**­­­Possible title:**

**Cell-type specific immunomethylomic predictors of colorectal cancer (CRC) metastasis based on Hierarchical Tumor Immune Microenvironment Epigenetic Deconvolution**

**Background/Intro:**

**From the Grant:**

***Burden of Colon Cancer and Assessment Challenges.*** Colorectal cancer (CRC) is the third leading cause of cancer both worldwide and in the United States and accounts for approximately 8 percent of cancer-related deaths13. CRC incidence is shifting towards younger demographics who are not included in established screening programs14,15. Disease management of CRC often includes lymph node resection to determine N-stage as a proxy for recurrence risk, after resection at the primary site at the time of diagnosis. This is followed by adjuvant therapy for patients with positive lymph nodes. Lymph node resection, histology and grossing is often suboptimal outside of a subspecialist-driven academic medical *Center of Excellence*. As an example, one population-based study concluded that only 37% of colon cancer cases had adequate assessment of the regional lymph nodes (at least 12 nodes assessed) 19–22. Inadequate resection and downstream analysis can impact prognostication and selection of relevant treatment options for clinical triage. But the same is true for upstream epigenetic changes. This work focuses on developing a practical, inexpensive reliable and valid DNA methylation (DNAm) assay that, if clinically validated, may be used as an adjunct screening and prognostication tool along other molecular markers of metastasis.

Current methods of prognosis of recurrence and of survival of CRC have limitations. Specimen inadequacy and batch effects can obstruct proper analysis, as well as potential variation at the host level, tumor level, and tumor microenvironment. There are many pathways leading to prognosis, making biomarker detection and reliance challenging. However, most of these findings are still in the discovery phase and have not been validated. Recent evidence suggests that both the origin and architectural location of cells play a significant role in prognostication. Gene Expression and DNA Methylation are two techniques used to estimate cell types in the primary sight and tumor microenvironment of many different cancers.

**Methylation studies in colorectal cancer**

Several methylation studies now exist, pointing to methylation differences in DNA that may potentially help in early detection or the prediction of recurrence or metastasis and poor prognosis. The general understanding is that DNA hypomethylation in repetitive sequences or long interspersed nucleotide element 1 sequences may mobilize the retrotransposable elements leading to genomic instabilities. At the same time, hypermethylation of promoters has been linked to silencing of tumor suppressor genes, DNA repair problems and increased angiogenesis.

Fatemi N. et al document several useful methylation signatures of CRC, some with FDA approval (REF **Fatemi N, Tierling S, Es HA, et al. DNA methylation biomarkers in colorectal cancer: Clinical applications for precision medicine. Int J Cancer. 2022;151(12):**

**2068‐2081. doi:10.1002/ijc.34186).** Some of the methylated and silenced genes include vimentin (VIM), cadherin-1 (CDH1), MLH1, TIMP3, SFRP1, and HIC-1. Other combinatorial examples include the biomarker panel established by Rademakers et al. testing a combination of three methylation markers (GDNF, SNAP91 and NDRG4) **(REF Rademakers G, Massen M, Koch A, et al. Identification of DNA methylation markers for early detection of CRC indicates a role for nervous system-related genes in CRC. Clin Epigenetics. 2021;13:1-12.)**; the TriMeth test (C9orf50, KCNQ5 and CLIP4) developed by Jensen et al. (**REF** **Jensen SØ, Øgaard N, Ørntoft M-BW, et al. Novel DNA methylation biomarkers show high sensitivity and specificity for blood-based detection of colorectal cancer—a clinical biomarker discovery and validation study. Clin Epigenetics. 2019;11:1-14.)**; and Cologuard, a multitarget stool DNA test approved by the FDA, with a sensitivity of 92.3% and specificity of 86.6%. This test assesses KRAS mutations and BMP3 and NDRG4 methylation levels in addition to an immunoassay for human hemoglobin. **(REF Ahlquist DA, Zou H, Domanico M, et al. Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. Gastroenterology. 2012;142:248-256.) (REF Bosch L, Melotte V, Mongera S, et al. Multitarget stool DNA test performance in an average-risk colorectal cancer screening population. Am J Gastroenterol. 2019;114:1909-1918. ).** In addition to diagnostic tests, several methylation markers have been implicated with CRC tumors’ metastasis potential and prognosis. For example, CDKN2A has been associated with increased risk of recurrence and metastasis; EVL, SEPT9 and IGFBP3 with shorter survival; HLTF and HPP1 with tumor recurrence; LINE-1, MGMT, TFAP2E, HIC1 and TIMP3 with responsiveness to chemotherapy. (REF **Fatemi N, Tierling S, Es HA, et al. DNA methylation biomarkers in colorectal cancer: Clinical applications for precision medicine. Int J Cancer. 2022;151(12): 2068‐2081. doi:10.1002/ijc.34186).**

