HHS Public Access

Author manuscript

Epigenomics. Author manuscript; available in PMC 2026 January 16.

Published in final edited form as:

Epigenomics. 2025 December ; 17(17): 1259–1266. doi:10.1080/17501911.2025.2570119.

DNA methylation-predicted protein differences between Yakutian and Central Russian populations

Jamaji C. Nwanaji-Enwerem[a], **Dennis Khodasevich**[b], **Andres Cardenas**[b,c]

[a]Department of Emergency Medicine, Center for Health Justice, and Center of Excellence in Environmental Toxicology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[b]Department of Epidemiology and Population Health, Stanford University, Palo Alto, CA, USA

[c]Department of Pediatrics, Stanford University, Stanford, CA, USA

Abstract

Background: Populations in subarctic regions, like Yakutia in the Russian Sakha Republic, have adapted to extreme environmental conditions, including intense cold, pronounced shifts in daylight, and variable food availability. However, the biological mechanisms underlying these adaptations remain poorly understood despite insights from genome-wide (GWAS) and epigenome-wide association studies (EWAS).

Methods: Since protein profiles may more directly reflect functional physiology, we analyzed DNA methylation data from 245 healthy Russian participants using methylation-based estimators of circulating protein levels to investigate estimated proteomic differences between residents of Yakutia and Central Russia.

Results: We identified regional variation in 25 protein surrogates enriched in pathways, including MET receptor activation and PI3K-Akt signaling. Some proteins mapped to previously identified GWAS genes. To our knowledge, none mapped to previously identified, differentially methylated in EWAS genes, suggesting that methylation-based protein estimation may capture distinct, complementary aspects of physiological regulation.

Conclusion: These findings align with prior -omics research by highlighting regional molecular differences possibly associated with cold adaptation. They also underscore the potential of

CONTACT Jamaji C. Nwanaji-Enwerem nwanajij@pennmedicine.upenn.edu Department of Emergency Medicine, University of Pennsylvania, Ground Ravdin, HUP 3400 Spruce Street, Philadelphia, PA, 19104 USA; Andres Cardenas andres.cardenas@stanford.edu Department of Epidemiology and Population Health, Stanford University, Palo Alto, CA, USA.

Author contributions

Jamaji Nwanaji-Enwerem conceived of the analyses, performed data analysis, visualization, original writing. Andres Cardenas and Dennis Khodasevich contributed to the data processing and analysis. Andres Cardenas supervised the work. All authors contributed to writing/editing of the manuscript. All authors reviewed and approved the final manuscript.

Disclosure statement

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript.

methylation-derived proteomic proxies as a useful, indirect approach for studying proteomic variation when direct protein measurements are unavailable. While promising, this method warrants further validation, particularly with respect to potential genetic confounding.

PLAIN LANGUAGE SUMMARY

People living in subarctic regions such as Yakutia in the Russian Far East have adapted to some of the harshest environments in the world. These include extreme cold, long winter nights, and limited access to food. Scientists believe these adaptations are reflected in the body's biology, but studies examining genes and other molecular traits have not fully explained how these changes occur. In this study, we used a new method that estimates protein levels in the body based on patterns of DNA methylation. DNA methylation is a type of chemical change on DNA that can show how the environment influences the body. We applied this method to blood samples from 245 healthy adults living in two regions of Russia, Yakutia and Central Russia. We found differences in 25 estimated protein levels between the two groups. These proteins are involved in important processes like managing energy use. To our knowledge, these proteins were not linked to genes found in previous epigenetic studies. This suggests the method may provide new insights. Our findings show that estimating protein levels from DNA can help uncover unique biological features in populations adapted to extreme environments. This approach could help guide future research on human adaptation and health.

Keywords

DNA methylation; proteome; temperature; thermoregulation; energy; metabolism

1. Introduction

A growing body of research has investigated how populations residing in the subarctic and arctic territories of Siberia, such as Yakutia in the Russian Sakha Republic, have adapted to their extreme environments, characterized by cold temperatures, dramatic seasonal fluctuations in daylight, and associated variability in food availability. Prior studies have implicated biological pathways related to energy metabolism in these adaptations, noting patterns such as lower serum lipid levels and higher blood pressure [1–3]. Building on this work, researchers have aimed to understand the genetics underlying these adaptations through genome-wide association studies (GWAS) [4]. Most recently, researchers expanded on these efforts by conducting the first epigenome-wide association study (EWAS) in this population, comparing residents of Yakutia to those living in Central Russia [5]. GWAS and EWAS findings have demonstrated different allele frequency and differentially methylated sites between the two regions corresponding to genes involved in basal metabolic rate, thermoregulation, energy metabolism and smooth muscle contraction among other processes. For instance, cg24649109, located in the *MYO5C* gene, was found to be hypermethylated in Yakutia residents compared to Central Russians [5], potentially implicating pathways related to vesicle transport and metabolic regulation [6,7]. In addition to identifying differentially methylated sites between the two groups, the previous EWAS tested seven epigenetic aging markers – DNA methylation-based biomarkers of biological

aging associated with morbidity and mortality [8] – and found evidence of greater age acceleration in Yakutian individuals compared to Central Russians [5].

