

# Transcriptomic Meta-Analysis Identifies Upregulated Clotting and Fibrinolysis Pathways in Colorectal Cancer Tumors Containing Hereditary PMS2 Mismatch Repair Deficiency

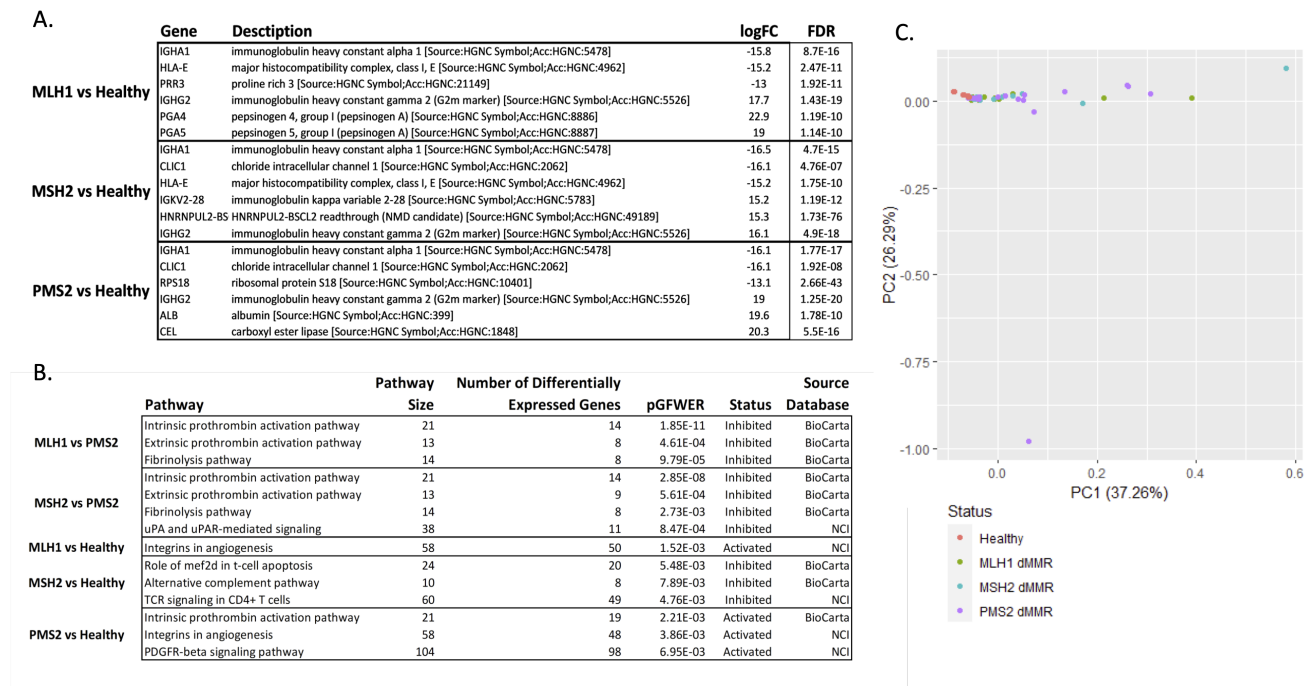
Trenton M Gibson<sup>1</sup>, Mauri D Spendlove<sup>1</sup>, Brett E Pickett<sup>1§</sup>

<sup>1</sup>Microbiology and Molecular Biology, Brigham Young University

§To whom correspondence should be addressed: brett\_pickett@byu.edu

## Abstract

Lynch Syndrome is characterized by deficient mismatch repair (dMMR) components. We performed a meta-analysis of multiple RNA-sequencing datasets from patients with different dMMR variants (PMS2, MLH1, and MSH2) to better characterize the unique transcriptional profiles. Our results reveal enriched signaling pathways from tumor samples with germline mutations in the PMS2 gene including upregulation in pathways related to intrinsic and extrinsic prothrombin activation, fibrinolysis, and uPA/uPAR-mediated signaling. These pathways have been associated with tumor growth, invasiveness, and metastasis. This work provides support for further exploration into the role of PMS2 in tumor development, and as a potential therapeutic mechanism.



