Inhoudsopgave

[Nota’s 15 november 2](#_Toc499374004)

[Voor cisplatin 4](#_Toc499374005)

[Nota’s 17 november 5](#_Toc499374006)

[Voor cisplatin 5](#_Toc499374007)

[Voor Doxorubicin 6](#_Toc499374008)

[Voor Gemcitabine 6](#_Toc499374009)

[Voor SN38 8](#_Toc499374010)

[Voor Vinorelbine 9](#_Toc499374011)

[Nota’s 20 november 10](#_Toc499374012)

[Nota’s 21 November 11](#_Toc499374013)

[Start test with SN38 (new result) 11](#_Toc499374014)

[Cisplatin (same result) 12](#_Toc499374015)

[Doxorubicin (same result) 13](#_Toc499374016)

[Gemcitabine (new result) 14](#_Toc499374017)

[Vinorelbine (new result) 14](#_Toc499374018)

[Nota’s 24 November 16](#_Toc499374019)

[Cisplatin 🡪 good comp. to literature and ok comp large dataset 17](#_Toc499374020)

[Doxorubicin -> very different comp. to large dataset; good to literature 19](#_Toc499374021)

[Gemcitabine 🡪 different to large dataset ; not ok comp to literature 21](#_Toc499374022)

[SN38 🡪 no similarity to large dataset ; don’t find info in literature 23](#_Toc499374023)

[Vinorelbine 🡪 many differences comp large dataset ; no info to comp with literature 25](#_Toc499374024)

# Nota’s 15 november

🡪 put new scripts on server

🡪 run processData script:

%% Main processing script

% Fill in the names of the file

mutfile = 'mutcadd15fathmmpath.csv';

gexfile = 'gexfulltable1000.csv';

cnvfile = 'cnv\_gene.csv';

% corrcnvth is a parameter to filter CNVs to only those that correlates to

% expression

corrcnvth = 0.01;

% Load molecular data

chdata = loadMolecularData(entrezmapfile, mutfile, gexfile, cnvfile, corrcnvth);

% Load Drug monotherapy data

chdata = drugMonoscore(chdata);

% Load gene network information

disp('Loading network...');

net = loadNet('KEGG-Rel-ACSN-PPI2.csv');

% Restrict to network genes

chdata\_net = restrictNet(chdata, net);

% Computing expression marginals

% thdiff = parameter discretizing expression

% thgexon = minimum for expression to be considered on

% thstdgex = minimum STD for gene expression to be considered varied enough

thdiff = 0.7;

thgexon = 5;

thstdgex = 2;

disp('Computing marginals...');

chdata\_net = marginalExp(chdata\_net, thdiff, thgexon, thstdgex);

% Now we can perform network-based feature selection on drugs

## Voor cisplatin

% idxd = drug id 🡪 idxd = 22

% idxpos = cell ids with positive response 🡪 see lower in script!

% idxneg = cell ids with negative response

***nr 83 is known to be intermediate***

drug = 'cisplatin';

fprintf('Performing network-based feature selection on drug %s...\n', drug);

idxd = find(strcmpi(chdata.allDrugs, drug)); 🡪 idxd moet 22 zijn (control) ☺

% Simple threshold of 0.5 for Activity Area 🡪 adjust threshold (see word doc); set threshold on < 0,5 and >0,54 🡪 so delete nr 83 that is intermediate

idxpos = find(chdata.AAMat(idxd, :)<= 0.5);

idxneg = find(chdata.AAMat(idxd, :)> 0.54);

dg = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

% Select genes from network scores

% topn = Number of genes to select among the top scoring in each feature set

topn = 5;

selectedGenes = topGenes(dg, chdata\_net, net, topn); 🡪 ERROR!!!!

>> selectedGenes = topGenes(dg, chdata\_net, net, topn);

To use 'quantilenorm', you might need:

  quantilenorm - Bioinformatics Toolbox

Error in topGenes (line 37)

    smut = quantilenorm(smut);

% Type selectedGenes to see the top selected genes

# Nota’s 17 november

🡪 restart script after changing 1 script + add 1

🡪 make sure in the end to save both selectedGenes as the dg as back-up files/results

## Voor cisplatin

🡪 Now no error ☺

selectedGenes =

      MUTP        MUTN        CNVP          CNVN         GEXP        GEXN        NETP        NETN

    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_

    'CDH1'      'PIK3CA'    'UGT2B17'    'EIF1AY'      'CAD'       'EIF1AY'    'NUP155'    'EIF1AY'

    'CLDN17'    'CDKN1B'    'GNAI3'      'CDKN2B'      'LAMA3'     'TNPO1'     'CDH1'      'TMSB4Y'

    'IL6ST'     'NOTCH3'    'OSMR'       'RBM5'        'SPOP'      'TPR'       'RYR3'      'TBL1Y'

    'NUP155'    'CYP2S1'    'LIFR'       'HYAL2'       'CREBBP'    'RBM5'      'SPOP'      'PIK3CA'

    'SNAP23'    'HTR4'      'IFNAR1'     'MAPKAPK3'    'NUP155'    'CDKN1B'    'CAD'       'RXRA'

>> selectedGenes\_cisplatin = topGenes(dg, chdata\_net, net, topn);

