05-hiernet-weak-robust-10uM-new-data

Compiled at 2023-01-19 23:16:27 UTC

```
here::i_am(paste0(params$name, ".Rmd"), uuid = "e1af8f1a-2477-40d6-8300-c9019db89425")
knitr::opts_chunk$set(dpi = 200, echo = T, warning = F)
```

We use hiernet -a Lasso for (weak) hierarchical interactions- with CPSS to robustly recover interaction effects between compounds on cell morphology features. In addition we add a Huber loss to the optimization problem to account for outlier in the data. We only focus on experiments with a concentration of 10 uM per compound and use the updated data in this analysis.

Read data

Input data X

```
dim(X)
```

```
## [1] 443  21
```

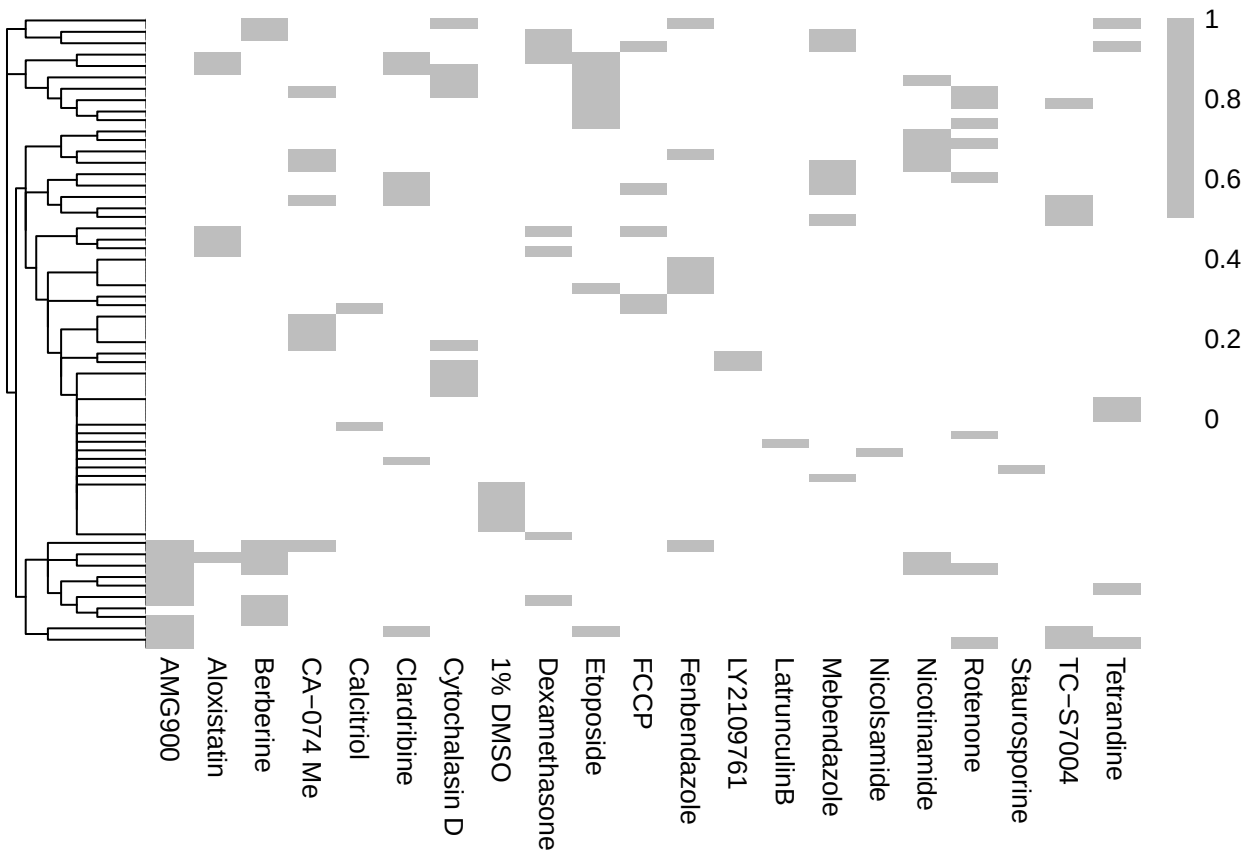
```
X[1:5, 1:5]
```

```
##       AMG900 Aloxistatin Berberine CA-074 Me Calcitriol
## 1-C03      0           0         0         0          0
## 1-C04      0           0         0         0          0
## 1-C05      0           1         0         0          0
## 1-C06      0           0         0         0          0
## 1-C07      0           0         0         0          0
```

```
colnames(X)[which(colnames(X) == "DMSO")] <- "1% DMSO"
```

Plot Design matrix X

```
pheatmap(X, cluster_rows = T, cluster_cols = F, color = c("white", "grey"),
         fontsize_row = 0.001)
```



Remove DMSO

```
# First, remove all rows (experiments) containing DMSO in X as well as in Y

# check if rows are ordered in the same way in X and Y
all(rownames(X) == rownames(Y))
```

```
## [1] TRUE
```

```
X_initital <- X
DMSO_experiments <- which(X[, "1% DMSO"] == 0)
X <- X[DMSO_experiments, ]

Y <- Y[DMSO_experiments,]
# Then remove DMSO columns (since only zeros remain)
ind_DMSO <- which(colnames(X) == "1% DMSO")
X <- X[,-ind_DMSO]

dim(X)
```

```
## [1] 408  20
```

The experimental design consists of 408 samples and 20 compounds.

Remove all experiments with a concentration of only 5 uM

It's already only the 10uM experiments in the updated data, so there's no need to remove experiments anymore.

Compute interactions

```
X_interactions <- cbind(X, hierNet::compute.interactions.c(X, diagonal = F))
dim(X_interactions)
```

```
## [1] 408 210
```

There are 210 interaction coefficients when considering all-pairs of interactions from 20 compounds.

How many replicates do exist?

```
X_unique <- X[!duplicated(X),]
dim(X_unique)
```

```
## [1] 50 20
```

There are 50 different experiments.

```
## index_replicate
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
##  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6 18 18 18 18  8  8  8  8  8  8
## 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
##  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8
```

The individual experiments were repeated 6, 8 or 18 times.

```
dim(Y)
```

```
## [1] 408 146
```

For each experiment 146 features were observed.

Scale Features Y

We scale the features to make our model coefficients comparable.

```
Ysc = scale(Y)
```

Let's have a look at a subset of scaled Y

```
Ysc[1:5, 1:4]
```

```
##       ValidObjectCount ObjectTotalAreaCh1 ObjectAvgAreaCh1 ObjectTotalIntenCh1
## 1-C05        1.4732718           1.459982       0.09479254           1.3242847
## 1-C06        1.9157636           2.004499       0.18774959           1.7112856
## 1-C07        0.4192962           0.143315      -0.89978854           0.2762874
## 1-C08        0.9284954           0.882009       0.34920378           0.8804891
## 1-C09        2.0447311           2.148380       0.17045864           1.8024286
```

Check if replicates do differ via heatmap plots. The heatmap plots can be found in a separate document compare_replicates_10uM_newData.pdf.

We account for outlier samples by using a robust loss. We do not exclude any experiments in advance anymore.

Robust HierNet

In this version the design matrix X includes duplicated rows representing replicates in the experiment.

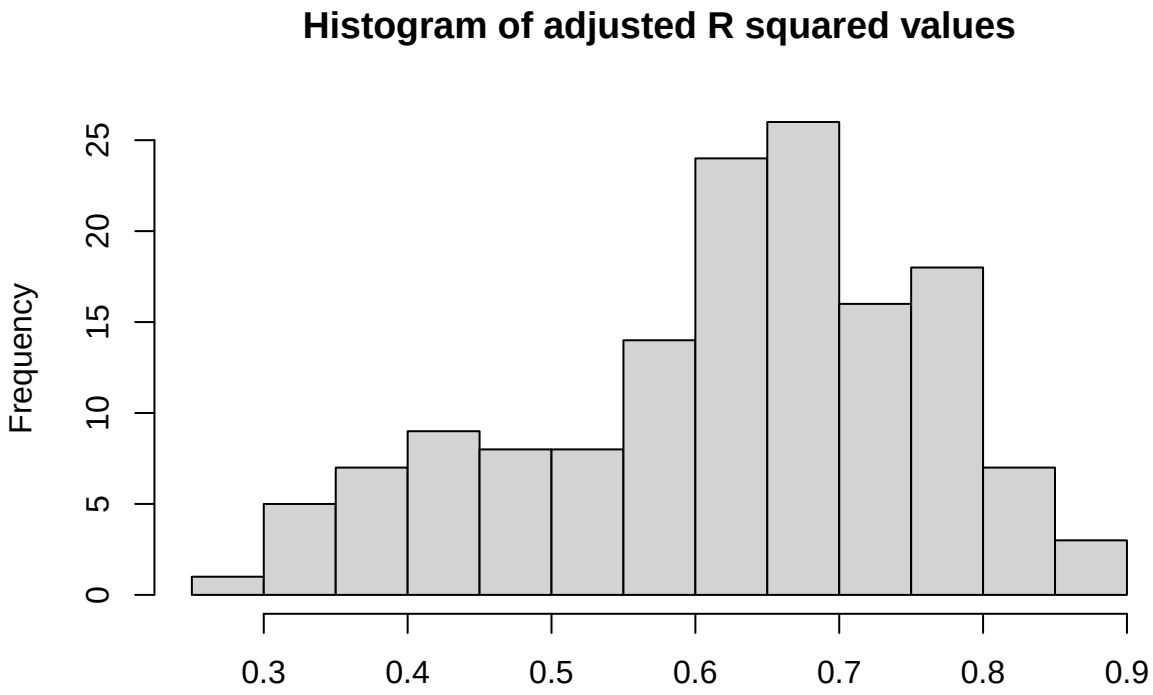
hierNet with Huber loss

```
fit.list.weak.10uM <- hiernet.stabsel.all(X = X, Y = Ysc,
                                          selection_probability = 0.6)
```