**Figure 1.**

**A. Table of Differentially Expressed Genes in Lynch Syndrome Tumors.** Depicts the top three upregulated and top three downregulated genes for various Lynch Syndrome dMMR variants (in the MLH1, MSH2, or PMS2 genes) compared to healthy colon samples. **B. Table of Most Significant Signaling Pathway Impact Analysis (SPIA) Results.** Multiple comparisons highlight the activated pathways in tumor samples vs. healthy control samples, characterized by PMS2 dMMR compared to MLH1 and MSH2 deficient patients. **C. Principal Component Analysis (PCA) of Gene Expression Levels.** A PCA plot indicating the relationships among RNA sequencing samples obtained from GEO. Samples were color-coded by categories including controls and dMMR germline mutations. Healthy, non-cancerous colorectal epithelial samples form a noticeable cluster, while patients with different MMR mutations exhibit heterogeneity in gene expression.

## Description

Lynch Syndrome is an autosomal dominant disorder in which the endogenous DNA mismatch repair (MMR) system becomes deficient (dMMR). Deleterious germline mutations in any gene product within the system (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) severely reduce the ability of the cell to find and repair spontaneous variants that occur across the somatic genome(1),(2). The protein products of these genes dimerize to form many of the components in the MMR protein complex. For example, *MSH2* dimerizes with *MSH6* to form the MutS $\alpha$  subunit and *MLH1* dimerizes with *PMS2* to form the MutL $\alpha$  subunit (3).

Individuals with dMMR have a higher prevalence of genomic instability since the inability to repair spontaneous variants increases the frequency of DNA nucleotide mismatches. These underlying issues cause increases in the occurrence of homologous recombination (4–6) and of developing cancer (7). Genomic stability in cancer patients is quantified by determining the relative abundance of microsatellite repeats in tumor cells compared to healthy somatic cells. Microsatellite instability (MSI) is a classification status given to tumors with differing numbers of microsatellite repeats than other somatic cells and is a defining characteristic of Lynch Syndrome (8).

The aim of this study was to identify differentially expressed genes and enriched signaling pathways among cancer tumors of patients with germline mutations in the relevant MMR genes. Particularly, we performed a transcriptomic meta-analysis to compare the expression from Lynch Syndrome samples where mutations were present in one of three dMMR genes both to each other and to non-cancerous colon samples. This study design was required since one study only contained samples from Lynch Syndrome patients who had mutations in dMMR genes, while a separate second study contained the healthy control samples.

We performed separate differential expression analyses of genes in the patients with mutations in the *MLH1*, *MSH2*, or *PMS2* genes, compared to healthy controls. Differential expression analyses were also performed between the different dMMR variants (e.g. *MLH1* vs *PMS2*, *MLH1* vs *MSH2*, and *MSH2* vs *PMS2* variants). Comparative analyses including *MSH6* were not performed because of the lack of a sufficient sample size to yield significant results. We focused our analysis on the six genes with the biggest log<sub>2</sub> fold-change values for each contrast (Figure 1A). Interestingly, when we compared these results, we noticed that the *IGHA1*, *IGHG2*, and *HLA-E* genes were differentially expressed across two or more comparisons.

We then used the DEGs from each comparison to calculate the intracellular signaling pathways that were most significantly affected using the signaling pathway impact analysis (SPIA) algorithm. We observed that tumor samples characterized by *PMS2* dMMR contain multiple inhibited pathways when compared to *MLH1*- and *MSH2*-deficient patients (Figure 1B). When comparing *MLH1* dMMR to *PMS2* dMMR, we found that the intrinsic and extrinsic prothrombin activation pathways were inhibited as well as the fibrinolysis pathway. The contrast of *MSH2* dMMR with *PMS2* dMMR showed these same pathways being inhibited as well as the uPA and uPAR-signaling pathway. These findings suggest a possible effect of *PMS2* on proteins in the prothrombin and fibrinolysis pathways (9–12). In fact, other studies have shown that these pathways are often upregulated in cancer and play a role in tumor growth, angiogenesis, and metastasis. Specifically, fibrin degradation products (FDP), small protein fragments resulting from the breakdown of fibrin clots, have been shown to promote malignancy in cancer cells (13–16). Interestingly, our contrast between *MLH1* and *MSH2* did not reveal any significant pathways. We next performed a principal component analysis (PCA) to gain a better understanding of the transcriptional relationships between the samples (Figure 1C). This analysis showed that healthy, non-cancerous colorectal epithelial samples were the only group that showed any noticeable clustering, suggesting at least some similarities in the gene expression profiles among the dMMR samples. Another potential interpretation for this lack of clustering among the patients with different MMR mutations is that it accurately reflects the wide heterogeneity of gene expression that is common in cancerous cells.