>> selectedGenes\_cisplatin

>> dg\_cisplatin = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

>> save dg\_cisplatin

## Voor Doxorubicin

drug = 'doxorubicin';

fprintf('Performing network-based feature selection on drug %s...\n', drug);

idxd = find(strcmpi(chdata.allDrugs, drug));

% Simple threshold of 0.5 for Activity Area 🡪 adjust threshold (see word doc)

idxpos = find(chdata.AAMat(idxd, :)<= 0.4500);

idxneg = find(chdata.AAMat(idxd, :)> 0.4500);

dg\_doxorubicin = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

% Select genes from network scores

% topn = Number of genes to select among the top scoring in each feature set

topn = 5;

selectedGenes = topGenes(dg, chdata\_net, net, topn);

% Type selectedGenes to see the top selected genes

selectedGenes =

MUTP MUTN CNVP CNVN GEXP GEXN NETP NETN

\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_

'EGFR' 'ZFHX3' 'BAP1' 'GNAI3' 'ADCY3' 'EIF1AY' 'TP53' 'EIF1AY'

'ACAT2' 'CHD6' 'CNOT1' 'CHD6' 'HMGCL' 'ASH1L' 'STXBP4' 'TMSB4Y'

'PIK3CA' 'SACS' 'ARID1A' 'ASH1L' 'ACAT2' 'SACS' 'ST6GAL2' 'CHD6'

'CREBBP' 'MYH2' 'STXBP4' 'MSH3' 'ELF1' 'KEAP1' 'HAO1' 'ZFHX3'

'STXBP4' 'STK11' 'ELF1' 'KEAP1' 'CREBBP' 'KRAS' 'MAP4K3' 'ASH1L'

🡪 save it:

selectedGenes\_doxorubicin = topGenes(dg, chdata\_net, net, topn);

save selectedGenes\_doxorubicin

>> dg\_doxorubicin = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

>> save dg\_doxorubicin

## Voor Gemcitabine

drug = 'gemcitabine';

fprintf('Performing network-based feature selection on drug %s...\n', drug);

idxd = find(strcmpi(chdata.allDrugs, drug));

% Simple threshold of 0.5 for Activity Area 🡪 adjust threshold (see word doc) ! not correct to choose according to literature

idxpos = find(chdata.AAMat(idxd, :)<= 0.69);

idxneg = find(chdata.AAMat(idxd, :)> 0.69);

dg\_gemcitabine = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

save dg\_gemcitabine

% Select genes from network scores

% topn = Number of genes to select among the top scoring in each feature set

topn = 5;

selectedGenes\_gemcitabine = topGenes(dg, chdata\_net, net, topn);

save selectedGenes\_gemcitabine

selectedGenes\_gemcitabine

 MUTP        MUTN        CNVP          CNVN         GEXP        GEXN        NETP        NETN

    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_

    'CDH1'      'PIK3CA'    'UGT2B17'    'EIF1AY'      'CAD'       'EIF1AY'    'NUP155'    'EIF1AY'

    'CLDN17'    'CDKN1B'    'GNAI3'      'CDKN2B'      'LAMA3'     'TNPO1'     'CDH1'      'TMSB4Y'

    'IL6ST'     'NOTCH3'    'OSMR'       'RBM5'        'SPOP'      'TPR'       'RYR3'      'TBL1Y'

    'NUP155'    'CYP2S1'    'LIFR'       'HYAL2'       'CREBBP'    'RBM5'      'SPOP'      'PIK3CA'

    'SNAP23'    'HTR4'      'IFNAR1'     'MAPKAPK3'    'NUP155'    'CDKN1B'    'CAD'       'RXRA'

## Voor SN38

drug = 'SN38';

fprintf('Performing network-based feature selection on drug %s...\n', drug);

idxd = find(strcmpi(chdata.allDrugs, drug));

% Simple threshold of 0.5 for Activity Area 🡪 adjust threshold (see word doc)

idxpos = find(chdata.AAMat(idxd, :)<= 0.86);

idxneg = find(chdata.AAMat(idxd, :)> 0.86);

dg\_SN38 = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

save dg\_SN38

% Select genes from network scores

% topn = Number of genes to select among the top scoring in each feature set

topn = 5;

selectedGenes\_SN38 = topGenes(dg, chdata\_net, net, topn);

save selectedGenes\_SN38

selectedGenes\_SN38

 MUTP        MUTN        CNVP          CNVN         GEXP        GEXN        NETP        NETN

    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_

    'CDH1'      'PIK3CA'    'UGT2B17'    'EIF1AY'      'CAD'       'EIF1AY'    'NUP155'    'EIF1AY'

    'CLDN17'    'CDKN1B'    'GNAI3'      'CDKN2B'      'LAMA3'     'TNPO1'     'CDH1'      'TMSB4Y'

    'IL6ST'     'NOTCH3'    'OSMR'       'RBM5'        'SPOP'      'TPR'       'RYR3'      'TBL1Y'

    'NUP155'    'CYP2S1'    'LIFR'       'HYAL2'       'CREBBP'    'RBM5'      'SPOP'      'PIK3CA'

    'SNAP23'    'HTR4'      'IFNAR1'     'MAPKAPK3'    'NUP155'    'CDKN1B'    'CAD'       'RXRA'