Although GWAS and DNA methylation marks are valuable for studying biological processes and can influence protein levels, proteins more directly reflect the functional outcomes of gene expression and disease [9]. However, proteomic data are often unavailable in GWAS and EWAS cohorts. DNA methylation-based protein scores offer a promising alternative by leveraging methylation patterns to estimate circulating protein levels. These scores are trained on multi-CpG models, often incorporating sites located far from the gene encoding the protein, and can capture systemic or downstream biological processes that may not be evident through traditional locus-specific methylation analyses [10]. This approach enables the indirect exploration of functional biology that may be distinct from or additive to canonical EWAS findings, particularly in cohorts lacking direct proteomic measurements. Motivated by this opportunity, we applied previously published methods to estimate epigenetic protein scores [10], which reflect circulating levels of 109 plasma proteins, to Yakutian and Russian populations – an approach not previously reported, to our knowledge. We then examine regional differences in these proteome surrogates between residents of Yakutia and Central Russia. We hypothesize that region-specific differences in these methylation-based proteome estimates will reflect biological processes previously implicated in the literature, thereby providing supportive evidence for using this approach as an indirect method to study protein levels when direct measurements are not available. Additionally, this DNA methylation – based protein analysis may not only highlight known biological processes but also uncover novel ones relevant to understanding regional aging and health disparities in these populations.

2. Methods

2.1. Study sample

This analysis used publicly available DNA methylation data downloaded from the NCBI Genome Expression Omnibus (GEO) from a cohort of 245 healthy participants recruited across two distinct regions of Russia between 2020 and 2022 (GSE234461). The Central Russian sample included 131 ethnically Russian participants (78 women and 53 men) from the Nizhny Novgorod, Vladimir, and Moscow regions. The Yakutian sample consisted of 114 indigenous Yakut (Sakha) individuals (63 women and 51 men) from Yakutsk and surrounding rural villages. All Yakut participants met strict ethnic criteria as they were required to have confirmed Yakut ancestry across three generations, with all ancestors born and residing within the Republic of Sakha. Participants with active chronic diseases, cancer diagnoses, and acute respiratory infections at the time of sampling were excluded. Female participants who were currently pregnant were also excluded. The demographic characteristics and recruitment protocols for this cohort have been previously described in detail [5].

2.2. DNA methylation analysis and protein levels

DNA methylation was measured using the Illumina Infinium MethylationEPIC BeadChip. DNA methylation IDAT files were accessed from GEO using accession number GSE234461

[5]. IDATs were imported into R for preprocessing using *minfi* [11]. We computed detection *P*-values relative to control probes and excluded probes with non-significant detection ($p > 0.01$) for 5% or more of the samples (124,062 probes removed). We preprocessed our data using functional normalization [12], adjusted for probe-type bias using the regression on correlated probes method [13], and used ComBat from the *sva* package to adjust for sample slide as a technical batch [14]. We then used the methscore function from the *ENmix* package to calculate epigenetic scores for 109 circulating protein biomarkers [10,15]. Methscore imputes any required CpGs that are missing using reference data before calculating epigenetic scores [15]. The degree of missingness of component CpG sites for each predictor is summarized in Table S1.

2.3. Primary statistical analysis

We used linear models to examine the relationship of Russian region (Yakutia versus Central) with the epigenetically-derived protein biomarkers. Models were adjusted for chronological age, sex, and DNA methylation-estimated leukocyte proportions (granulocytes, monocytes, CD4T, CD8T, NK, and B cells) from the Houseman algorithm [16]. We also performed a sensitivity analysis that did not include adjustments for leukocyte proportions to assess the robustness of our findings in the absence of cell composition adjustment. Several of the proteins explored in our analysis, such as PAPPA, PIGR, CRP, and LY9, are functionally related to immune processes and may be influenced by leukocyte composition. Although we primarily considered immune cell composition as a potential confounder, we recognize that it could also act as a mediator in the relationship between region and protein levels. Therefore, we included this sensitivity analysis to provide a more comprehensive assessment of our findings. Statistical significance in the main and sensitivity analysis was defined as a false discovery rate (FDR)-adjusted *P*-value < 0.05 using the Benjamini-Hochberg (BH) procedure to account for multiple testing. We then used the STRING database to perform protein pathway enrichment analyses of our statistically significant results from the main analysis and the sensitivity analysis [17]. STRING curates pathway data from two major databases: Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome. Given that our analysis included 109 DNA methylation – based protein surrogates, we first analyzed all 109 surrogates to generate a background list of enriched pathways. We then performed a separate enrichment analysis on the subset of proteins that were statistically significant in our main and sensitivity analyses, identifying pathways that were enriched beyond those observed in the background set.