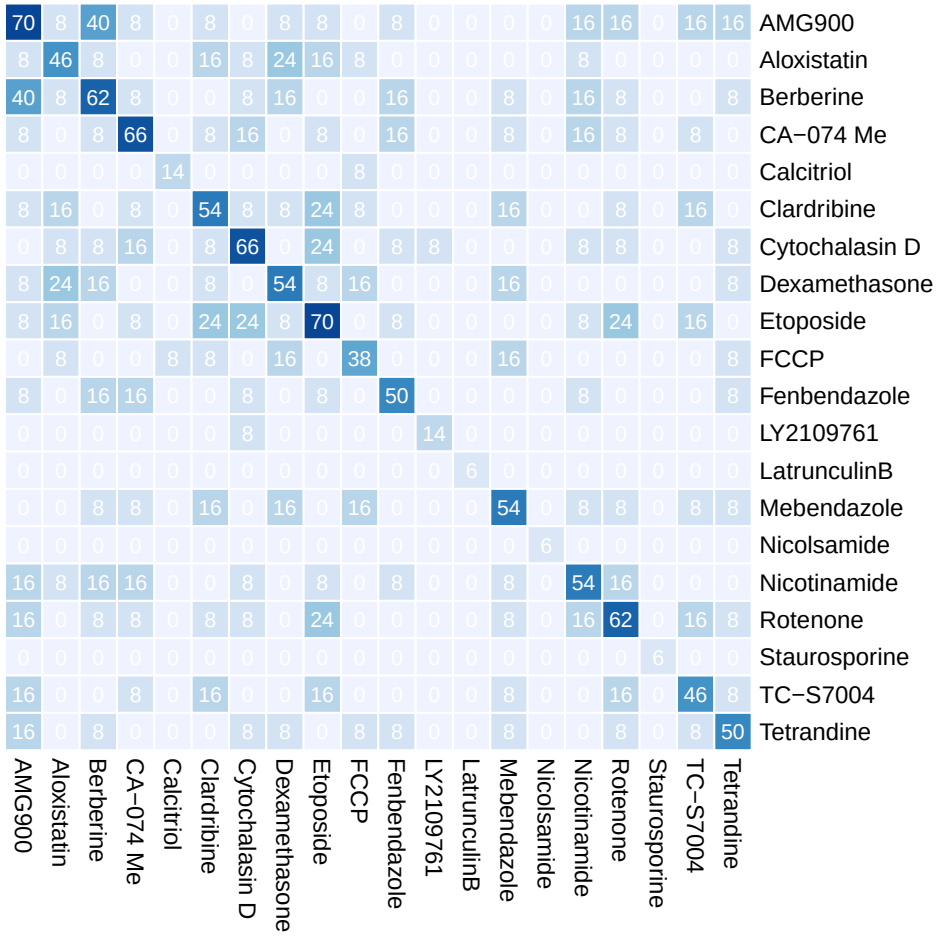
Refitting

```
refit_10uM <- refit.all(X = X, Y = Ysc, fit_list_weak =
                                    fit.list.weak.10uM)
refit_list_weak_10uM <- refit_10uM$refit_list
coef_list_weak_10uM <- refit_10uM$coef_list
```

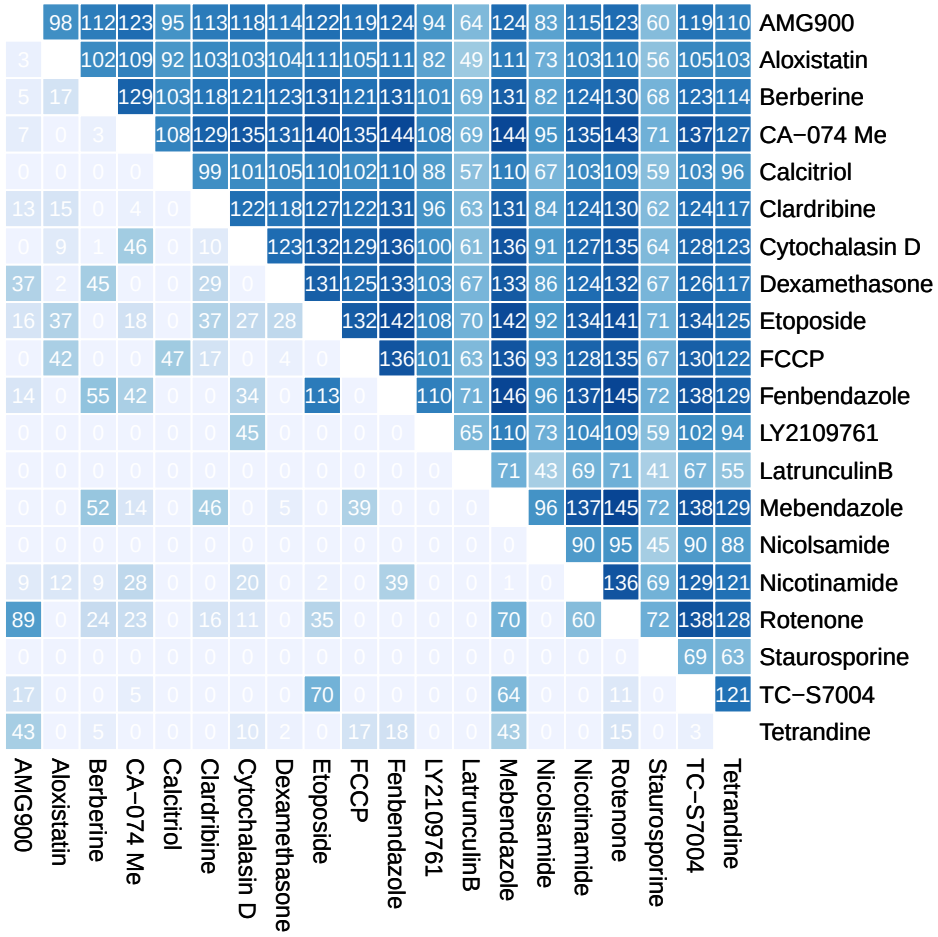
How good is the predictive accuracy?



How often do compounds occur together in the experimental design?



How often do we observe what?

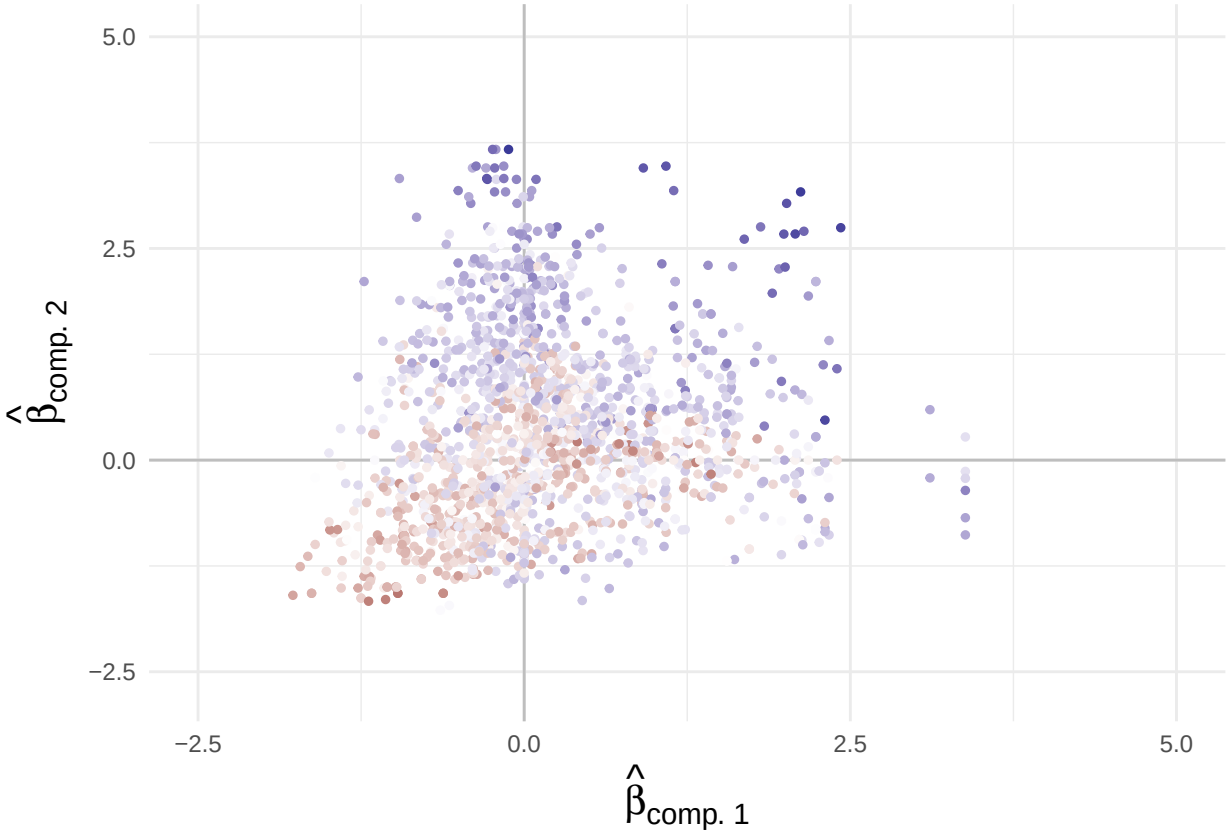


We can only find combinatorial effects for compounds that got measured together in an experiment. This means that out results are biased by the experimental design. What might be of speciel interest are combinations of compounds that occur often in the experiments, but which do not (or almost not) show combinatorial effects in our model.

For example Clardribine and Aloxistatin

Overall this looks similar to the results in the old data.

All features into one plot

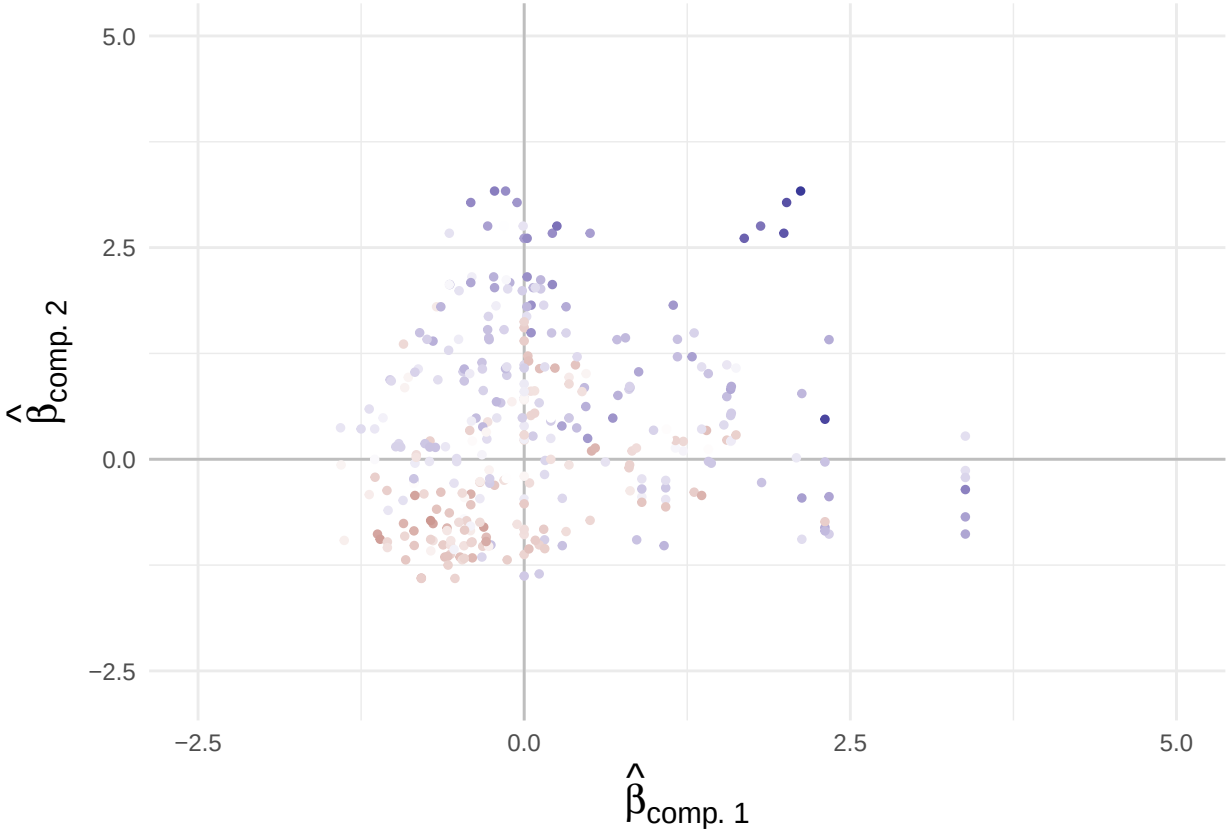


We can't label all features here, but we get a rough idea what kind of effects are most prevalent. We see, for instance, many antagonistic effects, where two compounds do have individual positive effects on a feature, but in combination this effect is not that strong anymore (upper right quadrant, blue points). We also see this behavior into the other direction, where tow individual effects are negative, but the additional combinatorial effect is positive (lower left quadrant, red points). We also see multiple competitive (or even cancelling out) effects between the compounds, where the two compounds show an opposite behavior and the additional combinatorial effect promotes on of the two behaviors.

