Predicting MCL1 inhibitor sensitivity in large cell line panels using a gene expression signature

Abstract # 3987

Andreas Wernitznig[1,4], Dorothea Rudolph[1], Matthias Samwer[1], Norbert Schweifer[1], Francesca Trapani[1], Tobias Wunberg[1], Heribert Arnhof[1], Teakyu Lee[2], John L. Sensintaffar[2], Edward T. Olejniczak[2], Cyril H. Benes[3], Stephen W. Fesik[2] and Norbert Kraut[1]

[1]Boehringer Ingelheim RCV GmbH & Co KG, 1120, Vienna, Austria, [2]Vanderbilt University, Nashville, TN, USA, [3]Massachusetts General Hospital, Boston, MA, USA, [4]andreas.wernitznig@boehringer-ingelheim.com

INTRODUCTION

MCL1, an anti-apoptotic member of the BCL-2 family of proteins, is a key regulator of cancer cell survival and a known resistance factor to anti-cancer drugs, making it a highly desirable target for therapeutic intervention. Recently several MCL1 inhibitors have entered Phase I clinical development. Data derived from large cancer cell line panels suggest that cell lines of hematopoietic origin are more broadly sensitive to MCL1 inhibition than cell lines derived from solid tumor types. Therefore in particular for the therapy of solid tumor patients with an MCL1 inhibitor, a patient selection biomarker would be highly desirable.

METHODS

3 solid tumor cell line panels from the Massachusetts General Hospital (682 cell lines), the shRNA screen Drive[1] (348 cell lines), and the CRISPR/Cas9 screen Avana[2] (430 cell lines, release Q1-2019) have been classified based on their sensitivity to MCL1 inhibition (IC_{50} of VU0810994), growth inhibition due to shRNA depletion (RSA score) or growth inhibition due to knock-out (CERES score), respectively.

	MGH (VU)	Drive	Avana
# cell lines	682	348	430
Sensitive	IC_{50} < 2µM	RSA < -3	top 100 CERES
Resistant	IC_{50} > 6µM	bottom 100 RSA	bottom 150 CERES

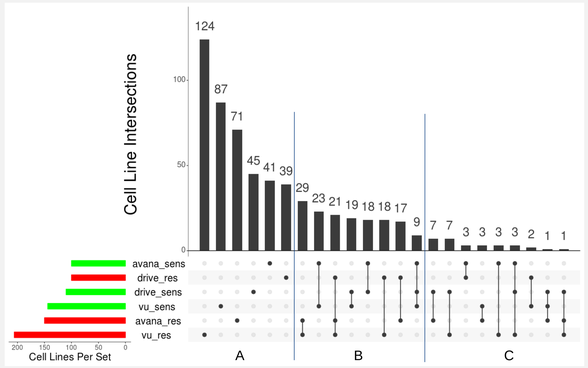
Gene expression data of these cell lines are based on Illumina TruSeq RNA-seq from either the Cancer Cell Line Encyclopedia or internal sequencing runs. RNA isolation, library preparation, and sequencing was performed as described in ref. 3. Data processing is based on an adapted ENCODE RNA-seq pipeline (GRCh38) and gene annotations from Ensembl 86.[4]
Differential gene expression of RNA-seq data is based on a modified t-test, implemented in the R package limma with voom extension.[5] This test has been applied with and without adding a strong confounding factor to the linear model.

The expression values (in TPM, transcripts per million) of differentially expressed genes have been utilized to train a support vector machine classifier (R statistical computing language[6], library e1071, kernel type: nu-classification). The training data set comprises two thirds and the test set the remaining third of the available cell lines.

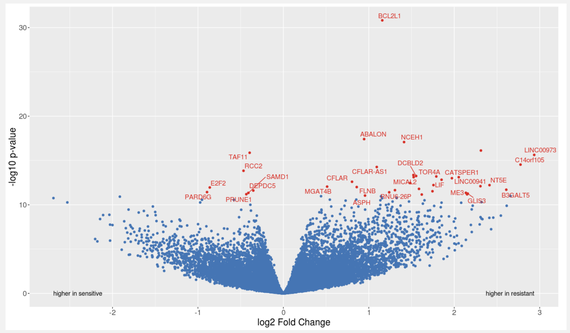
Another classifier has been derived from all sensitive and resistant solid tumor cell lines and was applied to tumor and normal tissues to estimate the response rates and putative toxicities of MCL1 inhibition.

