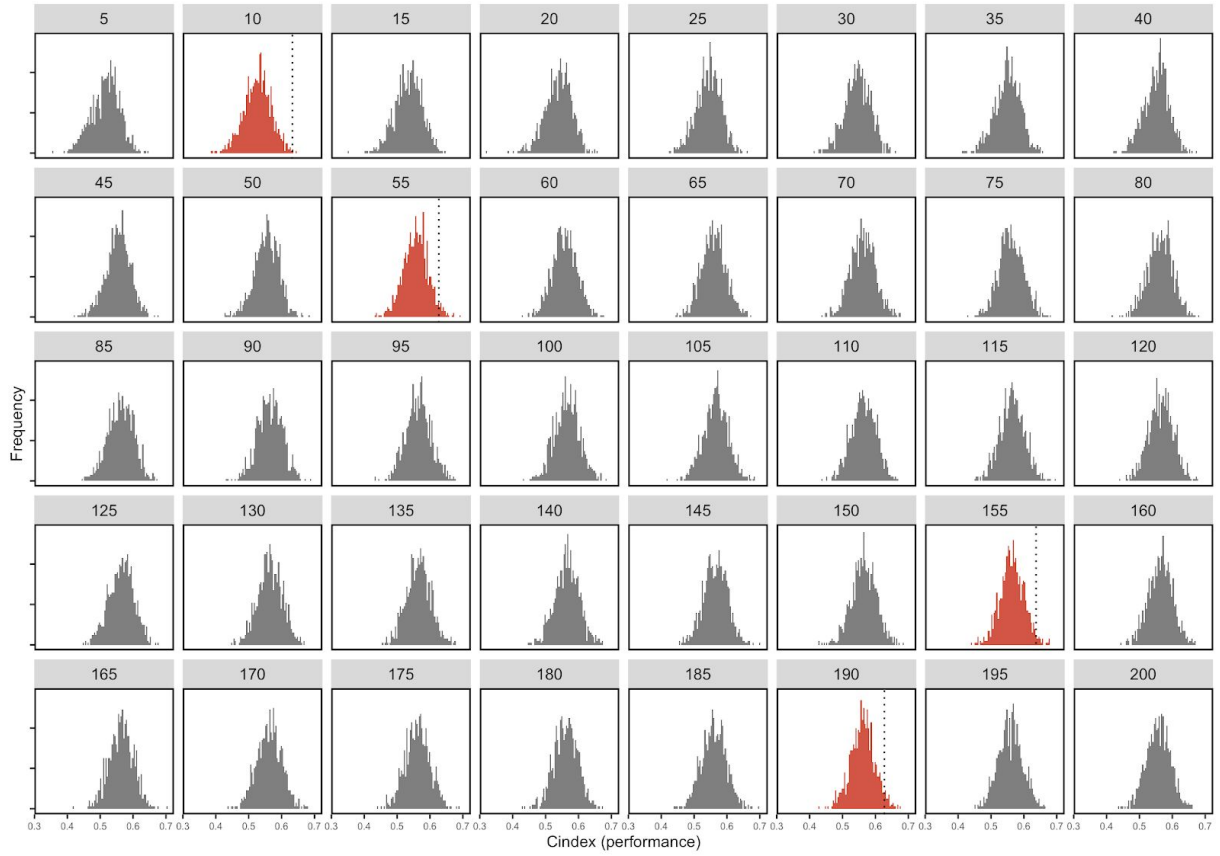
# CoRaS: Cox proportional hazards and random survival forest methods for identifying drug repurposing candidates to treat breast cancer.

Pradeep Mangalath[1], Stan Finkelstein[2], Artem Sokolov[3]