This plot looks a bit shifted compared to the old data.

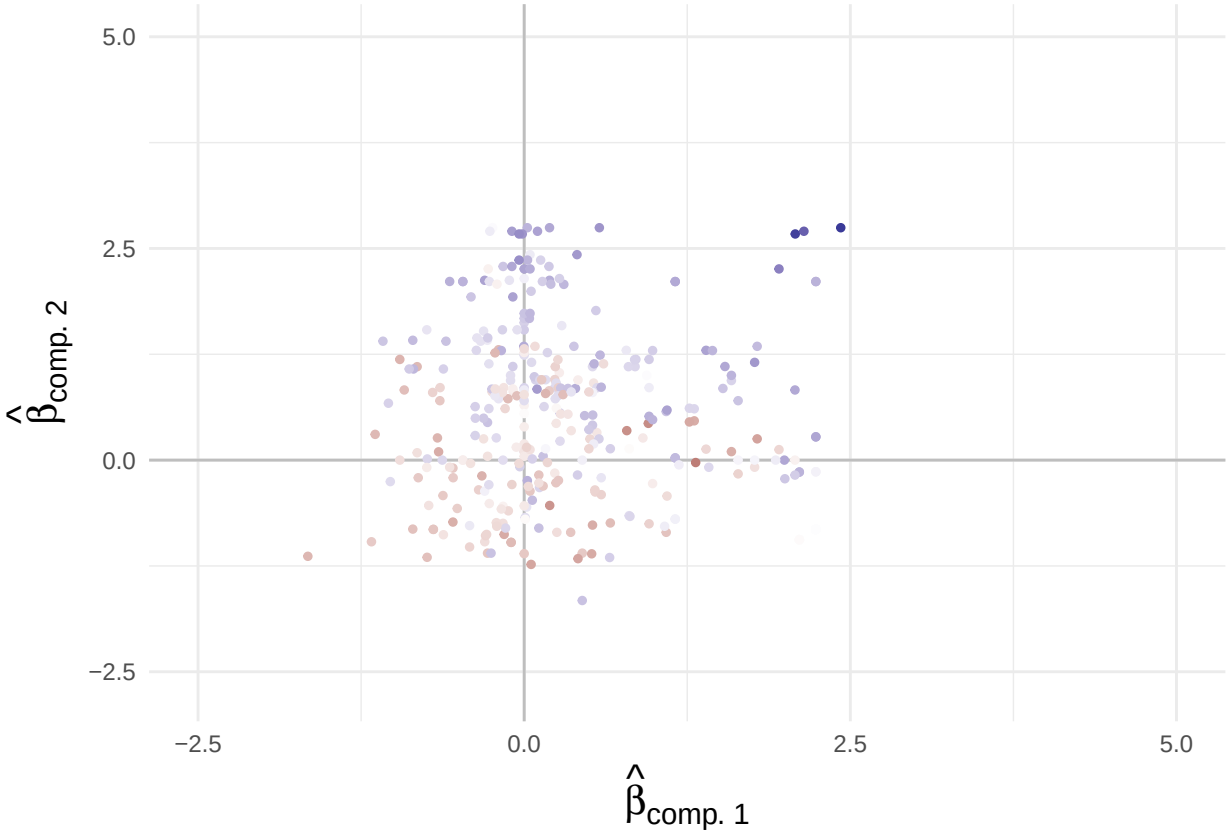
What does this figure look like if we split it into Channels?

Channel 1: Nucleus

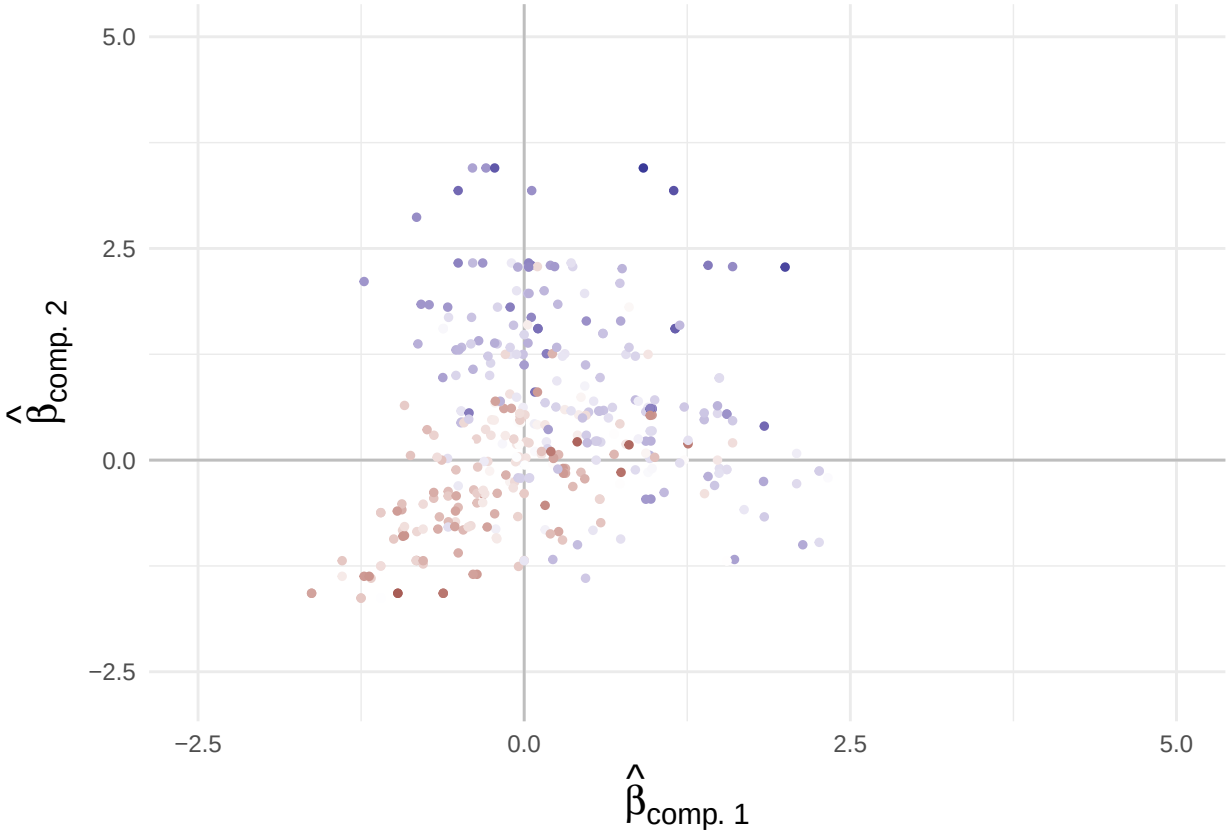


The behavior in the Nucleus is very similar to the overall behavior.

