Targeting EHMT2 sensitizes colorectal cancer to Vorinostat

Leonhard Valentin Bamberg et al.

2021-06-08

Abstract

Epigenetic dysregulation is an important feature of colorectal cancer (CRC). Combining epigenetic drugs with other antineoplastic agents is a promising treatment strategy for advanced cancers. Here, we exploited the concept of synthetic lethality to identify epigenetics targets that act synergistically with histone deacetylase (HDAC) inhibitors to reduce the growth of CRC. We applied a pooled CRISPR-Cas9 screen using a custom sgRNA library directed against 612 epigenetic regulators and discovered that knockout of the euchromatic histone lysine methyltransferases EHMT1 and EHMT2 strongly enhance the antiproliferative effect of clinically used HDAC inhibitors. Using tissue microarrays from 1066 CRC samples with different tumor stages, we show that low EHMT2 protein expression is predominantly found in advanced CRC and associated with poor clinical outcome. Co-targeting of HDAC and EHMT1/2 with specific small molecule inhibitors synergistically reduced proliferation of CRC cell lines. Mechanistically, we used a high-throughput Western blot assay to demonstrate that both inhibitors elicit distinct cellular mechanisms to reduce tumor growth, including cell cycle arrest and modulation of autophagy. Finally, we used a panel of patient-derived CRC organoids to show that HDAC and EHMT1/2 inhibitors synergistically reduce tumor viability in advanced models of CRC.

Contents

1 Overview

Working Hypothesis: Low EHMT2 expression predicts response to HDACi.

Analysis: Correlating expression level of EHMT2 with response to HDACi

Data sources: drug sensitivity data GDSC screens (Iorio et al. 2016) PRISM screens (Corsello et al. 2020) expression data ArrayExpress (accession number: E-MTAB-3610) annotations CCLE_DepMap sample_info (Barretina et al. 2012)

```
#first we read in the cell line annotations

cell_mutant_annotation <- bind_rows(
  list(
    "GDSC1"=read_delim(
      "source_data/COREAD_Genetic_feature_GDSC1_220221_commacorrected.csv",
                       delim = ";",col_types = "cffffcfcc",trim_ws = T),
    "GDSC2"=read_delim(
      "source_data/COREAD_Genetic_feature_GDSC2_220221_commacorrected.csv",
                       delim = ";",col_types = "cffffcfcc",trim_ws = T)),
  .id = "dataset")

#second we read in the actual screening data

screening_data <- bind_rows(
  list(
    "GDSC1"=read_excel("source_data/GDSC1_fitted_dose_response_25Feb20.xlsx"),
    "GDSC2"=read_excel("source_data/GDSC2_fitted_dose_response_25Feb20.xlsx")))

colnames(screening_data) <- tolower(colnames(screening_data))

screening_data <- screening_data %>% rename("cosmic_sample_id"=cosmic_id)

cell_expression_data <-
  read_delim("source_data/Cell_line_RMA_proc_basalExp_EHMT.txt",delim = "\t")%>%
  gather(cosmic_sample_id,normed_expression,-GENE_SYMBOLS,-GENE_title ) %>%
  extract(cosmic_sample_id,"cosmic_sample_id","DATA.(\\d+)") %>%
  filter(GENE_SYMBOLS %in% c("EHMT2"))

colnames(cell_expression_data) <- tolower(colnames(cell_expression_data))

cell_lines_of_interest <- cell_mutant_annotation %>%
                                filter(gdsc_desc2=="large_intestine") %>%
                                pull(cosmic_sample_id) %>%
                                unique()

total_expression_data  <-
  read_delim("source_data/Cell_line_RMA_proc_basalExp.txt",delim = "\t")%>%
  gather(cosmic_sample_id,normed_expression,-GENE_SYMBOLS,-GENE_title ) %>%
  extract(cosmic_sample_id,"cosmic_sample_id","DATA.(\\d+)")

colnames(total_expression_data) <- tolower(colnames(total_expression_data))
```