## Supporting Information.

## LINCS results - cox-proportional hazards (CPH) approach.

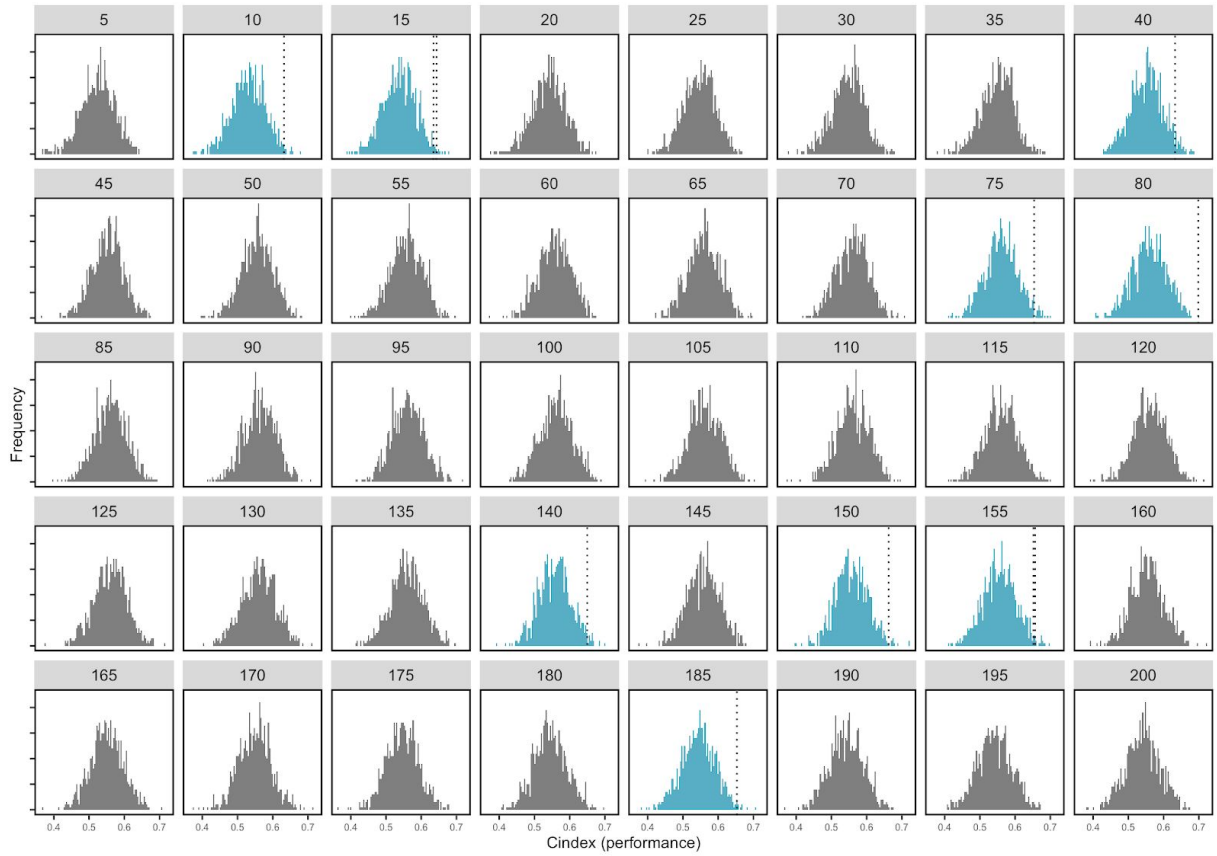
### Combined (1097 patients)




**CPH combined results** | Significant drugs were identified by comparing 1000 random gene signature models of signature lengths n by bins sized in increments of 5 genes, to drug models of equal signature lengths. Significant drugs were identified by considering performance (c-index) below 5% threshold of empirical p-values, computed by comparing to the background distribution. Bins with significant drugs are highlighted in color. Dotted lines indicate drug performance (c-index).

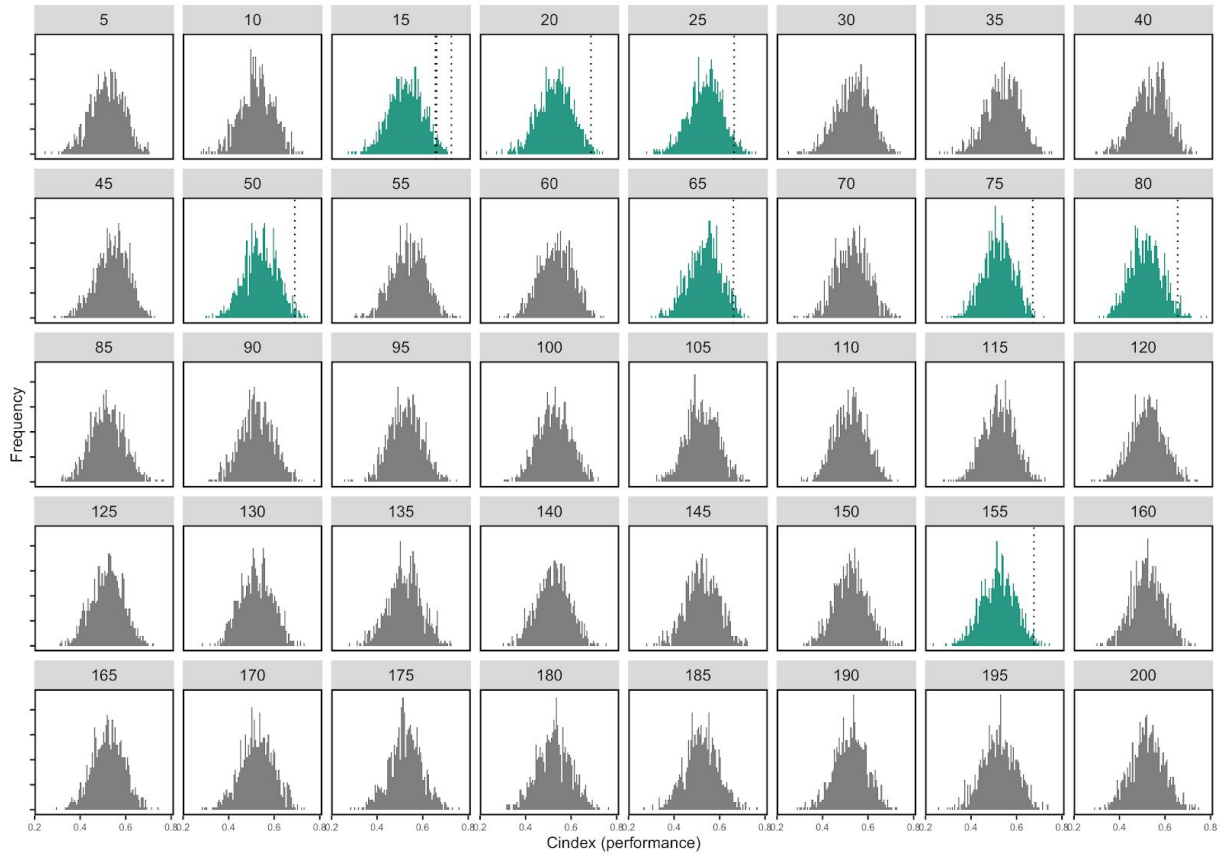
[1] Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, United States
[2] Massachusetts Institute of Technology, Cambridge, Massachusetts, United States
[3] Laboratory of Systems Pharmacology, Harvard Medical School, Boston, Massachusetts, United States