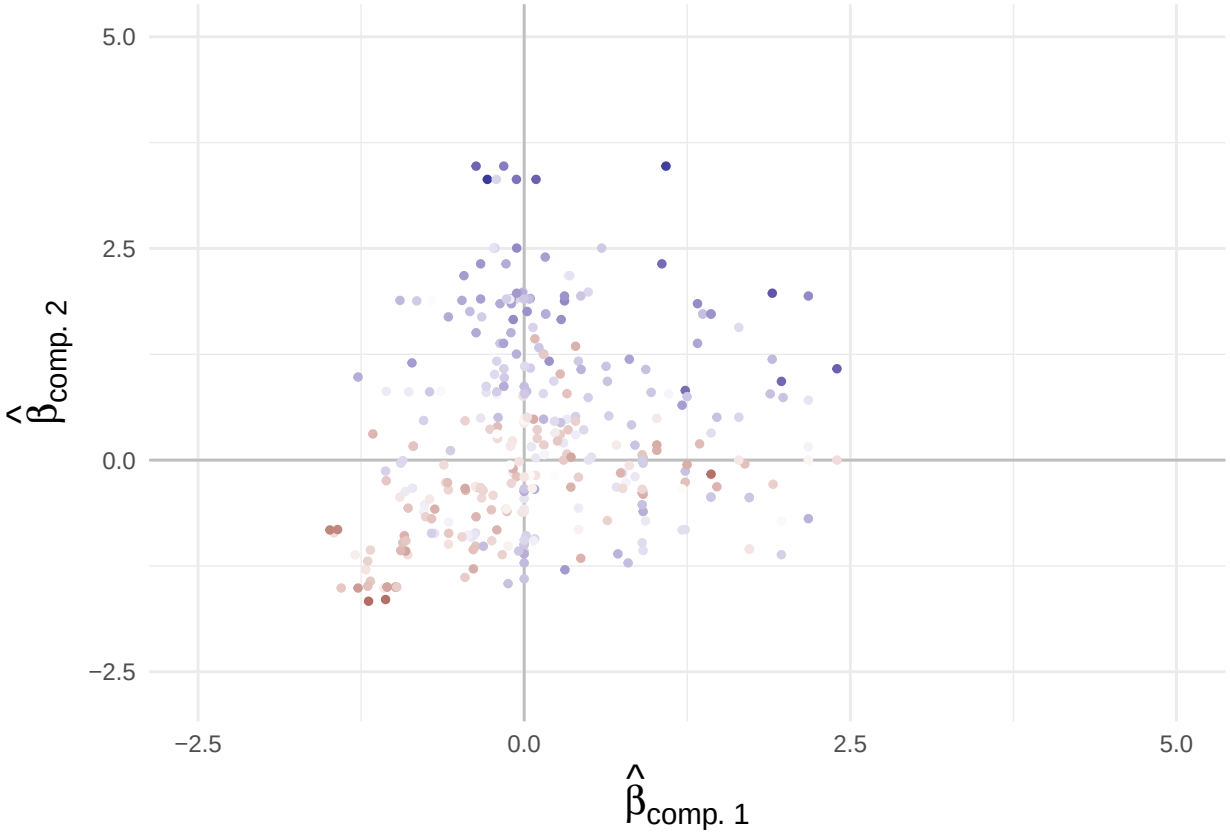
Channel 2: Endoplasmic reticulum



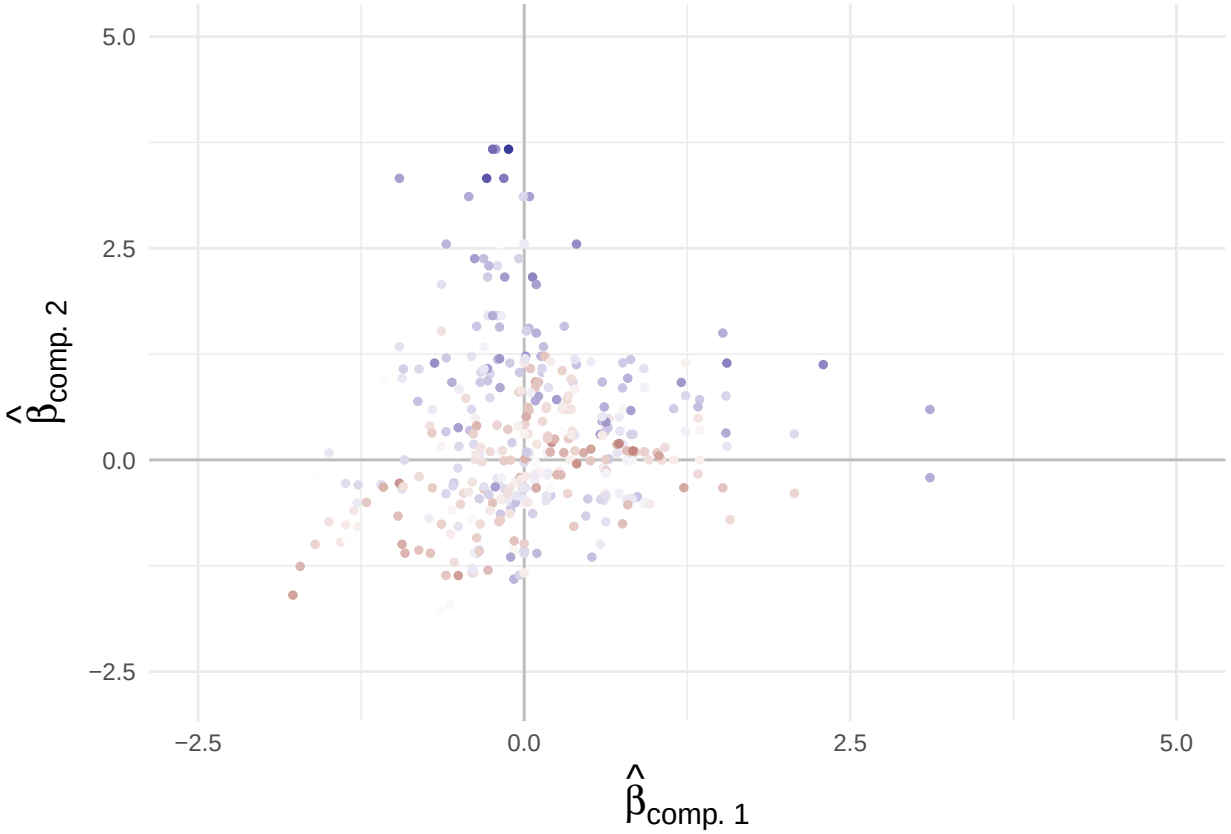
Channel 3: Nucleoli, cytoplasmic RNA



Channel 4: Golgi, plasma membrane, F-actin, cytoskeleton



Channel 5: Mitochondria



Specific compound combinations

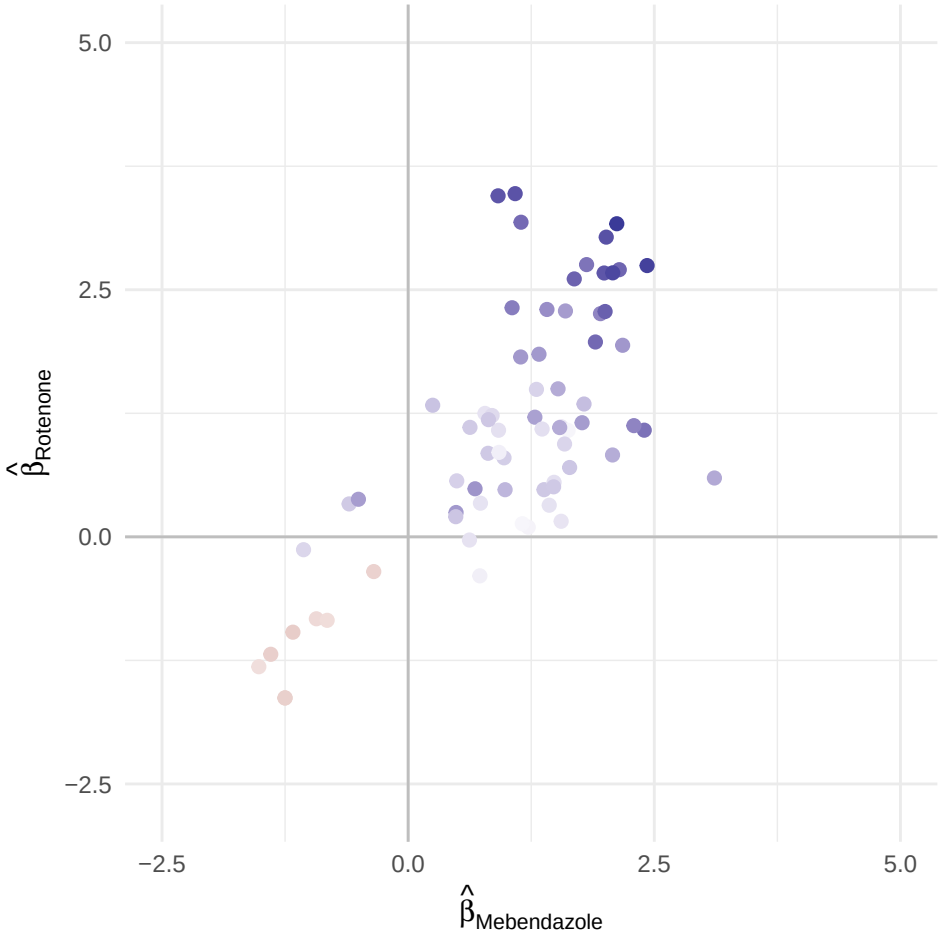
Instead of looking into different channels, we can also look into this representation by taking into account specific pairs of compounds. The Figures are in a separate document (see scatter_plt_each_combi_robust_10uM_all_newData.pdf)

There are different categories of behavior:

1. Antagonists: In this category two compounds show similar individual effects, but in combination their effect is lowered (less strong then expected under independence of the two individual effects or even cancelled out).

In this summary I'm only focusing on combinations that affect many features.

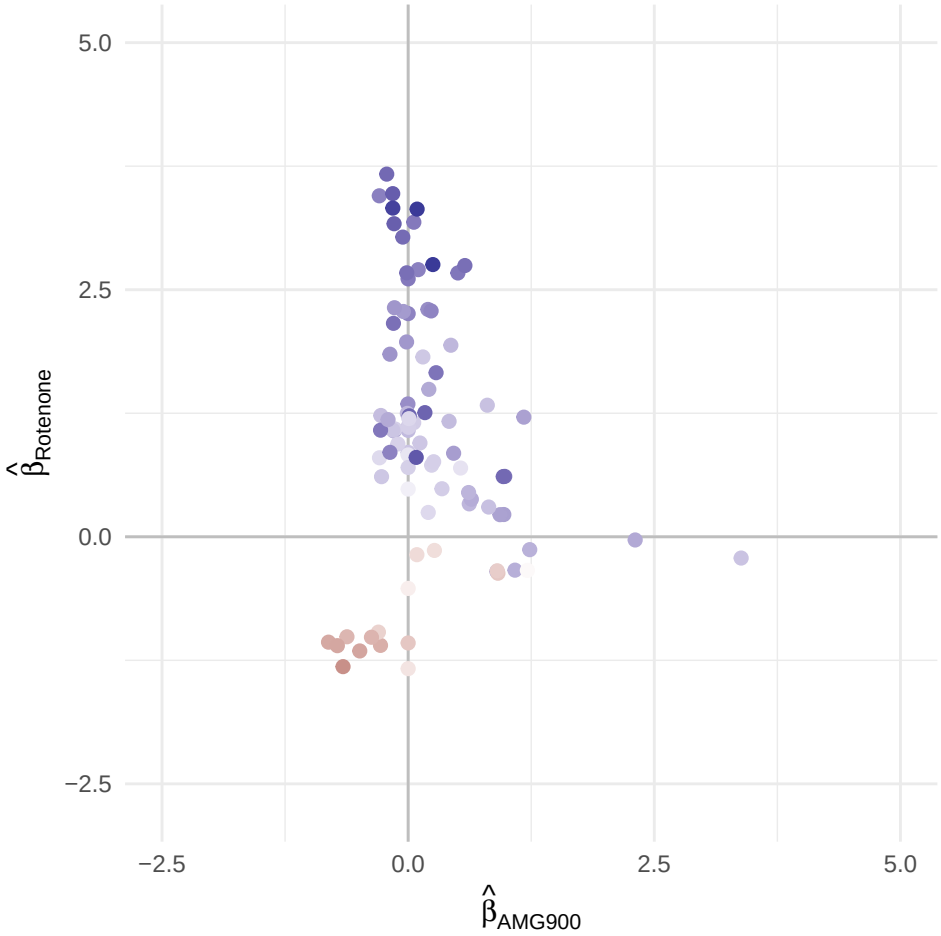
Here's an example:



- Rotenone and Mebendazole
- Tetrandine and Mebendazole
- Tetrandine and FCCP
- Mebendazole and FCCP
- Dexamethasone and Berberine
- Rotenone and CA-074
- Cladribine and Etoposide
- Aloxistatin and Berberine

2. Cancelling out effect (weak hierarchy): In this category a compound has a specific effect on the features, but this behavior gets reduced in combination with another compound that on its own does have no (or almost no) effect.