We analyzed the top DEGs from the cancer vs healthy control comparisons in the University of Alabama at Birmingham Cancer data analysis portal (UALCAN) to better understand the clinical importance of our findings (17). We focused this analysis on the *PMS2* vs healthy control comparison since it had some of the most statistically significant DEGs in our initial results. In particular, we found that the carboxyl ester lipase (*CEL*) gene product was significantly upregulated across all colorectal cancer stages, as well as at significantly higher transcriptional levels both before and after lymph node metastasis. The *CEL* gene product was also detected at higher levels in the CPTAC proteomics data (18). We observed that the *CEL* gene locus had a significantly broader range of methylation in tumor samples than in healthy controls, and that higher expression of this gene led to a non-significant increase in patient survival. It is difficult to filter for colorectal cancer samples in these platforms that were caused directly by MMR mutations. A similar analysis using the PanCancer Atlas data in cBioPortal revealed that missense mutations in the *PMS2* gene was associated with a much broader range of *CEL* expression (19). Consequently, additional validation experiments are justified, particularly given the relatively small numbers of public samples that could be included in this analysis.

We then tried to identify existing therapeutics that could be repurposed for Lynch Syndrome using Pathways2Targets; however, there were no known targets in those significant signaling pathways. This result is somewhat expected given the relatively low number of enriched signaling pathways and the lack of drugs targeting those pathways. However, this negative

result suggests that future work towards the development of drugs targeting proteins in the prothrombin and/or fibrinolysis pathways as potential therapeutics for colorectal cancer and Lynch syndrome could be beneficial. A potential limitation of this study is the inability to confirm the diversity of genetic backgrounds represented in the patient populations, which should be addressed in future studies. In the emerging era of personalized medicine, our results could support the exploration and development of therapeutics targeting these pathways only for individuals with mutations in the PMS2 locus.

## Methods

The RNA-sequencing datasets were found on the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database. The search term “Lynch syndrome” was used and only human RNA-sequencing datasets were selected for screening. Of the 11 results available as of January 17th, 2024; only one study met our filtering criteria (GSE146889). Specifically, these criteria involved including only case samples that characterized the specific dMMR variants (MLH1, MSH2, MSH2, and PMS2) in Lynch Syndrome patients in our analyses. The other samples were excluded for various reasons that included: data collected from cell lines, organoids, or patients in interventional clinical trials and were not conducive to the purpose of our study. Fifty-five of these publicly available bulk RNA-sequencing files were obtained from patients with colorectal cancer and high microsatellite instability. A separate study containing 20 non-cancerous colon tissue samples was used as a healthy control (GSE214695). Cluster visualization using principal component analysis was performed in R.

Fastq files containing the RNA-sequencing read data were downloaded from the Sequence Read Archive (SRA) using the sra-tools software. The Fastq files, the associated metadata (Supplementary Table S1), and a configuration file were used as inputs to the Automated Reproducible MODular workflow for preprocessing and differential analysis of RNA-seq data (ARMOR) workflow (20). The ARMOR workflow involved quality control, removing sequencing adapters and low-quality regions with TrimGalore! , read pseudomapping and quantification to the human GRCh38 transcriptome using Salmon (21), calculation of significant differentially expressed genes using edgeR (22), and gene ontology enrichment analysis using Camera (23).

The differentially expressed genes (Lynch Syndrome vs healthy colorectal tissue) generated by the ARMOR software were categorized as statistically significant if the false discovery rate adjusted p-value < 0.05. These significant genes became the input for the signaling pathway impact analysis (SPIA) software, which identified significantly enriched signaling pathways (Bonferroni-adjusted p-values < 0.05) from the following databases: KEGG, Panther, BioCarta, Reactome, and NCI(24–28)The output of SPIA became the input for a drug repurposing and target prioritization algorithm, Pathways2Targets(29), which uses more than 20 attributes from the OpenTargets database and pathway information to rank and prioritize therapeutic targets based on customizable criteria.

## Acknowledgements:

We acknowledge the BYU research computing center for their assistance with supporting the high-performance computing hardware.

