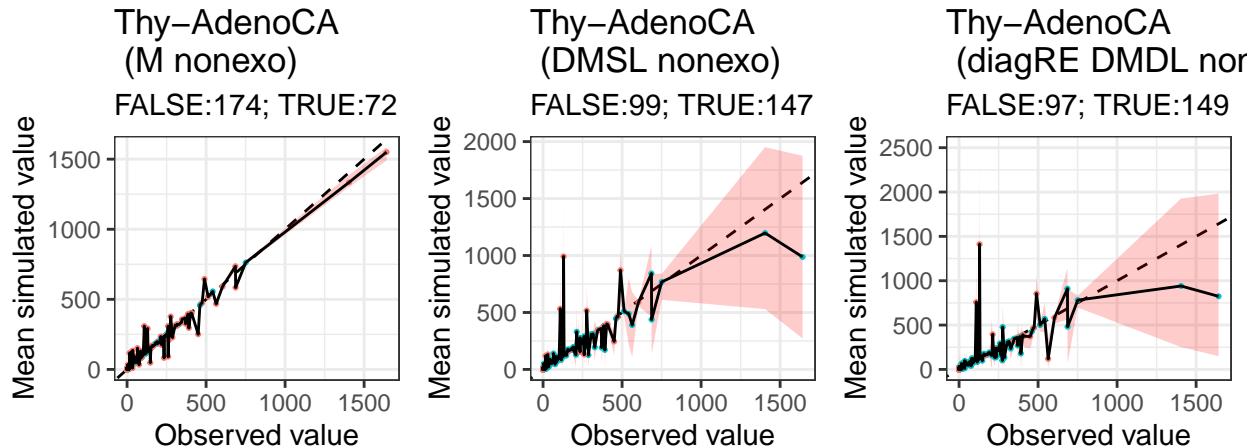












```

col   •  FALSE   •  TRUE
Uterus-AdenoCA
(M nonexo)
FALSE:741; TRUE:2

col   •  FALSE   •  TRUE
Uterus-AdenoCA
(DMSL nonexo)
FALSE:374; TRUE:58

col   •  FALSE   •  TRUE
Uterus-AdenoCA
(diagRE DMDL n
FALSE:368; TRUE:59

for(ct in enough_samples){
  try({
    grid.arrange(give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], ct=ct, typedata="signature",
                                         sigs_to_remove=unique(nonexogenous$V1),
                                         bool_give_PCA = T, path_to_data= ".../data/",
                                         obj_data=all_objects_nonexo_SP[[ct]],
                                         sig_of_interest=colnames(all_objects_nonexo_SP[[ct]]$Y)[1],
                                         bool_nonexo=T,
                                         model="fullRE_DMSL", nrow_pca_plot=1)[[2]]+
      ggttitle(paste0('Simulation of ', ct, ' fullRE_DMSL')),
    give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], ct=ct, typedata="signature",
                           sigs_to_remove=unique(nonexogenous$V1),
                           bool_give_PCA = T, path_to_data= ".../data/",
                           obj_data=all_objects_nonexo_SP[[ct]],
                           sig_of_interest=colnames(all_objects_nonexo_SP[[ct]]$Y)[1],
                           bool_nonexo=T,
                           model="diagRE_DMDL", integer_overdispersion_param=1000,
                           nrow_pca_plot=1)[[2]]+
  
```

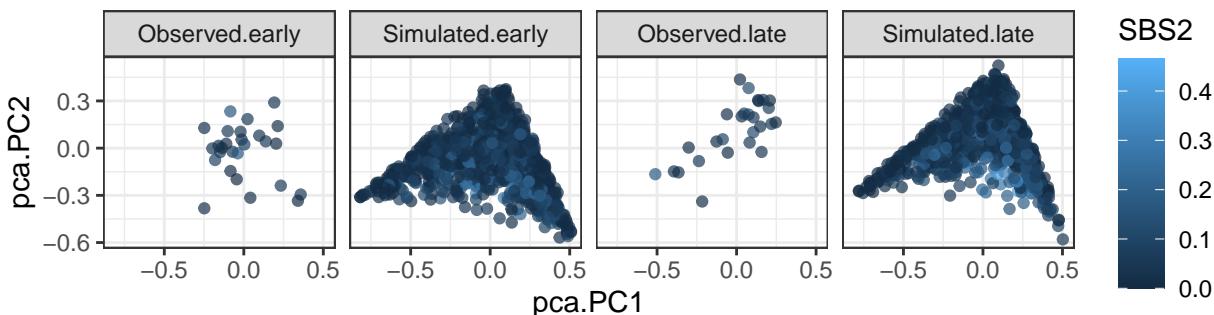
```

        ggtile(paste0('Simulation of ', ct, ' diagRE_DMDL')))

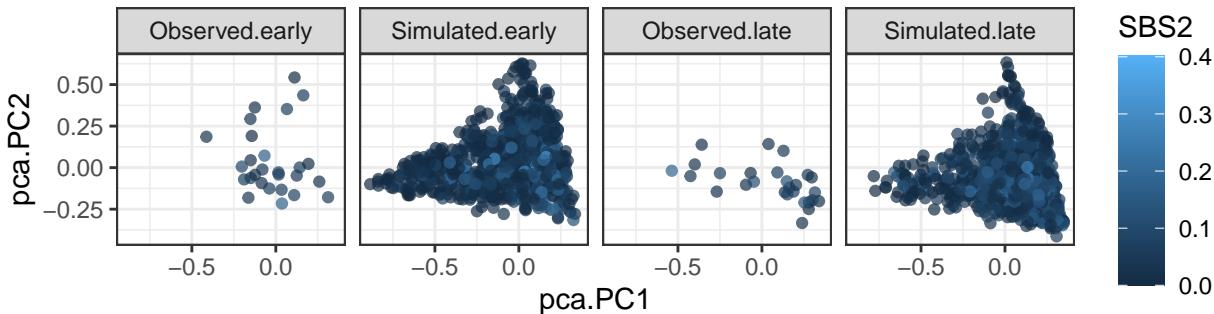
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably
## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), : This
## function had been incorrect until now (30 july 2021)
## Warning in mvtnorm::rmvnorm(n = n_sim, mean = rep(0, dmin1), sigma = cov_mat):
## sigma is numerically not positive semidefinite
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably
## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), : This
## function had been incorrect until now (30 july 2021)

```

Simulation of Bone–Osteosarc fullRE_DMSL



Simulation of Bone–Osteosarc diagRE_DMDL

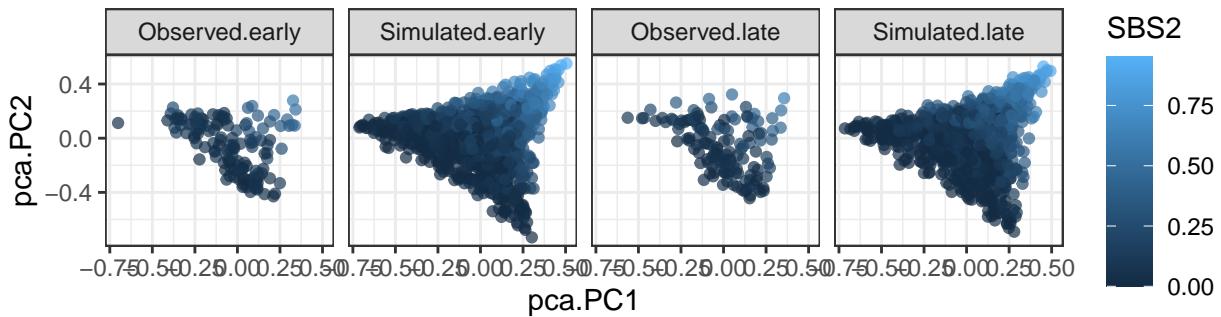


```

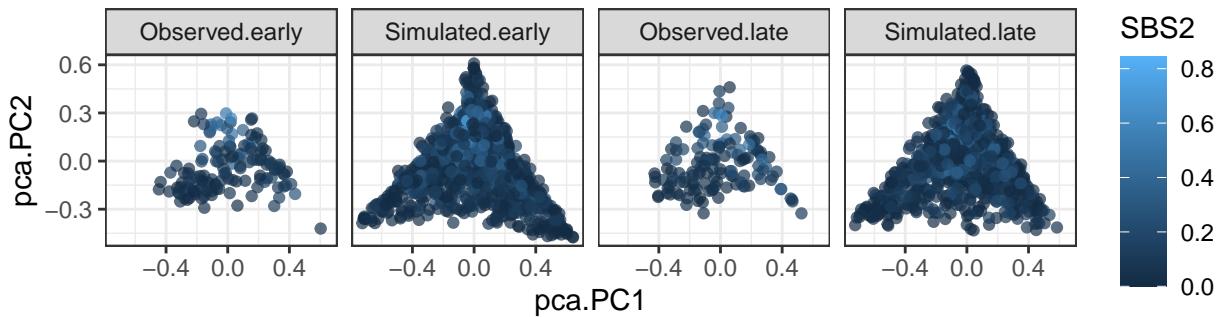
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably
## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), : This
## function had been incorrect until now (30 july 2021)

```

Simulation of Breast–AdenoCA fullRE_DMSL



Simulation of Breast–AdenoCA diagRE_DMDL



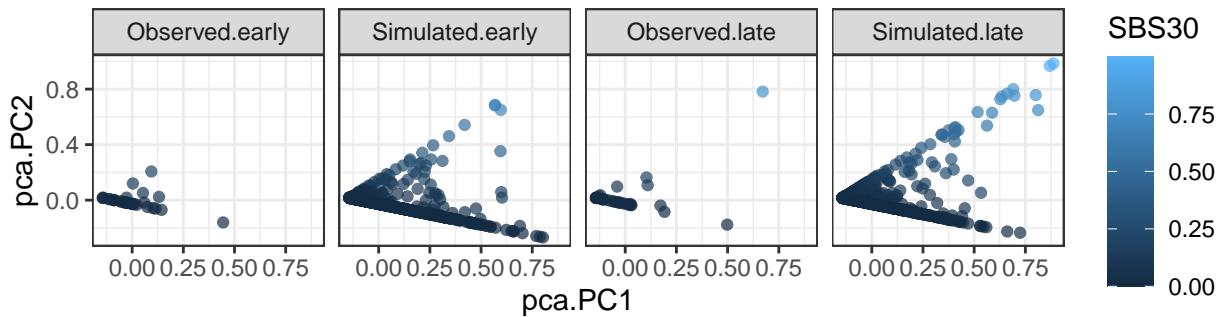
```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

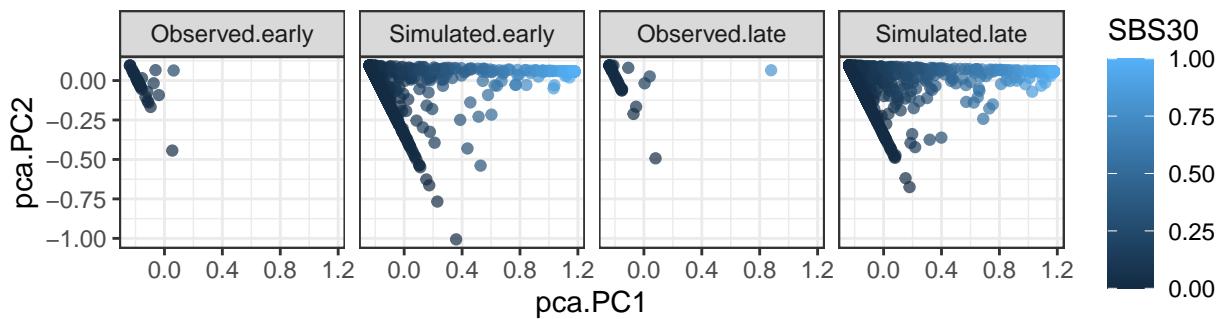
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)
```

Simulation of CNS–GBM fullRE_DMSL



Simulation of CNS–GBM diagRE_DMDL



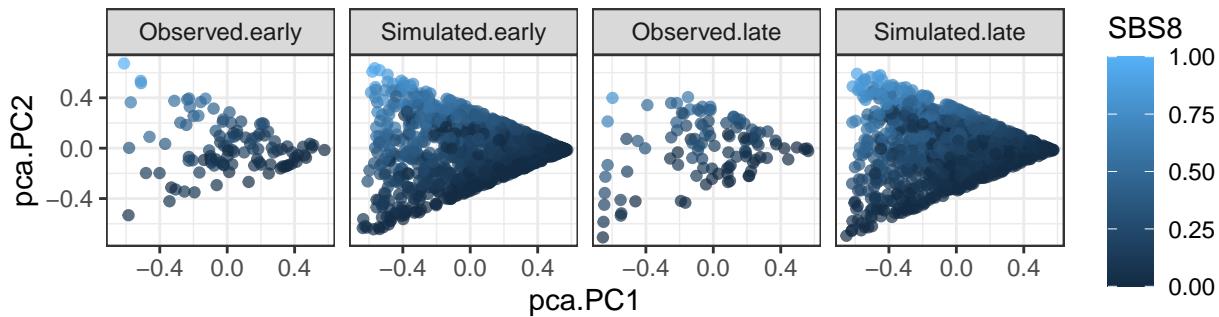
```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

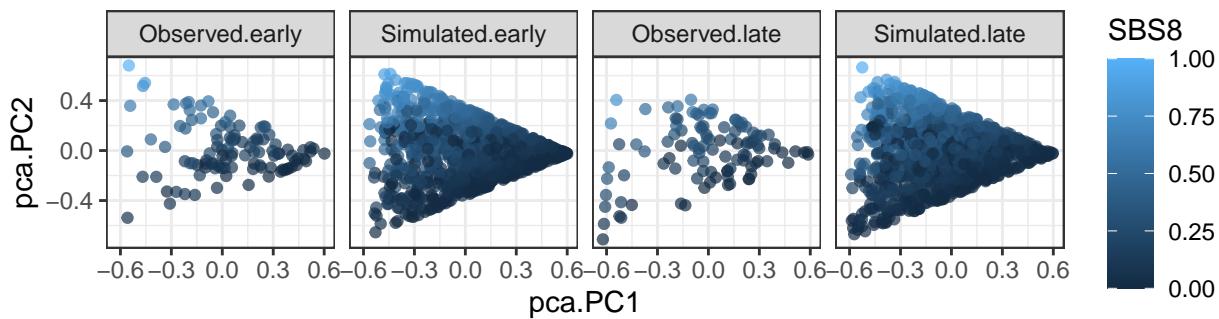
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)
```

Simulation of CNS–Medullo fullRE_DMSL



Simulation of CNS–Medullo diagRE_DMDL



```

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

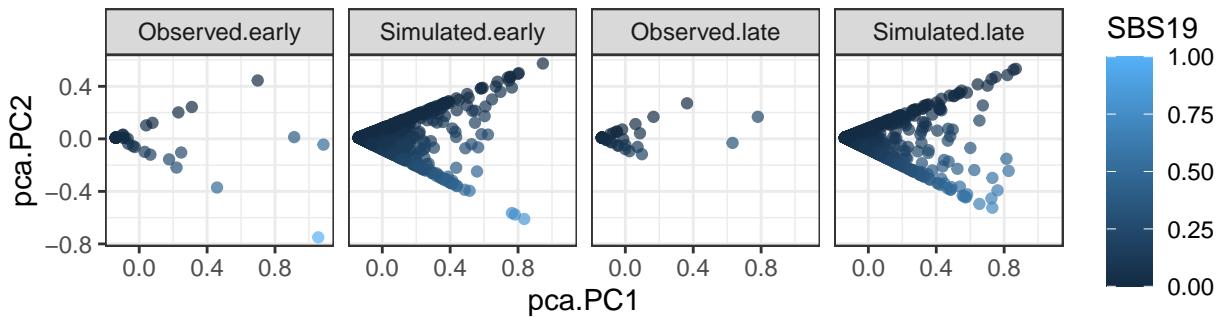
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

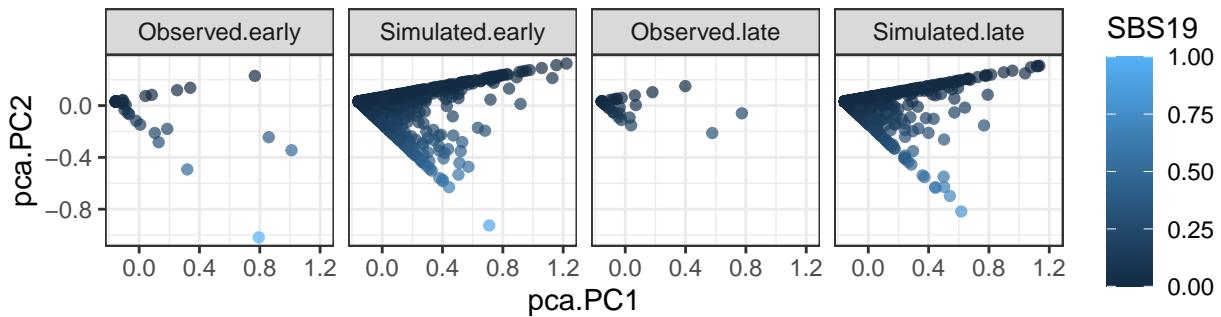
## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)

