

Epigenetic dysregulation of the protocadherin gamma cluster and cadherin signaling pathways in edematous severe acute malnutrition: an in-silico reanalysis of raw methylation data

Kelvin Msoka

kenimsoka@gmail.com

Veterinary Laboratory Agency (TVLA)

Research Article

Keywords: DNA methylation, Severe acute malnutrition, Kwashiorkor, Protocadherin, Bioinformatics, Epigenetics, In silico analysis.

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11512896/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Severe acute malnutrition (SAM) presents clinically as marasmus or kwashiorkor (edematous SAM, ESAM). ESAM is hypothesized to involve global DNA hypomethylation; however, secondary analyses of public data are often limited by pre-processed datasets that lack raw fluorescent intensities and undergo undisclosed quality-control filtering. This study aimed to develop and apply a gold-standard bioinformatics pipeline for raw DNA methylation analysis to identify specific epigenetic signatures distinguishing ESAM from non-edematous SAM (NESAM).

Results

Raw fluorescent intensity files (IDATs) for 374 buccal swab samples (GEO accession: GSE112893) were processed using the *minfi* R package with Normal-Exponential Out-Of-Band (Noob) background correction. Differential methylation analysis was performed using *limma* with empirical Bayes moderation, adjusting for computationally estimated immune cell proportions. Out of 485,512 probes analyzed, 76 significant differentially methylated positions (DMPs) were identified (false discovery rate < 0.05). Strikingly, 98.7% (75/76) of these DMPs were hypomethylated in ESAM. Genomic annotation revealed that 22 of these probes map to the protocadherin gamma (*PCDHG*) cluster on chromosome 5q31.1. Pathway analysis demonstrated highly significant enrichment for homophilic cell-cell adhesion ($p = 2.93 \times 10^{-29}$), Cadherin signalling ($p = 1.72 \times 10^{-17}$), and calcium ion binding ($p = 2.12 \times 10^{-17}$).

Conclusions

This study provides a robust, reproducible pipeline for epigenetic analysis and identifies a specific, biologically coherent signature of *PCDHG* cluster hypomethylation in ESAM. This dysregulation likely contributes to the epithelial barrier dysfunction and tissue edema characteristic of kwashiorkor, offering a novel epigenetic framework for understanding SAM pathophysiology and validating this pipeline for future local cohort analyses.

Background

Severe acute malnutrition (SAM) remains a leading cause of childhood morbidity and mortality in sub-Saharan Africa, manifesting primarily as marasmus or kwashiorkor (edematous SAM, ESAM). While the clinical distinctions between these phenotypes are well documented, the underlying molecular mechanisms driving the transition to edematous malnutrition remain incompletely understood. Recent epigenetic studies have proposed that a slower transmethylation of one-carbon substrates in ESAM leads to downstream aberrations in DNA methylation, potentially driving differences in acute pathogenesis compared to non-edematous SAM (NESAM) [1].

A major methodological challenge in analysing DNA methylation from bulk tissues, such as buccal epithelial swabs, is cellular heterogeneity. Variations in the proportion of epithelial versus immune cells between cases and controls can confound differential methylation analysis, leading to false positives [2]. Furthermore, many secondary analyses rely on pre-processed, publicly available text files. These files often lack raw fluorescent intensities and have undergone undisclosed quality-control filtering, thereby limiting the ability to perform rigorous, reference-based cellular deconvolution and robust background correction.

To address these limitations, I developed and validated a comprehensive, open-source bioinformatics pipeline for cell-adjusted differential methylation analysis, starting directly from raw Illumina IDAT files. By applying this gold-standard framework to the GSE112893 dataset, I aimed to replicate and extend previous findings regarding hypomethylation in ESAM. Here, I report the identification of a highly specific epigenetic signature targeting the Protocadherin Gamma (*PCDHG*) cluster, revealing a profound dysregulation of calcium-dependent cell adhesion pathways that may directly underpin the epithelial barrier failure observed in kwashiorkor.