## Voor Vinorelbine

drug = 'Vinorelbine';

fprintf('Performing network-based feature selection on drug %s...\n', drug);

idxd = find(strcmpi(chdata.allDrugs, drug));

% Simple threshold of 0.5 for Activity Area 🡪 adjust threshold (see word doc) ! no correct according to literature

idxpos = find(chdata.AAMat(idxd, :)<= 0.50);

idxneg = find(chdata.AAMat(idxd, :)> 0.50);

dg\_Vinorelbine = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

save dg\_Vinorelbine

% Select genes from network scores

% topn = Number of genes to select among the top scoring in each feature set

topn = 5;

selectedGenes\_Vinorelbine = topGenes(dg, chdata\_net, net, topn);

save selectedGenes\_Vinorelbine

selectedGenes\_Vinorelbine

MUTP        MUTN        CNVP          CNVN         GEXP        GEXN        NETP        NETN

    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_

    'CDH1'      'PIK3CA'    'UGT2B17'    'EIF1AY'      'CAD'       'EIF1AY'    'NUP155'    'EIF1AY'

    'CLDN17'    'CDKN1B'    'GNAI3'      'CDKN2B'      'LAMA3'     'TNPO1'     'CDH1'      'TMSB4Y'

    'IL6ST'     'NOTCH3'    'OSMR'       'RBM5'        'SPOP'      'TPR'       'RYR3'      'TBL1Y'

    'NUP155'    'CYP2S1'    'LIFR'       'HYAL2'       'CREBBP'    'RBM5'      'SPOP'      'PIK3CA'

    'SNAP23'    'HTR4'      'IFNAR1'     'MAPKAPK3'    'NUP155'    'CDKN1B'    'CAD'       'RXRA'

# Nota’s 20 november

🡪 4 drugs same selected genes

-> recheck if normal

**all same outcome**

**🡪 check on common cellines**

**vb. Cisplatin – SN38 no similar cellines**

# Nota’s 21 November

**🡪** Correction adaption script:

Perhaps this is the problem: when you run the topGenes script, you have to give it the 'dg' object  
from the corresponding result of drugGeneNetAA, for example:

dg\_SN38 = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);&nbsp;  
  topGenes(dg\_SN38, chdata\_net, net, topn);

and not

dg\_SN38 = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);&nbsp;  
  topGenes(dg, chdata\_net, net, topn);

Can you try again and see whether you get different results this time?

## Start test with SN38 (new result)

drug = 'SN38';

fprintf('Performing network-based feature selection on drug %s...\n', drug);

idxd = find(strcmpi(chdata.allDrugs, drug));

% Simple threshold of 0.5 for Activity Area 🡪 adjust threshold (see word doc)

idxpos = find(chdata.AAMat(idxd, :)<= 0.86);

idxneg = find(chdata.AAMat(idxd, :)> 0.86);

dg\_SN38 = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

save dg\_SN38

% Select genes from network scores

% topn = Number of genes to select among the top scoring in each feature set

topn = 5;

selectedGenes\_SN38 = topGenes(dg\_SN38, chdata\_net, net, topn);

save selectedGenes\_SN38

selectedGenes\_SN38

selectedGenes\_SN38 =

      MUTP        MUTN        CNVP        CNVN        GEXP        GEXN         NETP        NETN

    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_

    'LRRC4C'    'COL6A3'    'CDKN2A'    'IFNGR1'    'POLQ'      'SLC44A2'    'NT5C1B'    'TLN1'

    'POLQ'      'TLN1'      'MINPP1'    'IFNAR1'    'USP34'     'THBS2'      'LRRC4C'    'THBS2'

    'TP53'      'RASA2'     'TBL1Y'     'EIF1AY'    'NUP153'    'SMARCA1'    'POLQ'      'EIF1AY'

    'MPDZ'      'KSR2'      'POLQ'      'TLN1'      'NSFL1C'    'EMILIN1'    'MPDZ'      'RASA2'

    'NT5C1B'    'ADH1B'     'CHD9'      'ARID1A'    'WHSC1'     'AIM2'       'TP53'      'COL6A3'

🡪 NEW RESULT!!!!

🡪 recheck all drugs

## Cisplatin (same result)

drug = 'cisplatin';

fprintf('Performing network-based feature selection on drug %s...\n', drug);

idxd = find(strcmpi(chdata.allDrugs, drug)); 🡪 idxd moet 22 zijn (control) ☺

% Simple threshold of 0.5 for Activity Area 🡪 adjust threshold (see word doc); set threshold on < 0,5 and >0,54 🡪 so delete nr 83 that is intermediate

idxpos = find(chdata.AAMat(idxd, :)<= 0.5);

idxneg = find(chdata.AAMat(idxd, :)> 0.54);

dg\_cisplatin = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

save dg\_cisplatin

% Select genes from network scores

% topn = Number of genes to select among the top scoring in each feature set

topn = 5;

selectedGenes\_cisplatin = topGenes(dg\_cisplatin, chdata\_net, net, topn);

selectedGenes\_cisplatin

! get error Error: "idxd" was previously used as a variable, conflicting with its use here as the name of a function or command.