```

Simulation of CNS–PiloAstro fullRE_DMSL



Simulation of CNS–PiloAstro diagRE_DMDL



```

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

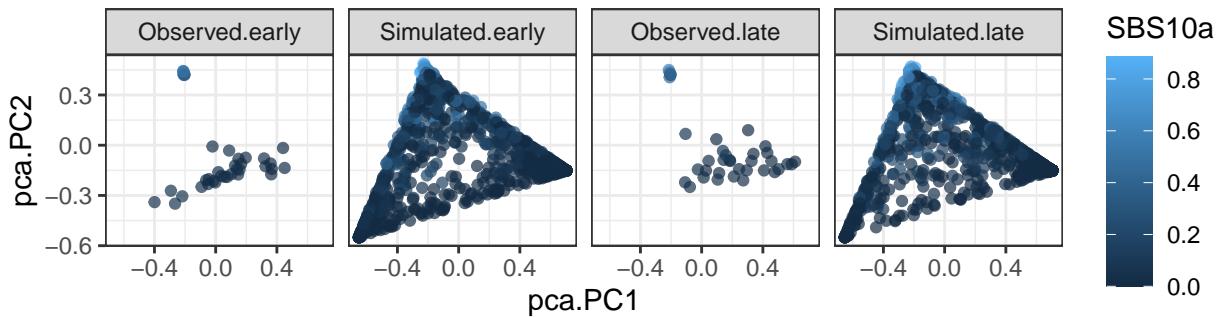
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

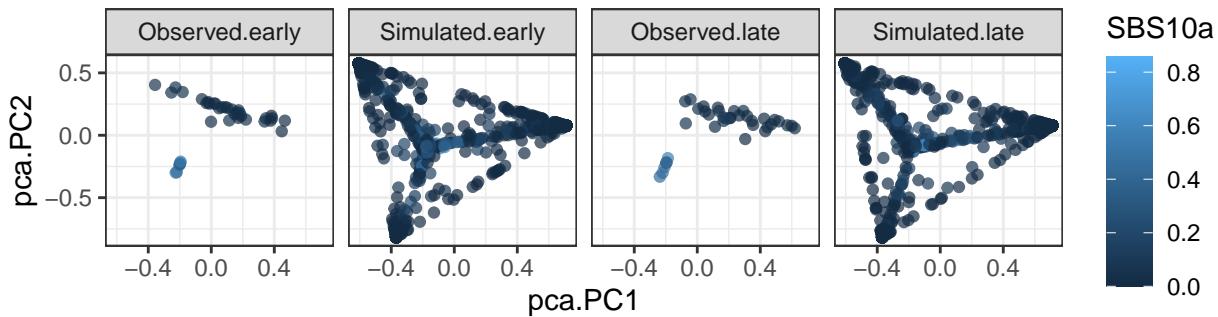
## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)

```

Simulation of ColoRect–AdenoCA fullRE_DMSL



Simulation of ColoRect–AdenoCA diagRE_DMDL



```

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

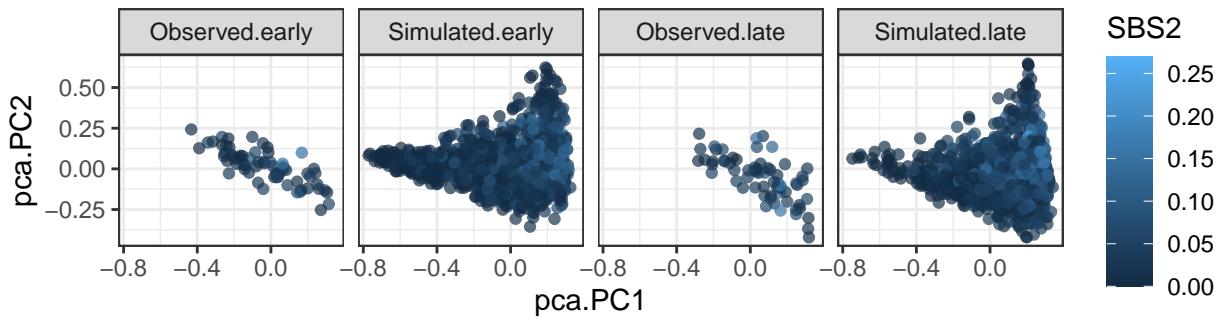
## Warning in mvtnorm:::rmvnorm(n = n_sim, mean = rep(0, dmin1), sigma = cov_mat):
## sigma is numerically not positive semidefinite

## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

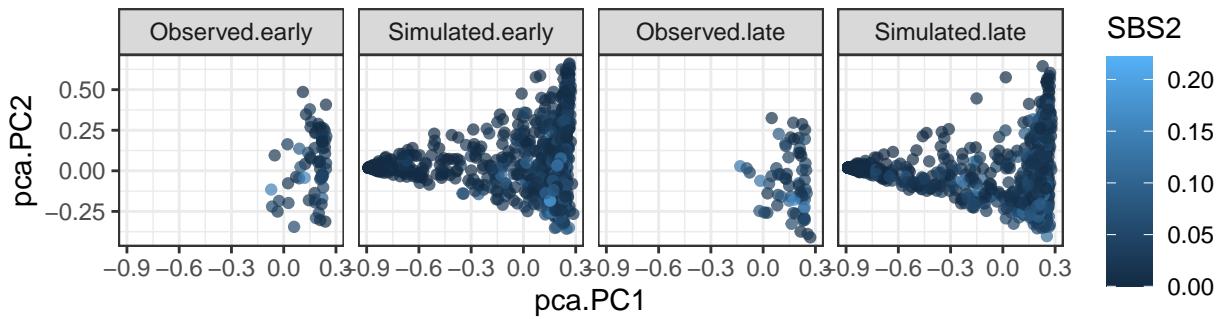
## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)

```

Simulation of Eso–AdenoCA fullRE_DMSL



Simulation of Eso–AdenoCA diagRE_DMDL



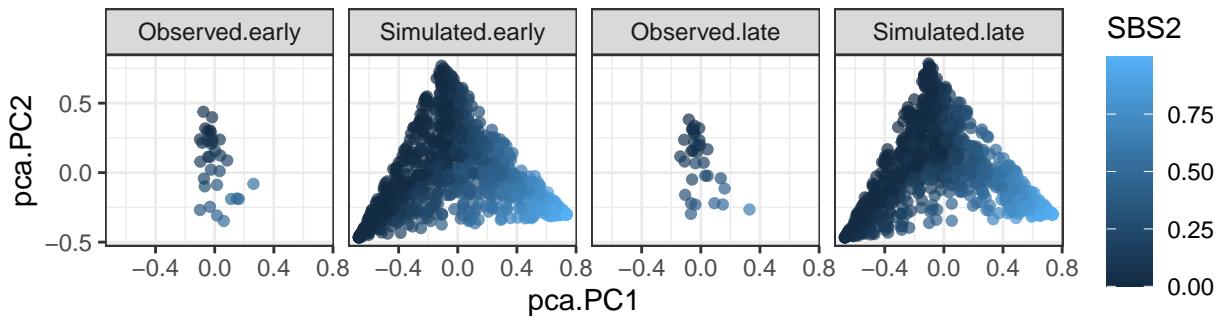
```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

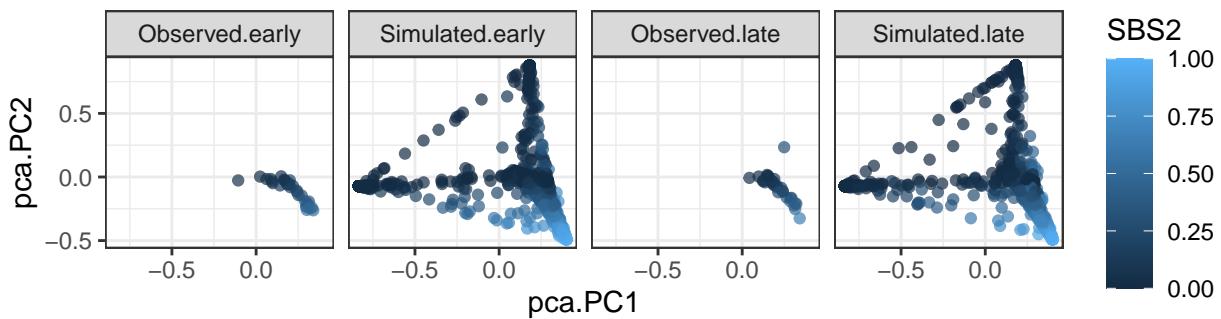
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)
```

Simulation of Head–SCC fullRE_DMSL



Simulation of Head–SCC diagRE_DMDL



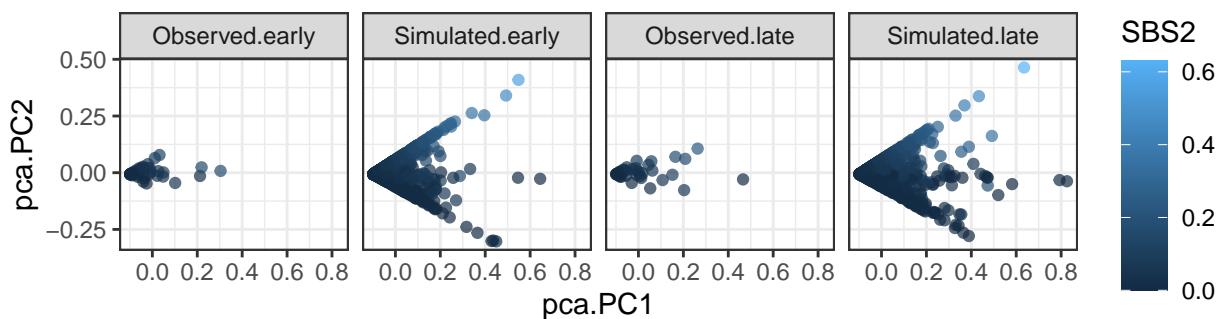
```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

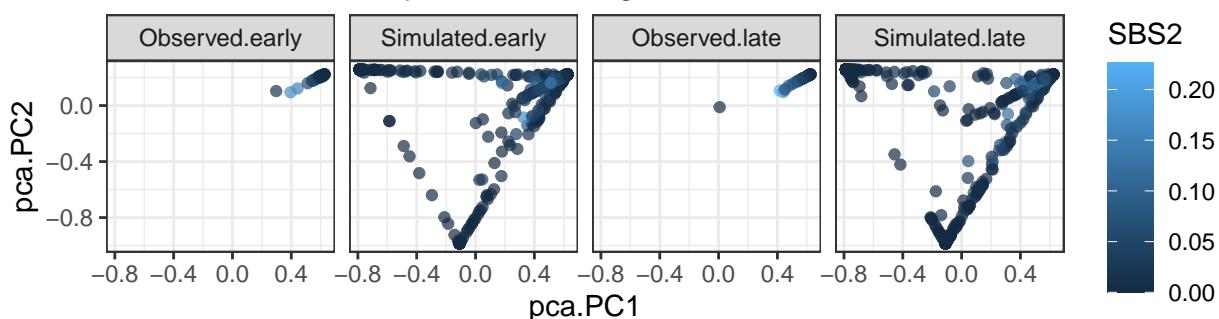
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)
```

Simulation of Kidney–ChRCC fullRE_DMSL



Simulation of Kidney–ChRCC diagRE_DMDL



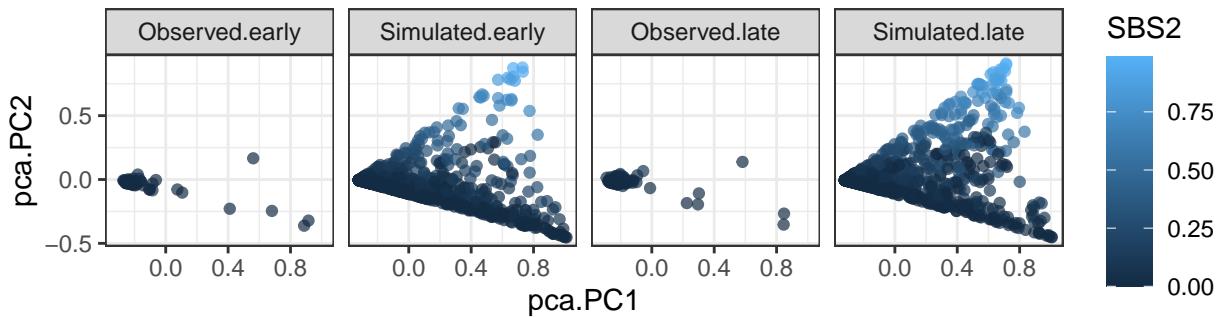
```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

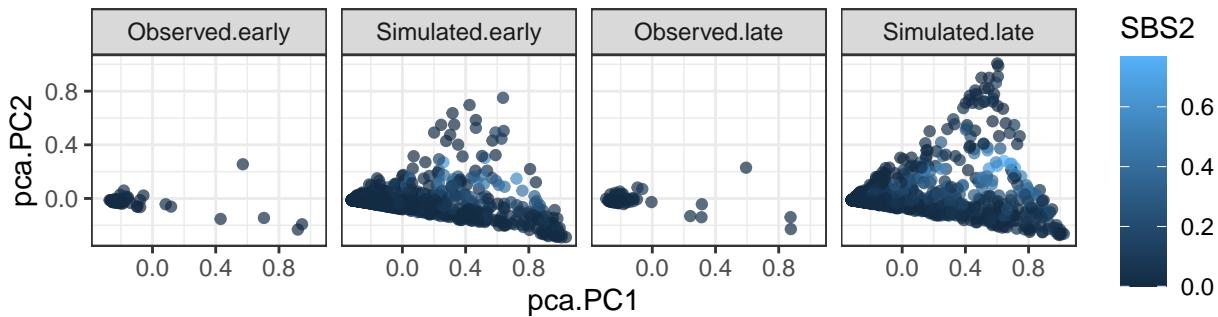
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), : This
## function had been incorrect until now (30 july 2021)
```

Simulation of Kidney–RCC.clearcell fullRE_DMSL



Simulation of Kidney–RCC.clearcell diagRE_DMDL



```

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

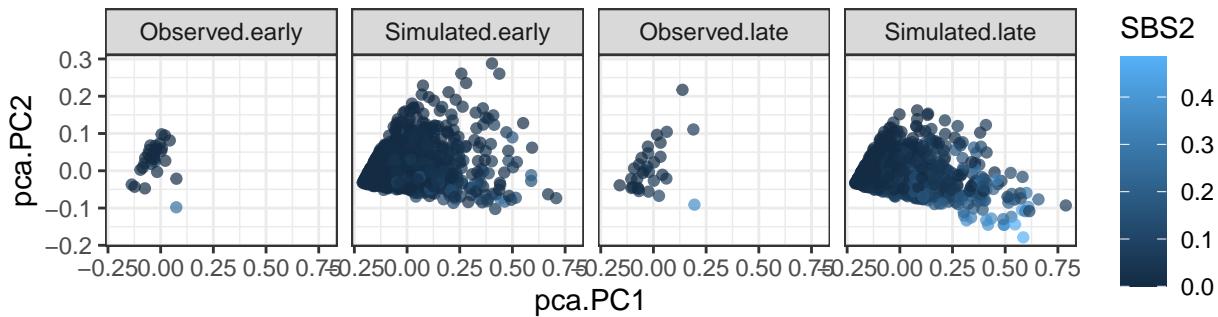
## Warning in mvtnorm:::rmvnorm(n = n_sim, mean = rep(0, dmin1), sigma = cov_mat):
## sigma is numerically not positive semidefinite

## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

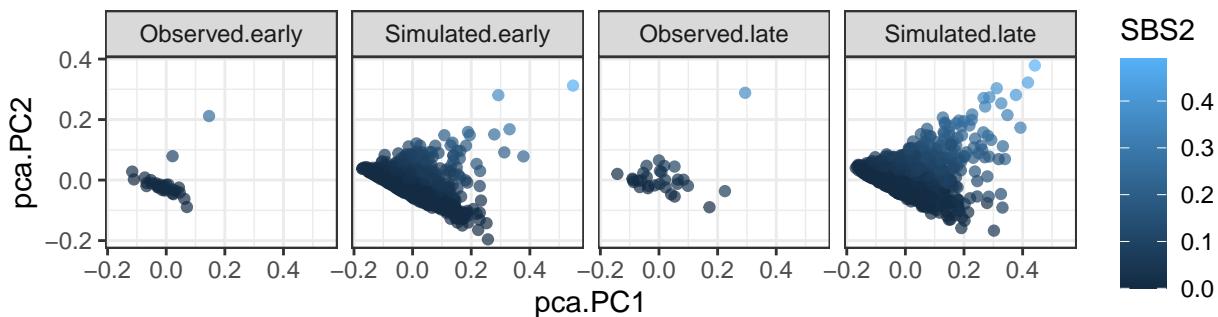
## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## This
## function had been incorrect until now (30 july 2021)

