Methylation Analysis of DepMap Data

Ashir Borah[1]

[1] *BMI PhD Program, University of California, San Francisco*

December 5, 2022

Methylation impacts the epigenomic state of cells and has been implicated in cancer. Epigenomic changes can impact cancer cell line viability, presenting new cancer targets. Downstream analysis of probe-level data is challenging to arrive at a biological hypothesis. GOMeth helps discover new targets previously not possible, presenting new avenues for cancer therapeutics development

1 Problem statement

The Dependency Map project has performed genome-scale CRISPR knockout in cancer cell lines (Tsherniak et al., 2017). This data presents an incredible resource for new therapeutic discovery if an observed clinically exploitable phenotype can be tied to molecular features. While such analysis has been performed for different data types like RNAseq, Mutation, Fusion, Metabolomics and Proteomics (Dempster et al., 2020), a systematic Methylation analysis is missing. DepMap stopped producing Infinium 450k Methylation array data after a few cycles.

A large Methylation dataset for cancer cell lines (Iorio et al., 2016) presents an opportunity to model the cancer cell lines data using Methylation data. Models with high predictive performance need further analysis to build confidence before further validation in the lab. GOMeth (Maksimovic, Oshlack, and Phipson, 2021) can help link probes predictive to a biological process.

The methylation data has 450,000 probes, each enabling the measurement of the methylation levels at these CpG sites. These measurements, along with other informative features like the lineage of the cell lines and confounders, are used to model the observed viability data after knockout.

Table 1: *Model performance*

Model	Mean ρ
Methylation	0.132
Methylation-permutated	0.00491

2 Results

Methylation data has a wide degree of predictive power based on the knockout being studied. Knockout of lineage-specific essential genes like SOX10 for skin lineages are not interesting targets, and they are easily spotted with a lineage feature having very high feature importance.

As shown in Figure 1A, the Pearson correlation of the methylation data is shifted to the right of the permuted methylation data indicating that the methylation data improved predictive performance overall. The medians of the distributions in Table 1 reflect this fact. When the feature importance of the statistically significant models are analysed using GOMeth (details in Methods), cancer pathways are seen to be enriched in the probes that were predictive of cell viability data.

Some of the top models are already implicated in cancer(EBF1 (Shen et al., 2020), FAM50A (Köferle et al., 2022)). Further analysis with biological experts might reveal new theraputics targets for cancer.

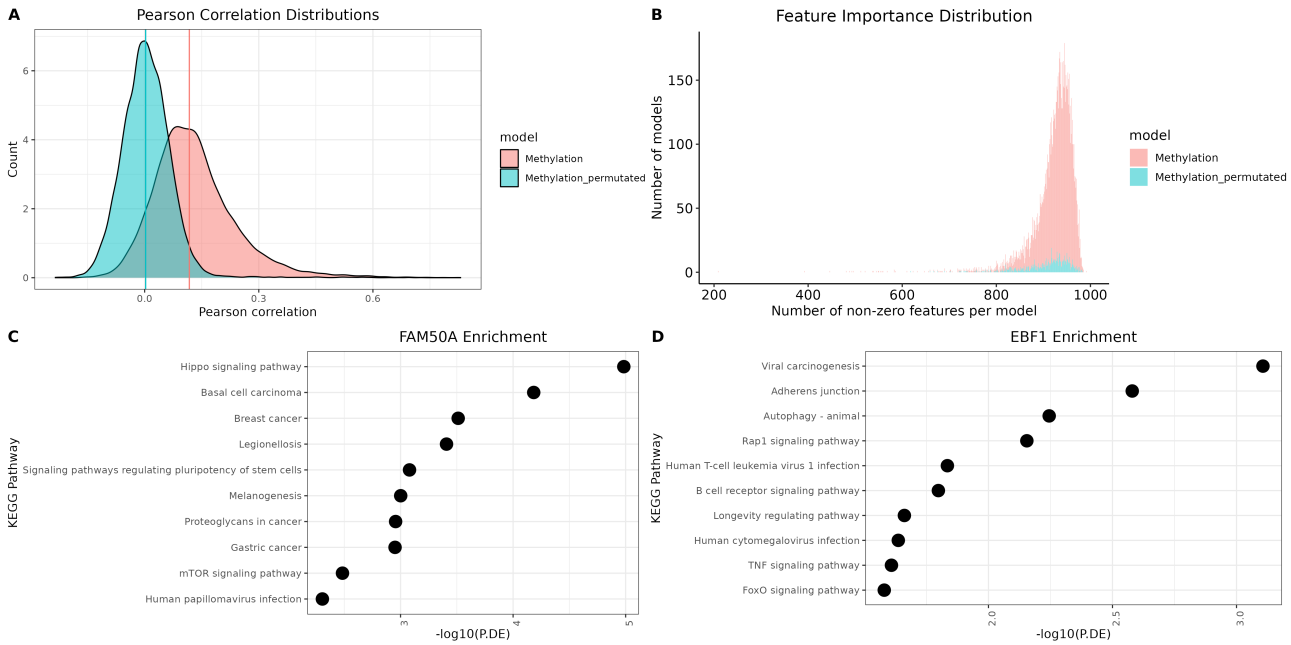

Figure 1: *Methylation Analysis of DepMap Data*

(A) Pearson correlation distribution between the observed and predicted values. The blue line represents the