🡪 doesn’t want to change idxd to 22 ☹ change it manually to 22

🡪 rerun with idxd = 22: same result as original:

selectedGenes\_cisplatin =

      MUTP        MUTN        CNVP          CNVN         GEXP        GEXN        NETP        NETN

    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_

    'CDH1'      'PIK3CA'    'UGT2B17'    'EIF1AY'      'CAD'       'EIF1AY'    'NUP155'    'EIF1AY'

    'CLDN17'    'CDKN1B'    'GNAI3'      'CDKN2B'      'LAMA3'     'TNPO1'     'CDH1'      'TMSB4Y'

    'IL6ST'     'NOTCH3'    'OSMR'       'RBM5'        'SPOP'      'TPR'       'RYR3'      'TBL1Y'

    'NUP155'    'CYP2S1'    'LIFR'       'HYAL2'       'CREBBP'    'RBM5'      'SPOP'      'PIK3CA'

    'SNAP23'    'HTR4'      'IFNAR1'     'MAPKAPK3'    'NUP155'    'CDKN1B'    'CAD'       'RXRA'

## Doxorubicin (same result)

drug = 'doxorubicin';

fprintf('Performing network-based feature selection on drug %s...\n', drug);

idxd = find(strcmpi(chdata.allDrugs, drug));

% Simple threshold of 0.5 for Activity Area 🡪 adjust threshold (see word doc)

idxpos = find(chdata.AAMat(idxd, :)<= 0.4500);

idxneg = find(chdata.AAMat(idxd, :)> 0.4500);

dg\_doxorubicin = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

save dg\_doxorubicin

% Select genes from network scores

% topn = Number of genes to select among the top scoring in each feature set

topn = 5;

selectedGenes\_doxorubicin = topGenes(dg\_doxorubicin, chdata\_net, net, topn);

selectedGenes\_doxorubicin

selectedGenes\_doxorubicin =

      MUTP       MUTN        CNVP       CNVN        GEXP        GEXN        NETP         NETN

    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_

    'EGFR'      'ZFHX3'    'BAP1'      'GNAI3'    'ADCY3'     'EIF1AY'    'TP53'       'EIF1AY'

    'ACAT2'     'CHD6'     'CNOT1'     'CHD6'     'HMGCL'     'ASH1L'     'STXBP4'     'TMSB4Y'

    'PIK3CA'    'SACS'     'ARID1A'    'ASH1L'    'ACAT2'     'SACS'      'ST6GAL2'    'CHD6'

    'CREBBP'    'MYH2'     'STXBP4'    'MSH3'     'ELF1'      'KEAP1'     'HAO1'       'ZFHX3'

    'STXBP4'    'STK11'    'ELF1'      'KEAP1'    'CREBBP'    'KRAS'      'MAP4K3'     'ASH1L'

**🡪 Same result as original**

## Gemcitabine (new result)

drug = 'gemcitabine';

fprintf('Performing network-based feature selection on drug %s...\n', drug);

idxd = find(strcmpi(chdata.allDrugs, drug));

% Simple threshold of 0.5 for Activity Area 🡪 adjust threshold (see word doc) ! not correct to choose according to literature

idxpos = find(chdata.AAMat(idxd, :)<= 0.69);

idxneg = find(chdata.AAMat(idxd, :)> 0.69);

dg\_gemcitabine = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

save dg\_gemcitabine

% Select genes from network scores

% topn = Number of genes to select among the top scoring in each feature set

topn = 5;

selectedGenes\_gemcitabine = topGenes(dg\_gemcitabine, chdata\_net, net, topn);

save selectedGenes\_gemcitabine

selectedGenes\_gemcitabine

selectedGenes\_gemcitabine =

      MUTP        MUTN         CNVP         CNVN        GEXP        GEXN        NETP         NETN

    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_

    'EP300'     'PIK3CA'    'TBL1Y'       'CDKN2A'    'CAD'       'NCOR1'     'RYR2'       'PIK3CA'

    'RYR2'      'NRAS'      'UGT2B17'     'IFNGR1'    'NFKBIZ'    'KEAP1'     'EP300'      'CHD6'

    'STXBP4'    'CREBBP'    'RBM5'        'IFNAR1'    'ALCAM'     'ASH1L'     'HAO1'       'NCOR1'

    'HMGCL'     'NCOR1'     'RASSF1'      'JAK2'      'SHMT2'     'PIK3CA'    'ST6GAL2'    'ASH1L'

    'RHOA'      'MLL3'      'MAPKAPK3'    'LIFR'      'SCIN'      'SPOP'      'TRIP6'      'ANK1'

## Vinorelbine (new result)

drug = 'Vinorelbine';

fprintf('Performing network-based feature selection on drug %s...\n', drug);

idxd = find(strcmpi(chdata.allDrugs, drug));