```

Simulation of Kidney–RCC.papillary fullRE_DMSL



Simulation of Kidney–RCC.papillary diagRE_DMDL



```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

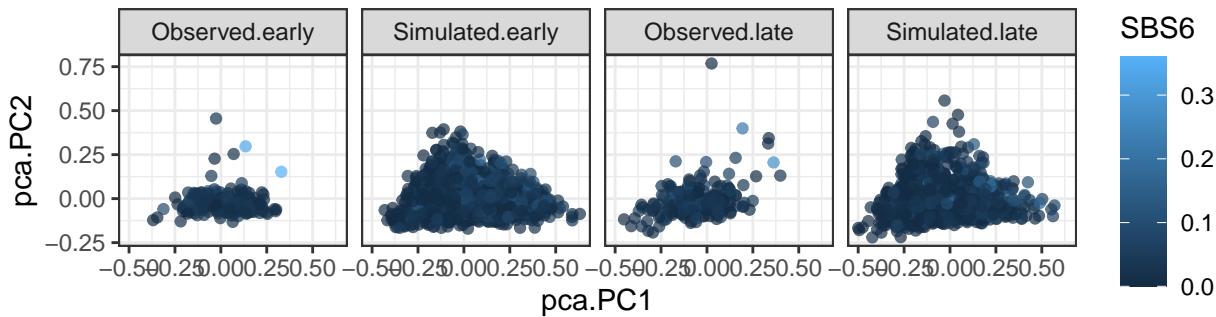
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

## Warning in mvtnorm:::rmvnorm(n = n_sim, mean = rep(0, dmin1), sigma = cov_mat):
## sigma is numerically not positive semidefinite

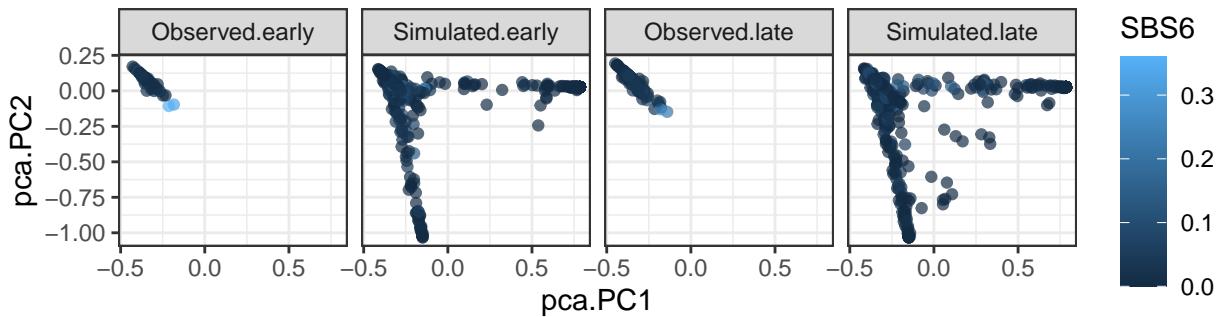
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)
```

Simulation of Liver–HCC fullRE_DMSL



Simulation of Liver–HCC diagRE_DMDL



```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

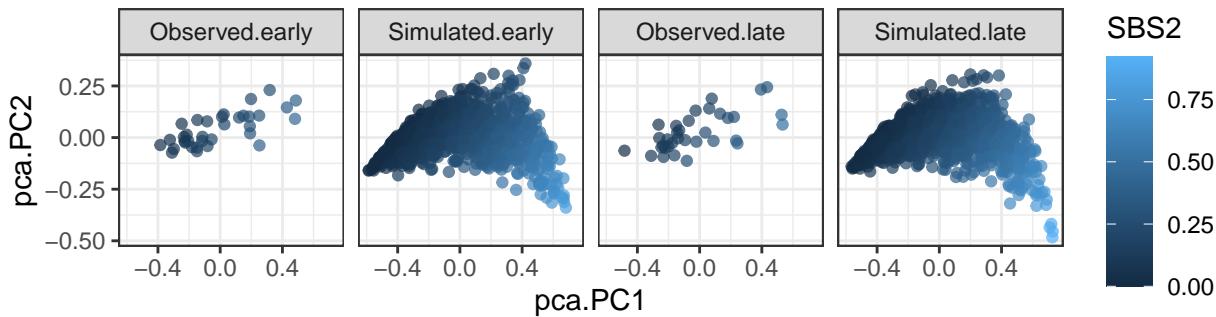
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

## Warning in mvtnorm:::rmvnorm(n = n_sim, mean = rep(0, dmin1), sigma = cov_mat):
## sigma is numerically not positive semidefinite

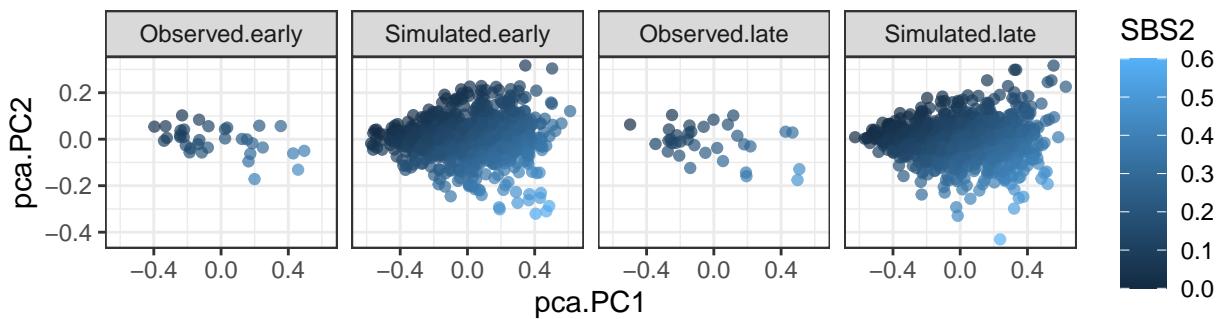
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)
```

Simulation of Lung–SCC fullRE_DMSL



Simulation of Lung–SCC diagRE_DMDL



```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

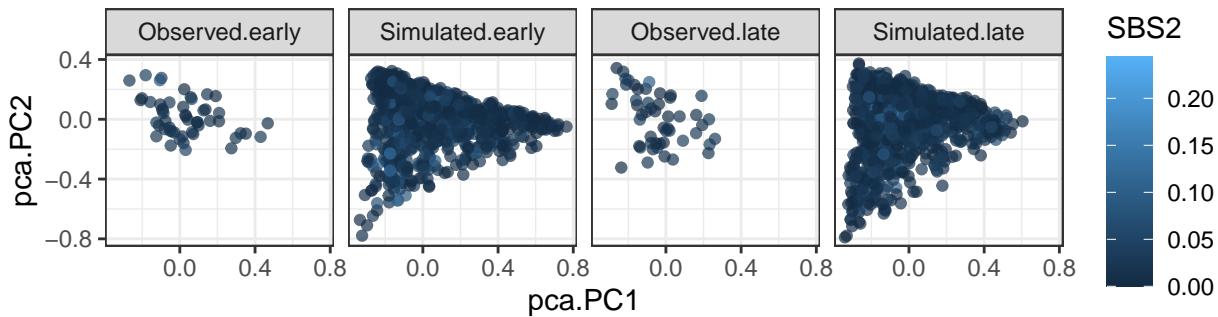
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

## Warning in mvtnorm:::rmvnorm(n = n_sim, mean = rep(0, dmin1), sigma = cov_mat):
## sigma is numerically not positive semidefinite

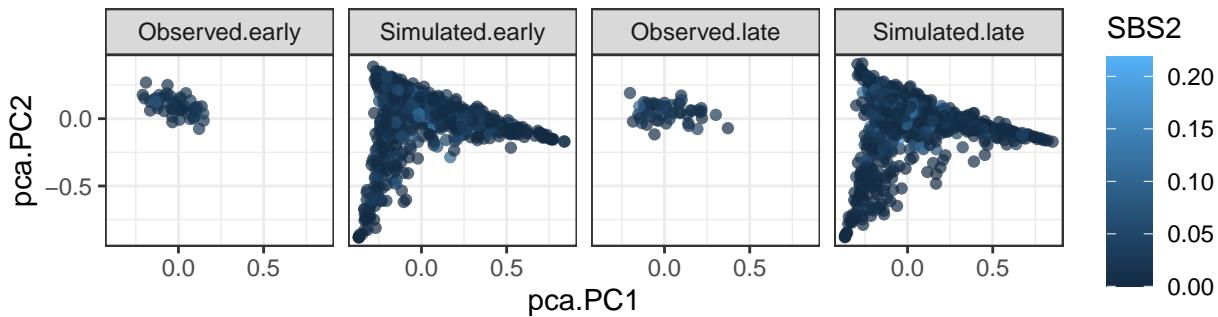
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)
```

Simulation of Lymph–BNHL fullRE_DMSL



Simulation of Lymph–BNHL diagRE_DMDL



```

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

## Error in m[unlist(sapply(2:nrow(m), function(rw) seq(from = rw, length.out = (rw -
## replacement has length zero

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

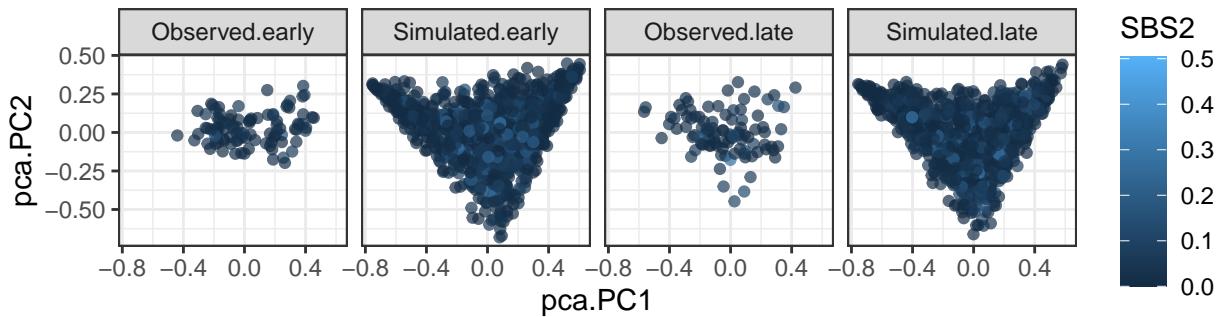
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

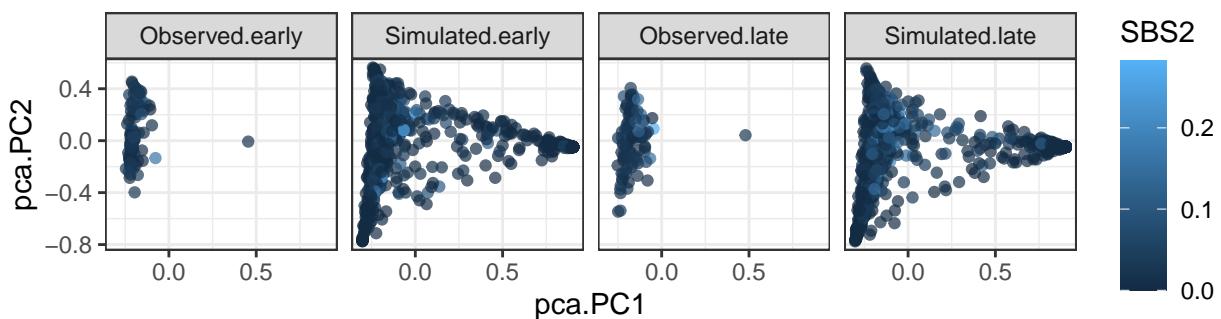
## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), : This
## function had been incorrect until now (30 july 2021)

```

Simulation of Ovary–AdenoCA fullRE_DMSL



Simulation of Ovary–AdenoCA diagRE_DMDL



```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

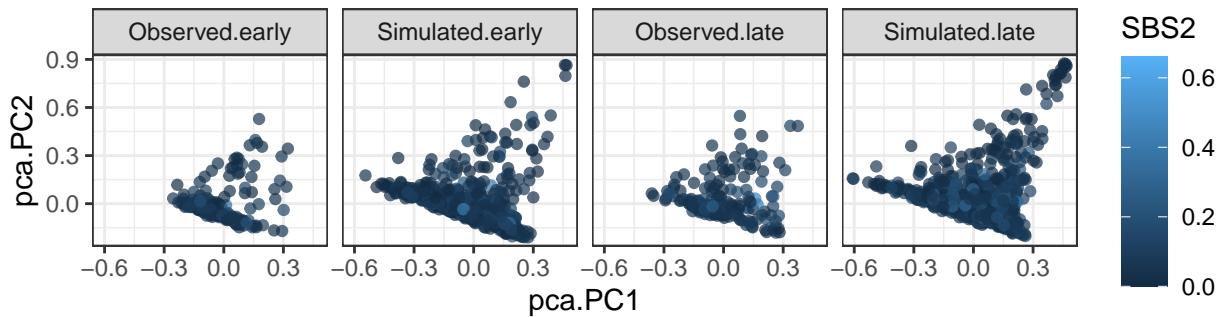
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

## Warning in mvtnorm:::rmvnorm(n = n_sim, mean = rep(0, dmin1), sigma = cov_mat):
## sigma is numerically not positive semidefinite

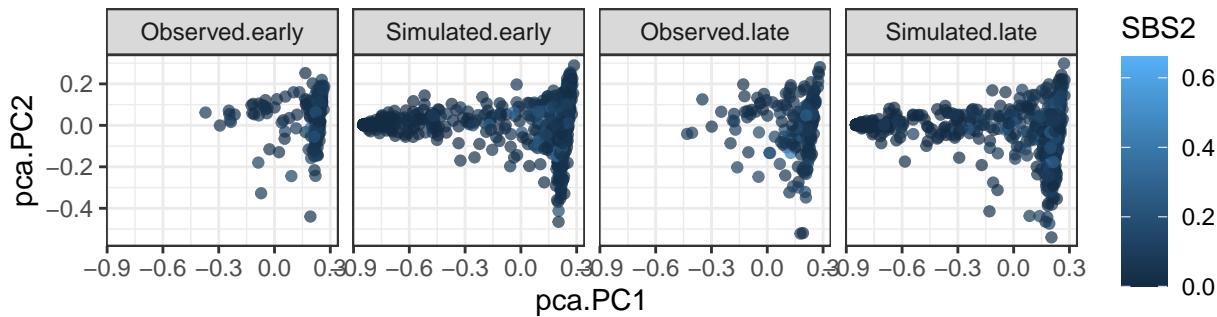
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## This
## function had been incorrect until now (30 july 2021)
```

Simulation of Panc–AdenoCA fullRE_DMSL



Simulation of Panc–AdenoCA diagRE_DMDL



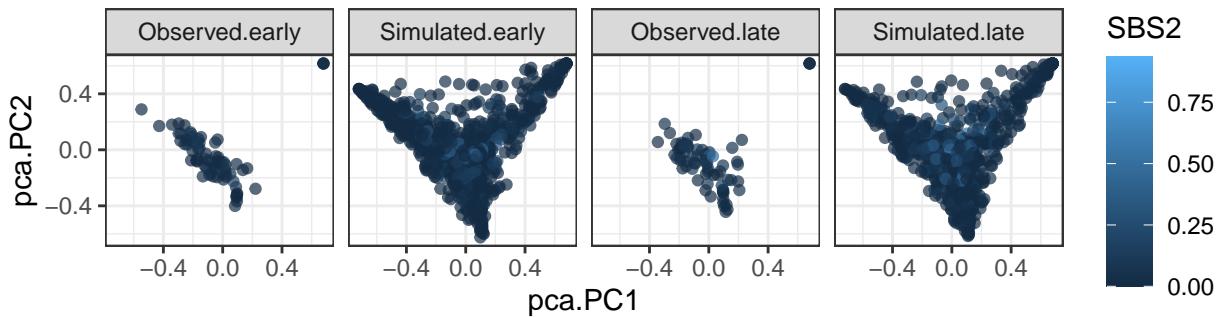
```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

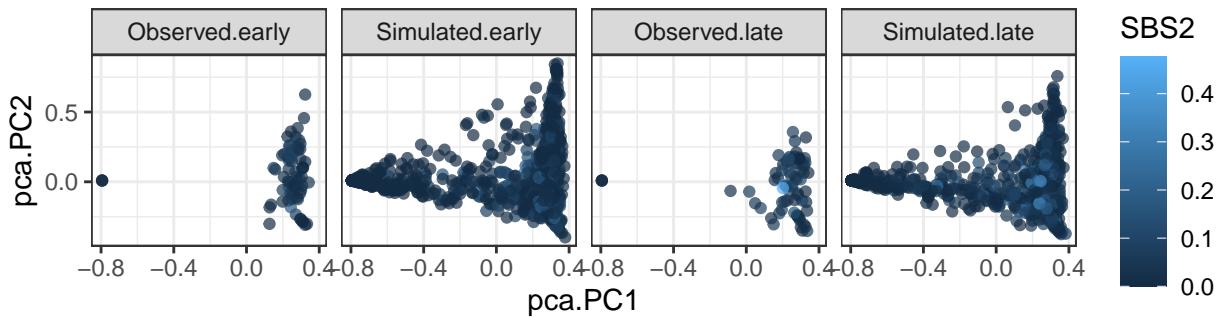
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), : This
## function had been incorrect until now (30 july 2021)
```