ER+ (808 patients)



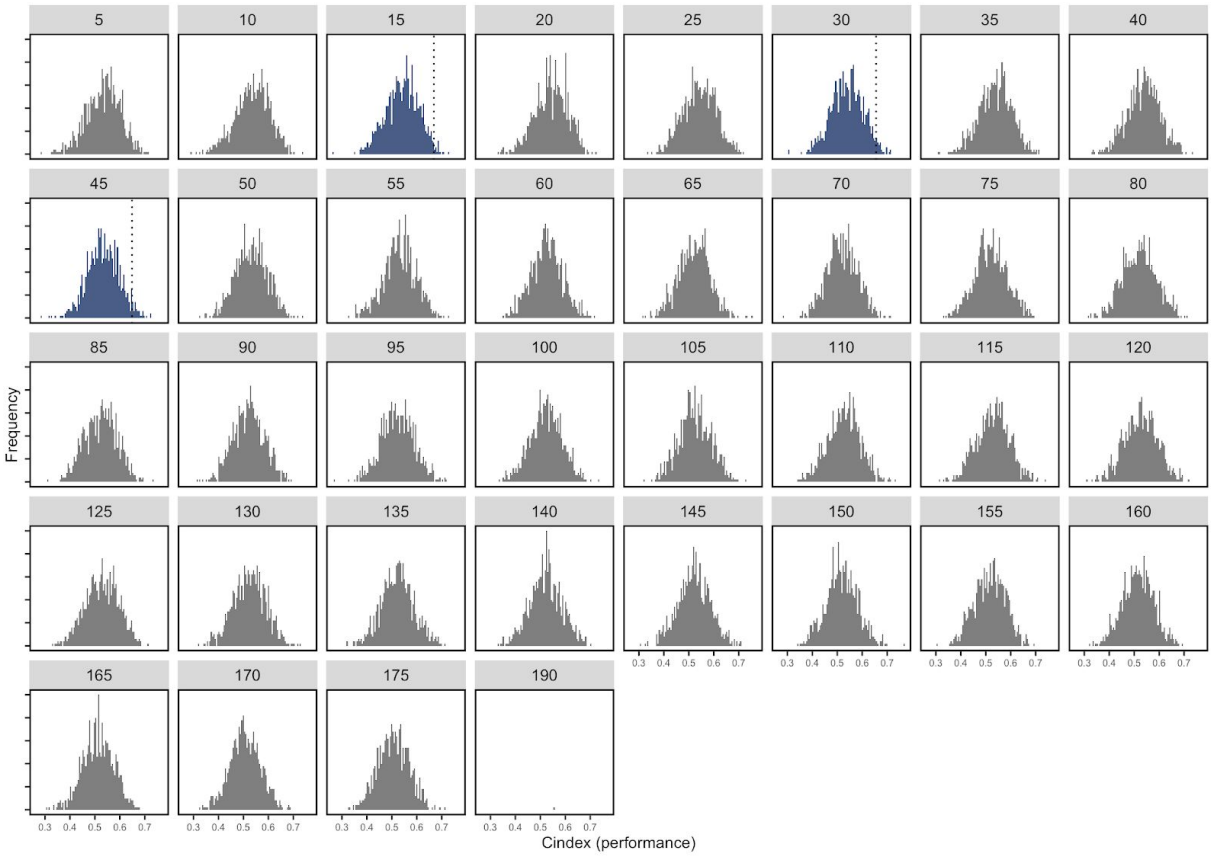

CPH ER+ results | Significant drugs were identified by comparing 1000 random gene signature models of signature lengths n by bins sized in increments of 5 genes, to drug models of equal signature lengths. Significant drugs were identified by considering performance (c-index) below 5% threshold of empirical p-values, computed by comparing to the background distribution. Bins with significant drugs are highlighted in color. Dotted lines indicate drug performance (c-index).