% Simple threshold of 0.5 for Activity Area 🡪 adjust threshold (see word doc) ! no correct according to literature

idxpos = find(chdata.AAMat(idxd, :)<= 0.50);

idxneg = find(chdata.AAMat(idxd, :)> 0.50);

dg\_Vinorelbine = drugGeneNetAA(chdata\_net, idxd, net, idxpos, idxneg);

save dg\_Vinorelbine

% Select genes from network scores

% topn = Number of genes to select among the top scoring in each feature set

topn = 5;

selectedGenes\_Vinorelbine = topGenes(dg\_Vinorelbine, chdata\_net, net, topn);

save selectedGenes\_Vinorelbine

selectedGenes\_Vinorelbine

selectedGenes\_Vinorelbine =

      MUTP         MUTN         CNVP        CNVN        GEXP        GEXN        NETP         NETN

    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_    \_\_\_\_\_\_\_\_\_

    'ZFHX3'      'HSPG2'      'BAP1'       'GNAI3'    'GLG1'       'KEAP1'    'ZFHX3'      'TP53'

    'PIK3CA'     'NTSR1'      'EIF3A'      'ASH1L'    'STIP1'      'ASH1L'    'TMSB4Y'     'HSPG2'

    'STK11'      'RAPGEF2'    'CPSF6'      'KEAP1'    'GSPT1'      'WNK1'     'CACNA1D'    'NTSR1'

    'CACNA1D'    'CHD6'       'PPIP5K2'    'PBRM1'    'CDKN2A'     'ELF1'     'STK11'      'KLRD1'

    'ZFYVE16'    'KLRD1'      'SPATA13'    'WNK1'     'CSNK2A1'    'ASXL1'    'EIF1AY'     'RAPGEF2'

# Nota’s 24 November

*That's great. I think we can even start* ***analyzing the results*** *now, by checking whether the* ***top mutation and CNV genes (top 5 or top 10****) contain the genes known in the* ***literature****, or by comparing them to the list I found with the* ***larger data set*** *(GDSC, attached). If you can find matches, then it make sense to continue, otherwise, then probably our method does not work well with this dataset (due to the size). Maybe we should discussed again how to proceed, based on your analysis of this result.*

op eerste zicht tussen lijst volledige data set en mijn resultaten geen gelijkenissen ☹

--> nu top 10 ook checken (grondig) + literatuur

MUT**N** --> mutation negative reacting cel lines

MUT**P** --> mutation positive reacting cel lines

## Cisplatin 🡪 good comp. to literature and ok comp large dataset

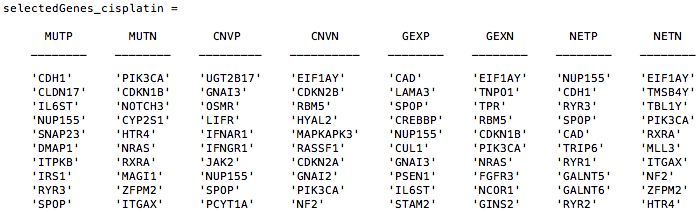
🡪 10 cell lines in analysis; drugscore results were similar to literature; threshold is OK

load dg\_cisplatin

topn = 10;

selectedGenes\_cisplatin = topGenes(dg\_cisplatin, chdata\_net, net, topn);

selectedGenes\_cisplatin



🡪 VGL large dataset:

|  |
| --- |
| CISPLATIN,APC-MUT,1,26.9185146988319 |
| CISPLATIN,MAP2K4-CNV,2,17.4788306475341 |
| CISPLATIN,JAK2-CNV,3,16.0501732791024 🡪 Nr. 7 bij CNVP |
| CISPLATIN,SDHC-CNV,4,13.4309882001767 |
| CISPLATIN,IRS2-CNV,5,8.10262162125008 🡪 IRS1 bij MUTP |
| CISPLATIN,TP53-MUT,6,6.39144779202793 |
| CISPLATIN,KRAS-CNV,7,5.66070906080542 🡪 NRAS nr. 6 MUTN |
| CISPLATIN,APC-CNV,8,3.38294968757442 |

🡪 Literature:

* CDH1-MUTP: role in E-cadherine 🡪 invasion & metastasis 🡪 VERY important in cancer! Ivm EMT/MET pathway and switch; cancer stem cells + link pos. cisplatin
* CLDN17\_MUTP: role in claudin-17 🡪 tight junctions ; also more a role in spread of cancer; members of this family are often downregulated in cancer ( prob. Most gastro-intest. cancers) 🡪 cisplatin sensitivity
* IL6ST-MUTP: interleukin 6 signal transductor 🡪 should have link to breast cancer and herpes viruses (inflammation aspect cancer; kaposi sarcoma skin cancer) 🡪 ALDHhi endometrial cells are resistant to cisplatin and sensitive to combination of cisplatin and IL6 signaling inhibitors
* NUP155-MUTP: role nuclear pore complex 🡪 Nup155 is linked to the p53 pathway in hepatocellular carcinoma 🡪 p53 is most mutated gen in cancer 🡪 rather linked to resistance of cisplatin
* SNAP23-MUTP: promotes the process of ovarian cancer 🡪 cisplatin sensitive
* PIK3CA-MUTN; CDKN1B-MUTN: associated with breast cancer 🡪 cisplatin resistance
* NOTCH3-MUTN: is tumor suppressor or promotor depending on cell tissue (if promoting again role in adhesion) 🡪 cisplatin resistance
* CYPS1-MUTN: marker in colorectal cancer; CYP2S1 has been suggested to be involved in the growth and/or spread of certain tumors of epithelial cell origin: its higher expression in breast or colorectal cancer tissues appears associated respectively with shorter survival times or poor prognoses, and it is more highly expressed in metastasis compared to primary tumor tissues of ovarian cancer.
* HTR4-MUTN: serotonine receptor; role in prostate cancer + neg. link cisplatin
* ugt2b17-CNVP prostate cancer by variants and gene deletions + pos link cisplatin
* GNAI3-CNVP inhibit tumor cell migration and invasion
* OSMR-CNVP 🡪 Oncostatin M receptor -> macrophages treated with cisplatin produce OSMR; is new therapeutic target in cervical cancer; when methylated 🡪 non-invasive cancer
* Methyation EIF1AY-CNVN 🡪 prostate cancer ; linked to cisplatin resistance
* CDKN2B-CNVN needs to be used in comb. With Gemcitabine to treat bladder cancer
* RB5M-CNVN 🡪 RBM5 reduces small cell lung cancer growth, increases cisplatin sensitivity and regulates key transformation-associated pathways. ([10.1016/j.heliyon.2016.e00204](https://doi.org/10.1016/j.heliyon.2016.e00204))