Simulation of Panc–Endocrine fullRE_DMSL



Simulation of Panc–Endocrine diagRE_DMDL



```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

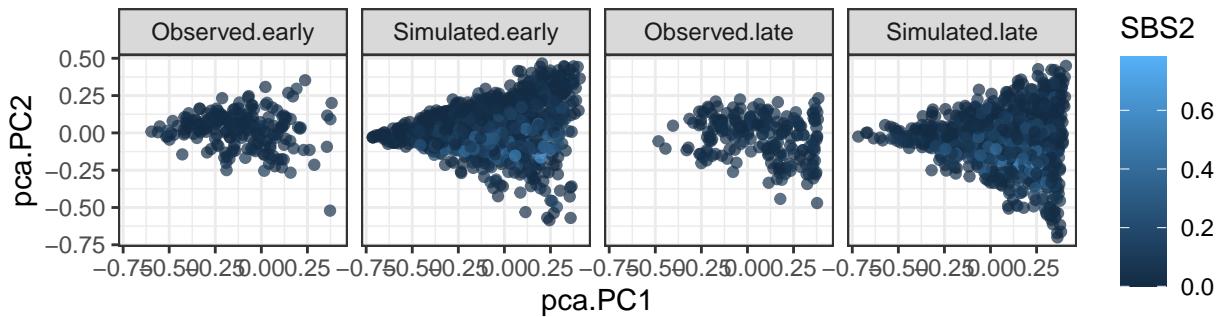
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

## Warning in mvtnorm:::rmvnorm(n = n_sim, mean = rep(0, dmin1), sigma = cov_mat):
## sigma is numerically not positive semidefinite

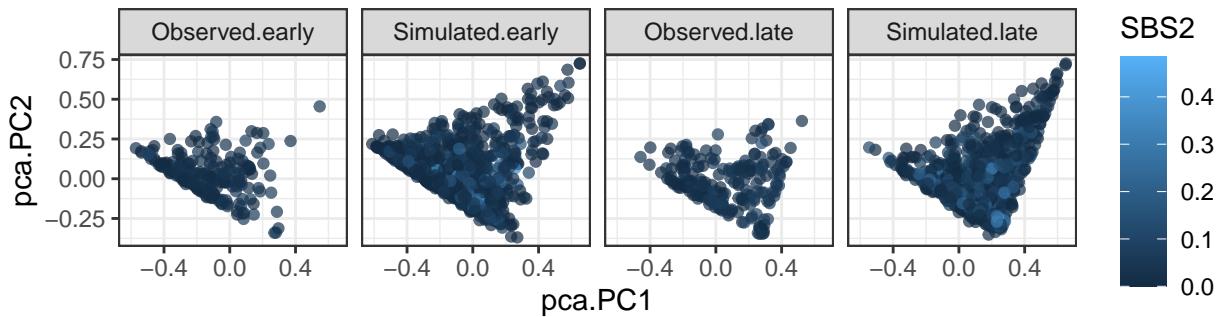
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)
```

Simulation of Prost–AdenoCA fullRE_DMSL



Simulation of Prost–AdenoCA diagRE_DMDL



```

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

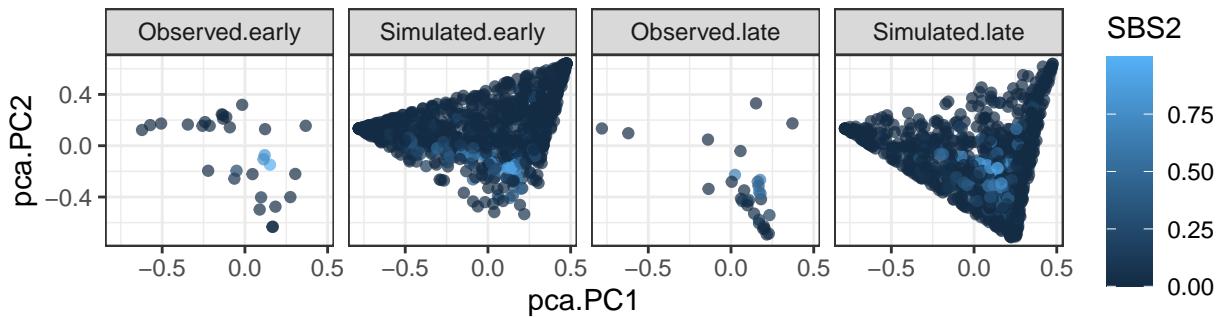
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

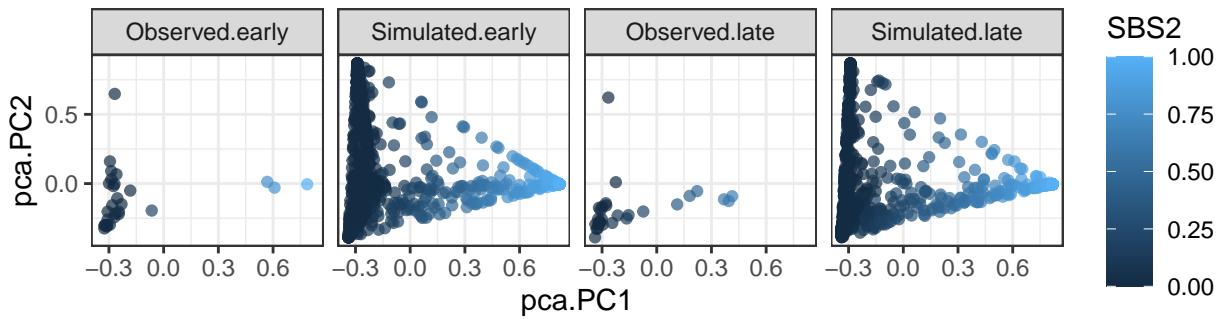
## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)

```

Simulation of Skin–Melanoma.cutaneous fullRE_DMSL



Simulation of Skin–Melanoma.cutaneous diagRE_DMDL



```

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

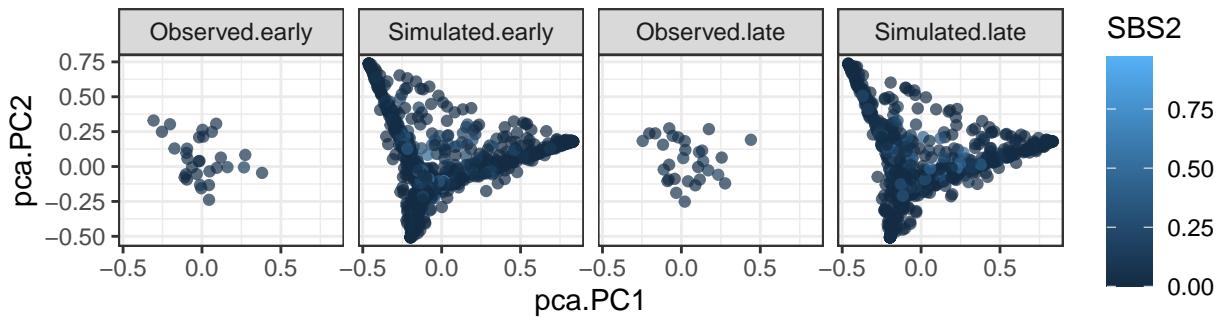
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

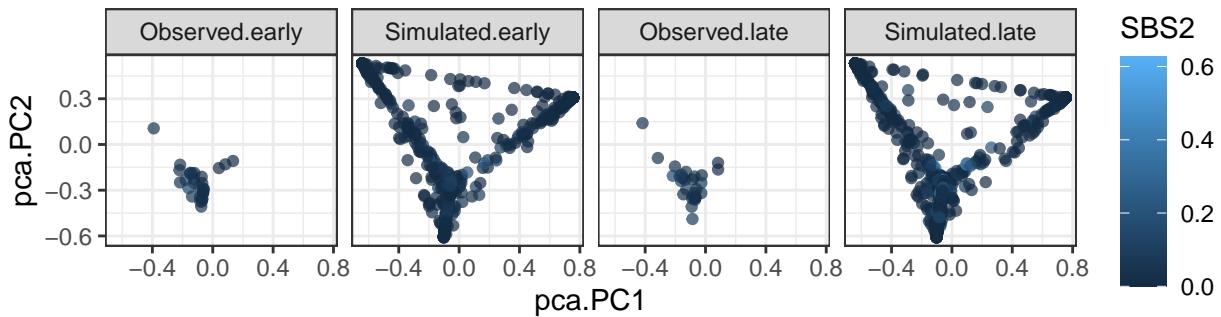
## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)

```

Simulation of Stomach–AdenoCA fullRE_DMSL



Simulation of Stomach–AdenoCA diagRE_DMDL



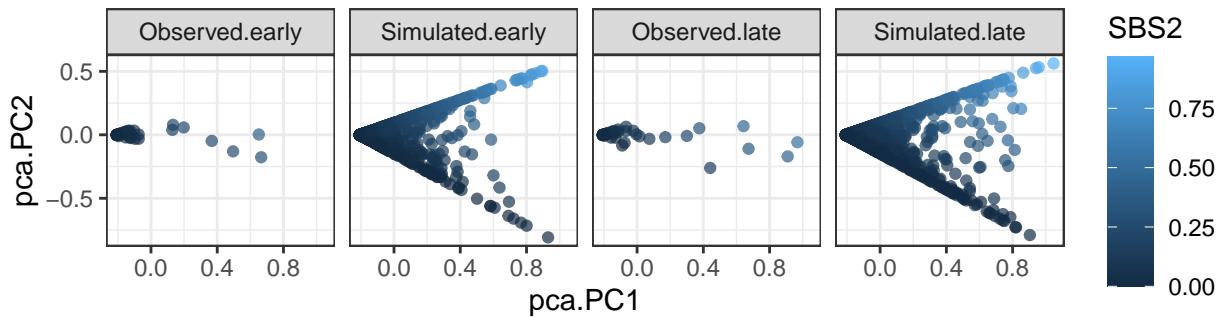
```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

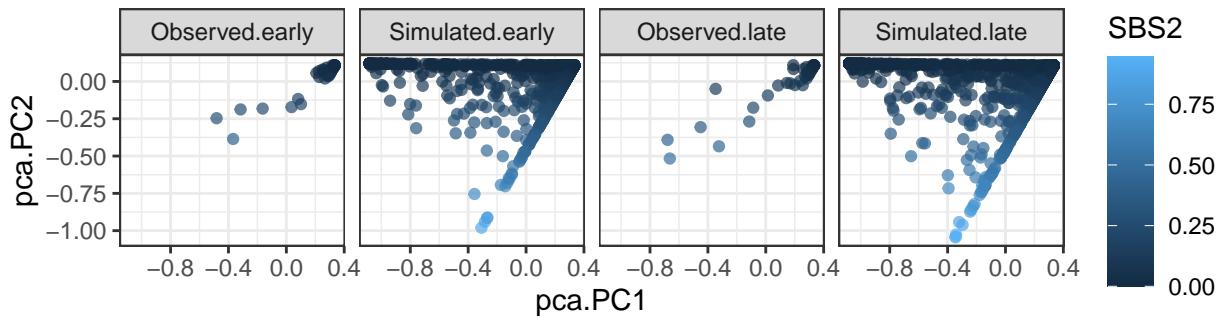
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)
```

Simulation of Thy–AdenoCA fullRE_DMSL



Simulation of Thy–AdenoCA diagRE_DMDL



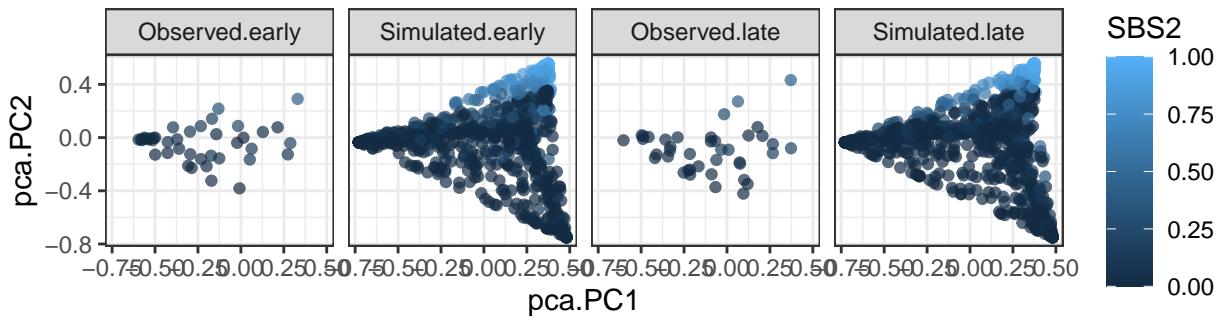
```
## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in give_sim_from_estimates(tmb_object = fullRE_DMSL_nonexo_SP[[ct]], :
## This function had been incorrect until now (30 july 2021)

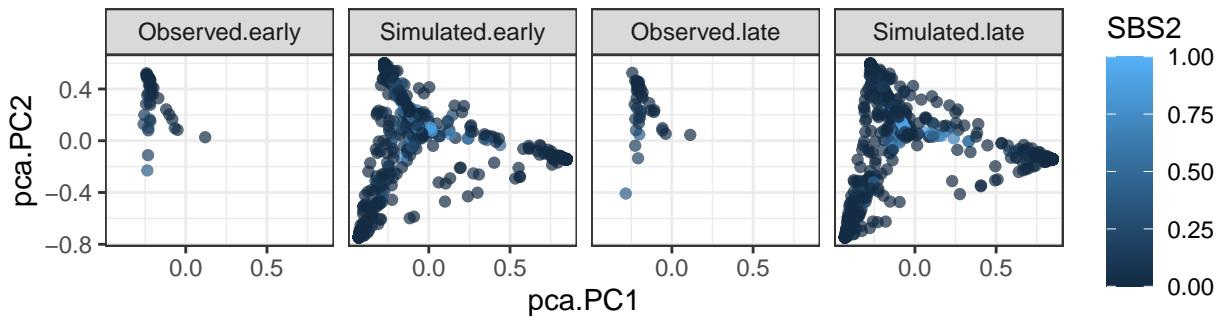
## Warning in give_sim_from_estimates(tmb_object = diagRE_DMDL_nonexo_SP[[ct]], :
## <fullRE_DMSL> and <fullRE_DM> used be to used interchangeably

## Warning in fill_covariance_matrix(arg_d = d, arg_entries_var = rep(1, d), :
## function had been incorrect until now (30 july 2021)
```

Simulation of Uterus–AdenoCA fullRE_DMSL

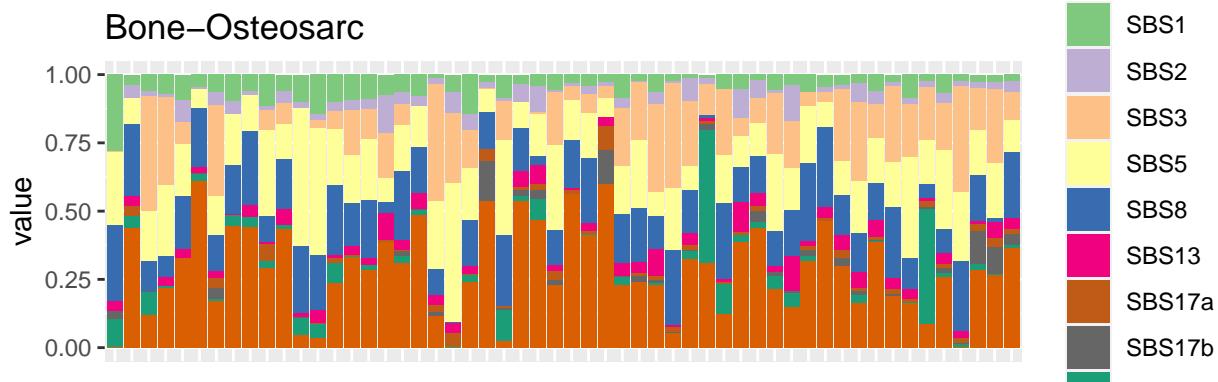


Simulation of Uterus–AdenoCA diagRE_DMDL

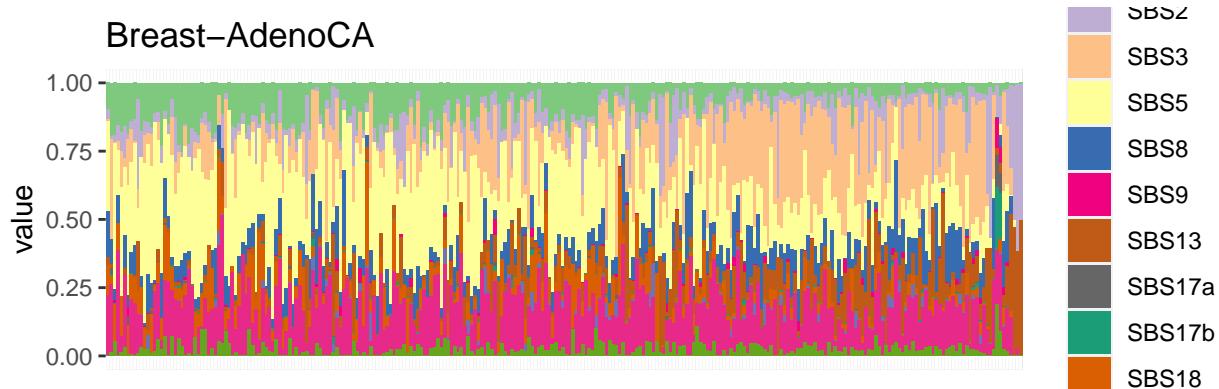


```
for(ct in enough_samples){
  print(createBarplot(normalise_rw(non_duplicated_rows(all_objects_SP[[ct]]$Y)),
    order_labels = names(sort(rowSums(non_duplicated_rows(all_objects_SP[[ct]]$Y)),
      decreasing = F)), remove_labels=T)+ggtitle(ct))
}
```