Her2 (343 patients)




CPH Her2 results | Significant drugs were identified by comparing 1000 random gene signature models of signature lengths n by bins sized in increments of 5 genes, to drug models of equal signature lengths. Significant drugs were identified by considering performance (c-index) below 5% threshold of empirical p-values, computed by comparing to the background distribution. Bins with significant drugs are highlighted in color. Dotted lines indicate drug performance (c-index).

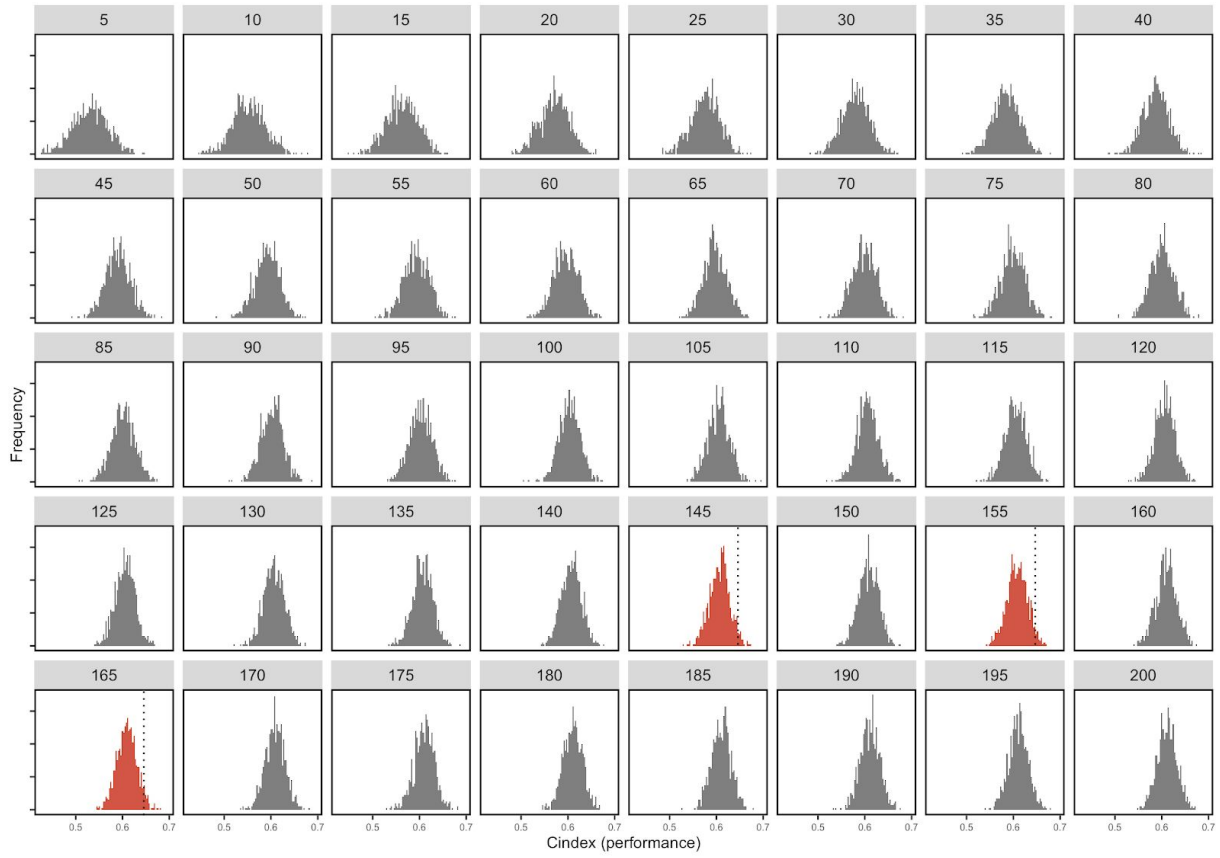
Triple Negative (247 patients)




CPH Triple negative results | Significant drugs were identified by comparing 1000 random gene signature models of signature lengths n by bins sized in increments of 5 genes, to drug models of equal signature lengths. Significant drugs were identified by considering performance (c-index) below 5% threshold of empirical p-values, computed by comparing to the background distribution. Bins with significant drugs are highlighted in color. Dotted lines indicate drug performance (c-index).