# ***From the Grant: The Tumor Immune Microenvironment (TIME).*** The importance of Tumor Infiltrating Lymphocytes (TIL) on characterizing and modulating the Tumor Microenvironment (TME) and Tumor Immune Microenvironment (TIME)23 to both prognosticate and establish novel immunotherapies cannot be understated but has been understudied. The TME is represented by an amalgamation of malignant and benign cells, blood vessels, and extracellular matrix, networked with complex communication patterns through the secretion of cytokine recruitment factors 23–25. Many recent studies have shown that T cell, B cell, NK cell, and other monocyte/lymphocyte immune infiltrates and their spatial distribution, density and relationships play an important role in providing a coordinated antitumoral response, modified by Microsatellite Instability (MSI) status26,27. For example, Zou Q et al. developed a PCR-based assay for quantitative analysis of DNA methylation at single-base resolution (QASM) to determine CD8 + MeTIL signature scores. The authors were able to identify three CD8 + T cell-specific differentially methylated positions that together constituted a score with very high discriminatory power. The low CD8 + MeTIL score (enriched CD8 + TILs) was associated with MSI-H tumors and predicted better survival in two different CRC cohorts (REF **Zou Q, *et al*. DNA methylation-based signature of CD8+ tumor-infiltrating lymphocytes enables evaluation of immune response and prognosis in colorectal cancer. *J Immunother Cancer* 2021;9:e002671. doi:10.1136/jitc-2021-002671).** Using a different approach of creating immune cell specific signatures for 17 different immune cell types across different cancer types, Mitra et al. attempted to study the role of specific immune cell lineages across cancer types. Interestingly the authors found that immune-deprived tumor types like CNS tumors (lower grade glioma, glioblastoma multiforme, etc.) colocalized with a group of similarly immune-poor tumor types from the developmental gastrointestinal tumors (pancreatic adenocarcinoma, liver hepatocellular carcinoma. The authors hypothesized that that the immune microenvironments of certain tumor types bear resemblance (based on their immunological activity) despite their diverse tissue of origin. Next, they analyzed the prognostic implications of immune cell type methylation scores and lineage across the cancer cohorts and found that methylation scores for CD56+ dim NK cells, and activated and effector memory CD8+ T cells, to be significantly prognostic across multiple cancer cohorts including melanoma, with low methylation levels inferring good prognosis in most cancer types). They concluded that low methylation scores of immune cell types are associated with positive outcome for the majority of cell types, particularly for immune cells from the lymphoid lineage, which might reflect the role of the adaptive immune system for an effective response to tumor neoantigens. (REF **Mitra S et al.** **Analysis of DNA methylation patterns in the tumor immune microenvironment of metastatic melanoma. Mol Oncol 2020;14(5):993-950)**

While there are limited studies documenting the effects of TIME in CRC, immunomethylomic information on cell-type specific epigenetic alterations within this spatial arrangement related to colon cancer metastasis has not been fully elucidated due to methodological limitations.

