Statistical analysis of the Inhibition Index from Touret et al (2020)

Exploration of the results

Jacques van Helden

2020-04-09

Contents

Supplementary data tables

```
## Directories
dir <- c(data = "data",
         results = "results",
         figures = "figures")
dir.create(dir["results"], showWarnings = FALSE, recursive = TRUE)

## Data file
supTableFile <- file.path(dir["data"], "suppl-table_Touret-2020.xlsx")
```

```
supTable <- read.xlsx(file = supTableFile, sheetIndex = 1, rowIndex = TRUE)
# dim(supTable)
# View(supTable)
# names(supTable)

## Suppress the last row (NA)
supTable <- supTable[!is.na(supTable$Plate.Number....Position.Number), ]
# dim(supTable)

## Assign row names for convenience
rownames(supTable) <- supTable[, 1]
# View(supTable)
```

```
## Extract plate number
supTable$plateNumber <- as.numeric(substr(supTable[, 1], start = 1, stop = 2))
# table(supTable$plateNumber)
plateNumbers <- unique(supTable$plateNumber)

## Assign a color to each molecule according to its plate number
plateColor <- rainbow(n = length(plateNumbers))
names(plateColor) <- unique(supTable$plateNumber)

supTable$color <- plateColor[supTable$plateNumber]
```

The supplementary table downloaded from bioRxiv contains 1520 molecules.

Inhibitiion index

Descriptive stats

```
ii <- supTable$Inhibition.Index
iiStat <- list(
  mean = mean(ii),
  sd = sd(ii),
  var = var(ii),
  min = min(ii),
  Q1 = as.vector(quantile(ii, probs = 0.25)),
  median = median(ii),
  Q3 = as.vector(quantile(ii, probs = 0.75)),
  max = max(ii)
)

kable(t(as.data.frame.list(iiStat)), col.names = "Stat")
```

	Stat
mean	0.2865268
sd	0.3758726
var	0.1412802
min	-0.4444311
Q1	0.0279599
median	0.1665643
Q3	0.4917926
max	2.0988878

Distribution

The distribution of inhibition index values is trongly asymmetrical. The mode is much lower than the mean and the median (robust estimator of central tendency). A normal fit will tuhs give a poor estimate of the p-values.

```
hist(supTable$Inhibition.Index, breaks = 100, col = "grey", border = "grey")
abline(v = iiStat$mean, col = "darkgreen")
abline(v = iiStat$median, col = "blue")

legend("topright", legend = c("mean", "median"),
       col = c("blue", "darkgreen"),
```

```
       lwd = 2)
```