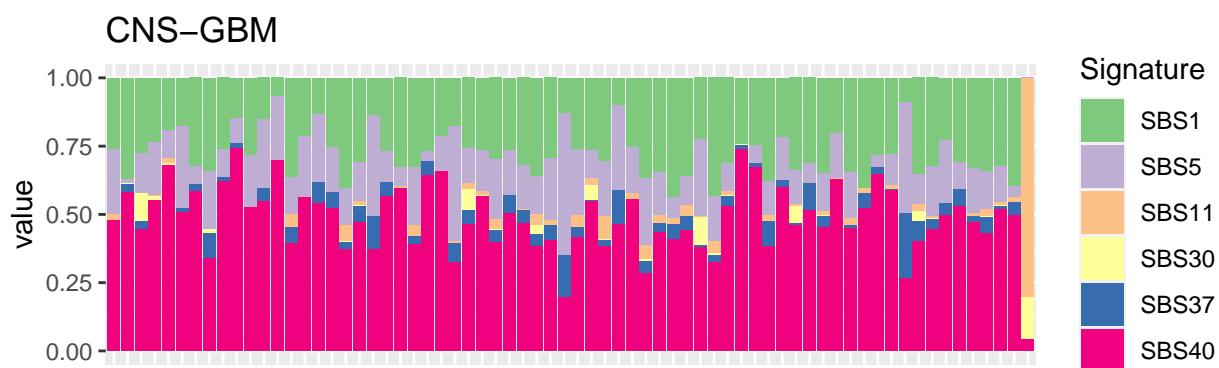
Creating plot... it might take some time if the data are large. Number of samples: 54



Creating plot... it might take some time if the data are large. Number of samples: 272



```
## Creating plot... it might take some time if the data are large. Number of samples: 68
```

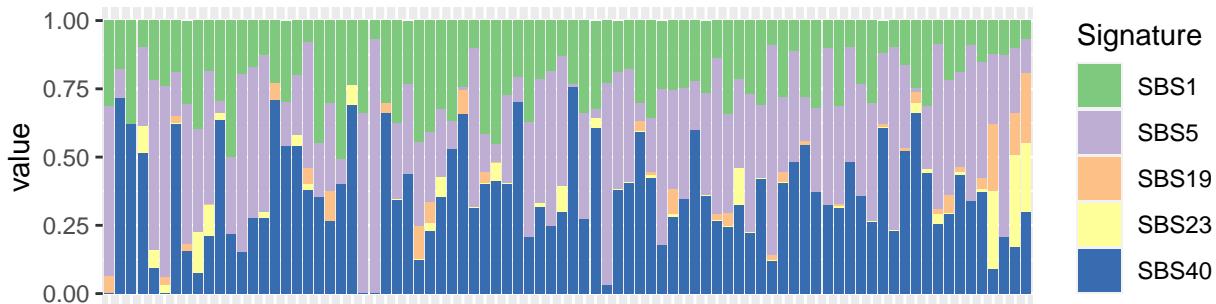


```
## Creating plot... it might take some time if the data are large. Number of samples: 212
```



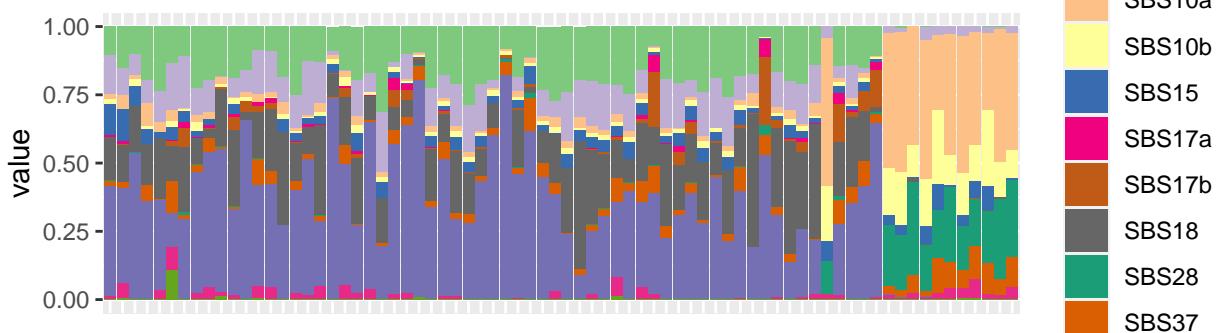
```
## Creating plot... it might take some time if the data are large. Number of samples: 84
```

CNS-PiloAstro



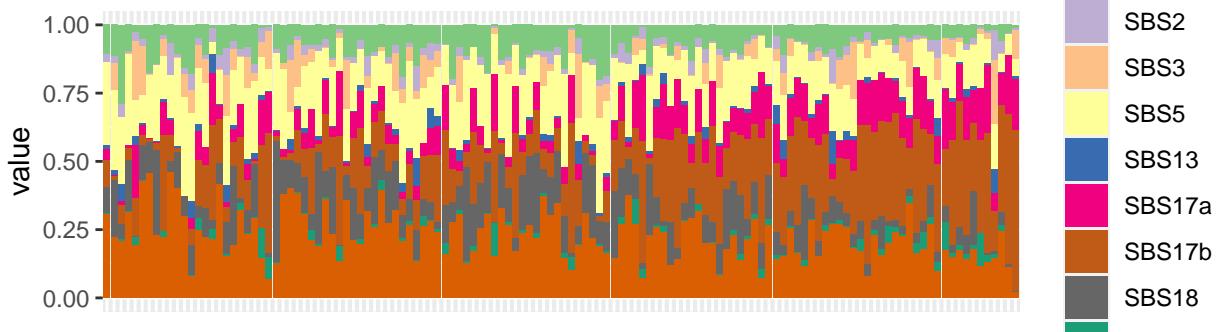
```
## Creating plot... it might take some time if the data are large. Number of samples: 74
```

ColoRect-AdenoCA

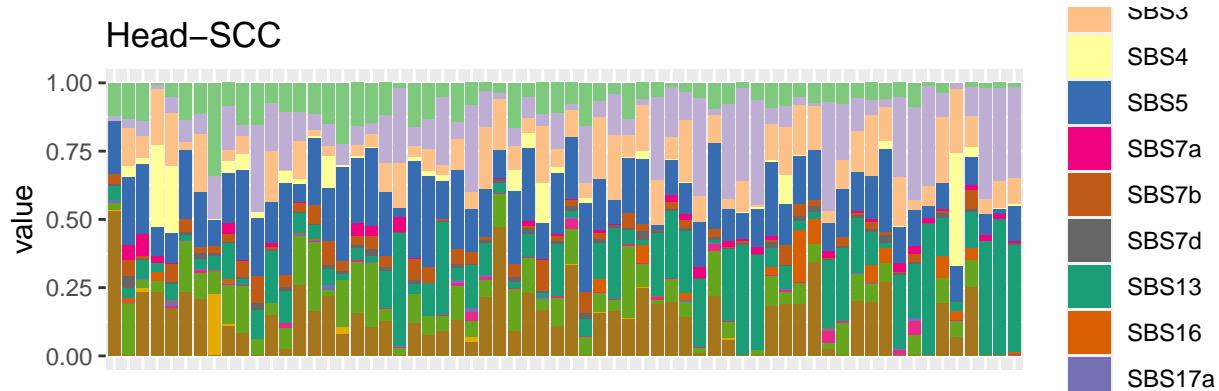


```
## Creating plot... it might take some time if the data are large. Number of samples: 130
```

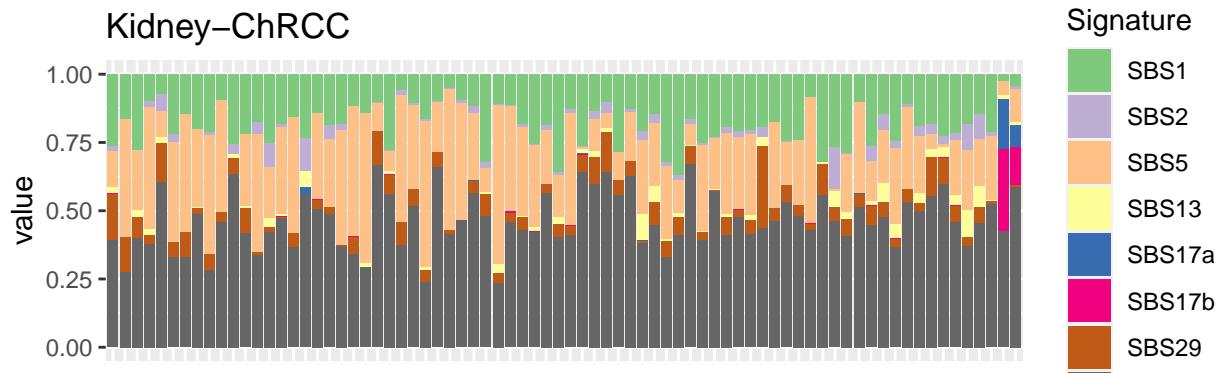
Eso-AdenoCA



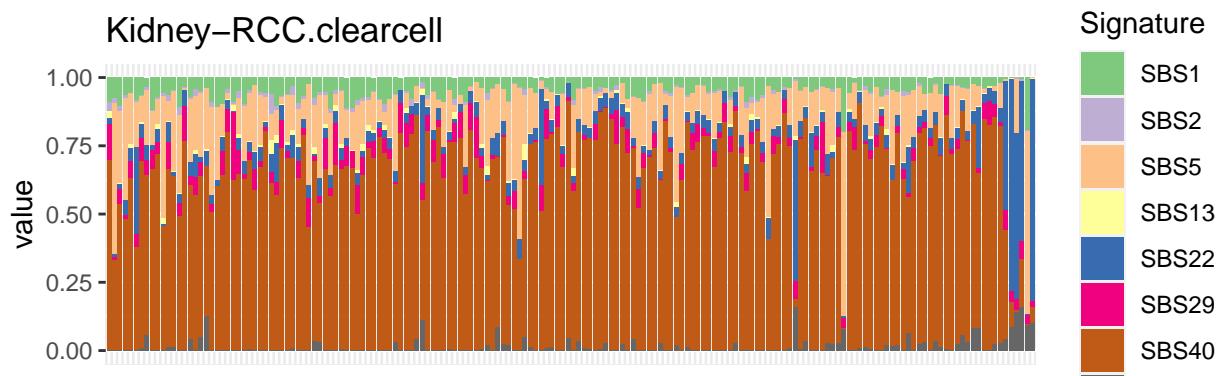
```
## Creating plot... it might take some time if the data are large. Number of samples: 64
```



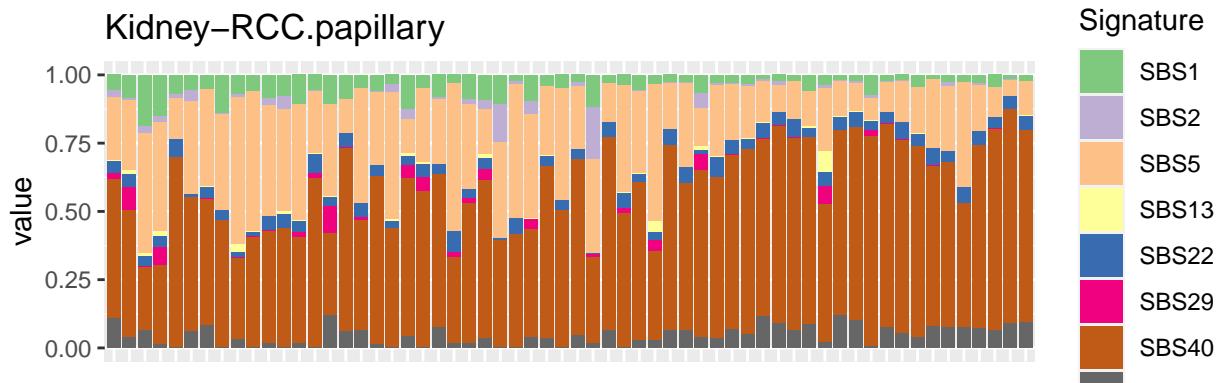
```
## Creating plot... it might take some time if the data are large. Number of samples: 76
```



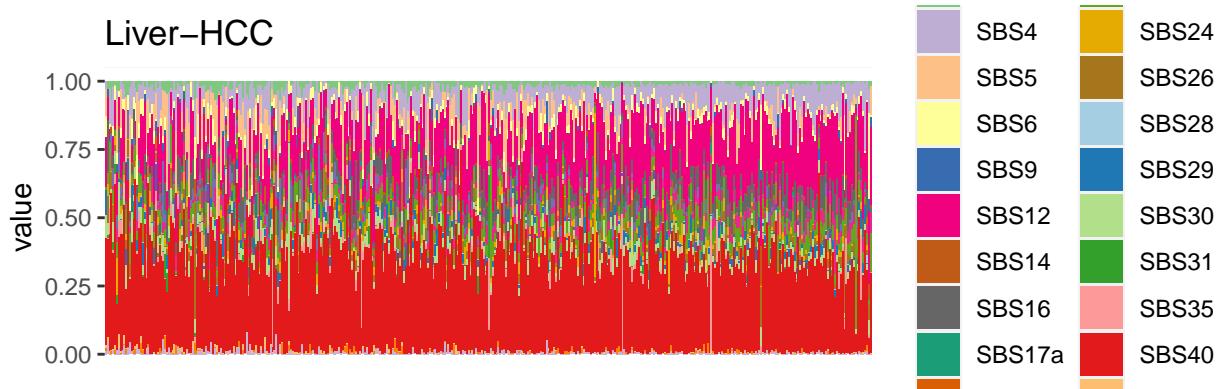
```
## Creating plot... it might take some time if the data are large. Number of samples: 172
```



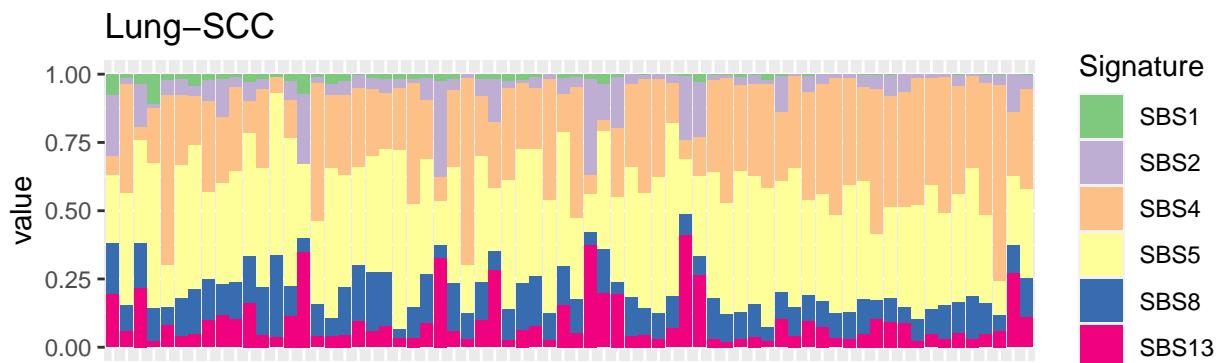
```
## Creating plot... it might take some time if the data are large. Number of samples: 60
```