**The need for deconvolution**

Deconvolution is the process of creating constituent cell-type data out of multiplexed studies obtained from bulk complex tissues and cell mixtures. While several canonical methods such as flow cytometry and single cell RNA-based assays exist to deconvolve samples, *in silico* methods present the advantage of cost- and time-efficiency and ease of use. In addition, DNA methylation appears to be superior to gene expression in estimating cell composition in complex mixtures due to several advantages. DNA methylation is molecularly more stable than RNA, less sensitive to technical variability, highly cell-type specific **(REF Arneson D, Yang X, Kang K. MethylResolver-a method for deconvoluting bulk DNA methylation profiles into known and unknwon cell contents. Communications Biology 2020;3:422)** and its covalent addition to a cytosine is binary, which makes it more accurate in tracking cell count. Additionally, the feature space of reference-based DNA methylation methods is at least 40 times larger and can be up to 2000 times higher than the typical gene expression feature space, leading to improved accuracy and performance. Libraries of reference-based DNA methylation deconvolution have been established and used to infer cell type composition in peripheral blood immune cells and in tissues such as the brain, breast, and skin. Although existing methods of DNA methylation such as MethylCIBERSORT have achieved some success in resolving cell types, they lack accuracy and specificity in more complex and heterogeneous environments like the TME, as they are based on data from cancer cell lines rather than primary cancer cells and use a universal standard reference. Other methods such as MethylResolver show significant promise in deconvolving cancer cells as compared with existing algorithms such as LLSR or nuSVR in both *in-silico* spike-in and *in-vitro* spike-in experiments, however, their development and validation has been again based on cancer cell lines. **(REF Arneson D, Yang X, Kang K. MethylResolver-a method for deconvoluting bulk DNA methylation profiles into known and unknwon cell contents. Communications Biology 2020;3:422)**

In cancer specifically, the need for in silico deconvolution is dictated by unknown cellular contents in resected primary tumors, the within-tumor-type heterogeneity of the microenvironment, tumor purity, or cells from the adjacent tissues. Across-tumor-type diversity in large reference studies such as TCGB, International Human Epigenome Consortium, the European BLUEPRINT increase the complexity of interpreting different studies **(REF Schreder et al. reference-free deconvolution, visualization and interpretation of complex DNA methylation data using DecompPipeline, MeDeCom and FactorViz. Nature Protocols 2020; 15:3240-3268)**

Recently a novel approach named HiTIMED has demonstrated significant deconvolution advantages in solid cancer tissue. **HiTIMED or** **H**ierarchical **T**umor **I**mmune **M**icroenvironment **E**pigenetic **D**econvolution uses deconvolution libraries specific to tumor type (12 libraries per type) to identify the most cell-discriminatory CpG sites for each cell type in each tumor type context. HiTIMED organizes deconvolution into the three major tumor microenvironment components (tumor, angiogenic, immune), resulting in the ability to resolve a total of 17 cell types in the TME: tumor, epithelial, endothelial, stromal, basophil, eosinophil, neutrophil, monocyte, dendritic cell (DC), B naive (Bnv), B memory (Bmem), CD4T naive (CD4nv), CD4T memory (CD4mem), CD8T naive (CD8nv), CD8T memory (CD8mem), T regulatory (Treg), and natural killer (NK) cells, in 20 carcinoma types. 71. Applying HiTIMED to various tumor type data, Zhang et al. (Journal of Translational Medicine (2022) 20:516) were able to identify distinct methylation signatures that were not identifiable previously in epigenome-wide association studies (EWAS). Interestingly, they applied HiTIMED to a EWAS of CRC, successfully increasing the specificity of DNAm markers when comparing normal tissue with adenocarninomas. The authors also identified distinct 5-year survival outcomes based on cell-type proportions across various solid tumors. Given these promising early pilot applications of the method, we intended to apply HiTIMED in the study of metastasis.

* **Short paragraph detailing the association with cell type and survival**

**Results**

Hitimed:  We observed worse 5-year survival outcomes with higher than median level endothelial cell proportions in lung adenocarcinoma (HR 1.83, 95% CI [1.13, 2.95]), head and neck squamous cell carcinoma (HR 1.57, 95% CI [1.07,2.29]), and kidney papillary carcinoma (HR: 3.48, 95% CI [1.27, 9.55]) (Fig. [3](https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-022-03736-6#Fig3)).