Tissues have been taken from The Cancer Genome Atlas (TCGA)[7] and the Genotype-Tissue Expression (GTEx)[8] data sets, which are collections of cancer and normal tissues. RNA-seq expression data are available for most samples and have been processed with the same analysis pipeline as the cell lines.

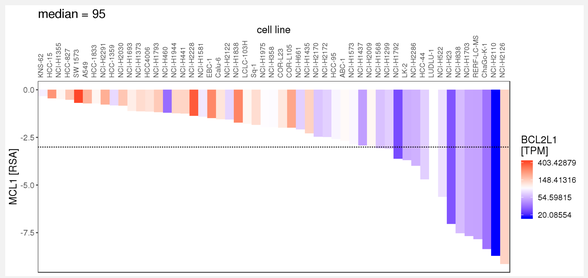
RESULTS CELL LINES



591 different solid tumor cell lines have been tested and classified (green = sensitive, red = resistant) in the three cell line panels. 407 cell lines (section A) have been tested once in either the Avana, Drive or MGH (VU) panel, respectively. 154 cell lines (section B) have been tested in 2 or more panels and show a consistent classification. 30 cell lines (section C) switch classes (sensitive/resistant) between panels. 561 cell lines (242 sensitive, 319 resistant) of section A and B have been used for further analysis.



Volcano plot of differentially expressed genes. BCL2L1 (BCL-XL) is by far the most significant differentially expressed gene between the sensitive and the resistant solid tumor cancer cell lines.



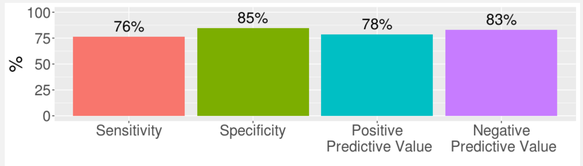
BCL2L1 (BCL-XL) expression of 16 sensitive and 37 resistant NSCLC cell lines of the Drive shRNA screen targeting MCL1. In this example as well as in all cell lines derived from solid tumors in general (see figure above), BCL2L1 is the most differentially expressed gene. Cell lines with a low expression of BCL2L1 are usually sensitive (→ RSA ≤ -3).

Symbol	Name	Ensembl gene	Location
BAK1	BCL2 antagonist/killer 1	ENSG00000030110	chr 6:33,572,547-33,580,293
BCL2L1	BCL2 like 1	ENSG00000171552	chr 20:31,664,452-31,723,989
FGFR3	fibroblast growth factor receptor 3	ENSG00000068078	chr 4:1,793,307-1,808,872
PRKCA	protein kinase C alpha	ENSG00000154229	chr 17:66,302,636-66,810,743
SLC22A23	solute carrier family 22 member 23	ENSG00000137266	chr 6:3,268,962-3,457,022
TULP1	tubby like protein 1	ENSG00000112041	chr 6:35,497,874-35,512,938

Taking BCL2L1 (BCL-XL) as a confounding factor in the linear model, five additional genes are consistently differentially expressed in sensitive versus resistant samples of each cell line panel. The expression of these genes has been utilized to train a supervised learning classifier (support vector machine).

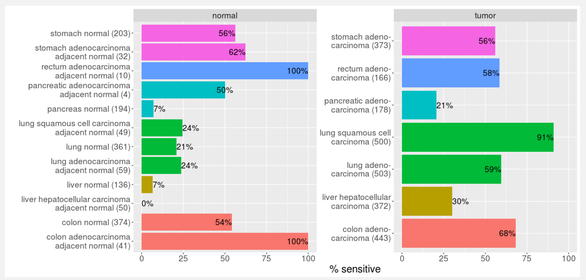
	Reference	
Prediction	resistant	sensitive
resistant	77	16
sensitive	14	51

The training set consists of 133 sensitive and 182 resistant cell lines. The derived predictor was applied to 67 sensitive and 91 resistant cell lines. The overall accuracy (correctly predicted cell lines) is 81% (128 of 158).