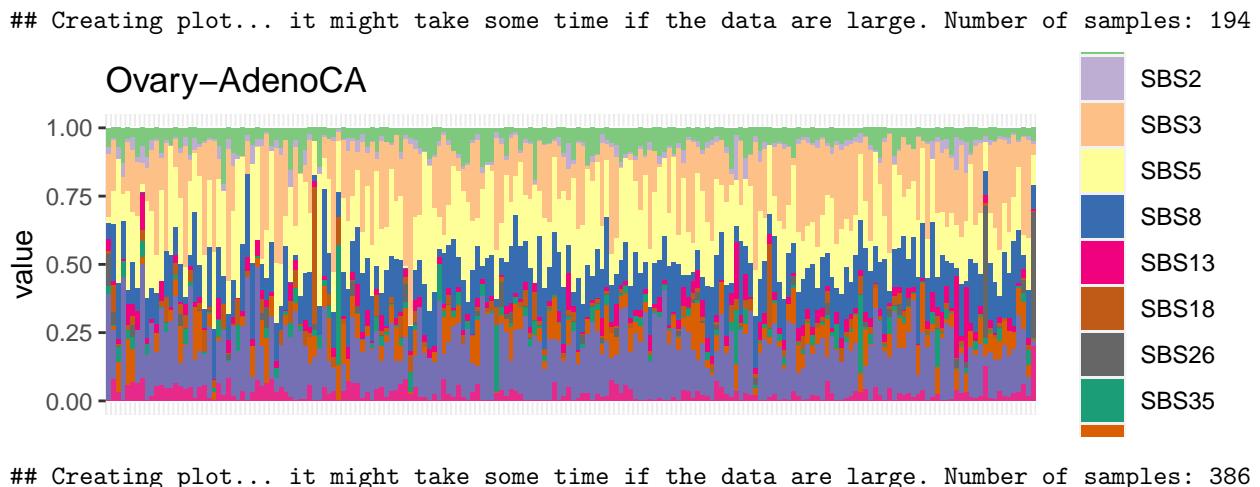
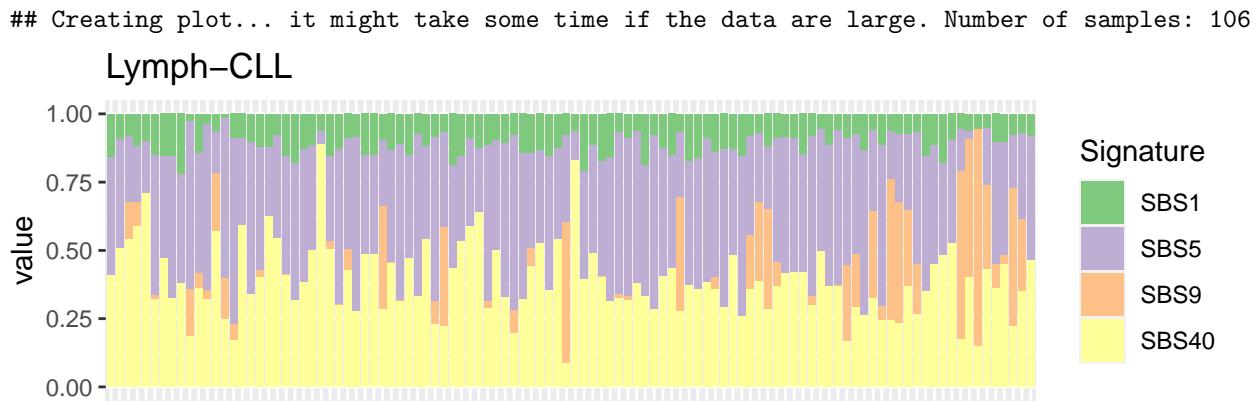
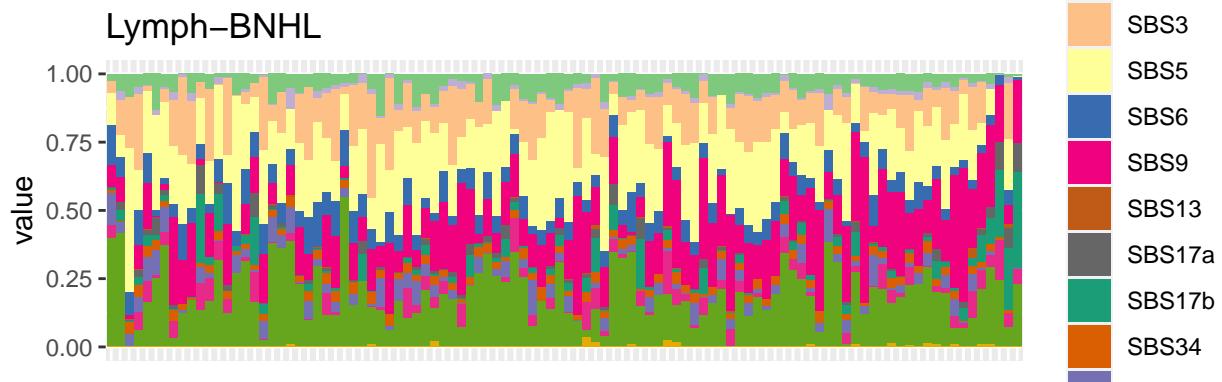
```
## Creating plot... it might take some time if the data are large. Number of samples: 414
```

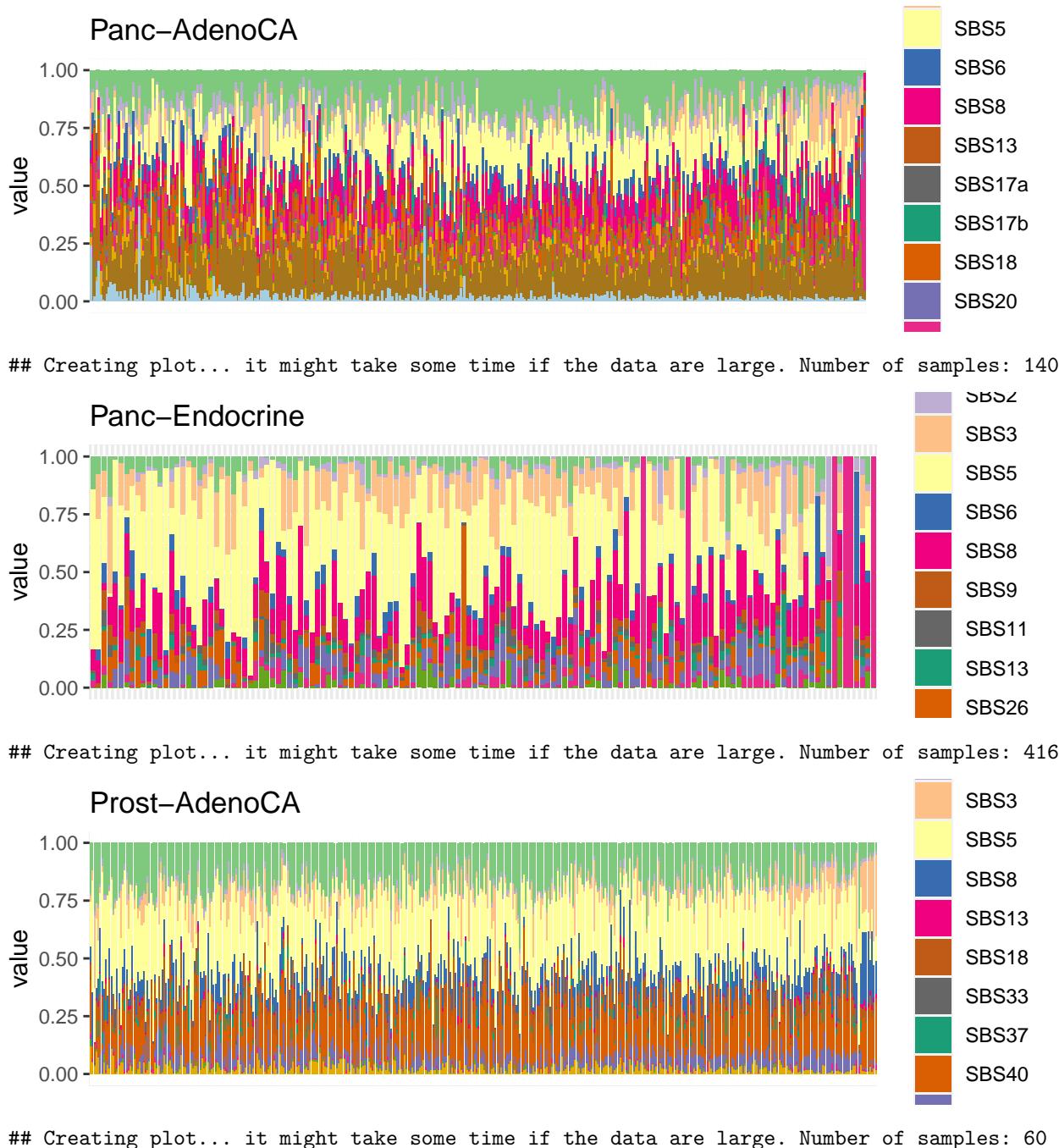


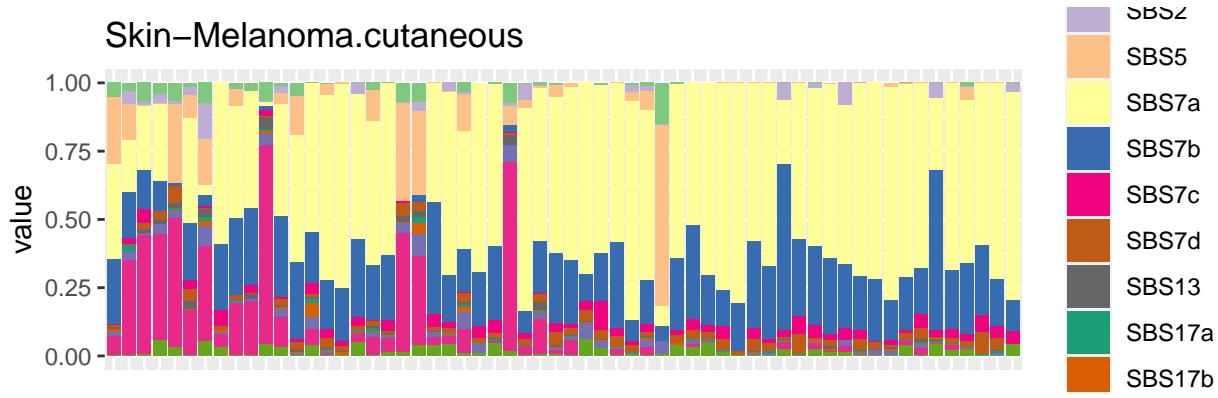
```
## Creating plot... it might take some time if the data are large. Number of samples: 68
```



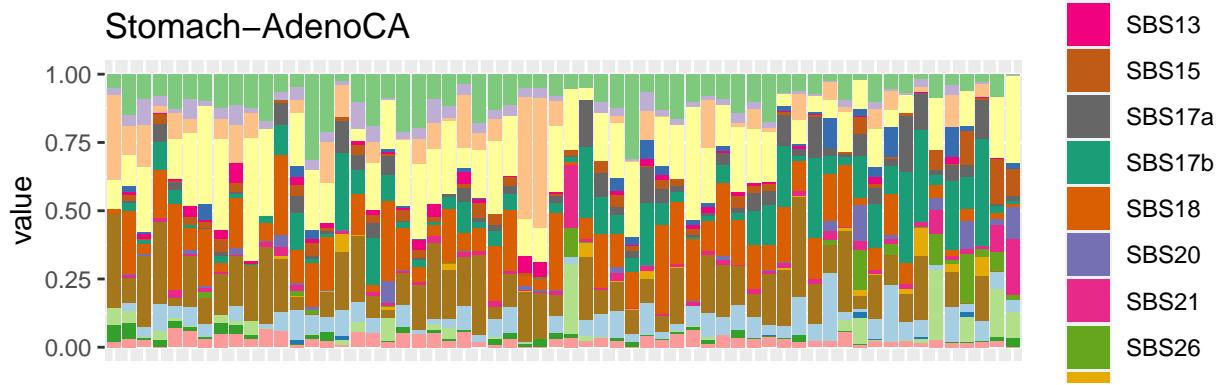
```
## Creating plot... it might take some time if the data are large. Number of samples: 102
```