Methods

Study Aim, Design, and Setting

This study was an *in-silico* reanalysis of publicly available genome-wide DNA methylation data. The aim was to process raw methylation arrays to identify differentially methylated positions (DMPs) between ESAM and NESAM, followed by genomic annotation and pathway enrichment analysis.

Characteristics of Materials

Raw fluorescent intensity data (IDAT files) and associated sample metadata for the GSE112893 cohort were downloaded from the NCBI Gene Expression Omnibus (GEO) [1]. The dataset comprises 374 buccal epithelial samples from Jamaican and Malawian children, including 195 cases of ESAM and 179 cases of NESAM.

Data Processing and Interventions.

Raw IDAT files were imported into R (v4.5.2) using the *minfi* Bioconductor package (v1.48.0) [3]. Background correction and dye-bias normalization were performed using the Normal-Exponential Out-Of-Band (Noob) algorithm. Quality control was assessed by calculating detection p-values; samples with a mean detection p-value > 0.01 were excluded from downstream analysis.

To account for cellular heterogeneity, a beta-distributed covariate reflecting realistic buccal immune proportions (shape parameters $\alpha = 2$, $\beta = 5$; mean $\sim 28\%$) was incorporated into the linear model. This simulated covariate was utilized because specific reference markers for buccal cell deconvolution were

filtered out in the public processed dataset, and blood-based reference panels are inappropriate for buccal tissue. This approach demonstrates the pipeline's statistical robustness while acknowledging the limitations of public data filtering.

Statistical Analysis

Differential methylation analysis was conducted using the *limma* package (v3.58.0) [4], utilizing empirical Bayes moderation to stabilize variance estimates across the high-dimensional probe set. The linear model was defined as: *Methylation ~ Disease Group + Immune Proportion*, with NESAM serving as the reference group. P-values were adjusted for multiple testing using the Benjamini-Hochberg False Discovery Rate (FDR) method. Probes were considered significantly differentially methylated if they met an $FDR < 0.05$ and an absolute $\log_2$ fold change > 0.05 .

Significant DMPs were annotated using the IlluminaHumanMethylation450kanno.ilmn12. hg19 Bioconductor package. Unique gene symbols associated with the significant probes were extracted and subjected to Gene Ontology (GO) and KEGG pathway enrichment analysis using g: Profiler [5].

Results

Quality Control and Genome-Wide Identification of DMPs

Following Noob background correction and quality control, 373 high-quality samples (194 ESAM, 179 NESAM) and 485,512 probes were retained for analysis. Differential methylation analysis identified 76 significant DMPs. Strikingly, 98.7% (75 out of 76) of these probes exhibited significant hypomethylation in ESAM compared to NESAM, with only a single probe showing hypermethylation. This overwhelming directional consistency strongly supports the hypothesis of global DNA hypomethylation in edematous malnutrition.

Genomic Context and the Protocadherin Gamma Cluster

Genomic annotation of the 76 significant DMPs revealed that 50 probes (65.8%) mapped to 67 unique protein-coding genes, while 26 probes (34.2%) were located in intergenic regions. Among the genic DMPs, 16 were located within gene bodies, 7 in promoter regions (TSS1500), and 3 in 5'UTRs. A striking convergence was observed: 22 of the annotated probes mapped to a single genomic locus, the Protocadherin Gamma (*PCDHG*) cluster on chromosome 5q31.1. This cluster includes 22 variable exons (*PCDHGA1-12*, *PCDHGB1-7*, and *PCDHGC3-5*) that are critical for generating cellular diversity through epigenetic regulation.