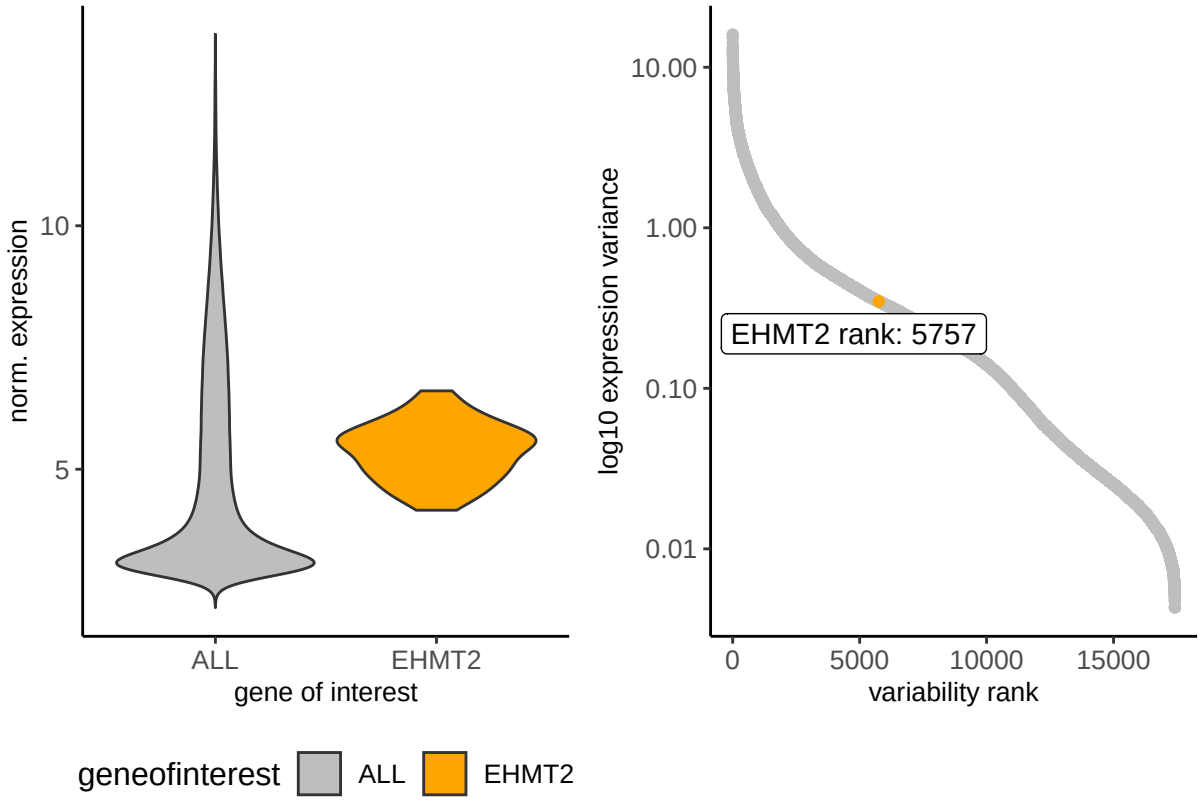
2 Basal gene expression analysis

Hypothesis: EHMT2 expression is variant enough throughout colorectal cancer cell lines to make assumptions on its impact on Vorinostat sensitivity

Thus we test the expression intensity of EHMT2 compared to all other genes and the variance of its expression when compared to the expression variance of all other genes.