2.4. Secondary mitotic clock analysis

Based on the results of our primary analysis, we evaluated the association between region of residence (Yakutia vs. Central Russia) and two epigenetic mitotic clocks, which are predictive of cancer risk: epiTOC2 and MiAge. The epiTOC2 biomarker, derived from 163 CpG sites, estimates the cumulative number of stem cell divisions (mitotic activity) within a tissue. Its acceleration has been linked to various cancers and pre-invasive lesions, including lung adenocarcinoma and carcinoma in situ [18]. Similarly, MiAge is based on 268 CpG sites and estimates the total number of lifetime cell divisions in a tissue. Studies comparing tumor and adjacent normal tissues have found that MiAge is consistently accelerated across 13 cancer types and is associated with poorer cancer survival outcomes

[19]. Both measures were calculated with the methscore function. Models were adjusted for chronological age, sex, and estimated leukocyte proportions. A sensitivity analysis was also conducted without adjusting for leukocyte proportions. Statistical significance was defined as an FDR – adjusted *P*-value < 0.05.

3. Results

Only one protein score, lysozyme C, had a percentage of missing required CpGs greater than 20% (Table S1). Figure S1 shows a histogram of missingness across all 109 protein scores and for the scores identified as statistically significant in our analyses. In models adjusted for chronological age, sex, and estimated leukocyte proportions, 18 proteins were significantly associated with Russian region (Table 1). Yakuts had significantly lower estimated levels of four proteins (Afamin, CD6, C5a, and GZMA) and significantly higher levels of the remaining 14 proteins (CCL17, CD163, FCER2, HGF, HGFA, HGF1, LTAB, MMP9, OSM, S100-A9, SKR3, Testican-2, TGFA, and THBS2). Sensitivity analysis models, not including leukocyte adjustments, demonstrated significant associations of 16 proteins with Russian region. 9 of these 16 proteins were observed in the main analysis and 7 were new (Table 1). The distinct associations in these sensitivity models included CLEC11Ae1, CLEC11Ae2, CRTAM, ENPP7, FGF21, Granulysin (GNLY), and TNFRSF17. Figure 1 depicts volcano plots of associations from the main analysis adjusted for leukocyte proportions (Figure 1(A)) and sensitivity analysis with no adjustment for leukocyte proportions (Figure 1(B)) models. The STRING database produced both KEGG pathway and Reactome pathway enrichment results. Compared to the background list, significant proteins from our analysis demonstrated statistically significant enrichment for MET receptor activation (FDR *P*-value = 0.03) in Reactome pathways and of phosphatidylinositol 3-kinase (PI3K)/protein kinase B (Akt) signaling (FDR *P*-value = 0.01) in the KEGG pathway enrichment results (Table 2).

Figure 2(A) displays a network map of protein-protein associations derived from experimental data, co-expression patterns, co-occurrence, and other interaction sources queried via the STRING database. The thickness of the connecting lines reflects the strength of evidence supporting each interaction. Noteworthy associations include those between LTB and CD6, as well as Testican-2 (SPOCK2); HGFA and its connections with HGF, OSM, and MMP9; and GNLY with GZMA. Figure 2(B) presents a heat map illustrating gene co-expression corresponding to significantly associated proteins. The strongest co-expression relationships were observed between genes related to GZMA and GNLY, CD6 and SPOCK2, and CD6 and LTB.

In the secondary mitotic clock analysis, Yakuts had lower estimated mitotic stem cell division in both epiTOC2 (β = −267, 95%CI: −414, −120, FDR *P*-value = 0.001) and MiAge (β = −12, 95%CI: −30, 6, FDR *P*-value = 0.20), but only the epiTOC2 results met the threshold for statistical significance. Neither epiTOC2 (β = −15, 95%CI: −192, −161, FDR *P*-value = 0.87) nor MiAge (β = 20, 95%CI: −3, 43, FDR *P*-value = 0.10) were associated with Russian region in models not adjusted for estimated leukocyte proportions.