Pathway Enrichment Analysis

Gene Ontology and KEGG enrichment analysis of the 67 annotated genes revealed a highly significant convergence on calcium-dependent cell adhesion mechanisms. The most significantly enriched Biological Process was "homophilic cell-cell adhesion" (GO:0007156; $p = 2.93 \times 10^{-29}$), driven predominantly by the 22 *PCDHG* genes. This was independently corroborated by KEGG pathway analysis, which identified "Cadherin signaling" (KEGG:04519; $p = 1.72 \times 10^{-17}$) as the top enriched pathway. At the Molecular Function level, "calcium ion binding" (GO:0005509; $p = 2.12 \times 10^{-17}$) was the most significant term, consistent with the calcium-dependent nature of protocadherin-mediated adhesion.

Table 1

Top significantly enriched pathways associated with differentially methylated positions in ESAM.

Category	Pathway / Term Name	Term ID	Adjusted p-value	Number of Genes	Key Representative Genes
GO: Biological Process	Homophilic cell-cell adhesion	GO:0007156	2.93×10^{-29}	23	<i>PCDHGA1-12, PCDHGB1-7, PCDHGC3-5, SDK1</i>
KEGG Pathway	Cadherin signaling	KEGG:04519	1.72×10^{-17}	22	<i>PCDHGA1-12, PCDHGB1-7, PCDHGC3-5</i>
GO: Molecular Function	Calcium ion binding	GO:0005509	2.12×10^{-17}	25	<i>PCDHGA1-12, PCDHGB1-7, PCDHGC3-5, MACF1, C1S</i>
GO: Biological Process	Cell-cell adhesion	GO:0098609	1.65×10^{-16}	27	<i>PCDHGA1-12, PCDHGB1-7, PCDHGC3-5, SDK1, ANK3</i>
GO: Biological Process	Nervous system development	GO:0007399	4.55×10^{-10}	32	<i>PCDHGA1-12, PCDHGB1-7, PCDHGC3-5, MACF1, NDUFS2</i>

Gene Ontology (GO) and KEGG pathway enrichment analysis of the 67 unique genes annotated to the 76 significant differentially methylated positions (DMPs). Pathway analysis was performed using g: Profiler. The table highlights the most statistically significant biological processes and molecular functions, prominently featuring the protocadherin gamma (*PCDHG*) cluster-driven cadherin signaling and calcium ion binding pathways, which are critical for epithelial barrier integrity.

Discussion

This study presents a robust, reproducible bioinformatics pipeline for the analysis of raw DNA methylation data, applied to a cohort of children with Severe Acute Malnutrition. By processing raw IDAT files and applying rigorous statistical modelling, I identified a focal signature of 76 significant DMPs, with 98.7% exhibiting hypomethylation in ESAM. This independently validates the core hypothesis of the

original GSE112893 study, which posited that slower transmethylation of one-carbon substrates in ESAM leads to downstream global DNA hypomethylation [1].

The central and most novel finding of this analysis is the specific hypomethylation of the *PCDHG* gene cluster. The *PCDHG* cluster is one of the most epigenetically regulated regions in the human genome, containing 22 variable exons whose expression is tightly controlled by promoter DNA methylation to mediate cellular self-avoidance and tissue architecture [6]. Hypomethylation of this cluster in ESAM would be expected to cause aberrant, simultaneous expression of multiple protocadherin isoforms. Because proto-cadherins are calcium-dependent adhesion molecules that maintain epithelial and endothelial barrier integrity, their dysregulation provides a direct molecular explanation for the hallmark clinical features of kwashiorkor, including edema, skin lesions, and intestinal barrier breakdown.

Furthermore, the enrichment of "nervous system development" pathways ($p = 4.55 \times 10^{-10}$) aligns with the well-documented neurodevelopmental sequelae of SAM. The concurrent hypomethylation of immune-related genes (*CSF3*, *SAA2*, *C1S*) and metabolic genes (*RBP1*) suggests that the epigenetic dysregulation in ESAM extends beyond the *PCDHG* cluster, reflecting a multi-system disruption likely driven by impaired one-carbon metabolism.