* HiTIMED associates cell type with survival - we will do this with colorectal cancer metastasis.

Chart, line chart

Description automatically generated

Kaplan-Meier plots for survival - could we adjust this for metastasis? Instead of survival population - could we include different levels for cancer development and metastasis. For example, the graph can have different Y axes.

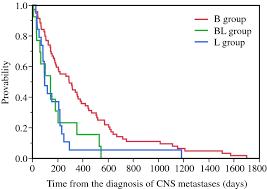
* Major limitation of the study - studying metastasis concurrent with resection to the primary site! Won’t be able to do that until we have more data. For the future study (discussion study) recurrence risk (predictor) for recurrence and death - we will have recurrence. Maybe not for the cohort collecting this year. Maybe another year out. A long enough time span with good recurrence data. Can already get that with the TCGA dataset

Other possible outcomes:

What about nodal spread vs. no spread?

Distant metastasis-free survival (DMFS) ?

Occurrence-free survival ?



example

Significance:

* Deconvolving the tumor microenvironment with DNA Methylation with HiTIMED’s tumor-specific hierarchical model that broadens the amount of immune cell types that are deconvolved
* Doing this analysis specific to colorectal cancer, which hasn’t been done before! There are studies that have detailed cell types of the tumor micronenvironment using gene expression approaches, but no one has used HiTIMED’s algorithm and DNA methylation specific to quantify cell types for colorectal cancer and associated it with metastasis.
* RNA vs DNA - DNA methylation is better

**Investigation into treatment response:**

lower levels of dendritic cell (Δ = 2.26%, p-value = 0.02), NK cell (Δ = 1.19%, p-value = 0.04), basophil (Δ = 0.53%, p-value = 0.01), neutrophil (Δ = 1.25%, p-value = 0.03), and a significantly higher tumor proportion (Δ = 7.74%, p-value = 0.03), in FOLFOX or FOLFIRI drug-sensitive patients compared to drug-resistant patients (Additional file [2](https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-022-03736-6#MOESM2): Figure S15).

* Could validate these with our data and then write about its application into treatment response
* Beyond the study scope - expand cohort
* DNA methylation over gene expression. Our tool leverages - passing mension. Methylation is on/off for a specific step type. Beta values is a proportion for a methylated allele - exact proportion of cells that are methylated.
* Gene expression is unbounded

Why DNA methylation? Why its better etc. how is it measured. How have folks done dna methylation and tumor metastasis? Deficiencies in the prior methods and papers that we aim to rectify using the hitimed libraries and we have two cohorts.

Methods overview section - before diving into the methods. Small outlines. We ran lima post halk adjustment. Compared proportion of cell types based on these tests.

Next steps - mention as a next step. Recurrence and metastasis as proxy of recurrence.

**METHODS**:

Two cohorts were used in this study. The first was conducted locally at Dartmouth Hitchcock Medical Center where thirty-six adenocarcinoma resections were performed from 2016 to 2019 with IRB approval. This cohort has been successfully profiled in previous studies, including a study involving a novel digital spatial profiling approach to spatial proteomic signature identification of colorectal cancer metastasis. As stated in the DSP study, “approximately half of the resections showed local invasion but no nodal or distant metastasis (no METS), and the other half showed nodal and/or distant metastasis (METS). Of the cases with concurrent metastasis, all cases exhibited local lymph node involvement– about half of these cases metastasized to distant sites.” Sample size was determined by a pilot study that utilized an empirical power analysis to simulate data from statistical models used in the digital spatial profiling study.