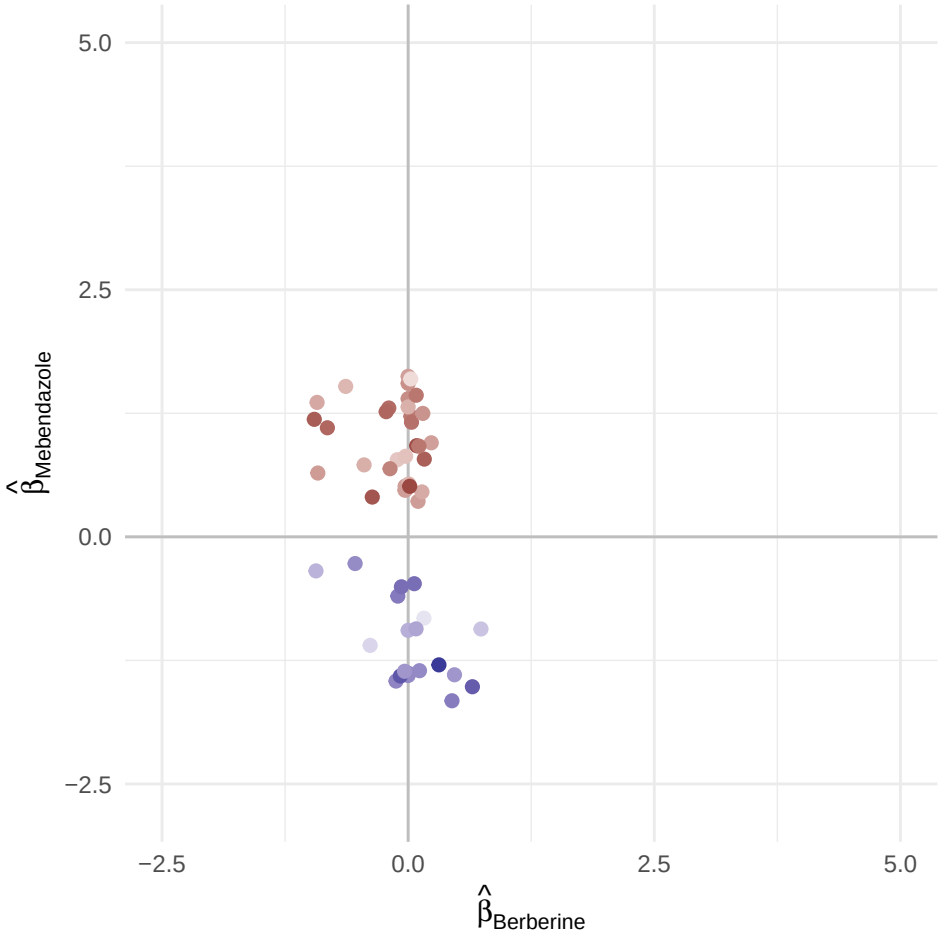
Here's an example:



- Positive (and negative) Rotenone effect suppressed by AMG900
- Positive and negative Rotenone effect suppressed by Berberine
- Positive and negative Mebendazole effect suppressed by Berberine
- Positive and negative TC-S7004 effect suppressed by Mebendazole
- Positive Mebendazole effect suppressed by Cladribine
- Positive Rotenone effect suppressed by Cladribine
- Positive Rotenone effect suppressed by Etoposide

3. Amplification effect

For Mebendazole and Berberine there's another interesting behavior. Berberine on it's own doesn't show a strong effect, but it amplifies the effect of Mebendazole.



We see such an effect for:

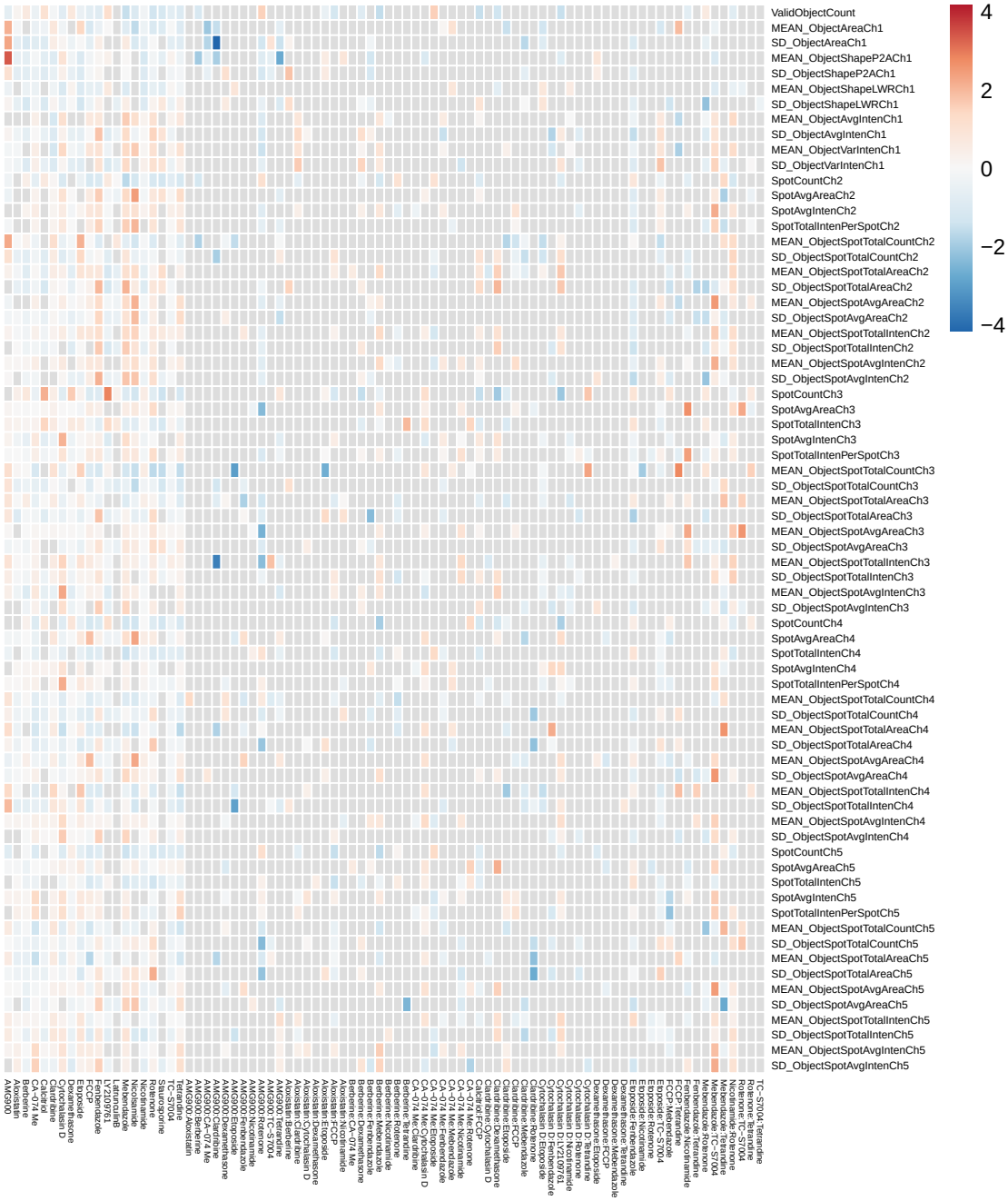
- Mebendazole and Berberine
- Mebendazole and TC-S7004
- (also a little bit for) Rotenone and Nicotinamide

70 most important features

Since 146 Features are a lot, we reduce the set to 70 Features

```
## New names:
## * `` -> `...1`
```

The following heatmap shows the estimated coefficients from the model starting with linear effects on the left followed by all interaction effects that are non zero for at least one morphological feature.


ValidObjectCount
MEAN_ObjectAreaCh1
SD_ObjectAreaCh1
MEAN_ObjectShapeP2ACh1
SD_ObjectShapeP2ACh1
MEAN_ObjectShapeLWRCh1
SD_ObjectShapeLWRCh1
MEAN_ObjectAvgIntenCh1
SD_ObjectAvgIntenCh1
MEAN_ObjectVarIntenCh1
SD_ObjectVarIntenCh1
SpotCountCh2
SpotAvgAreaCh2
SpotAvgIntenCh2
SpotTotalIntenPerSpotCh2
MEAN_ObjectSpotTotalCountCh2
SD_ObjectSpotTotalCountCh2
MEAN_ObjectSpotTotalAreaCh2
SD_ObjectSpotTotalAreaCh2
MEAN_ObjectSpotAvgAreaCh2
SD_ObjectSpotAvgAreaCh2
MEAN_ObjectSpotTotalIntenCh2
SD_ObjectSpotTotalIntenCh2
MEAN_ObjectSpotAvgIntenCh2
SD_ObjectSpotAvgIntenCh2
SpotCountCh3
SpotAvgAreaCh3
SpotTotalIntenCh3
SpotAvgIntenCh3
SpotTotalIntenPerSpotCh3
MEAN_ObjectSpotTotalCountCh3
SD_ObjectSpotTotalCountCh3
MEAN_ObjectSpotTotalAreaCh3
SD_ObjectSpotTotalAreaCh3
MEAN_ObjectSpotAvgAreaCh3
SD_ObjectSpotAvgAreaCh3
MEAN_ObjectSpotTotalIntenCh3
SD_ObjectSpotTotalIntenCh3
MEAN_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgIntenCh3
SpotCountCh4
SpotAvgAreaCh4
SpotTotalIntenCh4
SpotAvgIntenCh4
SpotTotalIntenPerSpotCh4
MEAN_ObjectSpotTotalCountCh4
SD_ObjectSpotTotalCountCh4
MEAN_ObjectSpotTotalAreaCh4
SD_ObjectSpotTotalAreaCh4
MEAN_ObjectSpotAvgAreaCh4
SD_ObjectSpotAvgAreaCh4
MEAN_ObjectSpotTotalIntenCh4
SD_ObjectSpotTotalIntenCh4
MEAN_ObjectSpotAvgIntenCh4
SD_ObjectSpotAvgIntenCh4
SpotCountCh5
SpotAvgAreaCh5
SpotTotalIntenCh5
SpotAvgIntenCh5
SpotTotalIntenPerSpotCh5
MEAN_ObjectSpotTotalCountCh5
SD_ObjectSpotTotalCountCh5
MEAN_ObjectSpotTotalAreaCh5
SD_ObjectSpotTotalAreaCh5
MEAN_ObjectSpotAvgAreaCh5
SD_ObjectSpotAvgAreaCh5
MEAN_ObjectSpotTotalIntenCh5
SD_ObjectSpotTotalIntenCh5
MEAN_ObjectSpotAvgIntenCh5
SD_ObjectSpotAvgIntenCh5
4
2
0
−2
−4
AM9300
Abrotasin
Berberine
CA-074 Me
Cladribine
Cytochalasin D
Dexamethasone
Etoposide
FCCP
Mebendazole
LY2109761
Latrunculin B
Meclofenamide
Metronidazole
Nicotinamide
Rotenone
Staurosporine
TC−S7004
Tetrandine
AM9300.Abrotasin
AM9300.Berberine
AM9300.CA−074 Me
AM9300.Cladribine
AM9300.Cytochalasin D
AM9300.Dexamethasone
AM9300.Etoposide
AM9300.Fenbendazole
AM9300.Nicotinamide
AM9300.Rotenone
AM9300.TC−S7004
AM9300.Tetrandine
Abrotasin.Cladribine
Abrotasin.Cytochalasin D
Abrotasin.Dexamethasone
Abrotasin.Etoposide
Abrotasin.FCCP
Abrotasin.Fenbendazole
Abrotasin.Nicotinamide
Abrotasin.Rotenone
Abrotasin.Tetrandine
Berberine.Cytochalasin D
Berberine.Dexamethasone
Berberine.Etoposide
Berberine.Fenbendazole
Berberine.Mebendazole
Berberine.Nicotinamide
Berberine.Rotenone
Berberine.Tetrandine
CA-074 Me.Cytochalasin D
CA-074 Me.Etoposide
CA-074 Me.Fenbendazole
CA-074 Me.Mebendazole
CA-074 Me.Nicotinamide
CA-074 Me.Rotenone
CA-074 Me.Tetrandine
Cladribine.FCCP
Cladribine.Cytochalasin D
Cladribine.Dexamethasone
Cladribine.Etoposide
Cladribine.FCCP
Cladribine.Mebendazole
Cladribine.Rotenone
Cytochalasin D.Fenbendazole
Cytochalasin D.Nicotinamide
Cytochalasin D.LY2109761
Cytochalasin D.Rotenone
Cytochalasin D.Tetrandine
Dexamethasone.Etoposide
Dexamethasone.FCCP
Dexamethasone.Mebendazole
Dexamethasone.Nicotinamide
Dexamethasone.Rotenone
Dexamethasone.Tetrandine
Etoposide.Nicotinamide
Etoposide.Rotenone
Etoposide.TC−S7004
Etoposide.Tetrandine
FCCP.Mebendazole
FCCP.Nicotinamide
FCCP.Rotenone
FCCP.Tetrandine
Fenbendazole.Nicotinamide
Fenbendazole.Rotenone
Fenbendazole.TC−S7004
Fenbendazole.Tetrandine
Mebendazole.Nicotinamide
Mebendazole.Rotenone
Mebendazole.TC−S7004
Mebendazole.Tetrandine
Nicotinamide.Rotenone
Nicotinamide.TC−S7004
Nicotinamide.Tetrandine
Rotenone.TC−S7004
Rotenone.Tetrandine
TC−S7004.Tetrandine