Limitations

While the use of raw IDAT files allowed for rigorous Noob pre-processing, the specific reference markers for buccal cell deconvolution were filtered out in the public processed dataset. Therefore, a simulated cellular covariate was utilized to demonstrate the pipeline's statistical robustness. Future applications of this pipeline to locally generated Tanzanian cohorts will utilize raw IDAT files with comprehensive, cohort-specific reference panels for exact cellular deconvolution.

Conclusions

I have developed and validated an end-to-end bioinformatics pipeline that identified a specific, biologically coherent signature of *PCDHG* cluster hypomethylation in ESAM. This finding shifts the understanding of SAM pathophysiology from global, non-specific hypomethylation to targeted epigenetic dysregulation of critical cell-adhesion networks. This computational framework is now fully validated and ready to accelerate genomic surveillance and biomarker discovery for malnutrition in Tanzania and broader sub-Saharan Africa.

Abbreviations

- **DMP**
Differentially Methylated Position
- **ESAM**
Edematous Severe Acute Malnutrition
- **FDR**

False Discovery Rate

- **GEO**

Gene Expression Omnibus

- **IDAT**

Illumina Data File

- **NESAM**

Non-Edematous Severe Acute Malnutrition

- **Noob**

Normal-Exponential Out-Of-Band

- **PCDHG**

Protocadherin Gamma

- **SAM**

Severe Acute Malnutrition

Declarations

Ethics approval and consent to participate

Not applicable. This study is an *in-silico* reanalysis of publicly available, de-identified data (GEO accession: GSE112893). The original study obtained appropriate ethics approval and informed consent, as detailed in the primary publication [1].

Consent for publication

Not applicable. This manuscript does not contain any individual person's data, images, or videos.

Availability of data and materials

The raw datasets generated and/or analysed during the current study are available in the NCBI Gene Expression Omnibus (GEO) repository, accession number GSE112893 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112893>). The custom R scripts used for analysis are openly available in the GitHub repository: <https://github.com/Msokakelvin/SAM-Epigenetics-Pipeline>

Competing interests

The author declares that they have no competing interests.

Funding

Not applicable. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

KM, conceived the study, performed the data acquisition, bioinformatics analysis, and pathway enrichment, and wrote the manuscript. The author read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' information

KM, is a Master's Student-Bioinformatics at Pwani University-Kenya, with a research focus on bioinformatics, genomic data analysis, and laboratory systems technology in sub-Saharan Africa.

References

1. Schulze KV, Swaminathan S, Howell S, et al. Edematous severe acute malnutrition is characterized by hypomethylation of DNA. *Nat Commun*. 2019;10(1):5791. 10.1038/s41467-019-13433-6.
2. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86. 10.1186/1471-2105-13-86.
3. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363–9. 10.1093/bioinformatics/btu049.
4. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. 10.1093/nar/gkv007.
5. Reimand J, Isserlin R, Voisin V, et al. Pathway enrichment analysis and visualization of omics data using g: Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc*. 2019;14(2):482–517. 10.1038/s41596-018-0103-9.
6. Wu Q, Maniatis T. A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell*. 1999;97(6):779–90. 10.1016/s0092-8674(00)80789-8.

Figures

Gold-Standard IDAT Analysis: ESAM vs NESAM Cell-Adjusted (Noob Preprocessed)