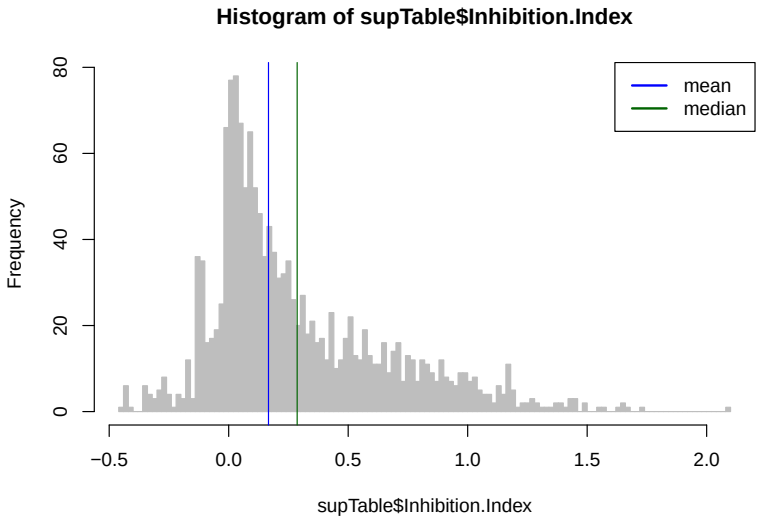

Figure 1: Distribution of the inhibition index

Normalisation

Log transform

A classical method for normalisation is to take the log of the values. We first had to shift the data in order for all of them to take positive values.

```
# fit.gamma <- fitdist(data = ii - iiStat$min + 1, distr = "gamma", method = "mge")
# summary(fit.gamma)
#
# plot(fit.gamma)

#### Compute a normalised distribution from inhibition indices ####
iiPositive <- ii - iiStat$min + 1 ## shift the distrib to achieve non-negative numbers
logII <- log(iiPositive)

logIIStat <- list(
  mean = mean(logII),
  sd = sd(logII),
  var = var(logII),
  min = min(logII),
  Q1 = as.vector(quantile(logII, probs = 0.25)),
  median = median(logII),
  Q3 = as.vector(quantile(logII, probs = 0.75)),
  max = max(logII)
)

kable(t(as.data.frame.list(logIIStat)), col.names = "Stat", caption = "Parameters of the log-normalised
```

Table 2: Parameters of the log-normalised inhibition index distribution

	Stat
mean	0.5272798
sd	0.2027703
var	0.0411158
min	0.0000000
Q1	0.3868877
median	0.4768523
Q3	0.6607395
max	1.2650638

However, even after log transformation the distribution remains highly asymmetrical, with a mode much smaller than the median and mean.

```
#### Histogram of log-normalised values ####
hist(logII, breaks = 100, col = "grey", border = "grey")
abline(v = mean(logII), col = "blue")
abline(v = median(logII), col = "darkgreen")

legend("topright", legend = c("mean", "median"),
       col = c("blue", "darkgreen"),
       lwd = 2)
```

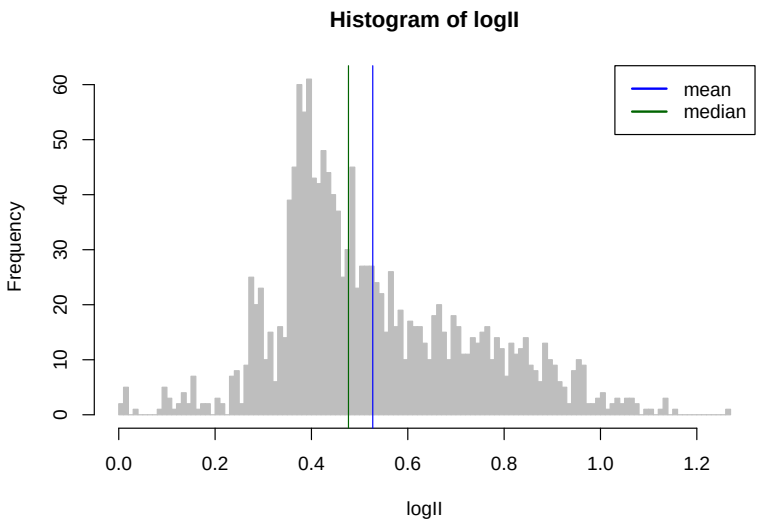

Figure 2: Distribution of the inhibition index

Evidence of a plate bias

Ranked values

We plot the inhibition index values ordered by plate and position number (top) or ranked by decreasing value (bottom). In both cases, the color denotes the plate number.

```
par(mfrow = c(2,1))
plot(ii,
     panel.first = grid(),
```

```
     main = "Inhibition index values",
     xlab = "Molecules (ranked by inhibition index)",
     ylab = "Inhibition index",
     col = supTable$color,
     cex = 0.5,
     xlim = c(0, length(ii)*1.05)
     )
legend("topright",
       legend = names(plateColor),
       col = plateColor, pch = 1,
       cex = 0.7)

sortedTable <- supTable[order(supTable$Inhibition.Index, decreasing = TRUE), ]
plot(sortedTable$Inhibition.Index,
     panel.first = grid(),
     main = "Ranked inhibition index values",
     xlab = "Molecules (ranked by inhibition index)",
     ylab = "Inhibition index",
     col = sortedTable$color,
     cex = 0.5,
     xlim = c(0, length(ii)*1.05)
     )
legend("topright",
       legend = names(plateColor),
       col = plateColor, pch = 1,
       cex = 0.7)
```

```
par(mfrow = c(1, 1))
```

The molecule-wised colored plots of inhibition index suggest a plate-wise effect.