EHMT2 expression variance expressed in percent expression variance of the average gene

```
## # A tibble: 1 x 3
##   `FALSE` `TRUE` percentVar
##     <dbl>  <dbl>      <dbl>
## 1   0.460  0.348       75.7
```



EHMT2 per cent expression range of all genes

```
## # A tibble: 1 x 3
##   `FALSE` `TRUE` perc_range
##     <dbl>  <dbl>      <dbl>
## 1    11.8   2.45       20.8
```

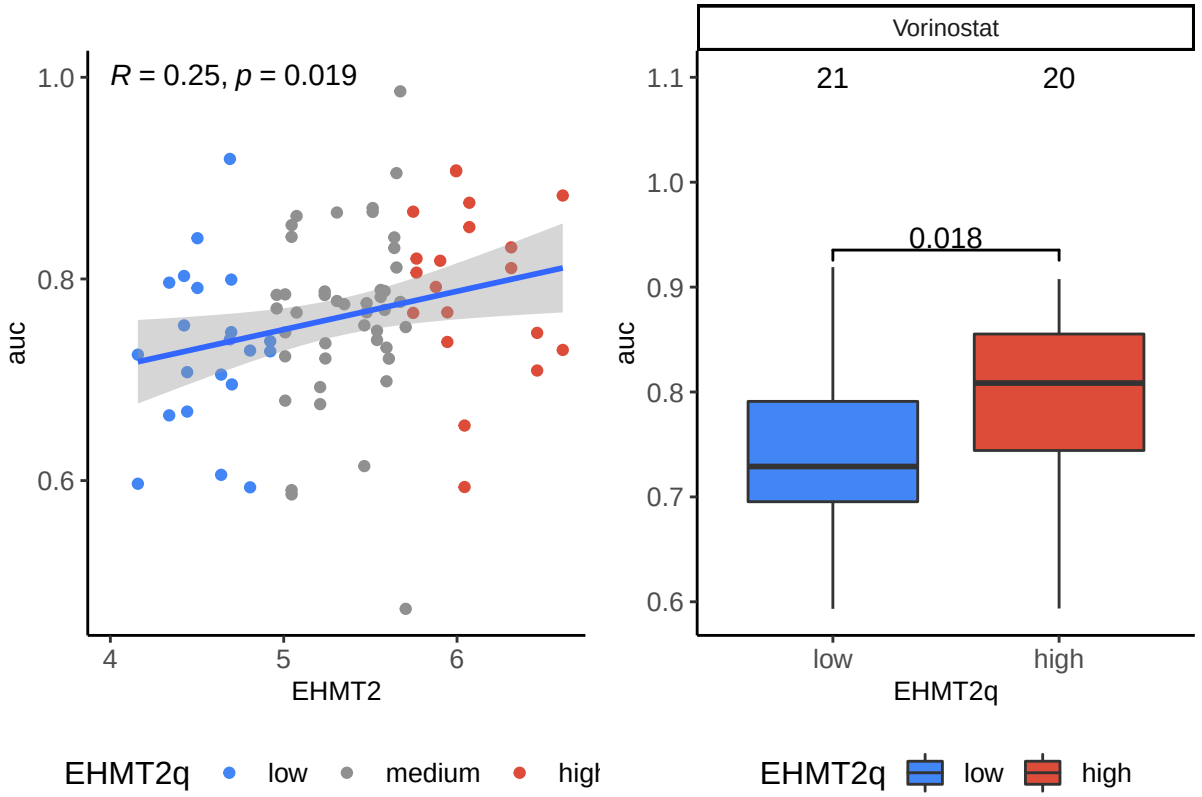
EHMT2 per cent variance of all genes

```
## # A tibble: 1 x 3
##   `FALSE` `TRUE` perc_range
##     <dbl>  <dbl>      <dbl>
## 1    4.58  0.348       7.59
```

From this figure and analysis we can conclude that compared to other genes, EHMT2 is rather highly expressed in colorectal cancer cell lines and that the variability of EHMT2 expression between different cell lines is on par with the average variability of expression of any other gene.