The cohort was restricted to stage pT3 assignments under the pTNM staging system. The pTNM system includes a balance between local invasion and metastasis as a prognostic indicator. Restricting stage assignments to the T stage, which characterizes invasion at the local site, allowed for additional prognostic utility above the pTNM system. Cases were characterized as non-metastatic vs. metastatic based on tissue-size (achieved through connected component analysis of whole slide images), tumor grade, mismatch repair (MMR) dysregulation status (*dMMR*– deficiency, *pMMR*– proficient; as assessed through IHC), tumor site (e.g., left or right colon), age, and sex.  Fisher's exact tests and two-sample t-tests after iterative resampling were used for randomization. As indicated in the DSP study, “MMR deficiencies (*dMMR*) reflect the loss of staining in at least one of four mismatch repair genes (MLH1, PMS2, MSH2, MSH6). As MSH2 and MSH6 alterations were relatively rare for cases within the queried time periods, *dMMR* status was reported from alterations to either MLH1 or PMS2 (MSH2 and MSH6 alterations were not present in this cohort)” 25,26

The second cohort used in this study was pulled from TCGA (The Cancer Genome Atlas), which is a novel cancer genomics program that has characterized over 22,000 primary cancers across 33 different cancer types. The TCGA cohort was incorporated into the analysis using Bayesian modeling approaches (e.g. empirical bayes adjustment for differential expression) that allowed us to set informative priors from posterior parameter estimates based on similar patient characteristics. Informative priors where then compared to data from non-informative priors for extra validity.

TCGA pretraining allowed for ML parameters to be closely initialized as with the DH cohort. pwrEWAS was used to collect 64 samples to increase statistical power from 0.62 (n=32) to 0.76 in order to find  2,500 differentially methylated CpGs from a set of 100,000 CpGs (79).

EWAS stands for Epigenome-Wide Association Study and measures the epigenetic modifications (epigenetic modifications to DNA that result in genes being over or under expressed) associated with a particular disease. DNAm and Histone modification are two examples of epigenetic modifications. For this study, we focused on DNAmethylation as the sole epigenetic modification to identify relevant epigenetic predictors of colon cancer metastasis. Genes were filtered out based on variation between the genetic profile of patients using GoMeth and GOregion, which are two novel methods to perform unbiased gene set testing post-methylation analysis of CpG sites and regions. These methods enrich the data set by accounting for biases that relate to the annotation process of differentially methylated probes into differentially methylated genes, which can then be assigned to a particular gene set.

Ultimately GoRegion was used instead of GoMeth for increased power and Type-I error thresholds were loosened. The MSigDB hallmark gene set database was consulted instead of GO and KEGG due to its accuracy in revealing relevant pathways such as epithelial to mesenchymal transition. gsameth was used to download custom gene sets.

Further tests to identify enrichment of specific genomic regions, such as CpG island promoters, promoter, enhancer, island, shore, shelf, sea, etc. were conducted via GREAT and LOLA. GREAT is a short for Genomic Regions Enrichment of Annotations Tool and was used to improve the significance of cis-regulatory regions which are measured by localized measurements of DNA binding events. LOLA stands for locus overlap analysis and provides enrichment for genomic region sets, making it easier to interpret epigenomics data. Both are available as R packages in Bioconductor and on the internet. <https://www.nature.com/articles/nbt.1630> and <https://academic.oup.com/bioinformatics/article/32/4/587/1743969>

Genome-wide DNAm profiling was accomplished through Epi COBRE Biorepository and CQB COBRE SCGC, which used DNA for all cases (n=64) for bisulfite conversion and DNAm profiling by the Illumina EPIC methylation array (69). Profiling yielded proportions of methylated alleles across the mixture (beta-value) at 850k CpG islands. The DAC and Salas labs at Dartmouth College assisted in DNAm preprocessing (70).

We incorporated the TCGA data, Bayesian approaches, relaxing FDR adjustments, variance filtering, and assessing coarse cellular hierarchy to strengthen the power of our findings.