In general: all linked to cancer, mostly the invasive and metastasis aspect; some genes are even known to play a role with cisplatin (some studies talk about “resistance” when in occurs in positive and reverse situation occurs as well)

## Doxorubicin -> very different comp. to large dataset; good to literature

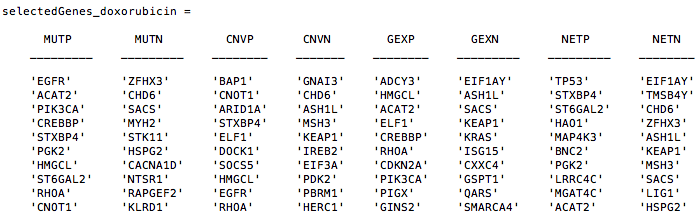
🡪 22 cell lines in analysis; drugscore results and threshold OK comp. to literature

load dg\_doxorubicin

topn = 10;

selectedGenes\_doxorubicin = topGenes(dg\_doxorubicin, chdata\_net, net, topn);

selectedGenes\_doxorubicin



🡪 VGL large dataset:

|  |
| --- |
| DOXORUBICIN,FGFR1-CNV,1,27.6217719749527 |
| DOXORUBICIN,LARP4B-CNV,2,23.84558397567 |
| DOXORUBICIN,CUL2-CNV,3,9.08259206413629 |
| DOXORUBICIN,IL7R-CNV,4,7.56631575013165 |
| DOXORUBICIN,ZMYM2-CNV,5,6.15439207966483 |
| DOXORUBICIN,PIP5K1A-CNV,6,5.44130807813353 |
| DOXORUBICIN,FAT1-CNV,7,5.42188128779742 |
| DOXORUBICIN,FBXW7-MUT,8,5.31194864466289 |
| DOXORUBICIN,MLL2-MUT,9,2.8072587023789 |
| DOXORUBICIN,NRAS-MUT,10,2.44314144739315 |
| DOXORUBICIN,TERT-CNV,11,1.28007553650537 |
| DOXORUBICIN,RAD21-CNV,12,1.10955912223294 |
| DOXORUBICIN,TP53BP1-CNV,13,0.654198191131813 |
| DOXORUBICIN,MECOM-CNV,14,0.60173173423444 |
| DOXORUBICIN,SMARCA4-MUT,15,0.325166329354635 |
| DOXORUBICIN,PCDH18-CNV,16,0.231004520785911 |
| DOXORUBICIN,SACS-MUT,17,0.102070560833566 🡪 NR.3 bij MUTN  🡪 VGL literature:   * EGFR-MUTP 🡪 EGFR-targeted multifunctional polymersomal doxorubicin induces selective and   potent suppression of orthotopic human liver cancer in vivo  (<http://www.sciencedirect.com/science/article/pii/S1742706117306293)>   * ACAT2-MUTP <https://www.researchgate.net/publication/280294458> \_Prediction\_of\_doxorubicin\_sensitivity\_in\_gastric\_cancers\_based\_on\_a\_set\_of\_novel\_markers * BRCA1 overexpression and phosphoinositide 3-kinase (***PIK3CA***-MUTP) pathway activation are involved in the **resistance to DNA damaging agents** like doxorubicin.   🡪 Except for first MUTP gene, other genes should be linked to resistance instead of sensitivity   * ZFHX3-MUTN 🡪 no link with drug found * CHD6-MUTN 🡪 linked to drug resistance of doxorubicin * SACS-MUTN 🡪 no link with drug found * BAP1-CNVP 🡪 BAP1 dependent expression of long non-coding RNA NEAT-1 contributes to sensitivity to gemcitabine in cholangiocarcinoma 10.1186/s12943-017-0587-x * CNOT1-CNVP 🡪 Co‐administration phenoxodiol with doxorubicin synergistically inhibit the activity of sphingosine kinase‐1 (SphK1), a potential oncogene of osteosarcoma, to suppress osteosarcoma cell growth both in vivo and in vitro. Mol Oncol 6, 392–404 ; CNOT1 cooperates with LMNA to aggravate osteosarcoma tumorigenesis through the Hedgehog signaling pathway * ARID1A-CNVP 🡪 Decreased ARID1A expression is correlated with chemoresistance in epithelial ovarian cancer * GNAI3-CNVN 🡪 no link found * CHD6-CNVN 🡪 linked to drug resistance of doxorubicin * ASH1L-CNVN 🡪 no link found |