4. Discussion

In this cross-sectional analysis of Yakutia and Central Russian residents, we examined DNA methylation-based surrogates for circulating proteins to explore biological processes that may underlie regional differences in adaptation to their respective environments. In our main analysis, we identified 18 proteins that were significantly associated with Russian region. 7 additional proteins were identified in our sensitivity analysis that adjusted for estimated leukocyte proportions. There was evidence of known protein-protein interactions and expression of genes associated with proteins from our STRING database analyses. A few proteins were previously identified in GWAS analyses (CD163, FCER2, GZMA, LTA, and LTB), and none mapped to genes identified as differentially methylated in previous EWAS analyses, suggesting that the methylation-based protein approach may offer complementary value by uncovering novel aspects of physiological processes. Again, this may reflect the nature of methylation-based protein scores, which are derived from multi-CpG models that often incorporate sites located far from the gene encoding the target protein. As a result, they can capture systemic or downstream biological processes that may not be evident through traditional locus-specific methylation analyses. Pathway enrichment analysis demonstrated enrichment of biological processes relevant to energy metabolism and thermoregulation including, PI3K-Akt signaling and MET receptor activation.

We conducted analyses both with and without adjusting for estimated leukocyte proportions, recognizing that leukocytes can be considered either confounders or mediators depending on the research perspective and specific questions being asked. Of note, we identified immune-related proteins in both models with and without leukocyte adjustments. Immune-related proteins such as CD163, an anti-inflammatory receptor expressed on macrophages that facilitates the clearance of hemoglobin complexes from circulation [20]; FCER2, a B-cell-specific surface antigen [21]; and LTA/B, which are involved in the formation of secondary lymphoid structures like lymph nodes [22], were among those linked to differentially expressed genes in previous GWAS [4]. Further exploration of these protein relationships is warranted, particularly considering our findings of well-established protein-protein interactions in the network map and corresponding gene co-expression patterns involving proteins central to immune cell function. For example, GZMA is a serine protease specific to T cells and natural killer cells. It is believed to be a shared component required for the destruction of target cells by cytotoxic T lymphocytes and natural killer cells [23]. Granulysin, another protein found in the cytotoxic granules of T cells and released upon antigen stimulation [24], showed strong interaction and co-expression with GZMA. CD6, a co-receptor and adhesion molecule on T cells, helps regulate immune responses by stabilizing cell-to-cell contact and influencing T cell activation and signaling [25]. CD6 also demonstrated strong associations with Testican-2 (SPOCK2), a protein that has been implicated in immune protection [26]. Further research is needed to investigate how immune-related proteins may contribute to regional differences between Yakutian residents and Central Russians, particularly in the context of environmental adaptation.

With respect to our formal pathway analysis results, the PI3K-Akt signaling pathway plays a critical role in maintaining metabolic homeostasis. PI3K is involved in a broad range of cellular processes, including cell growth, survival, and metabolism. Downstream

of PI3K, AKT serves as a key component of the insulin signaling pathway, where it regulates proteins that control lipid and glucose metabolism [27]. Together, this pathway coordinates essential metabolic functions and may be particularly relevant in understanding physiological adaptations to environmental stressors. For example, the PI3K-Akt signaling pathway plays a role in thermoregulation with prior work implicating activation PI3K signaling as a response to therapeutic hypothermia [28]. The enrichment of pathways related to PI3K signaling among proteins elevated in Yakutia residents is biologically plausible, as colder temperatures in this region may require greater metabolic (e.g., muscular, mitochondrial) and PI3K-related activity to support increased thermoregulatory demands.

Five proteins (OSM, HGF, THBS2, TGFA, and FGF21) were identified as mapping to the PI3K-Akt signaling pathway. OSM is a secreted cytokine and growth regulator that has been shown to inhibit the proliferation of several tumor cell lines [29]. It also contributes to obesity-related inflammation and may play a role in the development of insulin resistance [30]. HGF binds to the hepatocyte growth factor receptor and regulates cell growth, motility, and morphogenesis across a variety of tissues. Like OSM, HGF has been implicated in metabolic processes and insulin resistance [31,32]. FGF21 is a secreted endocrine factor that acts as a key regulator of metabolism [33]. It has also been shown to influence thermoregulation independently of adipose tissue, particularly in the context of bacterial inflammation [34]. TGFA serves as a ligand for the epidermal growth factor receptor, activating signaling pathways involved in cell proliferation, differentiation, and development [35]. THBS2 facilitates cell-to-cell and cell-to-matrix interactions and has been identified as a potent inhibitor of tumor growth and angiogenesis [36]. Compared to OSM, HGF, and FGF21, the evidence linking TGFA and THBS2 to metabolic regulation and thermoregulation is limited.