TME-related cell-types, including tumor, epithelial, endothelial, stromal, basophil, eosinophil, neutrophil, monocyte, dendritic cell, B naïve and memory, CD4T naïve and memory, CD8T naïve and memory, T regulatory, and natural killer cells, were accurately resolved using HiTIMED - **H**ierarchical **T**umor **I**mmune **M**icroenvironment **E**pigenetic **D**econvolution (71). HiTIMED, as mentioned in the introduction, is a novel technique that produces cell-types with DNAm rather than protein/RNA and gene expression methods, and has been validated in many cancer assays for its improved accuracy among older methods. In concordance with the HiTIMED algorithm, we deployed a DNAm assay for cell-type specific metastatic alterations that is more feasible and less expensive than an RNA-based assay. We then tied in findings from our Digital Spatial Profiling approach by correlating deconvolved DNAm cell types meanwhile binning ST genes due to potential DNAm dysregulation. The ​​Metastasis-related differentially methylated CpGs that were identified through an epigenome wide association study (EWAS)(72), used the statistical model:

*M* is the beta-value (CpG *j*) transformed to a normal distribution73 to address heteroskedasticity. Models adjusted for and conditioned on cell type proportions (; *K* cell-types; *K-1* assessed due to simplex constraint ). Effect estimates were compared to unadjusted models. Conditional cell-type specific effects used the CellDMC and Tensor Composition Analysis (TCA) approaches74,75, which estimate differentially methylated cell types through the addition of interactions () to the statistical model. MEML models incorporating public data identified additional CpG-cell-type (e.g., ) and CpG-covariate interactions (e.g., ), reported via: for identified combinations of CpG *j* and cell-type *k* or covariate *l*. ML models were compared to MethylCapsNet (developed by PI to report pathways) and Elastic Net 76. CpGs were associated with genes for pathway analysis after FDR adjustment77,78

Quality Control was run on SeSAME on R and used to create QC plots. SeSAME accepts IDATs and Beta values for input date. As a means of further processing, single nucleotide polymorphisms were identified using MethylToSNP – an R Bioconductor package which detects to predict sites confounded by polymorphisms. MethylToSNP only considers SNPs specific to to individuals in a study, has the potential to identify new polymorphisms in an uknown genomic landscape, and identifies variants in functional genomic regions, making it a valuable technique to accurately clean methylation data while preventing excessive data loss. Relevant SNP crosstalk was also noted in the results section. (<https://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/s13072-019-0321-6>)

All analyses were run with and without Mismatch-Repair, which is a term to describe a biochemical process that corrects errors in DNA replication. MMR deficiency is very common in colorectal cancer and leads to MSI, or microsatellite instability, in which short repeating sequences in the DNA are prone to mutation.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5841008/#:~:text=Microsatellite%20instability%20(MSI)%20is%20a,%2Drepair%20(MMR)%20genes>

Sex-related CPGs were removed on the X and Y chromosome using manifest.

DNA Methylation Is used to suppress the mobility of repetitive elements and is often observed in tumors. Repetitive elements are remnants of transposons and can proliferate and mobilize throughout the genome. Repeat element methylation is a more global type of methylation that can detect viral insertions or DNA outside the genome over large periods of time in addition to methylated CpG islands in the promoter region. If these elements are methylated, they can affect DNA methylation methods. While DNA methylation and other epigenetic modifications are frequently associated with gene silencing, repeat element methylation is mainly utilized to maintain genomic stability and prevent the transcription of viral insertions.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5587781/>

* Rewrite this section – not just junk DNA
* Hypomethylation – global marker
* There are broad families. If we can detect enrichment we can find broad functions of these lineages of the repeat elements. Different functionality.

Loci were ranked in one of two ways. First, they were ranked by median absolute deviation. Second, they were ranked via variance. Variance was first plotted from high to low to help identify the optimal cutoff. According to its creators, MethylMasteR – a DNA Methylation-based copy number variation (CNV) calling software – provides both copy number and methylation state information in a single library format, which is vary advantageous in comparing performance and CNV event identification among four methylation-based CNV callers.

[**https://pubmed.ncbi.nlm.nih.gov/35573871/**](https://pubmed.ncbi.nlm.nih.gov/35573871/)

Before running the Methylation array, two types of batch effects, named “row and chip” by Price and Robinson, were adjusted for using an R function ComBat which adjusts for non-biological signals.