3 EHMT - HDAC relationship (CancerXGene)

First we plot the RMA normalized expression of EHMT2 against the effect Vorinostat have on each cell line. For visualization of trends we also plot a linear models for the data. The slope denotes correlation and the gray confidence interval denotes the certainty on this slope. Further we color the lower and upper quantile of EHMT expression and plot the distribution of points within these quantiles to compare them statistically.



4 Statistical testing (CancerXGene)

In a next step we can use ANOVA on the linear model to assess the impact of the compound used (Vorinostat) and EHMT2 expression on the area under the curve (auc). ANOVA analysis on the model will then measure how much EHMT2 expression adds to the prediction of the auc.

ANOVA of EHMT2 influence on drug auc unfiltered quantitative data

```
##
## Call:
## lm(formula = auc ~ EHMT2, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30372 -0.04045  0.00804  0.05665  0.21073
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.55996    0.08515   6.576 4.03e-09 ***
## EHMT2        0.03797    0.01586   2.394   0.0189 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08514 on 83 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.06458,    Adjusted R-squared:  0.05331
## F-statistic:  5.73 on 1 and 83 DF,  p-value: 0.01893
```

We can observe significant p-value (<0.05) for EHTM2 as a predictive variable for Vorinostat sensitivity. This means that EHTM2 expression predicts HDAC inhibitor sensitivity. From the graphics and the direction of the estimate we can further conclude that there is a positive correlation between EHTM expression and HDAC inhibitor efficacy (auc), the higher EHMT expression, the less potent HDAC inhibitors are.

The same trends are found and even more pronounced when just comparing the cell lines with the highest and the lowest quantile of EHMT2 expression.

#Testing of EHMT2 on filtered stratified data

EHTM2, Vorinostat

```
##
##  Welch Two Sample t-test
##
## data:  auc by EHMT2q
## t = -2.4653, df = 38.772, p-value = 0.01822
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.11439891 -0.01127186
## sample estimates:
##  mean in group low mean in group high
##          0.7308477          0.7936831
```

5 Validate on PRISM data set

Next we validate our observations on a secondary drug screening dataset.

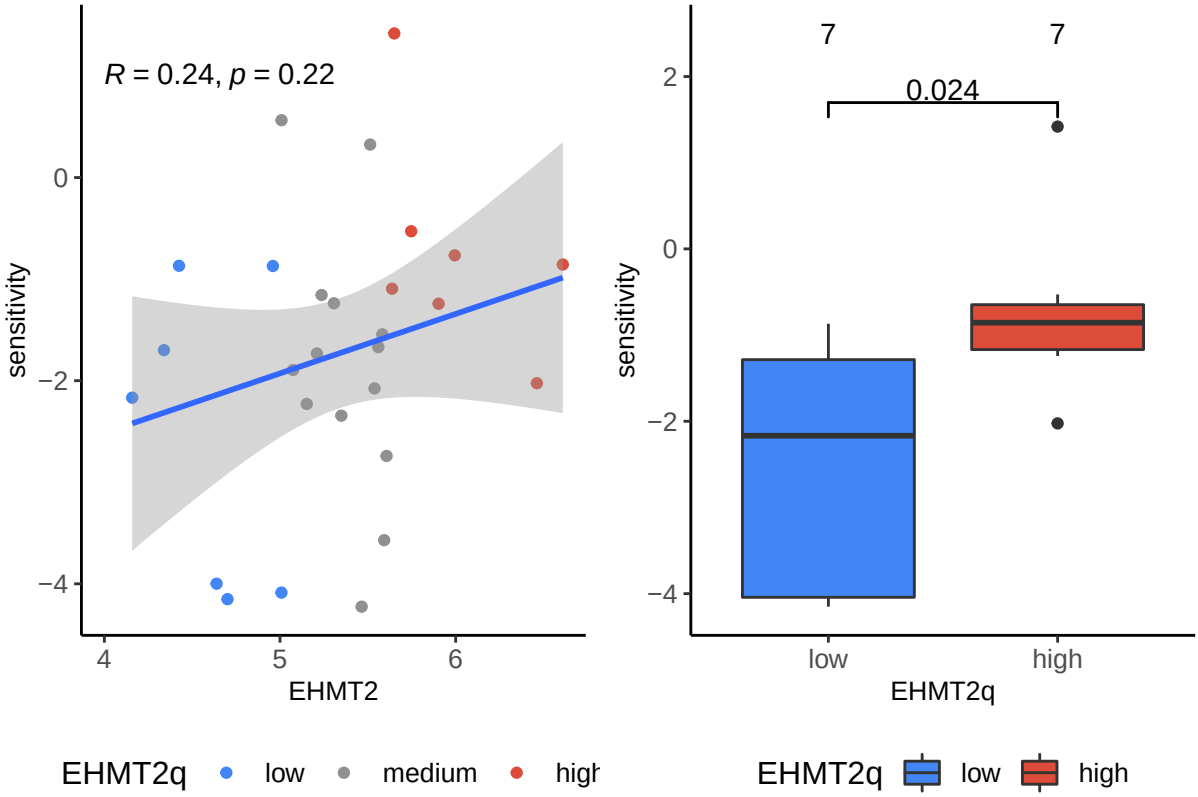
First, we get the conversion of Cosmic to CCLE to PRISM identifiers for cell lines

```
#rename mapping columns so they can be joined with others
CCLE_line_mapping <- read_delim("source_data/sample_info.csv",delim = ",") %>%
  dplyr::select(ccle_id=`CCLE Name`,cosmic_sample_id=COSMIC_ID,lineage) %>%
  mutate(cosmic_sample_id=as.character(cosmic_sample_id))
```