MET receptor signaling plays a key role in cellular energy metabolism and can activate several downstream pathways, including the PI3K-Akt signaling cascade. These pathways are particularly relevant in the context of cancer progression, where they have been implicated in tumor proliferation and metastasis [37–39]. HGF and HGFA were the two proteins identified as part of MET receptor signaling. HGF has been associated with promoting cancer cell proliferation, migration, invasion, and survival, as well as contributing to resistance against various therapies [40]. Similarly, HGFA has been found to be elevated in cancer and within cancer cells [41]. Although not implicated in MET receptor signaling, MMP9, known for its role in degrading the extracellular matrix during normal physiological remodeling and in pathological processes such as cancer metastasis [42,43], was found to have interactions with HGF, HGFA, and OSM in our network map. Given this background, we examined associations between region of residence and two DNA methylation – based measures of mitotic cell divisions that are linked to cancer risk: epiTOC2 and MiAge [18,19]. Our results showed that individuals from Yakutia had lower estimated rates of cell divisions, with the epiTOC2 measure reaching statistical significance. As our analysis was based on leukocyte DNA, these findings pertain specifically to blood-based cancer risks. Previous research has shown that rates of several malignancies are higher among indigenous populations compared to Central Russians, likely due to harsher environmental exposures and delays in healthcare access [44]. However, we were unable to find specific data on the incidence of leukemias and lymphomas. Our finding of potentially lower mitotic activity,

which may indicate reduced blood-based cancer risk among Yakuts, is speculative at best and warrants further investigation.

Our study has several strengths, including the use of validated DNA methylation-based scores for circulating serum protein levels to explore biological differences between residents of Yakutia and Central Russia. However, several limitations should be acknowledged. First, while these methylation-based protein scores are novel and informative, they may not fully reflect directly measured protein levels. Moreover, the protein scores are unitless and do not reflect absolute protein concentrations, which limits direct biological interpretation of effect sizes. Although higher scores correspond to higher predicted protein levels, the magnitude of these differences cannot be directly translated into physiological terms. Therefore, our findings should be confirmed using direct protein measurements. Still, the consistency of our results with pathways identified in prior GWAS and EWAS lends credibility to our approach. Furthermore, our findings represent associations and there remains a need for mechanistic experimental work, when possible, to truly understand these biological and physiological relationships. Second, we relied on a publicly available dataset with limited information on participant demographics and lifestyle factors, which may also contribute to regional differences. As such, residual confounding cannot be entirely ruled out. Similarly, we did not have access to genetic background differences that could contribute to ethnicity-related confounding. This should be an important area of focus in future research to fully characterize regional protein differences. Fourth, our cross-sectional design captures a single timepoint, limiting our ability to assess temporal trends or causal relationships. Thus, generalizability to other time periods may be limited. Fifth, it is unclear from the original manuscript whether samples from Yakutia and Central Russia were collected in the same season. Without this information, we cannot rule out the impact of seasonal effects on both DNA methylation and protein expression. Finally, some CpGs required for calculating protein scores were missing. To address this, methscore imputed the missing values using reference data prior to score computation. While we cannot completely rule out bias introduced by this imputation, we are reassured by the low degree of missingness: for 108 of 109 proteins, fewer than 20% of required CpGs were missing, and for proteins identified as statistically significant, missingness was less than 12%. Again, confirming results with direct protein measurements would further strengthen these findings. Despite these constraints, our study demonstrates the utility of applying novel epigenetic tools in population research and provides a framework that can inform future investigations in similar contexts.

5. Conclusion

In conclusion, we identified differences in DNA methylation – based estimates of circulating protein levels between residents of Yakutia and Central Russia. These differences were enriched for biological pathways that align with previously published GWAS and EWAS findings related to metabolism, such as PI3K-Akt signaling, which may help explain how these populations have adapted to their respective climates. Our findings also demonstrate the potential utility of using methylation-derived protein estimates as an initial approach to investigate proteomic patterns relevant to environmental adaptation and other research

questions. Nonetheless, there remains a critical need to validate these results with directly measured protein levels and within experimental frameworks.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the authors (Kalyakulina A. et al.) of the initial study [5] for making their data available.

Funding

Jamaji C. Nwanaji-Enwerem and Andres Cardenas are supported by National Institutes of Health grant [R01ES031259]. JCN was also supported by the Leonard Davis Institute of Health Economics. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Data availability statement

The datasets analyzed in the current study are available from the NCBI Genome Expression Omnibus (GEO) website (GSE234461).