## Gemcitabine 🡪 different to large dataset ; not ok comp to literature

🡪 28 cell lines in analysis; drugscores were completely different to literature

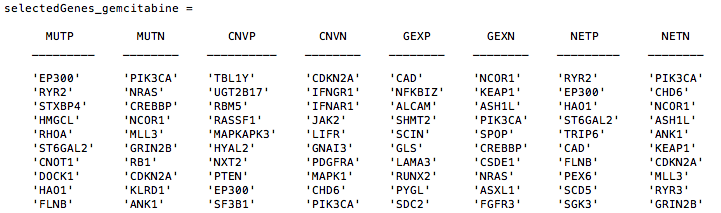
load dg\_gemcitabine

topn = 10;

selectedGenes\_gemcitabine = topGenes(dg\_gemcitabine, chdata\_net, net, topn);

save selectedGenes\_gemcitabine

selectedGenes\_gemcitabine



🡪 VGl large dataset

|  |
| --- |
| GEMCITABINE,ARFGAP3-CNV,2,14.6029017009355 |
| GEMCITABINE,STK4-CNV,3,11.3189496059983 |
| GEMCITABINE,NF1-MUT,4,7.85369483021555 |
| GEMCITABINE,LARP4B-CNV,5,6.67961382057083 |
| GEMCITABINE,ANK3-CNV,6,5.66481078774525 🡪 ANK1 nr. 10 MUTN |
| GEMCITABINE,CTTN-CNV,7,5.27751926103863 |
| GEMCITABINE,GNAS-CNV,8,4.34098884552644 |
| GEMCITABINE,MLL2-MUT,9,2.01249709742496 🡪 MLL3 nr. 5 MUTN |
| GEMCITABINE,ARID4B-CNV,10,1.97575588792142 |
| GEMCITABINE,NF1-CNV,11,1.97488186798592 |
| GEMCITABINE,EP300-MUT,12,1.84248214615972 🡪 NR. 1 MUTP ; NR. 9 CNVP |
| GEMCITABINE,PABPC1-CNV,13,1.60654922589604 |
| GEMCITABINE,NRAS-MUT,14,1.31718842107512 🡪 NR.2 MUTN |
| GEMCITABINE,FOXP1-CNV,15,1.08185420096165 |
| GEMCITABINE,KRAS-MUT,16,1.03575813155523 🡪 NRAS nr. 2 MUTN |

🡪 VGL literature

* EP300-MUTP 🡪 P300—a miRNA-regulated metastasis suppressor gene in ductal adenocarcinomas of the pancreas ; Bioinformatic analyses were also carried out on HGS ovarian cancer and PDAC patient datasets (TGCA and ICGC) as well as cell line datasets (Sanger Institute) aimed at determining single genes within DDR pathways whose aberrations in copy number and gene expression impacts upon patient survival or drug response to cisplatin or gemcitabine. 6 genes, TSTA3, RECQL4, ESRP1, NBN, SUMO3 and EP300, were identified that affected one of the above parameters, of which, EP300 functionally validated as a potential therapeutic target. SiRNA-mediated silencing of this was found to induce apoptosis in cell line models of HGS ovarian cancer (SKOV3) and PDAC (Panc-1). In addition, loss of EP300 increased the apoptotic response of SKOV3 cells to treatment with cisplatin, gemcitabine, doxorubicin, paclitaxel and the DNA-PKcs inhibitor NU7441. In Panc1 cells, only response to gemcitabine and paclitaxel was significantly increased with EP300 loss. (https://spiral.imperial.ac.uk/handle/10044/1/49207 )
* RYR2-MUTP 🡪 no link found
* STXBP4-MUTP 🡪 no link found
* PIK3CA-MUTN 🡪 PIK3CA mutations can initiate pancreatic tumorigenesis and are targetable with PI3K inhibitors + should be SENSITIVE to the drug
* NRAS-MUTN 🡪 no clear link with drug found ; plays role in metastatic pancreas cancer
* CREBBP-MUTN 🡪 Long non-coding RNA UCA1 promotes cisplatin/gemcitabine resistance through CREB modulating miR-196a-5p in bladder cancer cells.

🡪 STOP looking up genes ; CNV not checked ; IS not ok comp. with literature but drugscores were also not good comp. to literature

## SN38 🡪 no similarity to large dataset ; don’t find info in literature

🡪 10 cell lines in analysis; not much to find in literature; threshold needed to be set in the middle