Second, we get the PRISM data

```
PRISM_cell_line <-
  read_delim("source_data/primary-screen-cell-line-info.csv",delim = ",")

PRISM_drug <-
  read_delim("source_data/primary-screen-replicate-collapsed-treatment-info.csv"
             ,delim = ",")

PRISM_data <-
  read_delim("source_data/primary-screen-replicate-collapsed-logfold-change.csv",
             delim = ",") %>%
  dplyr::rename(row_name=X1) %>%
  gather(col_name,sensitivity,-row_name)

drugs_of_interest <-
  PRISM_drug %>%
  filter(grepl("vorinostat",name)) %>%
  dplyr::select(col_name=column_name,name)

cells_of_interest <- PRISM_cell_line %>%
  filter(grepl("color",primary_tissue)) %>%
  dplyr::select(row_name,ccle_id=ccle_name) %>%
  left_join(CCLE_line_mapping)

PRISM_data_filtered_anno_join <-  PRISM_data %>%
  filter(col_name %in% drugs_of_interest$col_name) %>%
  filter(row_name %in% cells_of_interest$row_name) %>%
  left_join(cells_of_interest) %>%
  left_join(cell_expression_data) %>%
  left_join(drugs_of_interest) %>%
  drop_na()

data_to_plot <- PRISM_data_filtered_anno_join %>%
    dplyr::select(sensitivity,ccle_id,gene_symbols,normed_expression,name) %>%
    spread(gene_symbols,normed_expression) %>%
    mutate(EHMT2q = if_else(EHMT2>quantile(EHMT2,probs = 0.75,na.rm = T),"high",
    if_else(EHMT2<quantile(EHMT2,probs = 0.25,na.rm = T),"low","medium"))) %>%
    mutate(EHMT2q=factor(EHMT2q,levels = c("low","medium","high")))
```

6 Plot the trend between expression and sensitivity (PRISM)



7 ANOVA trend between expression and sensitivity (PRISM)

```
##
## Call:
## lm(formula = sensitivity ~ EHMT2, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5691 -0.7036  0.1009  0.6201  2.9662
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.8559     2.5021  -1.941   0.0632 .
## EHMT2         0.5855     0.4662   1.256   0.2203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.393 on 26 degrees of freedom
## Multiple R-squared:  0.0572, Adjusted R-squared:  0.02094
## F-statistic: 1.577 on 1 and 26 DF,  p-value: 0.2203
```