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Snodgrass JJ, Leonard WR, Sorensen MV, et al. The influence of basal metabolic rate on blood pressure among indigenous Siberians. Am J Phys Anthropol. 2008;137(2):145–155. doi: 10.1002/ajpa.20851 [PubMed: 18470897]
2. Bjerregaard P, Dewailly E, Young TK, et al. Blood pressure among the inuit (Eskimo) populations in the Arctic. Scand J Public Health. 2003;31(2):92–99. doi: 10.1080/14034940210133924 [PubMed: 12745758]
3. Leonard WR, Snodgrass JJ, Sorensen MV. Metabolic adaptation in indigenous Siberian populations. Annu Rev Anthropol. 2005;34 (1):451–471. doi: 10.1146/annurev.anthro.34.081804.120558
4. Cardona A, Pagani L, Antao T, et al. Genome-wide analysis of cold adaptation in indigenous Siberian populations. PLOS ONE. 2014;9 (5):e98076. doi: 10.1371/journal.pone.0098076 [PubMed: 24847810] •• This is an important prior GWAS in the population of interest.
5. Kalyakulina A, Yusipov I, Kondakova E, et al. Epigenetics of the far northern Yakutian population. Clin Epigenet. 2023;15(1):189. doi: 10.1186/s13148-023-01600-y •• This is an important prior EWAS in the population of interest
6. Sladewski TE, Krementsova EB, Trybus KM. Myosin Vc is specialized for transport on a secretory superhighway. Curr Biol. 2016;26 (16):2202–2207. doi: 10.1016/j.cub.2016.06.029 [PubMed: 27498562]
7. Jacobs DT, Weigert R, Grode KD, et al. Myosin VC is a molecular motor that functions in secretory granule trafficking. Mol Biol Cell. 2009;20(21):4471–4488. doi: 10.1091/mbc.e08-08-0865 [PubMed: 19741097]
8. Kusters CDJ, Horvath S. Quantification of epigenetic aging in public health. Annu Rev Public Health. 2025;46(1):91–110. doi: 10.1146/annurev-publhealth-060222-015657 [PubMed: 39681336]

9. Cominetti O, Dayon L. Unravelling disease complexity: integrative analysis of multi-omic data in clinical research. Expert Rev Proteomics. 2025;22(4):149–162. doi: 10.1080/14789450.2025.2491357 [PubMed: 40207843]
10. Gadd DA, Hillary RF, McCartney DL, et al. Epigenetic scores for the circulating proteome as tools for disease prediction. Elife. 2022;11:e71802. doi: 10.7554/eLife.71802 [PubMed: 35023833] •• This paper describes the DNA methylation-based protein scores.
11. 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):1363–1369. doi: 10.1093/bioinformatics/btu049 [PubMed: 24478339]
12. Fortin J-P, Labbe A, Lemire M, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15(11):503. doi: 10.1186/s13059-014-0503-2 [PubMed: 25599564]
13. Niu L, Xu Z, Taylor JA. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. Bioinformatics. 2016;32 (17):2659–2663. doi: 10.1093/bioinformatics/btw285 [PubMed: 27153672]
14. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in micro-array expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118–127. doi: 10.1093/biostatistics/kxj037 [PubMed: 16632515]
15. Xu Z, Niu L, Kresovich JK, et al. Methscore: a comprehensive R function for DNA methylation-based health predictors. Bioinformatics. 2024;40 (5):btae302. doi: 10.1093/bioinformatics/btae302 •• This paper describes the methscore R function that can be used to compute the protein scores.
16. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinf. 2012;13(1):86. doi: 10.1186/1471-2105-13-86
17. Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(D1):D638–D646. [PubMed: 36370105] •• This paper describes the STRING database used for protein pathway enrichment analyses.
18. Teschendorff AE. A comparison of epigenetic mitotic-like clocks for cancer risk prediction. Genome Med. 2020;12(1):56. doi: 10.1186/s13073-020-00752-3 [PubMed: 32580750] • This paper describes a mitotic clock explored in the secondary analysis.
19. Youn A, Wang S. The MiAge calculator: a DNA methylation-based mitotic age calculator of human tissue types. Epigenetics. 2018;13 (2):192–206. doi: 10.1080/15592294.2017.1389361 [PubMed: 29160179] • This paper describes a mitotic clock explored in the secondary analysis.
20. Fabriek BO, Dijkstra CD, van den Berg TK. The macrophage scavenger receptor CD163. Immunobiology. 2005;210(2):153–160. doi: 10.1016/j.imbio.2005.05.010 [PubMed: 16164022]
21. Aranda CJ, Gonzalez-Kozlova E, Saunders SP, et al. IgG memory B cells expressing IL4R and FCER2 are associated with atopic diseases. Allergy. 2023;78(3):752–766. doi: 10.1111/all.15601 [PubMed: 36445014]
22. McDevitt H, Munson S, Ettinger R, et al. Multiple roles for tumor necrosis factor-α and lymphotoxin α/β in immunity and autoimmunity. Arthritis Res. 2002;4(Suppl 3):S141–S152. doi: 10.1186/ar570 [PubMed: 12110133]
23. Lieberman J Granzyme A activates another way to die. Immunol Rev. 2010;235(1):93–104. doi: 10.1111/j.0105-2896.2010.00902.x [PubMed: 20536557]
24. Clayberger C, Krensky AM. Granulysin. Curr Opin Immunol. 2003;15 (5):560–565. doi: 10.1016/S0952-7915(03)00097-9 [PubMed: 14499265]
25. Henriques SN, Oliveira L, Santos RF, et al. CD6-mediated inhibition of T cell activation via modulation of Ras. Cell Commun Signal. 2022;20(1):184. doi: 10.1186/s12964-022-00998-x [PubMed: 36414966]
26. Ahn N, Kim W-J, Kim N, et al. The interferon-inducible proteoglycan Testican-2/SPOCK2 functions as a protective barrier against virus infection of lung epithelial cells. J Virol. 2019;93(20):e00662–19. doi: 10.1128/JVI.00662-19 [PubMed: 31341044]
27. Savova MS, Mihaylova LV, Tews D, et al. Targeting PI3K/AKT signaling pathway in obesity. Biomed Pharmacother. 2023;159:114244. doi: 10.1016/j.biopha.2023.114244 [PubMed: 36638594]