Plate-wise normalisation

We perform a plate-wise normalisation using robust estimators, in order to avoid the effect of outliers (in this case, the suspected outliers are the molecules having an actual inhibitory effect).

To this purpose, we use: - plate-wise median to estimate the mean - plate-wise standardised inter-quantile range (IQR) to estimate the standard deviation

```
#### Copute plate-wise statistics ####
plateStat <- data.frame(
  plate = plateNumbers,
  mean = as.vector(by(data = supTable$Inhibition.Index, INDICES = supTable$plateNumber, FUN = mean)),
  sd = as.vector(by(data = supTable$Inhibition.Index, INDICES = supTable$plateNumber, FUN = sd)),
  median = as.vector(by(data = supTable$Inhibition.Index, INDICES = supTable$plateNumber, FUN = median)),
  min = as.vector(by(data = supTable$Inhibition.Index, INDICES = supTable$plateNumber, FUN = min)),
  max = as.vector(by(data = supTable$Inhibition.Index, INDICES = supTable$plateNumber, FUN = max)),
  IQR = as.vector(by(data = supTable$Inhibition.Index, INDICES = supTable$plateNumber, FUN = IQR))
)
rownames(plateStat) <- plateStat$plate

kable(plateStat, caption = "Plate-wise statistics")
```

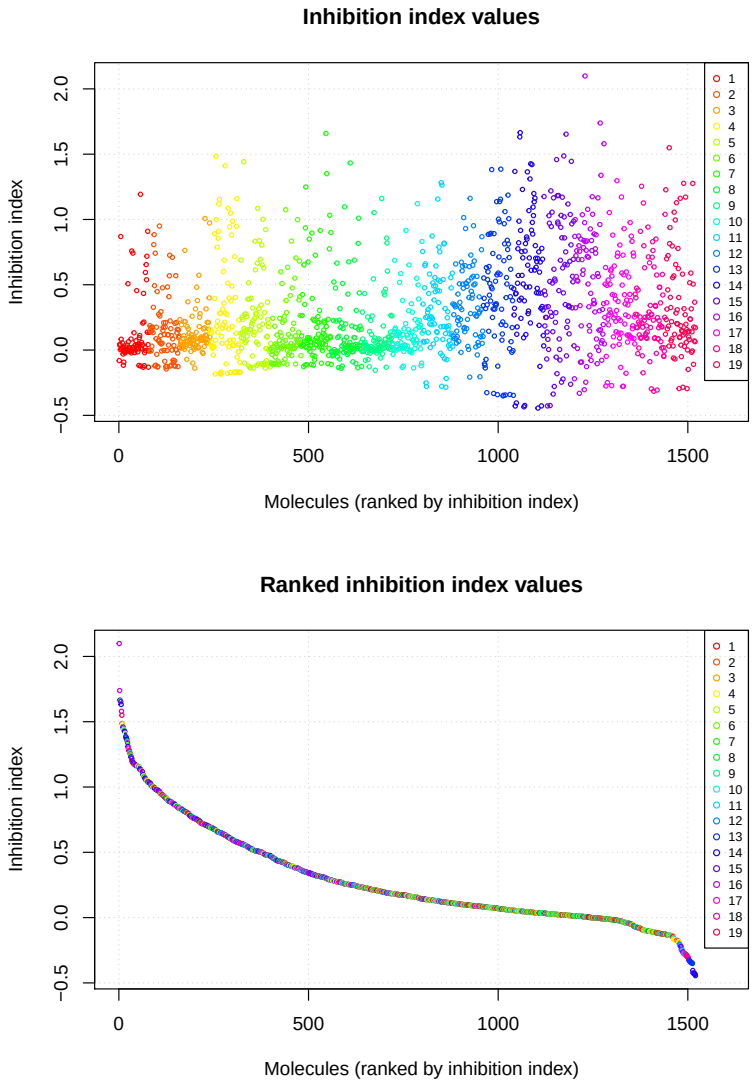

Figure 3: Values of the inhibition index for all the tested molecules. Molecules are colored according to the plate number.