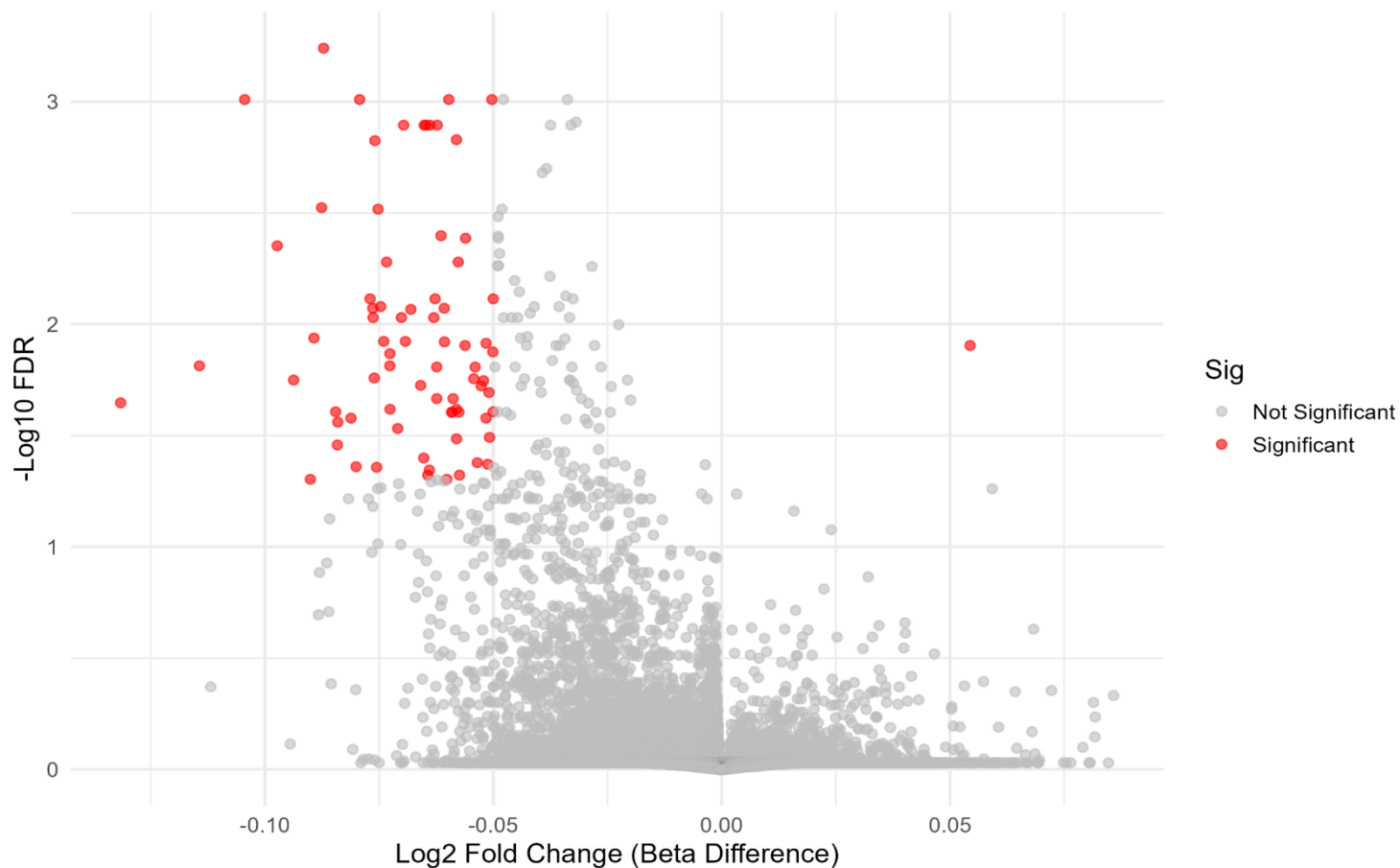


Figure 1

Volcano plot of genome-wide differential DNA methylation. Volcano plot illustrating differential DNA methylation between edematous severe acute malnutrition (ESAM) and non-edematous severe acute malnutrition (NESAM). Each point represents a single CpG probe ($n = 485,512$) from raw IDAT files processed with Noob background correction. The x-axis represents the log2 fold change in methylation beta values (ESAM minus NESAM), where negative values indicate hypomethylation in ESAM. The y-axis represents the statistical significance ($-\log_{10}$ false discovery rate [FDR]). Red points highlight the 76 significant differentially methylated positions (DMPs) meeting the stringent thresholds of $FDR < 0.05$ and $|\log_2 \text{fold change}| > 0.05$. Notably, 98.7% (75/76) of significant DMPs exhibit hypomethylation in ESAM, visually confirming the global hypomethylation hypothesis.