28. Li Y, Chen Y, Yu P, et al. Mild therapeutic hypothermic protection activates the PI3K/AKT signaling pathway to inhibit TRPM7 and suppress ferroptosis induced by myocardial ischemia-reperfusion injury. Mol Med Rep. 2024;30(6):220. doi: 10.3892/mmr.2024.13345 [PubMed: 39364741]
29. Wolf CL, Pruett C, Lighter D, et al. The clinical relevance of OSM in inflammatory diseases: a comprehensive review. Front Immunol. 2023;14. doi: 10.3389/fimmu.2023.1239732
30. Piquer-Garcia I, Campderros L, Taxerås SD, et al. A role for oncostatin M in the impairment of glucose homeostasis in obesity. J Clin Endocrinol Metab. 2020;105(3):e337–348. [PubMed: 31606738]
31. Perdomo G, Martinez-Brocca MA, Bhatt BA, et al. Hepatocyte growth factor is a novel stimulator of glucose uptake and metabolism in skeletal muscle cells*. J Biol Chem. 2008;283 (20):13700–13706. doi: 10.1074/jbc.M707551200 [PubMed: 18362143]
32. Oliveira AG, Araújo TG, Carvalho B de M, et al. The role of hepatocyte growth factor (HGF) in insulin resistance and diabetes. Front Endocrinol (lausanne). 2018;9:503. doi: 10.3389/fendo.2018.00503 [PubMed: 30214428]
33. Gliniak CM, Gordillo R, Youm Y-H, et al. FGF21 promotes longevity in diet-induced obesity through metabolic benefits independent of growth suppression. Cell Metab. 2025;37(7):1547–1567.e6. doi: 10.1016/j.cmet.2025.05.011 [PubMed: 40527315]
34. Huen SC, Wang A, Feola K, et al. Hepatic FGF21 preserves thermoregulation and cardiovascular function during bacterial inflammation. J Exp Med. 2021;218(10):e20202151. doi: 10.1084/jem.20202151
35. Bascom CC, Sipes NJ, Coffey RJ, et al. Regulation of epithelial cell proliferation by transforming growth factors. J Cell Biochem. 1989;39(1):25–32. doi: 10.1002/jcb.240390104 [PubMed: 2654145]
36. Wang X, Zhang L, Li H, et al. THBS2 is a potential prognostic biomarker in colorectal cancer. Sci Rep. 2016;6(1):33366. doi: 10.1038/srep33366 [PubMed: 27632935]
37. Zhang Y, Xia M, Jin K, et al. Function of the c-Met receptor tyrosine kinase in carcinogenesis and associated therapeutic opportunities. Mol Cancer. 2018;17(1):45. doi: 10.1186/s12943-018-0796-y [PubMed: 29455668]
38. Osaki M, Oshimura M, Ito H. PI3K-Akt pathway: its functions and alterations in human cancer. Apoptosis. 2004;9(6):667–676. doi: 10.1023/B:APPT.0000045801.15585.dd [PubMed: 15505410]
39. Fontana F, Giannitti G, Marchesi S, et al. The PI3K/Akt pathway and glucose metabolism: a dangerous liaison in cancer. Int J Biol Sci. 2024;20(8):3113–3125. doi: 10.7150/ijbs.89942 [PubMed: 38904014]
40. Owusu BY, Galemmo R, Janetka J, et al. Hepatocyte growth factor, a key tumor-promoting factor in the tumor microenvironment. Cancers (Basel). 2017;9(4):35. doi: 10.3390/cancers9040035 [PubMed: 28420162]
41. Jiang WG. Hepatocyte growth factor and the hepatocyte growth factor receptor signalling complex as targets in cancer therapies. Curr Oncol. 2007;14(2):66–69. doi: 10.3747/co.2007.108 [PubMed: 17576468]
42. Jiang H, Li H. Prognostic values of tumoral MMP2 and MMP9 overexpression in breast cancer: a systematic review and meta-analysis. BMC Cancer. 2021;21(1):149. doi: 10.1186/s12885-021-07860-2 [PubMed: 33568081]
43. Augoff K, Hryniewicz-Jankowska A, Tabola R, et al. MMP9: a tough target for targeted therapy for cancer. Cancers (basel). 2022;14(7):1847. doi: 10.3390/cancers14071847 [PubMed: 35406619]
44. Burtseva TE, Uvarova TE, Tomsky MI, et al. The health of populations living in the indigenous minority settlements of northern Yakutia. Int J Circumpolar Health. 2014;73(1):25758. doi: 10.3402/ijch.v73.25758