## References

- Abildgaard AB, Nielsen SV, Bernstein I, Stein A, Lindorff-Larsen K, Hartmann-Petersen R. 2022. Lynch syndrome, molecular mechanisms and variant classification. *British Journal of Cancer* 128: 726-734. DOI: [10.1038/s41416-022-02059-z](https://doi.org/10.1038/s41416-022-02059-z)
- Boonstra MC, Verbeek FP, Mazar AP, Prevoo HA, Kuppen PJ, van de Velde CJ, Vahrmeijer AL, Sier CF. 2014. Expression of uPAR in tumor-associated stromal cells is associated with colorectal cancer patient prognosis: a TMA study. *BMC Cancer* 14: 10.1186/1471-2407-14-269. DOI: [10.1186/1471-2407-14-269](https://doi.org/10.1186/1471-2407-14-269)
- Bootle-Wilbraham CA, Tazzyman S, Thompson WD, Stirk CM, Lewis CE. 2001. . *Angiogenesis* 4: 269-275. DOI: [10.1023/a:1016076121918](https://doi.org/10.1023/a:1016076121918)
- Dobbs Spendlove M, M. Gibson T, McCain S, Stone BC, Gill T, Pickett BE. 2023. Pathway2Targets: an open-source pathway-based approach to repurpose therapeutic drugs and prioritize human targets. *PeerJ* 11: e16088. DOI: [10.7717/peerj.16088](https://doi.org/10.7717/peerj.16088)
- Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al., D'Eustachio. 2017. The Reactome Pathway Knowledgebase. *Nucleic Acids Research* 46: D649-D655. DOI: [10.1093/nar/gkx1132](https://doi.org/10.1093/nar/gkx1132)
- Friedhoff P, Li P, Gotthardt J. 2016. Protein-protein interactions in DNA mismatch repair. *DNA Repair* 38: 50-57. DOI: [10.1016/j.dnarep.2015.11.013](https://doi.org/10.1016/j.dnarep.2015.11.013)
- Kołodziejczyk J, Ponczek MB. 2013. Reviews The role of fibrinogen, fibrin and fibrin(ogen) degradation products (FDPs) in tumor progression. *Współczesna Onkologia* 2: 113-119. DOI: [10.5114/wo.2013.34611](https://doi.org/10.5114/wo.2013.34611)