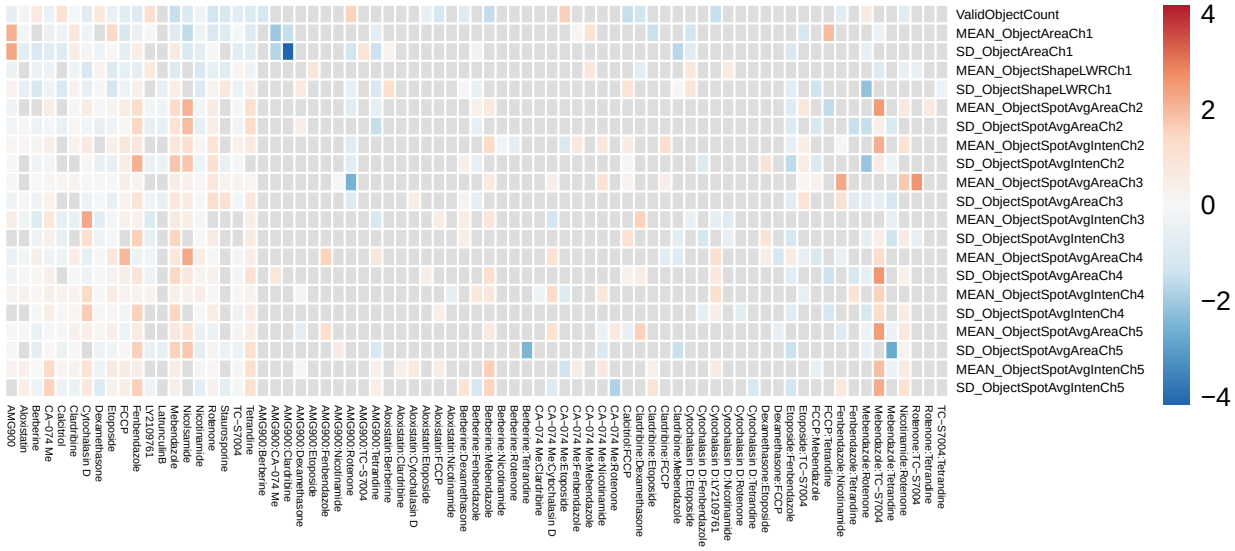
In this heatmap we can see some combinations that affect multiple features of a certain channels in the same way. For instance, a positive interaction effect between Mebendazole and TC-S7004 or between Nicotinamide and Rotenone on most of the Channel 5 features and Channel 2 features.

This is still a lot. So we reduce the set of features even further.

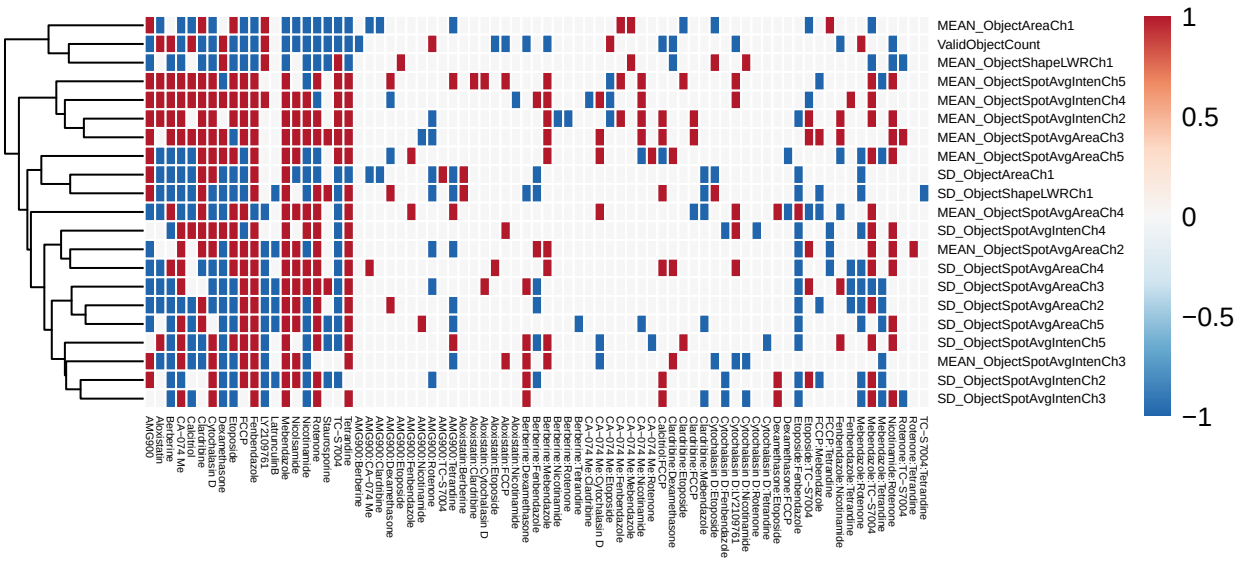
21 most important features

```
## New names:
## * `` -> `...1`
```

Reduced coefficient matrix for only 21 features



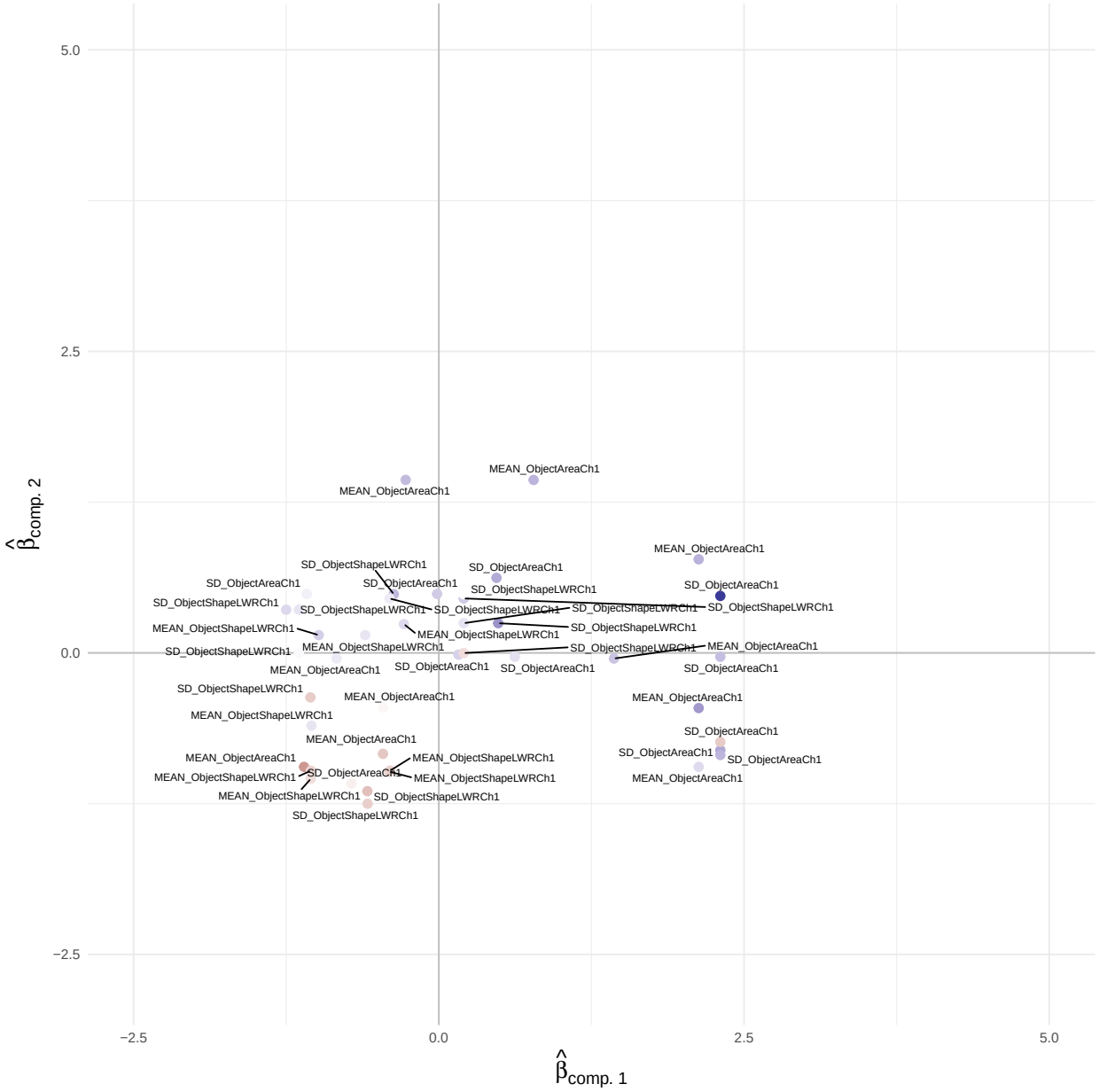
Representation with clustered rows and sign only for 21 features