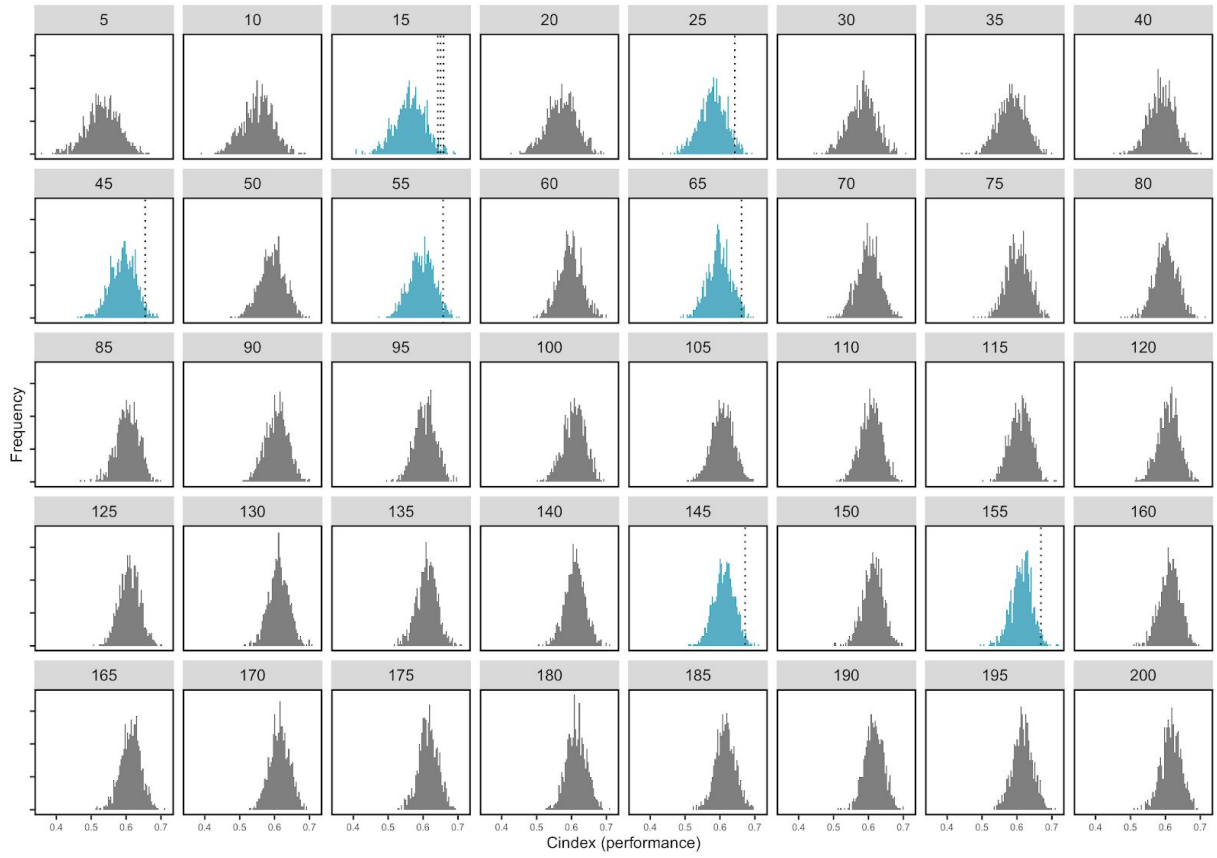
# LINCS results - Random survival forests (RSF) approach.