- Li GM. 2007. Mechanisms and functions of DNA mismatch repair. *Cell Research* 18: 85-98. DOI: [10.1038/cr.2007.115](https://doi.org/10.1038/cr.2007.115)
- Lin Y, Liu Z, Qiu Y, Zhang J, Wu H, Liang R, et al., Zou. 2018. Clinical significance of plasma D-dimer and fibrinogen in digestive cancer: A systematic review and meta-analysis. *European Journal of Surgical Oncology* 44: 1494-1503. DOI: [10.1016/j.ejso.2018.07.052](https://doi.org/10.1016/j.ejso.2018.07.052)
- Mahmood N, Rabbani SA. 2021. Fibrinolytic System and Cancer: Diagnostic and Therapeutic Applications. *International Journal of Molecular Sciences* 22: 4358. DOI: [10.3390/ijms22094358](https://doi.org/10.3390/ijms22094358)
- Mi H, Muruganujan A, Casagrande JT, Thomas PD. 2013. Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols* 8: 1551-1566. DOI: [10.1038/nprot.2013.092](https://doi.org/10.1038/nprot.2013.092)
- Nishimura D. 2001. BioCarta. *Biotech Software & Internet Report* 2: 117-120. DOI: [10.1089/152791601750294344](https://doi.org/10.1089/152791601750294344)
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 27: 29-34. DOI: [10.1093/nar/27.1.29](https://doi.org/10.1093/nar/27.1.29)
- Okholm M, Iversen LH, Thorlacius-Ussing O, Ejlersen E, Boesby S. 1996. Fibrin and fibrinogen degradation products in plasma of patients with colorectal adenocarcinoma. *Diseases of the Colon & Rectum* 39: 1102-1106. DOI: [10.1007/BF02081408](https://doi.org/10.1007/BF02081408)
- Olave MC, Graham RP. 2021. Mismatch repair deficiency: The what, how and why it is important. *Genes, Chromosomes and Cancer* 61: 314-321. DOI: [10.1002/gcc.23015](https://doi.org/10.1002/gcc.23015)
- Orjuela S, Huang R, Hembach KM, Robinson MD, Sonesson C. 2019. ARMOR: An Automated Reproducible Modular Workflow for Preprocessing and Differential Analysis of RNA-seq Data. *G3 Genes|Genomes|Genetics* 9: 2089-2096. DOI: [10.1534/g3.119.400185](https://doi.org/10.1534/g3.119.400185)
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 14: 417-419. DOI: [10.1038/nmeth.4197](https://doi.org/10.1038/nmeth.4197)
- Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140. DOI: [10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616)
- Salem ME, Bodor JN, Puccini A, Xiu J, Goldberg RM, Grothey A, et al., Hall. 2020. Relationship between MLH1, PMS2, MSH2 and MSH6 gene-specific alterations and tumor mutational burden in 1057 microsatellite instability-high solid tumors. *International Journal of Cancer* 147: 2948-2956. DOI: [10.1002/ijc.33115](https://doi.org/10.1002/ijc.33115)
- Südhoff T, Schneider W. 1992. Fibrinolytic mechanisms in tumor growth and spreading. *The Clinical Investigator* 70: 10.1007/bf00180278. DOI: [10.1007/BF00180278](https://doi.org/10.1007/BF00180278)
- Spies M, Fishel R. 2015. Mismatch Repair during Homologous and Homeologous Recombination. *Cold Spring Harbor Perspectives in Biology* 7: a022657. DOI: [10.1101/cshperspect.a022657](https://doi.org/10.1101/cshperspect.a022657)
- Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim Js, et al., Romero. 2008. A novel signaling pathway impact analysis. *Bioinformatics* 25: 75-82. DOI: [10.1093/bioinformatics/btn577](https://doi.org/10.1093/bioinformatics/btn577)
- Vieira MLC, Santini L, Diniz AL, Munhoz CdF. 2016. Microsatellite markers: what they mean and why they are so useful. *Genetics and Molecular Biology* 39: 312-328. DOI: [10.1590/1678-4685-GMB-2016-0027](https://doi.org/10.1590/1678-4685-GMB-2016-0027)
- Wei Dai YY. 2014. Genomic Instability and Cancer. *Journal of Carcinogenesis & Mutagenesis* 05: 10.4172/2157-2518.1000165. DOI: [10.4172/2157-2518.1000165](https://doi.org/10.4172/2157-2518.1000165)
- Wu D, Smyth GK. 2012. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Research* 40: e133-e133. DOI: [10.1093/nar/gks461](https://doi.org/10.1093/nar/gks461)
- Zhai BT, Tian H, Sun J, Zou JB, Zhang XF, Cheng JX, et al., Guo. 2022. Urokinase-type plasminogen activator receptor (uPAR) as a therapeutic target in cancer. *Journal of Translational Medicine* 20: 10.1186/s12967-022-03329-3. DOI: [10.1186/s12967-022-03329-3](https://doi.org/10.1186/s12967-022-03329-3)
- Chandrashekar DS, Karthikeyan SK, Korla PK, Patel H, Shovon AR, Athar M, et al., Varambally S. 2022. UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia* 25: 18-27. PubMed ID: [35078134](https://pubmed.ncbi.nlm.nih.gov/35078134/)
- Whiteaker JR, Halusa GN, Hoofnagle AN, Sharma V, MacLean B, Yan P, et al., Paulovich AG. 2014. CPTAC Assay Portal: a repository of targeted proteomic assays. *Nat Methods* 11(7): 703-4. PubMed ID: [24972168](https://pubmed.ncbi.nlm.nih.gov/24972168/)
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al., Schultz N. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6(269): p11. PubMed ID: [23550210](https://pubmed.ncbi.nlm.nih.gov/23550210/)

**Funding:**

No external funding was obtained to perform this work.

**Author Contributions:** Trenton M Gibson: data curation, formal analysis, visualization, writing - original draft, writing - review editing. Mauri D Spendlove: formal analysis, writing - original draft, writing - review editing. Brett E Pickett: methodology, software, writing - original draft, writing - review editing, supervision.

**Reviewed By:** Anonymous

**History:** **Received** February 16, 2024 **Revision Received** July 23, 2024 **Accepted** July 24, 2024 **Published Online** July 27, 2024 **Indexed** August 10, 2024

**Copyright:** © 2024 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Citation:** Gibson, TM; Spendlove, MD; Pickett, BE (2024). Transcriptomic Meta-Analysis Identifies Upregulated Clotting and Fibrinolysis Pathways in Colorectal Cancer Tumors Containing Hereditary PMS2 Mismatch Repair Deficiency. microPublication Biology. [10.17912/micropub.biology.001159](https://doi.org/10.17912/micropub.biology.001159)