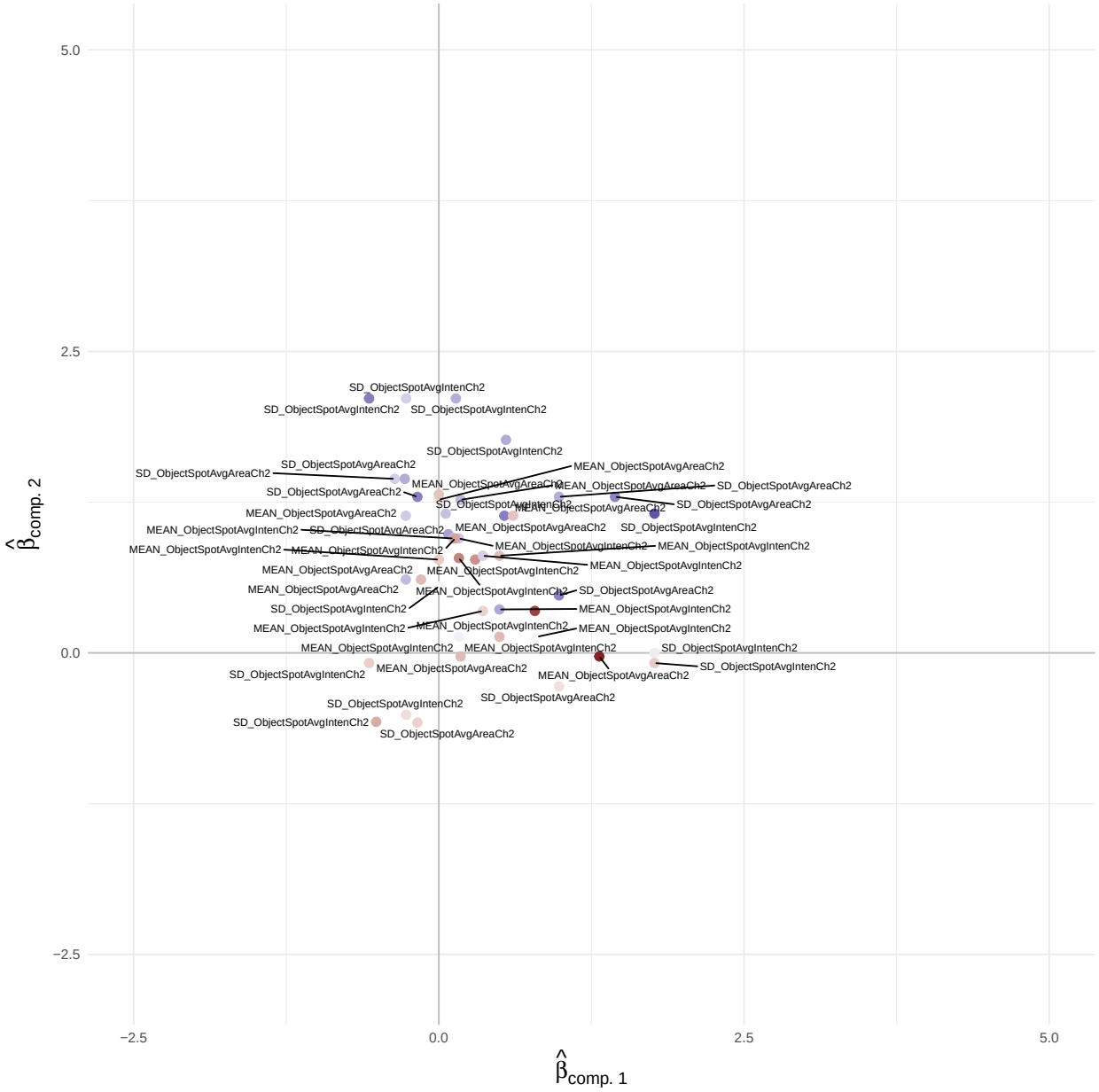
Scatterplot showing linear vs. interaction coefficients (per Channel)

Now we look again into the scatterplots, but for the reduced set with 21 features. The reduced set allows us to include the names of the morphological features.

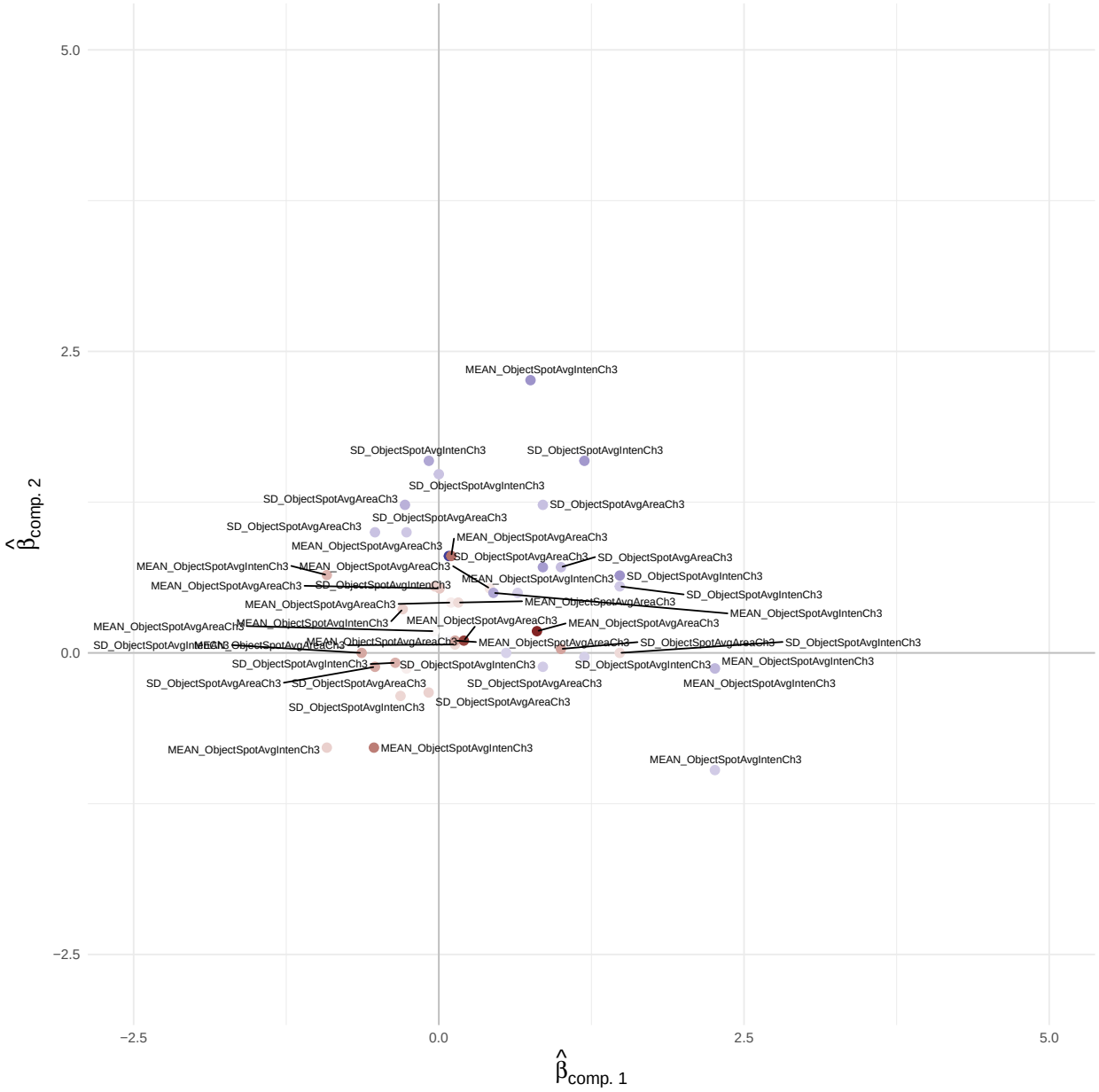
Channel 1


5.0
2.5
0.0
−2.5
−2.5
0.0
2.5
5.0
$\hat{\beta}_{comp.\ 2}$
$\hat{\beta}_{comp.\ 1}$
MEAN_ObjectAreaCh1
MEAN_ObjectAreaCh1
MEAN_ObjectAreaCh1
SD_ObjectShapeLWRCh1
SD_ObjectAreaCh1
SD_ObjectAreaCh1
SD_ObjectShapeLWRCh1
SD_ObjectAreaCh1
SD_ObjectShapeLWRCh1
SD_ObjectShapeLWRCh1
SD_ObjectShapeLWRCh1
SD_ObjectShapeLWRCh1
SD_ObjectShapeLWRCh1
MEAN_ObjectShapeLWRCh1
MEAN_ObjectShapeLWRCh1
MEAN_ObjectShapeLWRCh1
SD_ObjectShapeLWRCh1
SD_ObjectShapeLWRCh1
MEAN_ObjectAreaCh1
MEAN_ObjectAreaCh1
SD_ObjectAreaCh1
SD_ObjectAreaCh1
SD_ObjectAreaCh1
SD_ObjectShapeLWRCh1
MEAN_ObjectAreaCh1
MEAN_ObjectShapeLWRCh1
MEAN_ObjectAreaCh1
MEAN_ObjectAreaCh1
SD_ObjectAreaCh1
SD_ObjectAreaCh1
SD_ObjectAreaCh1
MEAN_ObjectAreaCh1
MEAN_ObjectShapeLWRCh1
MEAN_ObjectShapeLWRCh1
MEAN_ObjectShapeLWRCh1
SD_ObjectAreaCh1
MEAN_ObjectShapeLWRCh1
SD_ObjectShapeLWRCh1
SD_ObjectShapeLWRCh1