Table 3: Plate-wise statistics

plate	mean	sd	median	min	max	IQR
1	0.1165661	0.2665169	0.0184570	-0.1318422	1.1928831	0.0914910
2	0.1541782	0.2523084	0.1133156	-0.1450055	0.9498831	0.2156783
3	0.1443563	0.1960778	0.0863385	-0.1387721	1.0073009	0.1349143
4	0.2982710	0.4307195	0.1601301	-0.1844422	1.4836918	0.4739120
5	0.2345274	0.3194481	0.1628049	-0.1383338	1.4421869	0.2814504
6	0.1393632	0.2580634	0.0412022	-0.1109600	1.0748024	0.2271337
7	0.2241196	0.3673457	0.0873107	-0.1267828	1.6589036	0.2073556
8	0.1439492	0.2894590	0.0331407	-0.1396088	1.4326962	0.1976995
9	0.1253725	0.2449408	0.0420476	-0.1543792	1.1609812	0.1107754
10	0.1541551	0.2335481	0.0948601	-0.1365916	1.1209955	0.1617424
11	0.3071834	0.3265338	0.2478965	-0.2846758	1.2816167	0.4140013
12	0.3499997	0.2757037	0.3159377	-0.2074233	1.1619058	0.3761552
13	0.4314013	0.4397504	0.4146992	-0.3511397	1.3851376	0.5813343
14	0.5857602	0.4775020	0.6038263	-0.4444311	1.6653052	0.5380693
15	0.5187760	0.4972562	0.4842358	-0.4279441	1.6531873	0.8015369
16	0.5253860	0.4768292	0.5615571	-0.3309767	2.0988878	0.6858842
17	0.3590275	0.3386716	0.3050932	-0.2784932	1.2965981	0.4539432
18	0.2912052	0.3201378	0.2263674	-0.3152611	1.2542724	0.3426682
19	0.3404116	0.4026624	0.1982078	-0.2950106	1.5490336	0.4724498

```
## Centering: substract the median
## Scaling: divide by IQR
## Standardise: multiply by IQR of the normal distribution
normII <- (supTable$Inhibition.Index - plateStat[supTable$plateNumber, "median"]) / plateStat[supTable$p
# IQR(normII)
# IQR(rnorm(n = 1000000))

normIQR <- qnorm(p = 0.75) - qnorm(p = 0.25)
normII <- normII * normIQR
# sd(normII)
# IQR(normII)

supTable$normInhibIndex <- normII
```

```
### Descriptive statistics on the normalised Inhibition Index ####
normIIStat <- list(
  mean = mean(normII),
  sd = sd(normII),
  IQR = IQR(normII),
  var = var(normII),
  min = min(normII),
  Q1 = as.vector(quantile(normII, probs = 0.25)),
  median = median(normII),
  Q3 = as.vector(quantile(normII, probs = 0.75)),
  max = max(normII)
)

kable(t(as.data.frame.list(normIIStat)), col.names = "Stat", caption = "Statistics of the plate-wise no
```

Table 4: Statistics of the plate-wise normalised inhibition index

	Stat
mean	0.4144692
sd	1.8129741
IQR	1.3183994
var	3.2868752
min	-2.6280587
Q1	-0.5094508
median	0.0000000
Q3	0.8089486
max	17.3161990

The histogram of plate-wise normalised values shows a clear improvement : the median is much closer to the mode than with the raw or log-transformed II values.

```
hist(normII, breaks = 100, col = "grey", border = "grey")
abline(v = mean(normII), col = "blue")
abline(v = median(normII), col = "darkgreen")

legend("topright", legend = c("mean", "median"),
       col = c("blue", "darkgreen"),
       lwd = 2)
```