* ComBat (preprocessing phase) – moves everything to the middle and removes batch effect
* Could also do a random intercept model and will still report the global slope w minimal loss of power in the analysis phase
* Higherarchical modeling

In addition, we studied the overlap between sites of interest and key transcriptional factor binding sites and/or histone modifications. Data on other key epigenetic modifications (including DNA methylation) was sourced from the NIH Roadmap Epigenomics Mapping Consortium, which is a public resource of human epigenomic data.

<https://egg2.wustl.edu/roadmap/web_portal/>

Tumor purity was validated using RF\_Purity, a novel tool for tumor-purity in methylation array data. Given the fact that many array-based techniques methylation techniques of primary tumor samples measure bulk tumor sample DNA rather than individual cell methylation, tumor purity must be assessed due to the varying amount of stromal components and infiltrating immune components in addition to the tumor cells in the primary site. Since matching to control samples were not available, RF\_PURITY allowed us to use a reference free method to quantify tumor purity using two random forest classifiers which were trained on purity values of TCGA tumor samples. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6697926/>

Lastly, ToppGene Suite, which is a candidate gene prioritization portal and enrichment analysis, was used to detect gene enrichment of our gene list based on Regulome (miRNA) interactions.

<https://toppgene.cchmc.org/>

Discussion/results

## **SFRP2 (secreted frizzled-related protein 2) - literature review:**

Methylation changes in faecal DNA: a marker for colorectal cancer screening Lancet

2004 Apr 17;363(9417):1283-5.

doi: 10.1016/S0140-6736(04)16002-9

**Description from gene atlas**:

*Wnt (wingless-type)/β-catenin signaling is a major regulator of cell proliferation, migration and differentiation, controlling tissue homeostasis and tumor progression (Klaus et al., 2008). A) The binding of a canonical Wnt ligand to its cell-surface receptor complex, consisting of Frizzled (FZD) and one of two low-density-lipoprotein- receptor-related proteins (LRP-5 and LRP-6), initiates a signaling cascade that activates disheveled (DVL), which releases β-catenin from an inhibitory complex consisting of Axin, APC and glycogen synthase kinase 3b (GSK3B). On dephosphorylation and release, β-catenin translocates to the nucleus, where it interacts with members of the T-cell factor/lymphoid enhancer factor (TCF/LEF) families of transcription factors to stimulate expression of genes involved in cell survival, proliferation and osteoblastic differentiation (e.g., MMPs, CCND1, PTGS2, MYC, JUN and VEGFR) (Reya et al., 2005). B) Wnt signaling is regulated by several classes of negative regulators. The Secreted Frizzled-Related Protein (SFRP) class comprises SFRP1-SFRP5, Wnt inhibitory factor 1 (WIF1) and Cerberus. SFRPs are a family of soluble glycoproteins that possess a cysteine- rich domain (CRD) structurally similar to the extracellular Wnt-binding domain of the FZD receptors. SFRPs can thus modulate Wnt signaling by sequestering Wnts through their CRD or by acting as dominant-negative inhibitors, forming inactive complexes with the FZD receptors.*

**Function**

SFRP2 contains several domains which potentially govern protein-protein interactions.

- Beginning at the N-terminus there is a 24 aa hydrophobic signal domain, which presumably governs the targeting of SFRP2 to the secretory pathway.

- The CRD/FZ (cysteine-rich/Frizzled) domain allows SFRP proteins to antagonise Wnt/Frizzled binding at the plasma membrane either by sequestration of the Wnt ligand or dominant-negative binding to complimentary regions within the Frizzled protein.

- The C345C/Netrin domain is an accessory binding domain for SFRP2-Wnt interaction.