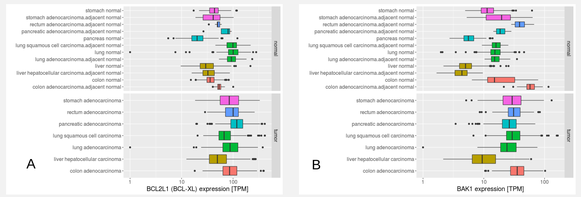


Performance parameters for the sensitivity/resistance prediction of solid tumor cell lines from the training set are in the range of 80%. This is a good performance for a heterogeneous data set.

RESULTS TISSUES



Predicted percentage of sensitive tumor and normal tissue samples from colon, liver, lung, pancreas, rectum and stomach. The total number of samples is shown in brackets. Only lung and hepatocellular carcinomas have a higher predicted sensitivity rate than their respective normal tissues.



Expression profile of two out of six predictor genes. (A) Many normal tissues have a low expression of BCL2L1 (BCL-XL). (B) BAK1 gene expression is correlated with sensitivity. Liver and pancreas normal tissues have a very low expression of BAK1 and are therefore classified mostly as resistant.

SUMMARY

Published *in vitro* data using various MCL1 inhibitors, siRNA or CRISPR/Cas9 technology show that tumor cell lines with low BCL-XL gene expression are mostly sensitive to MCL1 inhibition, down-regulation or inactivation. Here we demonstrate that by adding the gene expression data of additional five genes (including the MCL1 binding partner BAK1) to BCL-XL, a supervised learning predictor reached a performance of about 80% correctly classified solid tumor cell lines. An additional predictor based on all available cell line data was applied to either tumor samples, adjacent normal tissues or normal tissue samples from TCGA[7] and GTEx[8], showing that many normal tissue samples are categorized as being sensitive. Moreover, solid tumor samples in contrast to solid tumor cell lines are predicted to be broadly sensitive to MCL1 inhibition, in some tumor types reaching over 90% of all cases. In summary, our work describes the translational challenges when applying cell line-derived predictors on tumor samples and thus comprehensive studies are required.

CONCLUSIONS

- An shRNA screen, a CRISPR/Cas9 screen and a pharmacological screen of large solid tumor-derived cancer cell lines targeting MCL1 lead to a consistent classification into sensitive and resistant samples to growth inhibition.
- Cell lines with low BCL2L1 (BCL-XL) gene expression are mostly sensitive to MCL1 inhibition.
- A supervised learning predictor using the gene expression of BCL2L1 (BCL-XL), of the MCL1 binding partner BAK1 and of four additional genes can reliably predict sensitivity and resistance to MCL1 inhibition in cancer cell lines.
- However, the application of this cell line predictor to develop patient stratification strategies for clinical trial design suffers from clear limitations.

REFERENCES

1. McDonald et al., Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening, Cell, Vol 170, Issue 3, (2017)
2. Meyers et. al., Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells, Nature Genetics volume 49, pages 1779–1784 (2017)
3. Gerlach et al., The novel BET bromodomain inhibitor BI 894999 represses super-enhancer-associated transcription and synergizes with CDK9 inhibition in AML. Oncogene (2018).
4. Ensembl gene annotation, European Molecular Biology Laboratory's European - Bioinformatics Institute (EMBI-EBI), http://www.ensembl.org
5. Ritchie et al., limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research, 43(7), e47, (2015)
6. The R Project for Statistical Computing, https://www.r-project.org/
7. The Cancer Genome Atlas (TCGA): https://cancergenome.nih.gov
8. The Genotype-Tissue Expression (GTEx) project: https://gtexportal.org

ACKNOWLEDGEMENTS

Thanks to the Boehringer Ingelheim Vienna Computational Biology team, in particular to Daniel Gerlach for processing and providing gene expression data as well as helpful discussions.

Presented at AACR Annual Meeting 2019. March 29 - April 3, 2019. Atlanta, Georgia, USA.