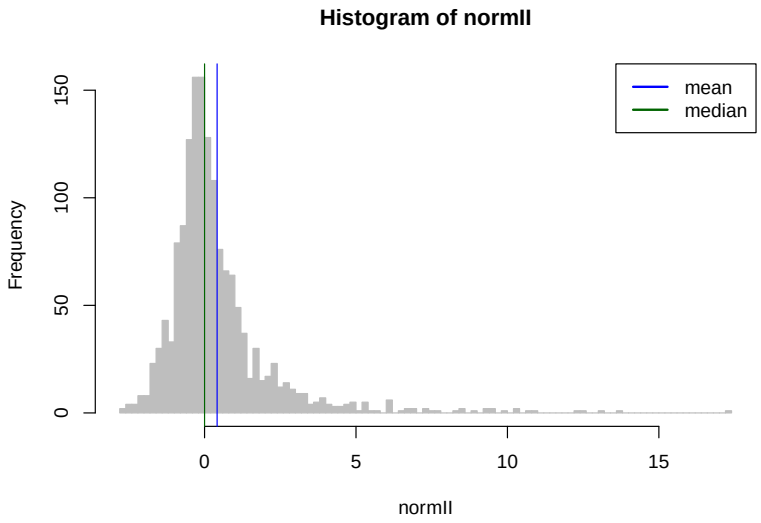

Figure 4: Distribution of the plate-wise normalised inhibition index

Normalised II plots

The plot of normalised II values (top panel) clearly shows that the plate-wise normalisation suppressed the background bias.

```
par(mfrow = c(2,1))
plot(normII,
     panel.first = grid(),
     main = "Inhibition index values",
     xlab = "Molecules (ranked by inhibition index)",
```

```
     ylab = "Inhibition index",
     col = supTable$color,
     cex = 0.5,
     xlim = c(0, length(normII)*1.05)
     )
legend("topright",
       legend = names(plateColor),
       col = plateColor, pch = 1,
       cex = 0.7)

# names(supTable)
normIIrank <- order(supTable$normInhibIndex, decreasing = TRUE)
plot(supTable[normIIrank, "normInhibIndex"],
     panel.first = grid(),
     main = "Ranked inhibition index values",
     xlab = "Molecules (ranked by inhibition index)",
     ylab = "Inhibition index",
     col = supTable[normIIrank, "color"],
     cex = 0.5,
     xlim = c(0, length(normII)*1.05)
     )
legend("topright",
       legend = names(plateColor),
       col = plateColor, pch = 1,
       cex = 0.4)
```

```
par(mfrow = c(1,1))
```

P-value computation

We compute the p-value as the upper tail of the normal distribution (rigth-side test) in order to detect significantly high values of the plate-wise normalised index.

```
#### Compute P-value for the inhibition index ####
supTable$p.value <- pnorm(normII, mean = 0, sd = 1, lower.tail = FALSE)
supTable$log10Pval <- log10(supTable$p.value)
supTable$e.value <- supTable$p.value * length(normII)
supTable$padj <- p.adjust(supTable$p.value, method = "fdr")
supTable$log10Padj <- log10(supTable$padj)
```

```
hist(supTable$p.value, breaks = 20,
     col = "grey",
     main = "P-value histogram after plate-wise normalisation",
     xlab = "Nominal P-value (unadjusted)",
     ylab = "Frequency")
```

Selection of candidate molecules

```
#### Select significant normalised II values ####
alpha <- 0.05
# table(supTable$padj < alpha)
selected <- subset(supTable, supTable$padj < alpha)

## Sort by decreasing adjusted p-value
```