- The PDZ ligand domain is a short sequence which is recognised by PDZ proteins such as Dishevelled (Dvl), and other proteins which interact with the cytoplasmic portion of Frizzled (Schulte and Bryja, 2007). Notably, a recent study (Zhang et al., 2009) has used synthetic PDZ ligands to interfere with Dvl/FZD cytoplasmic interaction, and thus antagonise canonical [Wnt signalling](https://atlasgeneticsoncology.org/deep-insight/20042/the-wnt-signaling-pathway-and-its-role-in-human-solid-tumors). Additionally, another group subsequently showed that the NSAID Sulindac inhibits canonical Wnt signaling by blocking the PDZ domain of Dvl (Lee et al., 2009). These data suggest that the presence of a PDZ ligand in SFRP2 may indicate a previously-unexplored role for SFRP2 as a cytoplasmic antagonist of Wnt signalling.

- A phosphoproteomic study has revealed phosphorylation of SFRP2 at serine-289 in response to growth factor stimulation (Olsen et al., 2006). Bioinformatic analysis suggests that this site is a motif for recognition and phosphorylation by PKA. Additionally, bioinformatic analysis of the SFRP2 sequence suggests a site for phosphorylation by [GSK3β](https://atlasgeneticsoncology.org/gene/40761/gsk3b-(glycogen-synthase-kinase-3-beta)) (itself an inhibitor of [β-catenin](https://atlasgeneticsoncology.org/gene/71/ctnnb1-(catenin-beta-1))) at either serine-34 or serine-38. Although the potential significance of phosphorylation at this site is unclear, it may overlap with the [Nec 1](https://atlasgeneticsoncology.org/gene/41671/pcsk1-(proprotein-convertase-subtilisin-kexin-type-1))/Nec 2 cleavage site and prevent removal of the N-terminal signal domain. Typically, Nec 1/2 is responsible for the cleavage of pro-proteins into their active form, and perhaps phosphorylation in this region is a mechanism by which GSK3β may regulate SFRP2 activity.

#### Expression

Widely expressed.

#### Localisation

Nucleus, cytoplasm and secreted.

#### Function

SFRP2 is a member of the secreted Frizzled-related protein (SFRP) family of soluble extracellular Wnt antagonists, which act in conjunction with the Dickkopf (DKK) class of Wnt antagonists. SFRP2 is thought to act primarily by binding directly to and sequestering Wnt ligands, but may also act by direct binding to the Wnt-receptor complex. Binding occurs primarily via the cysteine-rich domain, which bears a high degree of homology to similar domains in the Frizzled (Fzd) receptor. SFRP activity may conversely promote Wnt pathway signalling in some contexts, as a consequence of SFRP proteins interacting with each other and titrating each others activity, or by SFRP binding and stabilising Wnt-Fzd complexes.

#### Homology

SFRP2 has a high degree of homology to both other SFRP family members, as well as Fzd receptors via the cysteine-rich domain.

## **SND1 (staphylococcal nuclease and tudor domain containing 1)**

Literature review: <https://pubmed.ncbi.nlm.nih.gov/17909068/>

## **ZNF606 (zinc finger protein 606)**

Literature review: <https://pubmed.ncbi.nlm.nih.gov/25500544/>

## **GAS7 (growth arrest-specific 7)**

Literature review: Genetic analysis in a Dutch study sample identifies more ulcerative colitis susceptibility loci and shows their additive role in disease risk. Am J Gastroenterol. 2010 Feb;105(2):395-402.

doi: 10.1038/ajg.2009.576.<https://pubmed.ncbi.nlm.nih.gov/19861958/>

**Expression**

in quiescent/terminally differenciated cells, like other GAS (herien the name Growth Arrest-Specific); selective expression in growth-arrested fibroblasts; GAS7 is predominantly expressed in the brain (in particular, in the mouse, in Purkinje cells of the cerebellum, at lower level in other tissues such as the heart and the testes); gas7 expression is associated with neurite formation (and overproduction of gas7 in neuroblastoma cells promotes neurite-like outgrowth) which suggests that GAS7 may have a role in promoting, and possibly maintaining, maturation and morphological differentiation of cerebellar neurons