8 Statistical testing of separate combinations

Vorinostat EHMT2 separate (t-test)

```
##
##  Welch Two Sample t-test
##
## data:  sensitivity by EHMT2q
## t = -2.6209, df = 10.791, p-value = 0.02413
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.354807 -0.288216
## sample estimates:
##  mean in group low mean in group high
##         -2.5489533         -0.7274417
```

9 Session info

```
writeLines(capture.output(sessionInfo()), paste("results/SessionInfo.txt",sep = ""))
sessionInfo()
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] ggpubr_0.4.0.999 ggsignif_0.6.1   ggrepel_0.9.1    patchwork_1.1.1
##  [5] readxl_1.3.1     forcats_0.5.1    stringr_1.4.0    dplyr_1.0.5
##  [9] purrr_0.3.4      readr_1.4.0      tidyr_1.1.3      tibble_3.1.0
## [13] ggplot2_3.3.3    tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.6         lattice_0.20-41    lubridate_1.7.10   assertthat_0.2.1
##  [5] digest_0.6.27      utf8_1.2.1         R6_2.5.0           cellranger_1.1.0
##  [9] backports_1.2.1    reprex_2.0.0       evaluate_0.14      highr_0.8
## [13] httr_1.4.2         pillar_1.5.1       rlang_0.4.10       curl_4.3
## [17] rstudioapi_0.13    data.table_1.14.0 car_3.0-10         Matrix_1.3-2
## [21] rmarkdown_2.7      splines_4.0.3      labeling_0.4.2     foreign_0.8-81
## [25] munsell_0.5.0      broom_0.7.6        compiler_4.0.3     modelr_0.1.8
## [29] xfun_0.22          pkgconfig_2.0.3    mgcv_1.8-34        htmltools_0.5.1.1
## [33] tidyselect_1.1.0   rio_0.5.26         fansi_0.4.2        crayon_1.4.1
## [37] dbplyr_2.1.0       withr_2.4.1        grid_4.0.3         nlme_3.1-152
## [41] jsonlite_1.7.2     gtable_0.3.0       lifecycle_1.0.0    DBI_1.1.1
## [45] magrittr_2.0.1     scales_1.1.1       zip_2.1.1          carData_3.0-4
## [49] cli_2.4.0          stringi_1.5.3      farver_2.1.0       fs_1.5.0
## [53] xml2_1.3.2         ellipsis_0.3.1     generics_0.1.0     vctrs_0.3.7
## [57] openxlsx_4.2.3     tools_4.0.3        glue_1.4.2         hms_1.0.0
## [61] abind_1.4-5        yaml_2.2.1         colorspace_2.0-0   rstatix_0.7.0
## [65] rvest_1.0.0        knitr_1.31         haven_2.3.1
```

References

Barretina, Jordi, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A. Margolin, Sungjoon Kim, Christopher J. Wilson, et al. 2012. "The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity." *Nature* 483 (7391): 603–7. https://doi.org/10.1038/nature11003.

Corsello, Steven M., Rohith T. Nagari, Ryan D. Spangler, Jordan Rossen, Mustafa Kocak, Jordan G. Bryan, Ranad Humeidi, et al. 2020. "Discovering the Anticancer Potential of Non-Oncology Drugs by Systematic Viability Profiling." *Nature Cancer* 1 (2): 235–48. https://doi.org/10.1038/s43018-019-0018-6.

Iorio, Francesco, Theo A. Knijnenburg, Daniel J. Vis, Graham R. Bignell, Michael P. Menden, Michael Schubert, Nanne Aben, et al. 2016. "A Landscape of Pharmacogenomic Interactions in Cancer." *Cell* 166 (3): 740–54. https://doi.org/10.1016/j.cell.2016.06.017.