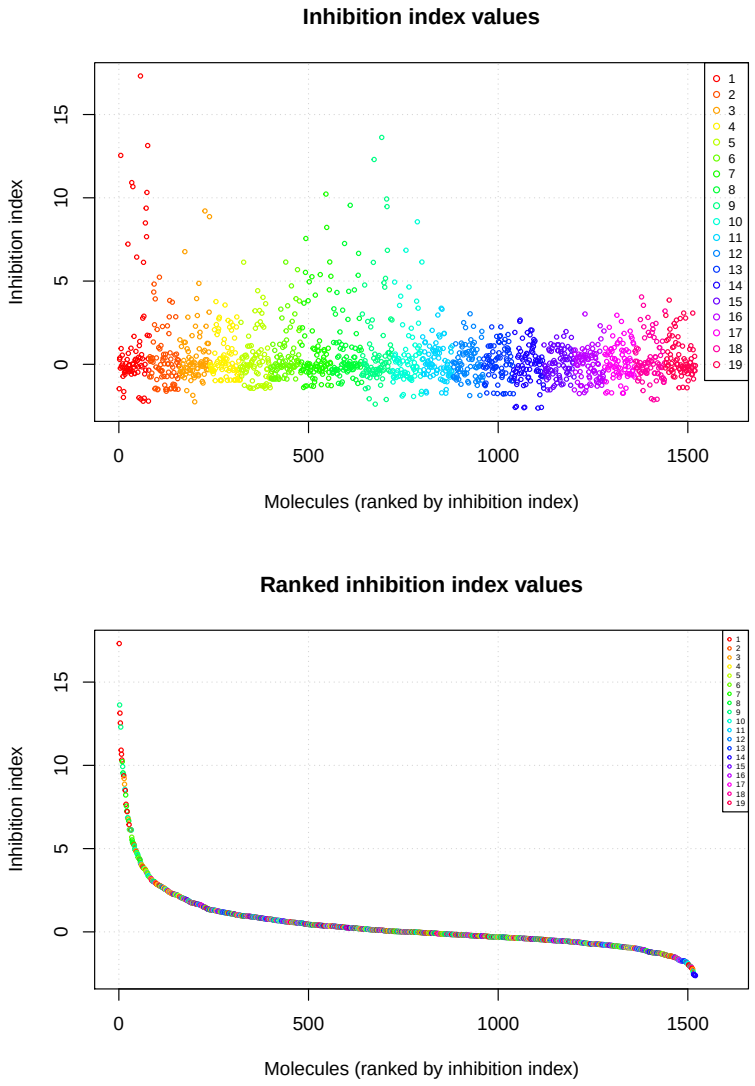

Figure 5: Values of the plate-wise normalised inhibition index for all the tested molecules. Molecules are colored according to the plate number.

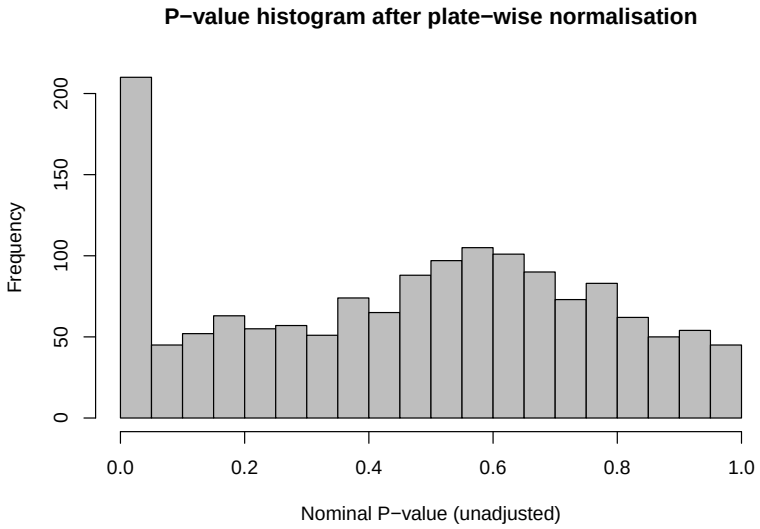

Figure 6: Histogram of the nominal (unadjusted) p-values derived from the plate-wise normalised inhibition index.

```
selected <- selected[order(selected$padj, decreasing = FALSE), ]
# kable(names(selected), row.names=TRUE)

## Print selected molecules
kable(selected[ , c(1:3, 5:7, 10, 14:15)],
      row.names = FALSE,
      digits = 4,
      caption = "Candidate moecules sorted by significance after plate-wise normalisation. ")
```