## Combined (1097 patients)



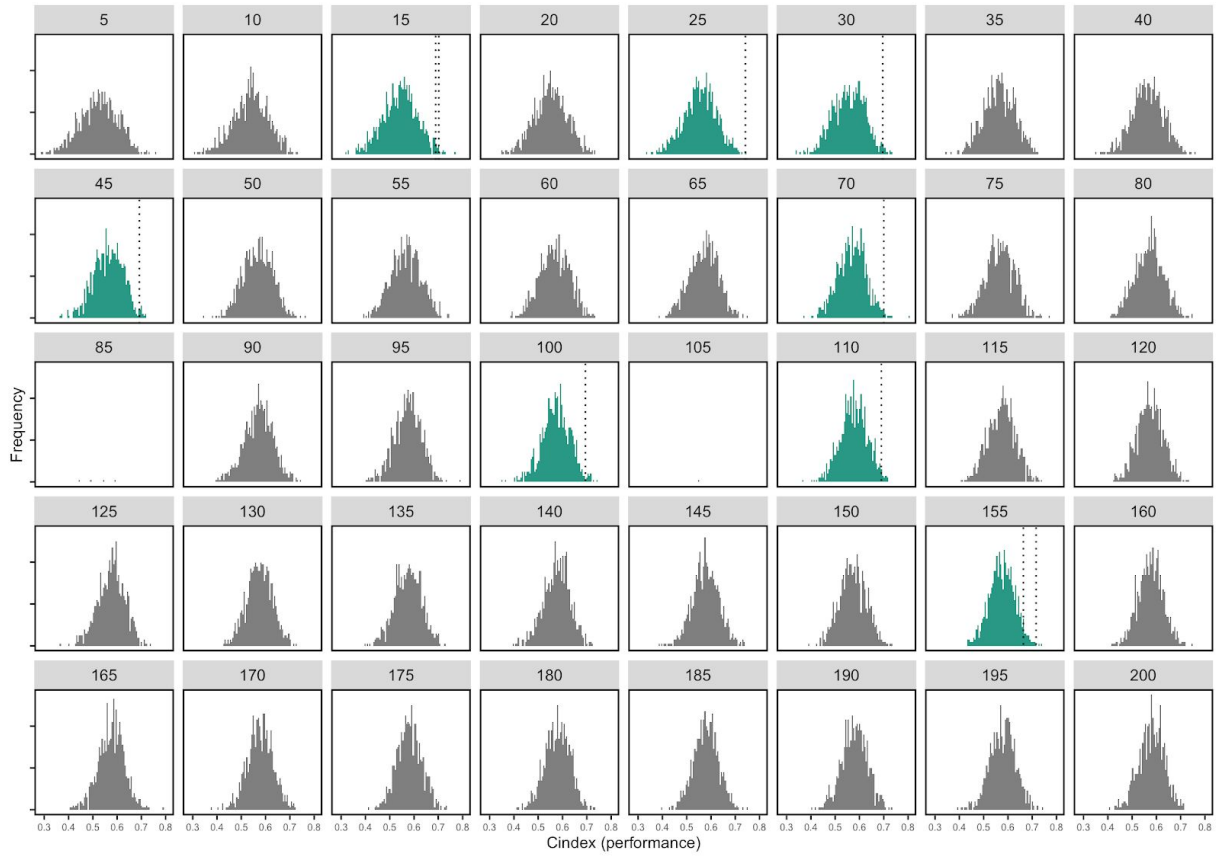

RSF combined results | Significant drugs were identified by comparing 1000 random gene signature models of signature lengths n by bins sized in increments of 5 genes, to drug models of equal signature lengths. Significant drugs were identified by considering performance (c-index) below 5% threshold of empirical p-values, computed by comparing to the background distribution. Bins with significant drugs are highlighted in color. Dotted lines indicate drug performance (c-index).

ER+ (808 patients)