3 Methods

3.1 Modeling

Methylation data from (Iorio et al., 2016) is processed with the SeSAMe (Zhou et al., 2018) to replicate the processing of TCGA to make values comparable. The other datasets used from modeling was downloaded from (DepMap, 2022).

Features:

- Infinium 450k Methylation beta values
- One hot encoded lineage
- Confounders

The predicted viability dataset is the Chronos CRISPR gene effects dataset (DepMap, 2022; Dempster et al., 2021). A random forest regression model from (Krill-Burger et al., 2022) was used to predict the gene effect data. To evaluate the effect of adding the Methylation data and differentiate the boost in performance merely from adding additional features, the methylation dataset was permutated while keeping other features like lineage and confounders unchanged.

Feature importances from these random forest models were extracted to further analyze the predictions. The feature importance from the permutated models were used to generate a null distribution enabling the calculation of p-values. This feature importance was used as a cut off for enrichment analysis using GOMeth (Maksimovic, Oshlack, and Phipson, 2021).

3.2 Analysis of models

4 Discussions

5 Conclusions

Bibliography

Tsherniak, Aviad et al. (July 2017). "Defining a cancer dependency map". en. In: *Cell* 170.3, 564–576.e16.

Dempster, Joshua M. et al. (2020). "Gene expression has more power for predicting in vitro cancer cell vulnerabilities than genomics". In: *bioRxiv*. DOI: 10.1101/2020.02.21.959627. eprint: https://www.biorxiv.org/content/early/2020/09/10/2020.02.21.959627.full.pdf. URL: https://www.biorxiv.org/content/early/2020/09/10/2020.02.21.959627.

Iorio, Francesco et al. (July 2016). "A landscape of pharmacogenomic interactions in cancer". en. In: *Cell* 166.3, pp. 740–754.

Maksimovic, Jovana, Alicia Oshlack, and Belinda Phipson (2021). "Gene set enrichment analysis for genome-wide DNA methylation data". In: *Genome Biology* 22.1, p. 173. ISSN: 1474-760X. DOI: 10.1186/s13059-021-02388-x. URL: https://doi.org/10.1186/s13059-021-02388-x.

Shen, Zhiqing et al. (June 2020). "Transcription factor EBF1 over-expression suppresses tumor growth in vivo and in vitro via modulation of the PNO1/p53 pathway in colorectal cancer". en. In: *Front. Oncol.* 10, p. 1035.

Köferle, Anna et al. (2022). "Interrogation of cancer gene dependencies reveals paralog interactions of autosome and sex chromosome-encoded genes". In: *Cell Reports* 39.2, p. 110636. ISSN: 2211-1247. DOI: https://doi.org/10.1016/j.celrep.2022.110636. URL: https://www.sciencedirect.com/science/article/pii/S2211124722003886.

Zhou, Wanding et al. (July 2018). "SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions". In: *Nucleic Acids Research* 46.20, e123–e123. ISSN: 0305-1048. DOI: 10.1093/nar/gky691. eprint: https://academic.oup.com/nar/article-pdf/46/20/e123/26578142/gky691.pdf. URL: https://doi.org/10.1093/nar/gky691.

DepMap, Broad (May 2022). "DepMap 22Q2 Public". In: DOI: 10.6084/m9.figshare.19700056.v2. URL: https://figshare.com/articles/dataset/DepMap_22Q2_Public/19700056.

Dempster, Joshua M. et al. (2021). "Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects". In: *Genome Biology* 22.1, p. 343. ISSN: 1474-760X. DOI: 10.1186/s13059-021-02540-7. URL: https://doi.org/10.1186/s13059-021-02540-7.

Krill-Burger, J. Michael et al. (2022). "Partial gene suppression improves identification of cancer vulnerabilities when CRISPR-Cas9 knockout is pan-lethal". In: *bioRxiv*. DOI: 10.1101/2022.03.02.482624. eprint: https://www.biorxiv.org/content/early/2022/03/03/2022.03.02.482624.full.pdf. URL: https://www.biorxiv.org/content/early/2022/03/03/2022.03.02.482624.