Ta
no

Plate.Number....Position.Number	Prestw.number	Chemical.name	Broad.Therapeutic
01F08	Prestw-57	Benoxinate hydrochloride	Neuromuscular
09F04	Prestw-693	Promazine hydrochloride	Central Nervous Sy
01H07	Prestw-76	Dibucaine	Neuromuscular
01A06	Prestw-5	Atracurium besylate	Neuromuscular
09D04	Prestw-1456	Opipramol dihydrochloride	Central Nervous Sy
01D05	Prestw-34	Triamterene	Metabolism
01D08	Prestw-37	Pyrimethamine	Infectiology
01H05	Prestw-74	Amitryptiline hydrochloride	Central Nervous Sy
07G07	Prestw-546	Pregnenolone	Endocrinology
09G07	Prestw-706	Chlorcyclizine hydrochloride	Allergology 'Centra
08E11	Prestw-1284	Hydroxychloroquine sulfate	Metabolism
09G08	Prestw-707	Diphenylpyraline hydrochloride	Allergology 'Centra
01H03	Prestw-72	Imipramine hydrochloride	Central Nervous Sy
03G08	Prestw-227	Clemizole hydrochloride	Allergology 'Derma
03H10	Prestw-239	Orphenadrine hydrochloride	Allergology 'Centra
10G08	Prestw-787	Merbromin disodium salt	Infectiology
01G11	Prestw-70	Tolnaftate	Infectiology
07G09	Prestw-548	Chloroquine diphosphate	Metabolism
01H04	Prestw-73	Sulindac	Central Nervous Sy
07B04	Prestw-493	Omeprazole	Gastroenterology

Plate.Number....Position.Number	Prestw.number	Chemical.name	Broad.Therapeutic
08D06	Prestw-1410	Exemestane	Endocrinology
01C05	Prestw-24	Norethynodrel	Endocrinology
09G09	Prestw-708	Benzethonium chloride	Infectiology
10D08	Prestw-757	Chlorotrianisene	Endocrinology
03B05	Prestw-174	Alverine citrate salt	Neuromuscular
08H03	Prestw-632	Dipivefrin hydrochloride	Ophthalmology
01E08	Prestw-47	Ticlopidine hydrochloride	Hematology
06D11	Prestw-440	Epiandrosterone	Endocrinology
07H07	Prestw-1144	Mirtazapine	Central Nervous Sy
10H10	Prestw-799	Pridinol methanesulfonate salt	Central Nervous Sy
05A10	Prestw-329	Tacrine hydrochloride	Central Nervous Sy
01G06	Prestw-65	Diphenhydramine hydrochloride	Allergology 'Centra
09D02	Prestw-671	Dydrogesterone	Endocrinology
06H02	Prestw-1358	Vatalanib	Oncology
07B03	Prestw-492	Nitrofural	Infectiology
07F02	Prestw-531	Pirenperone	Central Nervous Sy
08H02	Prestw-1210	Alendronate sodium	Metabolism
07H10	Prestw-1817	Tazarotene	Dermatology
07C10	Prestw-509	Bromperidol	Central Nervous Sy
02C08	Prestw-1314	Pioglitazone	Endocrinology
09G04	Prestw-703	Famprofazone	Central Nervous Sy
07C03	Prestw-502	Biperiden hydrochloride	Central Nervous Sy
10A08	Prestw-1140	Liranaftate	Infectiology
09F10	Prestw-699	Hexestrol	Endocrinology
03F02	Prestw-211	Piroxicam	Central Nervous Sy
02B04	Prestw-93	Azacyclonol	Central Nervous Sy
09A09	Prestw-1154	Nilvadipine	Cardiovascular
06F06	Prestw-455	Mebhydroline 1,5-naphtalenedisulfonate	Allergology
10E06	Prestw-765	Ethoxyquin	Metabolism
09G02	Prestw-701	Trihexyphenidyl-D,L hydrochloride	Central Nervous Sy
07H06	Prestw-555	Nifuroxazide	Infectiology 'Metab
08G02	Prestw-1506	Mizolastine	Allergology
05E07	Prestw-366	Ambroxol hydrochloride	Respiratory
08E07	Prestw-1331	Rimantadine hydrochloride	Infectiology
02B03	Prestw-92	Zimelidine dihydrochloride monohydrate	Central Nervous Sy
08B06	Prestw-1351	Tenatoprazole	Metabolism
07D09	Prestw-518	Budesonide	Endocrinology
10C04	Prestw-743	Medrysone	Metabolism
18B09	Prestw-1951	Eperisone HCl	Neuromuscular
05F11	Prestw-380	Clebopride maleate	Central Nervous Sy
03E06	Prestw-205	Tolfenamic acid	Central Nervous Sy
06G08	Prestw-1157	Rifapentine	Infectiology
02B07	Prestw-96	Guanabenz acetate	Central Nervous Sy
19B02	Prestw-1996	Budralazine	Cardiovascular
02F05	Prestw-134	Diltiazem hydrochloride	Cardiovascular 'He
07B10	Prestw-499	Propafenone hydrochloride	Cardiovascular
06H07	Prestw-476	Primaquine diphosphate	Infectiology
10G06	Prestw-785	Dicumarol	Hematology
04B07	Prestw-256	Isotretinoin	Dermatology
02G02	Prestw-141	Verapamil hydrochloride	Cardiovascular
07A09	Prestw-488	Dosulepin hydrochloride	Central Nervous Sy
05G08	Prestw-387	Carbetapentane citrate	Central Nervous Sy