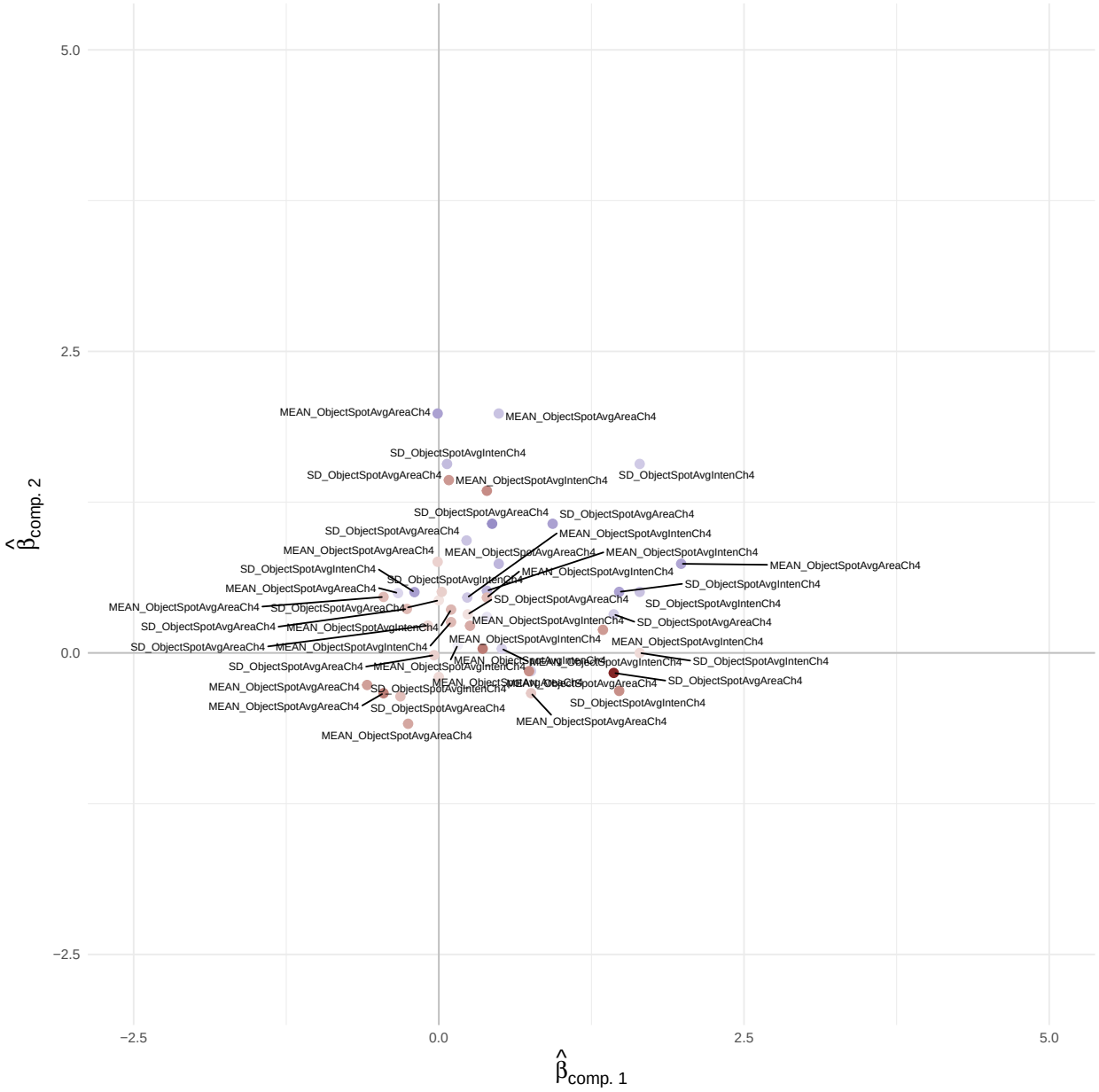
Channel 2


5.0
2.5
0.0
−2.5
−2.5
0.0
2.5
5.0
$\hat{\beta}_{comp.\ 2}$
$\hat{\beta}_{comp.\ 1}$
SD_ObjectSpotAvgIntenCh2
SD_ObjectSpotAvgAreaCh2
MEAN_ObjectSpotAvgAreaCh2
MEAN_ObjectSpotAvgIntenCh2

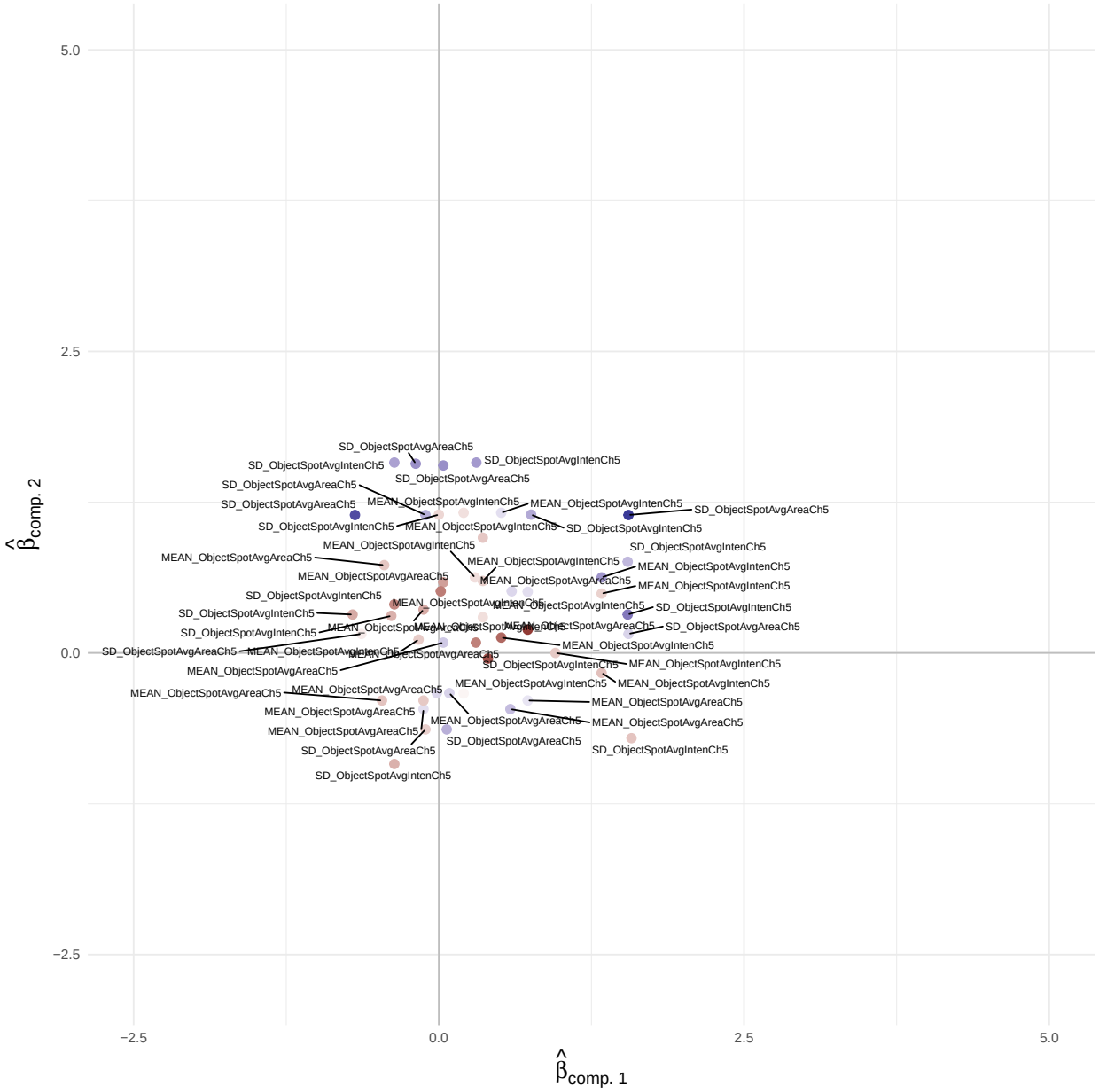
Channel 3


5.0
2.5
0.0
−2.5
−2.5
0.0
2.5
5.0
$\hat{\beta}_{comp.\ 2}$
$\hat{\beta}_{comp.\ 1}$
MEAN_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgAreaCh3
SD_ObjectSpotAvgAreaCh3
SD_ObjectSpotAvgAreaCh3
SD_ObjectSpotAvgAreaCh3
MEAN_ObjectSpotAvgAreaCh3
MEAN_ObjectSpotAvgAreaCh3
MEAN_ObjectSpotAvgIntenCh3
MEAN_ObjectSpotAvgAreaCh3
SD_ObjectSpotAvgAreaCh3
SD_ObjectSpotAvgAreaCh3
MEAN_ObjectSpotAvgAreaCh3
MEAN_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgIntenCh3
MEAN_ObjectSpotAvgAreaCh3
SD_ObjectSpotAvgIntenCh3
MEAN_ObjectSpotAvgAreaCh3
MEAN_ObjectSpotAvgIntenCh3
MEAN_ObjectSpotAvgAreaCh3
MEAN_ObjectSpotAvgAreaCh3
MEAN_ObjectSpotAvgAreaCh3
SD_ObjectSpotAvgAreaCh3
SD_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgIntenCh3
MEAN_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgAreaCh3
SD_ObjectSpotAvgAreaCh3
SD_ObjectSpotAvgAreaCh3
MEAN_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgIntenCh3
SD_ObjectSpotAvgAreaCh3
MEAN_ObjectSpotAvgIntenCh3
MEAN_ObjectSpotAvgIntenCh3
MEAN_ObjectSpotAvgIntenCh3

Channel 4


5.0
2.5
0.0
−2.5
−2.5
0.0
2.5
5.0
$\hat{\beta}_{comp.\ 2}$
$\hat{\beta}_{comp.\ 1}$
MEAN_ObjectSpotAvgAreaCh4
SD_ObjectSpotAvgIntenCh4
SD_ObjectSpotAvgAreaCh4
MEAN_ObjectSpotAvgIntenCh4
SD_ObjectSpotAvgIntenCh4

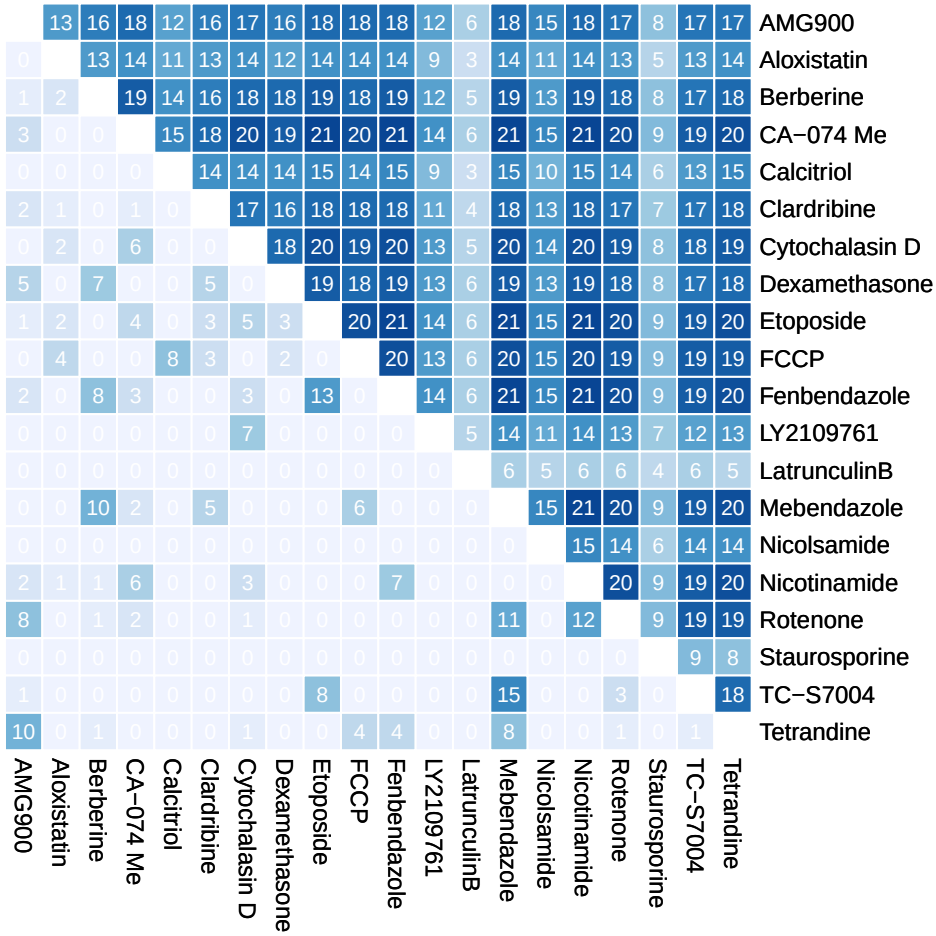
Channel 5



The document scatter_plt_each_combi_robust_10uM_21_newData.pdf shows the same figures but for concrete combinations of compounds. Some observations from this representation are summarized here:

- Most drug combinations affect multiple channels in a similar way

How often do we observe what for only 21 features?



Files written

These files have been written to the target directory, `data/05-hiernet-weak-robust-10uM-new-data`:

```
projthis::proj_dir_info(path_target())
```

```
## # A tibble: 0 x 4
## # ... with 4 variables: path <fs::path>, type <fct>, size <fs::bytes>,
## #   modification_time <dttm>
```