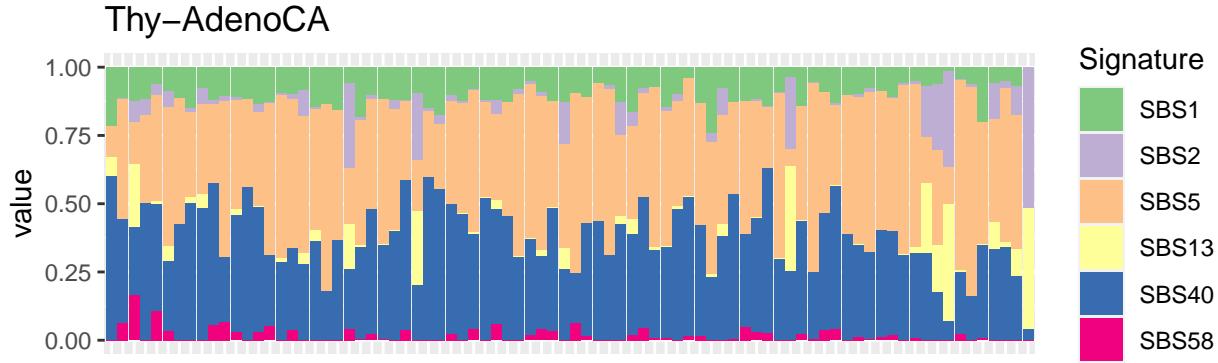




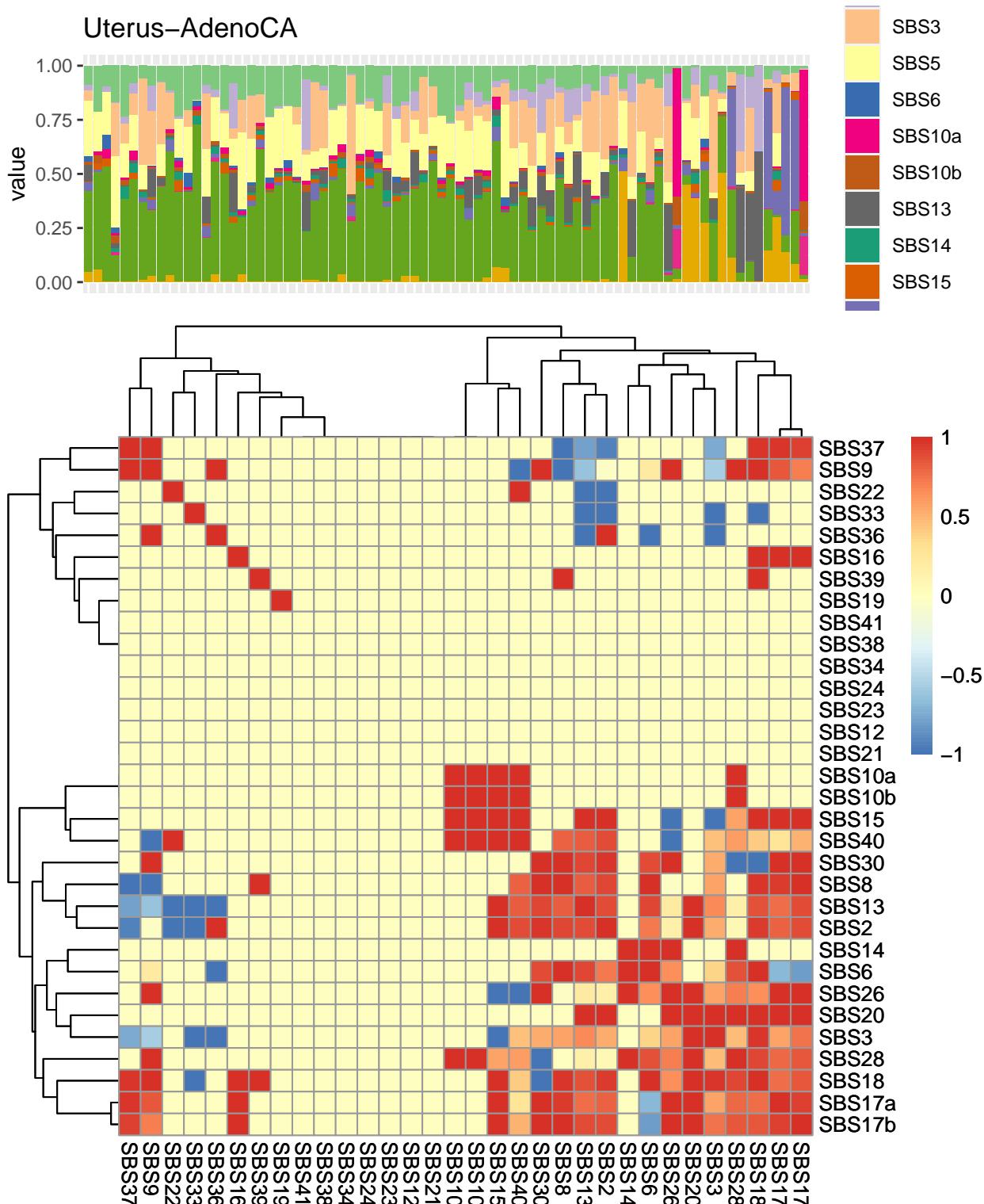
```
## Creating plot... it might take some time if the data are large. Number of samples: 60
```



```
## Creating plot... it might take some time if the data are large. Number of samples: 82
```



```
## Creating plot... it might take some time if the data are large. Number of samples: 80
```



```
## null device
##      1
```

```

## null device
##           1

cors_sigs_v2 <- outer(1:nrow(aaa_dcast), 1:nrow(aaa_dcast), Vectorize(function(i,j){
  if( sum(!is.na(aaa_dcast[i,]) & !is.na(aaa_dcast[j,])) <=2){
    NA ## if there are 2 or fewer points in common. if there are 2 the correlation is possible but always
  }else{
    try(cor(x = unlist(aaa_dcast[i,]), y = unlist(aaa_dcast[j,]), use = "pairwise.complete.obs"))
  }
}))
cors_sigs_v2 <- apply(cors_sigs_v2, 2, as.numeric)

rownames(cors_sigs_v2) <- colnames(cors_sigs_v2) <- paste0('SBS', rownames(aaa_dcast))

# rownames(cors_sigs) <- colnames(cors) <- rownames(aaa_dcast)

hclust_correlation_betas_logR <- hclust(dist(cors_sigs))

num_common_samples_logR <- outer(1:nrow(aaa_dcast), 1:nrow(aaa_dcast), Vectorize(function(i,j){
  try(sum(!is.na(unlist(aaa_dcast[i,])+unlist(aaa_dcast[j,])))))})
rownames(num_common_samples_logR) <- colnames(num_common_samples_logR) <- paste0('SBS', rownames(aaa_dcast))

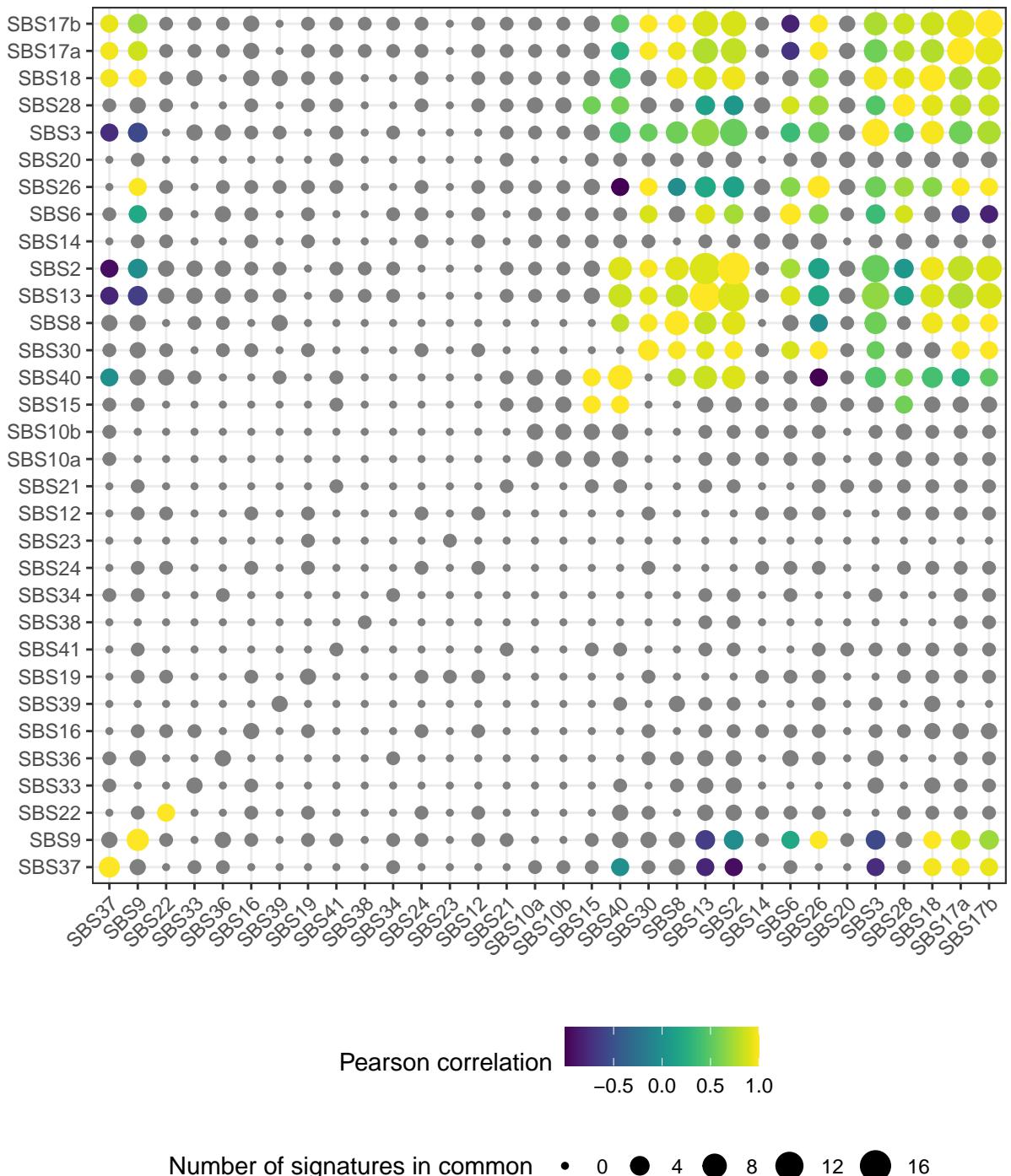
cors_melt_logR <- cbind.data.frame(cors=melt(cors_sigs_v2),
                                      num_common_samples=melt(num_common_samples_logR))

head(cors_melt_logR)

##   cors.Var1 cors.Var2 cors.value num_common_samples.Var1
## 1 SBS10a   SBS10a      NA        SBS10a
## 2 SBS10b   SBS10a      NA        SBS10b
## 3 SBS12    SBS10a      NA        SBS12
## 4 SBS13    SBS10a      NA        SBS13
## 5 SBS14    SBS10a      NA        SBS14
## 6 SBS15    SBS10a      NA        SBS15
##   num_common_samples.Var2 num_common_samples.value
## 1                      SBS10a                  2
## 2                      SBS10a                  2
## 3                      SBS10a                  0
## 4                      SBS10a                  1
## 5                      SBS10a                  1
## 6                      SBS10a                  2

ggplot(cors_melt_logR, aes(x=factor(num_common_samples.Var1,
                                         levels = hclust_correlation_betas_logR$labels[hclust_correlat
                                         y=factor(num_common_samples.Var2, levels=hclust_correlation_betas_logR$label
                                         col=cors.value, size=num_common_samples.value))+
  geom_point() + scale_color_viridis() + theme_bw() + theme(legend.position = "bottom", legend.box="vertical")
  theme(axis.text.x=element_text(angle = 45, hjust = 1, vjust=1))+
  labs(x=' ',y=' ', size='Number of signatures in common', col='Pearson correlation')

```



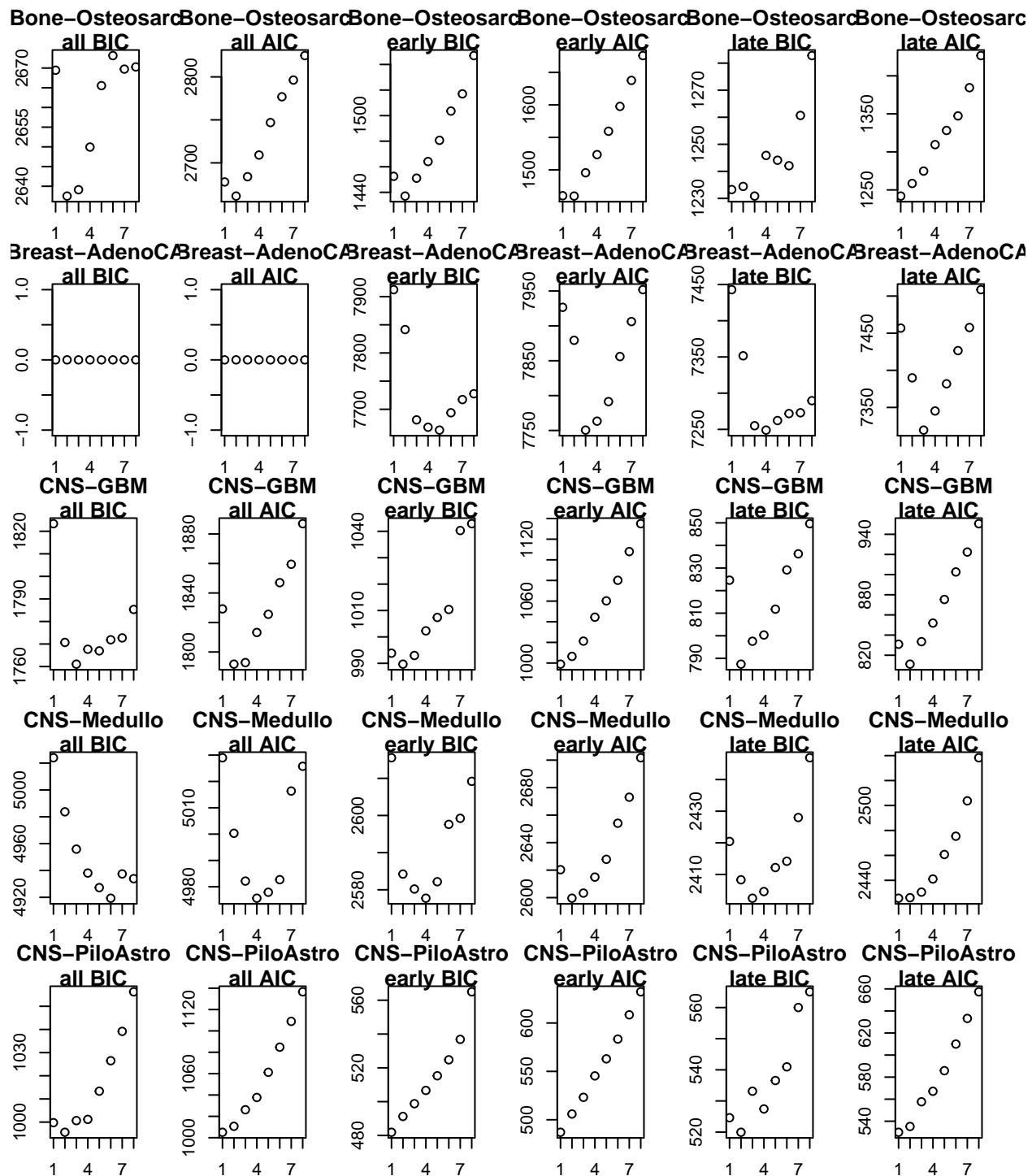
Dirichlet-Multinomial Mixtures

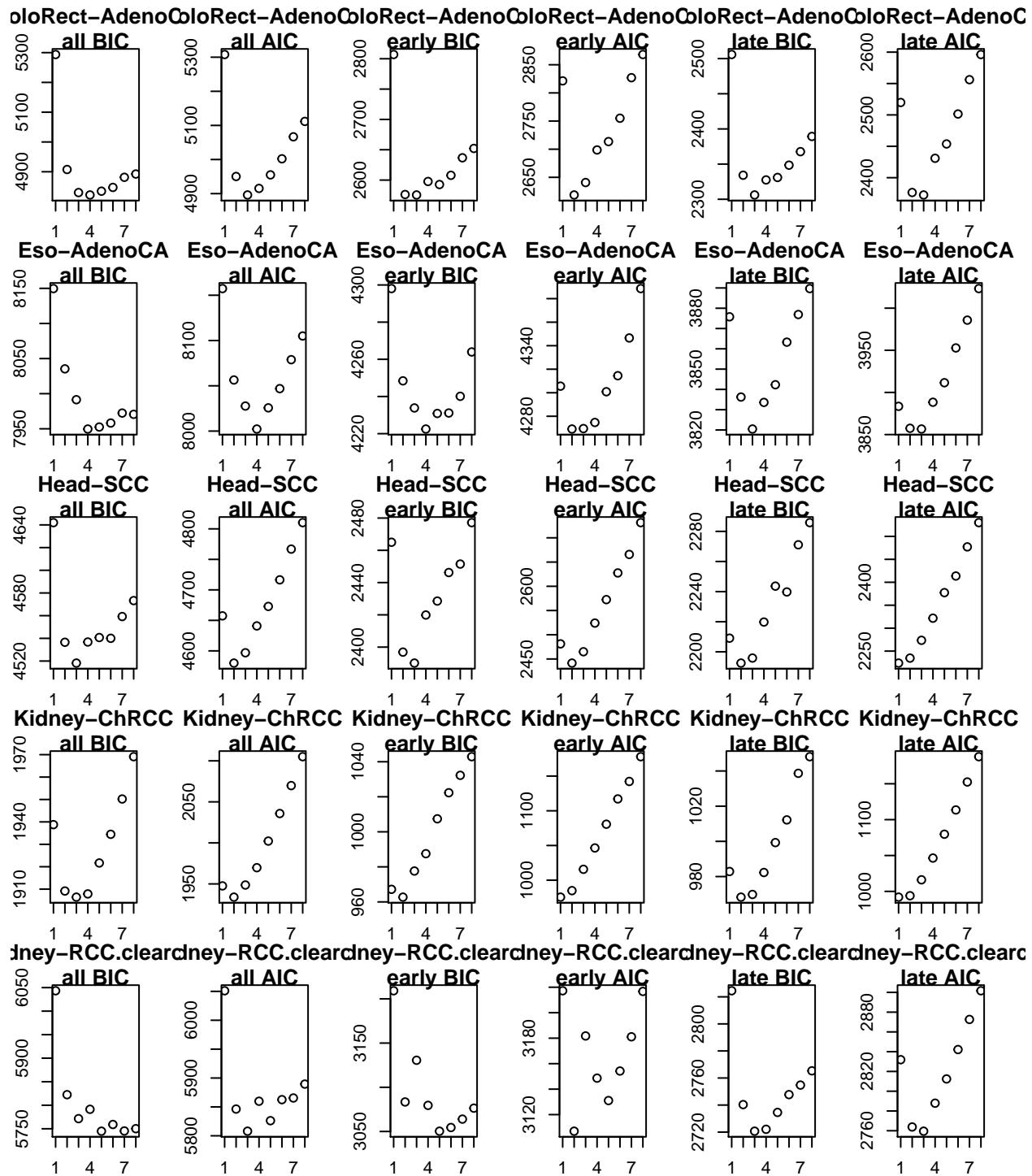
We run the software `MicrobeDMMv1.0` to determine whether we are facing DMM mixtures or not.

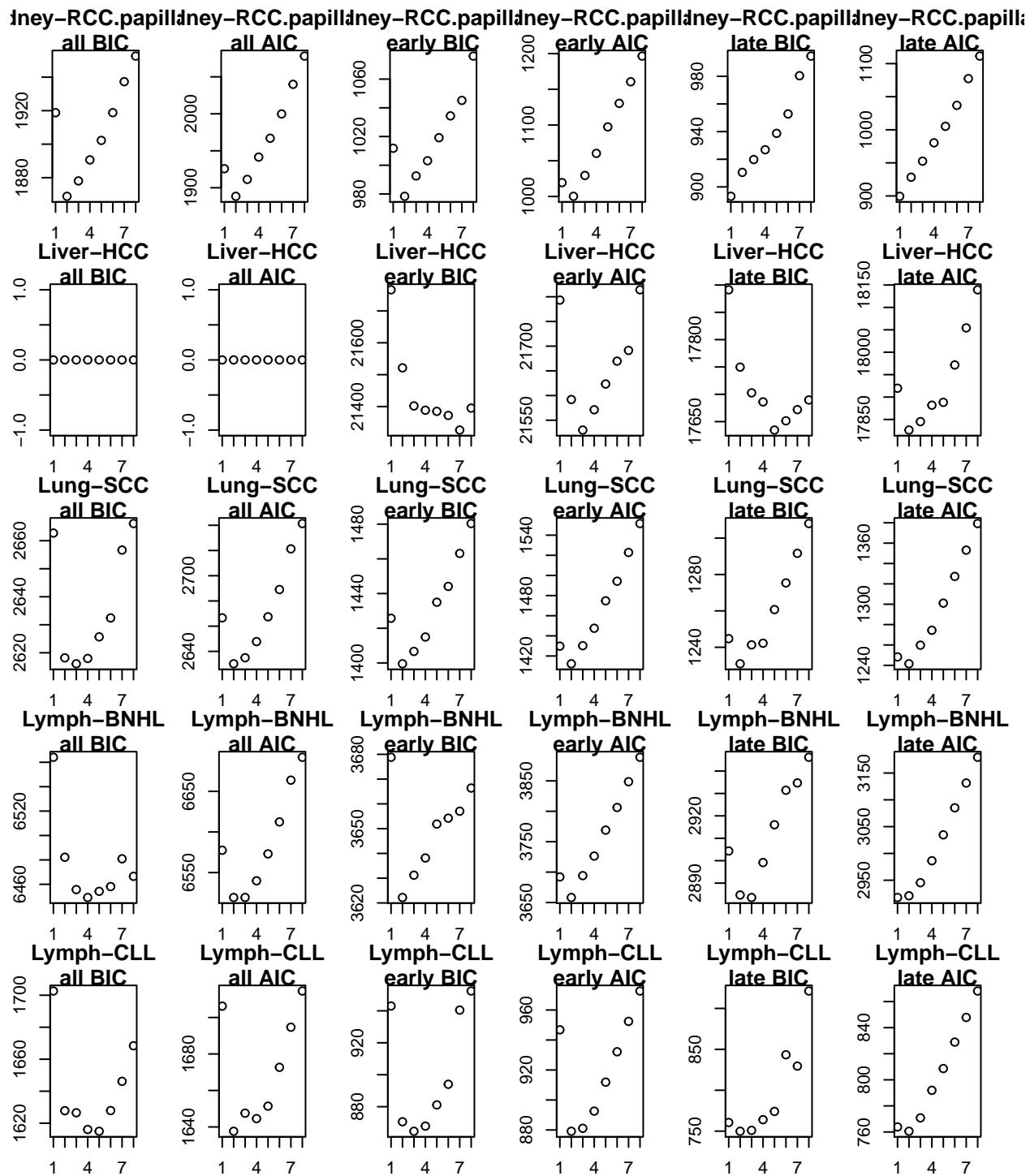
We save the files in two ways: all of the samples - early or not - together, and separately.