Plate.Number....Position.Number	Prestw.number	Chemical.name	Broad.Therapeutic
04D11	Prestw-280	Quinidine hydrochloride monohydrate	Cardiovascular 'Inf
18C04	Prestw-1961	Methandrostenolone	Endocrinology
10B11	Prestw-1820	Amprenavir	Infectiology
06E06	Prestw-445	Cyclobenzaprine hydrochloride	Neuromuscular
10G10	Prestw-789	Drofenine hydrochloride	Neuromuscular
11E11	Prestw-850	Equilin	Endocrinology
08A08	Prestw-1139	Itraconazole	Infectiology 'Metab
06C08	Prestw-1393	Dibenzepine hydrochloride	Central Nervous Sy
11F03	Prestw-1454	Nylidrin	Cardiovascular
08C11	Prestw-1409	Etretinate	Dermatology
05F02	Prestw-371	Ketotifen fumarate	Allergology
05D10	Prestw-359	Dextromethorphan hydrobromide monohydrate	Central Nervous Sy
18H11	Prestw-2043	Artenimol	Infectiology
03H09	Prestw-238	Lomefloxacin hydrochloride	Infectiology
19E10	Prestw-2052	Dilevalol	Cardiovascular
19H04	Prestw-1940	Acetyl spiramycin	Infectiology
05F08	Prestw-377	Nafronyl oxalate	Cardiovascular 'Ne
12E07	Prestw-926	Idazoxan hydrochloride	Central Nervous Sy
06G05	Prestw-1323	Quetiapine hemifumarate	Central Nervous Sy
09A03	Prestw-1270	Gefitinib	Oncology
16C10	Prestw-1710	Ethoxzolamide	Ophthalmology 'Ga
08B02	Prestw-571	Tetracaïne hydrochloride	Neuromuscular
11E02	Prestw-1455	Olanzapine	Central Nervous Sy
17D04	Prestw-1857	Oxiglutatione	Ophthalmology
19A03	Prestw-2045	Eletriptan	Central Nervous Sy
03E08	Prestw-1181	Tibolone	Endocrinology
01G07	Prestw-66	Minaprine dihydrochloride	Central Nervous Sy
02G11	Prestw-150	Dihydroergotamine tartrate	Central Nervous Sy
02F04	Prestw-133	Hydroxyzine dihydrochloride	Allergology 'Centra
04H02	Prestw-311	Ifenprodil tartrate	Cardiovascular
03C03	Prestw-182	Levamisole hydrochloride	Immunology 'Infec
04C06	Prestw-265	Dimenhydrinate	Allergology 'Centra
18D06	Prestw-2008	Azaribine	Oncology 'Dermato
06C11	Prestw-430	Cisapride	Gastroenterology
19F05	Prestw-2019	Vonoprazan	Gastroenterology
01G04	Prestw-63	Nifedipine	Cardiovascular
08H05	Prestw-1463	Tomoxetine hydrochloride	Central Nervous Sy
19D11	Prestw-2067	Cyclofenil	Endocrinology
04C05	Prestw-264	Dyclonine hydrochloride	Neuromuscular
09H08	Prestw-717	Finasteride	Endocrinology
08E05	Prestw-1252	Butenafine hydrochloride	Infectiology 'Metab
18B06	Prestw-1945	Exifone	Central Nervous Sy

Conclusions

I strongly recommend to use the plate-wise normalised inhibition index in order to select the candidate molecules.

With an adjusted p-value of 0.05, 114 molecules are declared significant and could be considered as candidate for further characterisation.