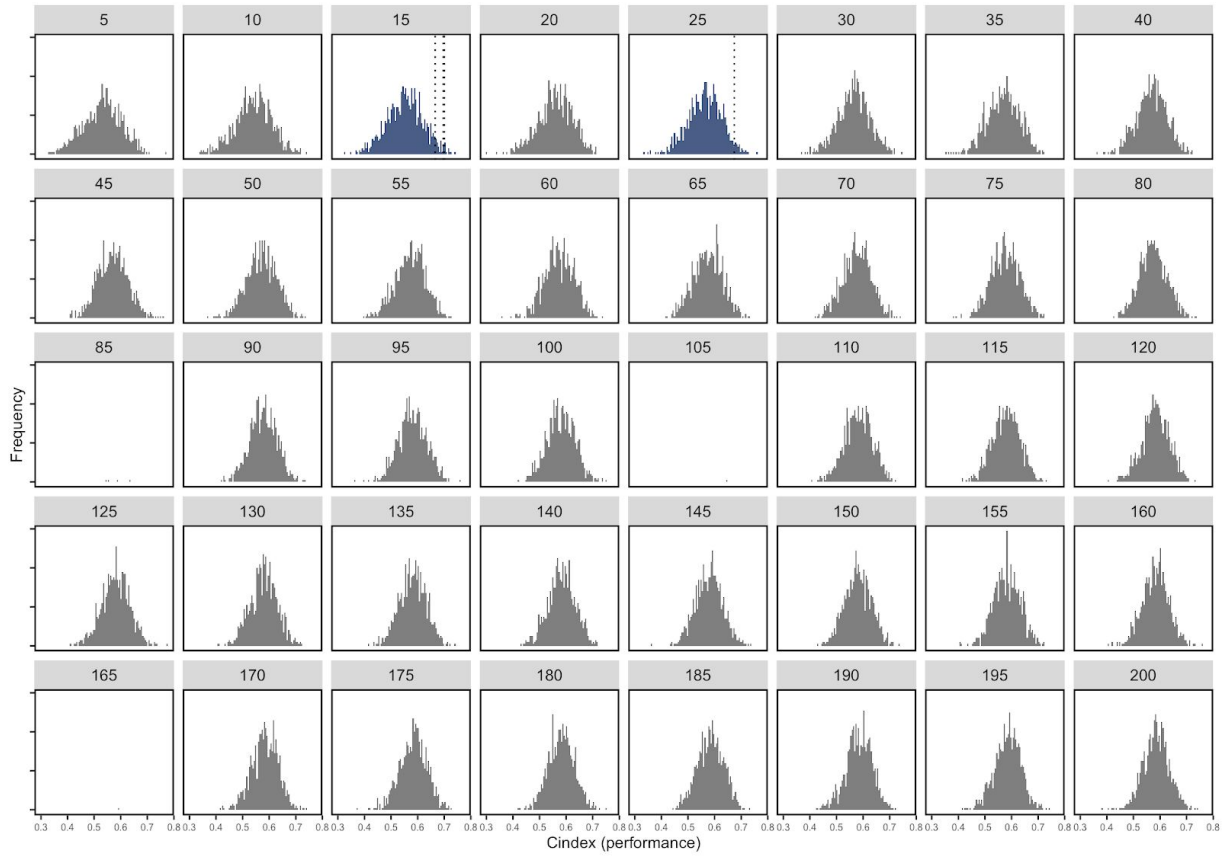

RSF ER+ results | Significant drugs were identified by comparing 1000 random gene signature models of signature lengths n by bins sized in increments of 5 genes, to drug models of equal signature lengths. Significant drugs were identified by considering performance (c-index) below 5% threshold of empirical p-values, computed by comparing to the background distribution. Bins with significant drugs are highlighted in color. Dotted lines indicate drug performance (c-index).

Her2 (343 patients)




RSF Her2 results | Significant drugs were identified by comparing 1000 random gene signature models of signature lengths n by bins sized in increments of 5 genes, to drug models of equal signature lengths. Significant drugs were identified by considering performance (c-index) below 5% threshold of empirical p-values, computed by comparing to the background distribution. Bins with significant drugs are highlighted in color. Dotted lines indicate drug performance (c-index).

Triple Negative (247 patients)




RSF Triple negative results | Significant drugs were identified by comparing 1000 random gene signature models of signature lengths n by bins sized in increments of 5 genes, to drug models of equal signature lengths. Significant drugs were identified by considering performance (c-index) below 5% threshold of empirical p-values, computed by comparing to the background distribution. Bins with significant drugs are highlighted in color. Dotted lines indicate drug performance (c-index).

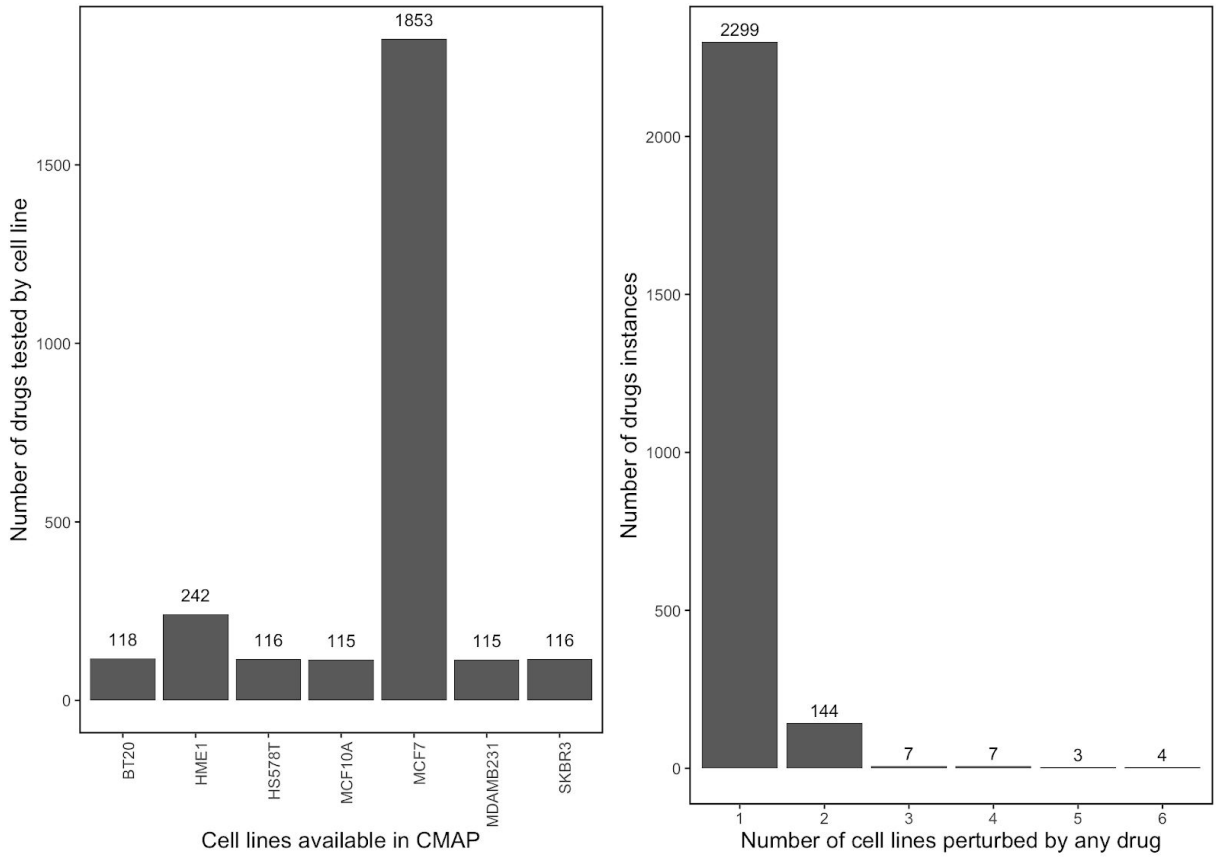
## LINCS Results - List of significant drugs using RSF approach