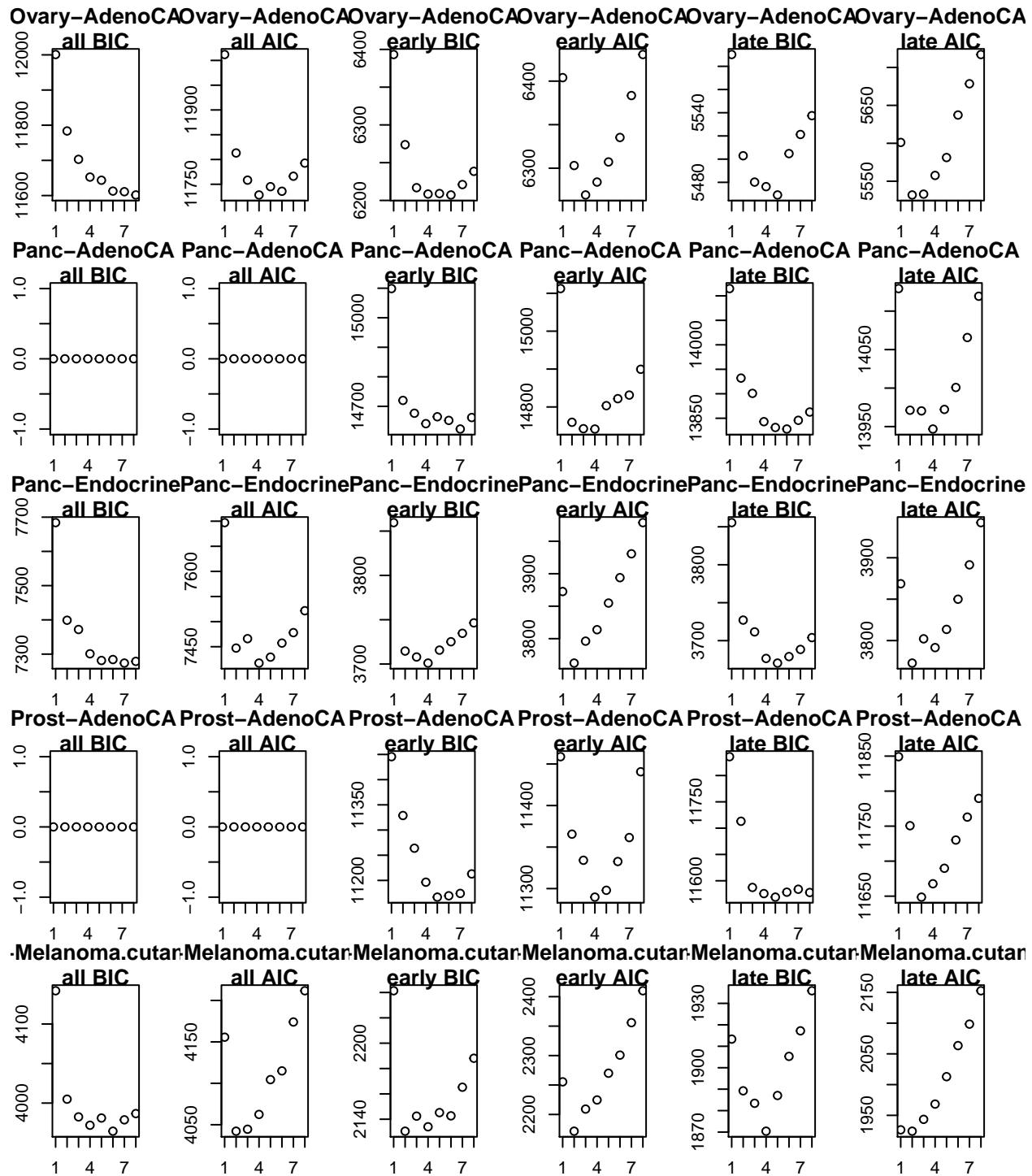
In some cases DMM says that there is an error with the input file - in this case the AIC or BIC is not plotted. If

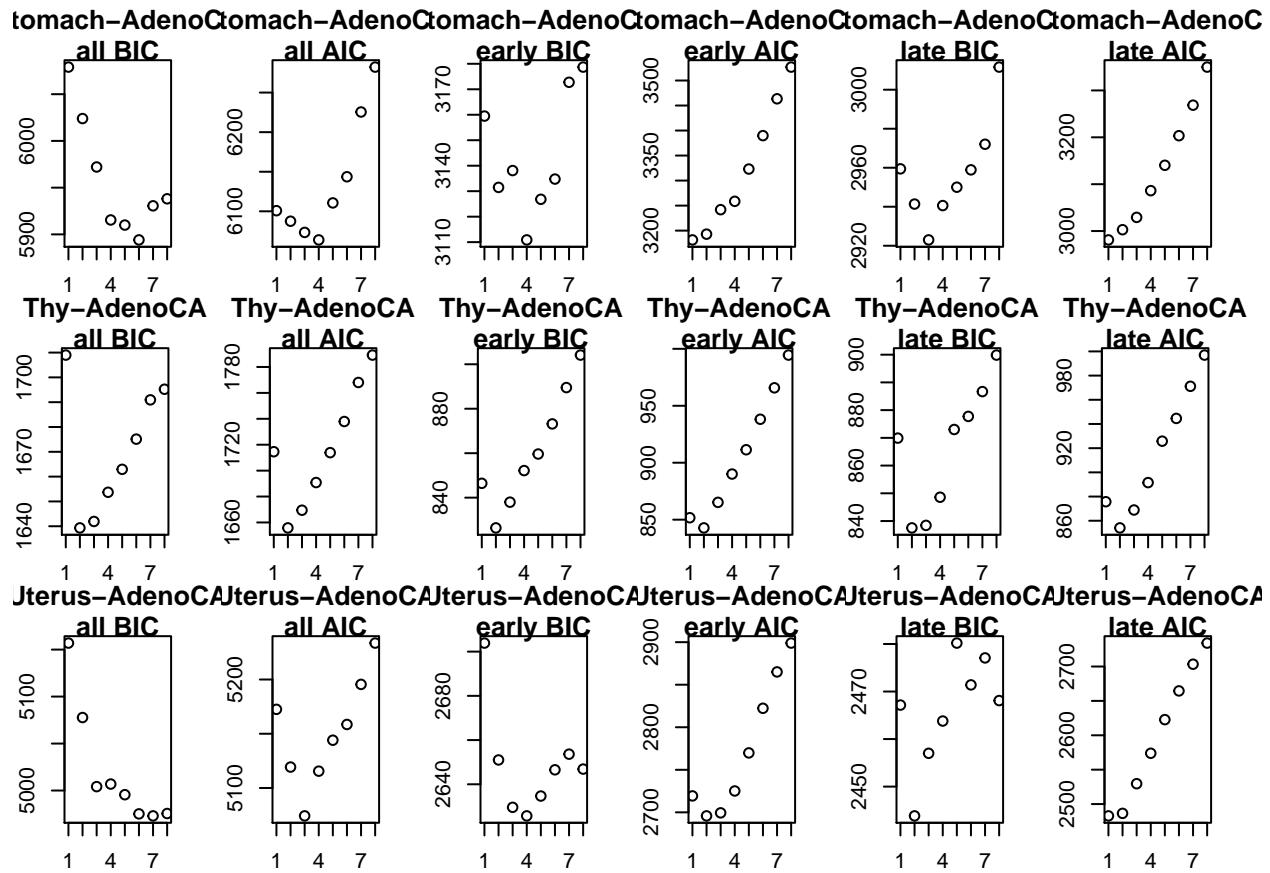
all of them are missing, all BIC and AIC are set to zero.







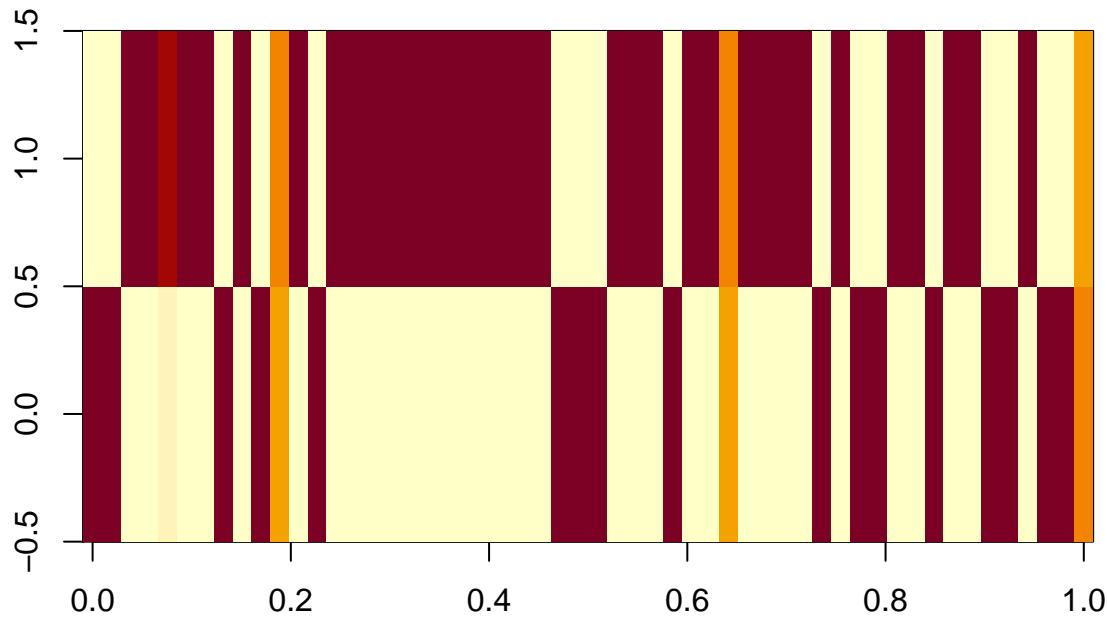




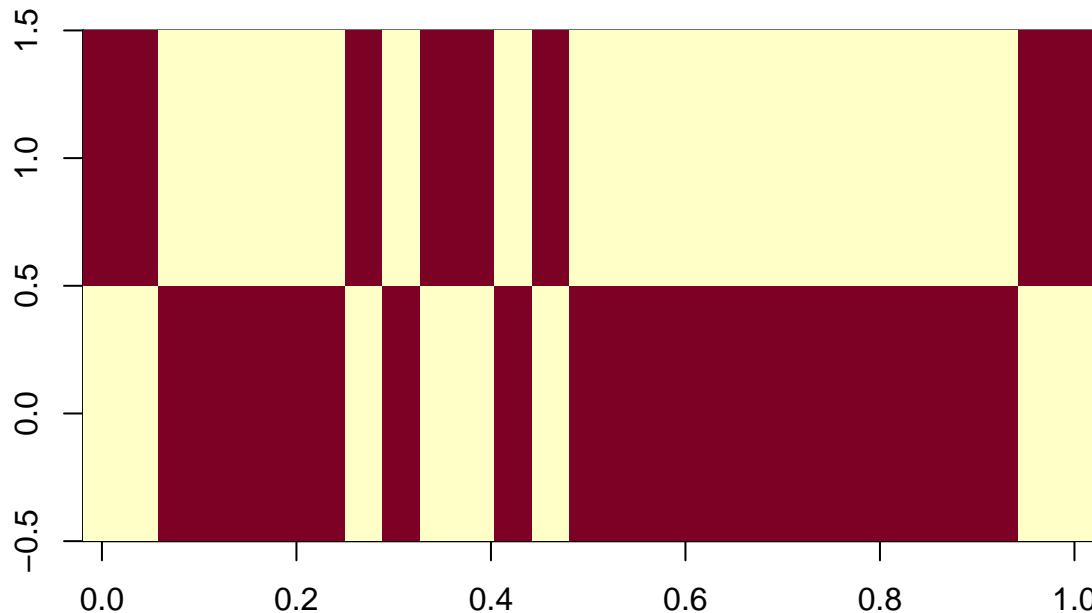
Now we have the optimal classes for each cancer type, both for early and late mutations separately, and for all of them together. We want to see if there is more than one group in general, and if the groups are maintained from early to late.

We look at the “all” mixtures, and look at the percentage of early/late paired observations that share group, and the percentage of early/late grouped samples that do.

```
image(as(z_DMM$`Bone-Osteosarc`$all[[2]][,-1], 'matrix'))
```



```
image(as(z_DMM$`Bone-Osteosarc`$early[[2]][,-1], 'matrix'))
```



```
give_matching_information_about_DMM <- function(ct, k){
  if(k!=2){stop('Only k==2 implemented')}
  .splitted_DMM <- split_matrix_in_half(as(z_DMM[[ct]]$all[[2]][,-1], 'matrix'))
  .whichmax1 <- apply(.splitted_DMM[[1]], 1, which.max)
  .whichmax2 <- apply(.splitted_DMM[[2]], 1, which.max)
  # percentage of early/late paired observations that share group
  .res1 <- max(c(mean(.whichmax1 == .whichmax2), mean(.whichmax1 != .whichmax2))) ## this is only for k==2
  #percentage of early/late grouped samples that share group
  .res2 <- mean(c(.whichmax1 == max(.whichmax1), .whichmax2 == max(.whichmax2)))
  return(list(pairedshared=.res1, groupshared=.res2))
}
```

```

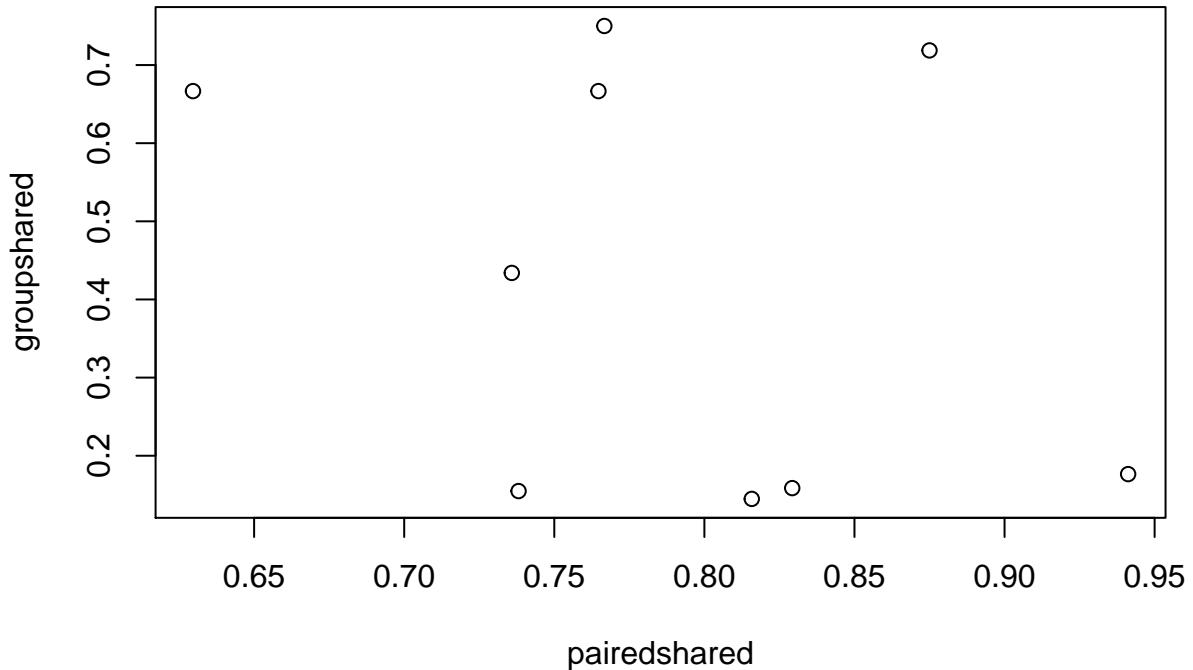
## for now, only the ones where there are two groups
matching_information_about_DMM_k2 <- sapply(c('Bone-Osteosarc', 'CNS-PiloAstro', 'Head-SCC', 'Kidney-ChRCC',
                                             'Lung-SCC', 'Lymph-BNHL', 'Lymph-CLL', 'Skin-Melanoma.cutaneous',
                                             'Thy-AdenoCA'), give_matching_information_about_DMM, k=2)

matching_information_about_DMM_k2

##          Bone-Osteosarc CNS-PiloAstro Head-SCC Kidney-ChRCC Lung-SCC
## pairedshared 0.6296296    0.7380952   0.875    0.8157895   0.9411765
## groupshared  0.6666667    0.1547619   0.71875   0.1447368   0.1764706
##          Lymph-BNHL Lymph-CLL Skin-Melanoma.cutaneous Thy-AdenoCA
## pairedshared 0.7647059   0.7358491   0.7666667   0.8292683
## groupshared  0.6666667   0.4339623   0.75        0.1585366

plot(t(matching_information_about_DMM_k2))

```



```

# ct = "Bone-Osteosarc"
# xxxx <- try(read.table(paste0("../data/roo_for_DMM/DMM_output/", ct, "_signatures_all.z"), sep = ',',
# length(xxxx)
# table(apply(xxxx[grep('early', xxxx$V1), -1], 1, which.max),
#       apply(xxxx[grep('late', xxxx$V1), -1], 1, which.max))

```