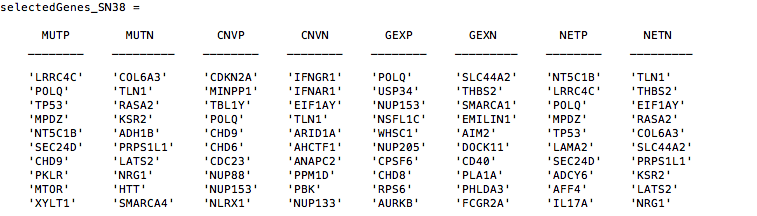
load dg\_SN38

topn = 10;

selectedGenes\_SN38 = topGenes(dg\_SN38, chdata\_net, net, topn);

save selectedGenes\_SN38

selectedGenes\_SN38



🡪 VGL large dataset:

|  |
| --- |
| SN-38,ARID4B-CNV,1,22.8180972320518 🡪 ARID1A nr. 5 bij CNVN |
| SN-38,BLM-CNV,2,19.6876368716556 |
| SN-38,APC-MUT,3,12.4382746461504 |
| SN-38,GNAS-CNV,4,11.5833914207852 |
| SN-38,RB1-MUT,5,10.775240599693 |
| SN-38,SDHC-CNV,6,7.47210061322612 |
| SN-38,MLL2-MUT,7,5.74148126749012 |
| SN-38,VIM-GEX,8,2.49776275020341 |
| SN-38,EWSR1-FLI1-MUT,9,2.32302599874596 |
| SN-38,SDHD-CNV,10,0.997281783094844 |
| SN-38,NRAS-MUT,11,0.973242159380218 |
| SN-38,STAG2-MUT,12,0.894974731051123 |
| SN-38,NF1-CNV,13,0.889239431183181 |
| SN-38,PCDH18-CNV,14,0.55115991405119 |
| SN-38,CREBBP-MUT,15,0.357090581237786 |

🡪 VGL Literature

* LRRC4C-MUTP 🡪 no link found
* POLQ-MUTP 🡪 no link found
* TP53-MUTP 🡪 creates cell cycle arrest
* COL6A3-MUTN 🡪 no link found
* TLN1-MUTN 🡪 no link found
* RASA1-MUTN 🡪 no link found

🡪 STOP search ; again in literature very few to find; often “patents” hits

## Vinorelbine 🡪 many differences comp large dataset ; no info to comp with literature

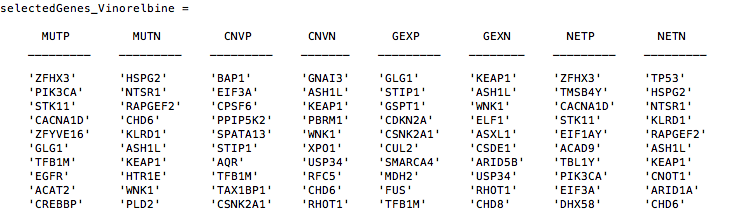
🡪 22 cell lines in analysis; not much to find in literature; threshold wss OK

load dg\_Vinorelbine

topn = 10;

selectedGenes\_Vinorelbine = topGenes(dg\_Vinorelbine, chdata\_net, net, topn);

save selectedGenes\_Vinorelbine



🡪 VGL large dataset:

|  |
| --- |
| VINORELBINE,LARP4B-CNV,1,17.0732117107962 |
| VINORELBINE,APC-MUT,2,13.1626824445369 |
| VINORELBINE,MLL2-MUT,3,10.9453279535709 |
| VINORELBINE,MDM2-CNV,4,10.7338763633039 |
| VINORELBINE,APC-CNV,5,7.34816831066844 |
| VINORELBINE,FGFR1-CNV,6,5.37781969771563 |
| VINORELBINE,ATRX-CNV,7,4.39063149884728 |
| VINORELBINE,CNOT1-CNV,8,4.31460610520618 |
| VINORELBINE,RAD21-CNV,9,3.49580246898428 |
| VINORELBINE,MYC-CNV,10,2.66862956871147 |
| VINORELBINE,GPS2-CNV,11,2.27453944597146 |
| VINORELBINE,GNAQ-CNV,12,2.1608403150669 |
| VINORELBINE,CHD9-CNV,13,2.10916576383762 🡪 CDH6 nr. 4 MUTN; nr. 9 CNVN |
| VINORELBINE,MYH11-MUT,14,1.92017245407955 |
| VINORELBINE,FOXA2-CNV,15,1.36013203344709 |
| VINORELBINE,ZNF292-CNV,16,1.2808245195308 |
| VINORELBINE,EGFR-CNV,17,1.20039708949379 🡪 nr. 8 MUTP |
| VINORELBINE,MACF1-MUT,18,1.08967866875102 |
| VINORELBINE,STK4-CNV,19,1.07264076842918 🡪 STK11 nr. 3 bij MUTP |
| VINORELBINE,CUL2-CNV,20,0.760381406432176 |
| VINORELBINE,SMAD4-CNV,21,0.668870693800043 |
| VINORELBINE,FOXP1-CNV,22,0.588724055638387 |

🡪 VGL Literature:

* ZFHX3-MUTP 🡪 no link found
* PIK3CA-MUTP 🡪 triple-negative breast cancer ; no link found
* STK11-MUTP 🡪 linked to magnoid subtype of lung cancer ; vinorelbine + cisplatin comb.
* HSPG2-MUTN 🡪 no link found ; breast cancer ; drug used by breast cancer
* NTSR1-MUTN 🡪 therapeutic target for lung cancer ; vinorelbine is given to cure lung cancer
* RAPGEF2-MUTN 🡪 no info found

🡪 STOP looking up genes; few info found, often “patent hits”