| Drug | Target | Approval | Combined | Er | Her2 | Tneg |
|---|---|---|---|---|---|---|
| Alvocidib | CDK1 | FDA | N | Y | Y | N |
| AT9283 | AURKA/B-JAK2 | Experimental | N | N | Y | N |
| AZ-628 | BRAF | Experimental | N | N | N | Y |
| AZD 5438 | CDK2 | Experimental | N | N | Y | N |
| BMS-536924 | IGF1R | Experimental | N | Y | N | N |
| CHIR-99021 | GSK3B | Experimental | N | N | Y | N |
| Crizotinib | c-Met | FDA | Y | Y | N | N |
| Dovitinib | FLT3 | FDA | Y | N | N | N |
| Epigallocatechin gallate | PP1A | FDA | N | Y | N | N |
| Gefitinib | EGFR | FDA | Y | N | Y | N |
| GSK 690693 | AKT1 | Experimental | N | N | Y | N |
| GSK1059615 | PIK3CA | Experimental | N | N | Y | N |
| HG-6-64-01 | ABL1 | Experimental | N | N | N | Y |
| Motesanib | VGFR1 | FDA | N | Y | N | N |
| Neratinib | ERBB2 | FDA | N | N | Y | N |
| PF-06463922 | ROS1 | Experimental | N | Y | N | N |
| PI103 | PIK3CA | Experimental | N | N | Y | N |
| PRT062607 | SYK | Experimental | N | Y | Y | N |
| Regorafenib | c-Kit | FDA | N | N | N | Y |
| S-Ruxolitinib |  | FDA | N | Y | N | N |
| Trichostatin A | HDAC1 | FDA | N | N | N | Y |
| VX-11e | ERK | Experimental | N | Y | N | N |

## LINCS Results - List of significant drugs using CPH approach

| Drug | Target | Approval | Combined | Er | Her2 | Tneg |
|---|---|---|---|---|---|---|
| Afatinib | ERBB2 | FDA | N | Y | N | N |
| Alvocidib | CDK1 | FDA | Y | Y | Y | N |
| AZD 5438 | CDK2 | Experimental | N | N | Y | N |
| Baicalein | CP2C9 | FDA | N | N | Y | N |
| BI-2536 | PLK1 | Experimental | N | Y | Y | N |
| BMS-536924 | IGF1R | Experimental | Y | N | N | N |
| Bosutinib | Src | FDA | Y | N | N | N |
| Cediranib | VEGF | FDA | N | Y | N | N |
| Gefitinib | EGFR | FDA | N | Y | N | N |
| Go 6976 | PKC | Experimental | N | N | Y | N |
| GSK1059615 | PIK3CA | Experimental | N | N | Y | N |
| HG-9-91-01 | SIK1 | Experimental | N | N | Y | N |
| LY2090314 | GSK3 | Experimental | N | N | Y | N |
| Masitinib | c-Kit | FDA | N | N | N | Y |
| Olomoucine II | CDK1 | FDA | N | Y | N | N |
| PF-06463922 | ROS1 | Experimental | N | Y | N | N |
| PI103 | PIK3CA | Experimental | N | N | N | Y |
| PRT062607 | SYK | Experimental | N | N | Y | N |
| Regorafenib | c-Kit | FDA | N | N | N | Y |
| SB 216763 | GSK3B | Experimental | Y | Y | N | N |
| Sotrastaurin | PKC | FDA | N | Y | N | N |
| XMD16-144 | AURKA | Experimental | N | N | Y | N |
| Y-27632 | ROCK1 | Experimental | N | Y | N | N |
| ZM-447439 | AURKA | Experimental | N | Y | N | N |

## cMAP drug instances by cell line




We considered only drugs that were tested on breast cancer (BRCA) cell lines in cMAP. There were 7 such cell lines. Distribution of drugs tested on each cell line is shown in the figure on the left. MCF7 (ER+) cell line had the greatest number of drugs tested. A distribution of number of drugs tested in multiple cell lines are shown on the figure on the right. Even though 2299 drugs were tested on at least 1 cell line, this number includes replicates. There were 1711 unique drugs in total.

## List of LINCS drugs available in CMAP drug instance list.

Link to excel file.

## List of FDA approved drugs for breast cancer.

Link to excel file.