Article highlights

- Methylation-based estimators of circulating protein levels were used to indirectly assess proteomic differences between Yakutian and Central Russian residents.
- In models adjusted for and unadjusted for estimated leukocyte proportions, we identified twenty-five unique protein surrogates that showed regional variation.
- The twenty-five proteins demonstrated enrichment in pathways of PI3K-Akt signaling and MET receptor activation.
- To the best of our knowledge, some protein surrogates were linked to genes identified in prior GWAS analyses but not EWAS analyses, suggesting complementary insights from proteomic estimation.
- Findings support the utility of methylation-derived proteomic proxies for studying regional differences possibly related to adaptation in population-level research.

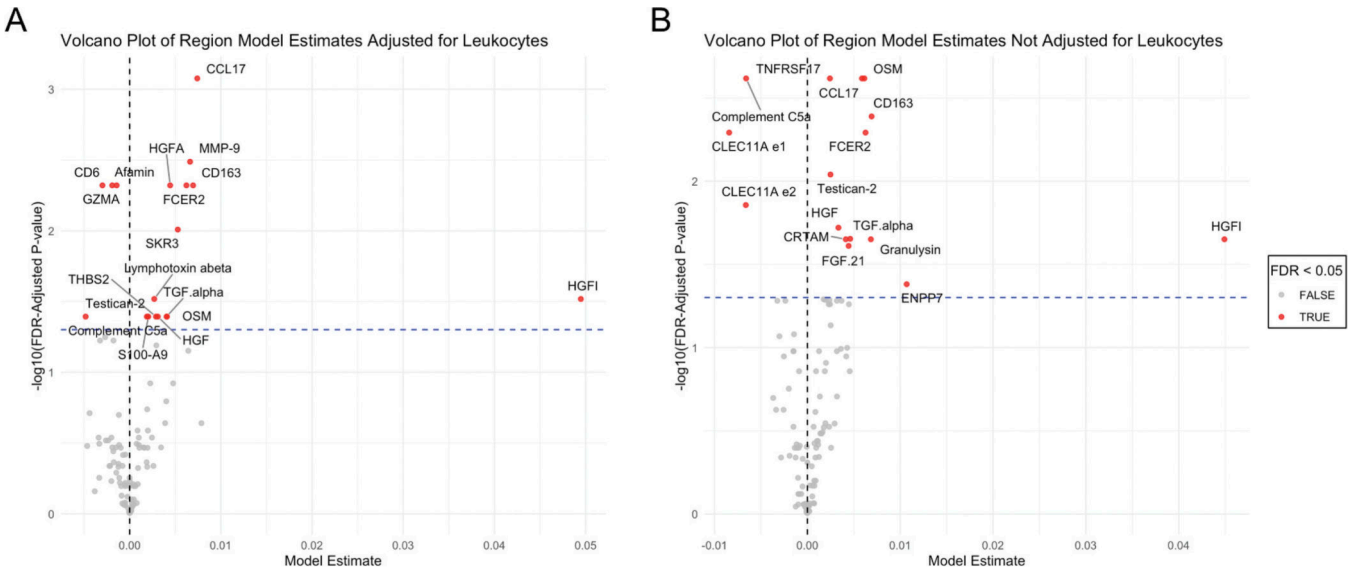

Figure 1.

Regional Model associations. Figure 1(A,B) presents the volcano plots examining the relationships of region with DNA methylation-estimated protein levels in models adjusted [A] and unadjusted [B] for estimated leukocyte proportions. Both models are adjusted for chronological age and sex.

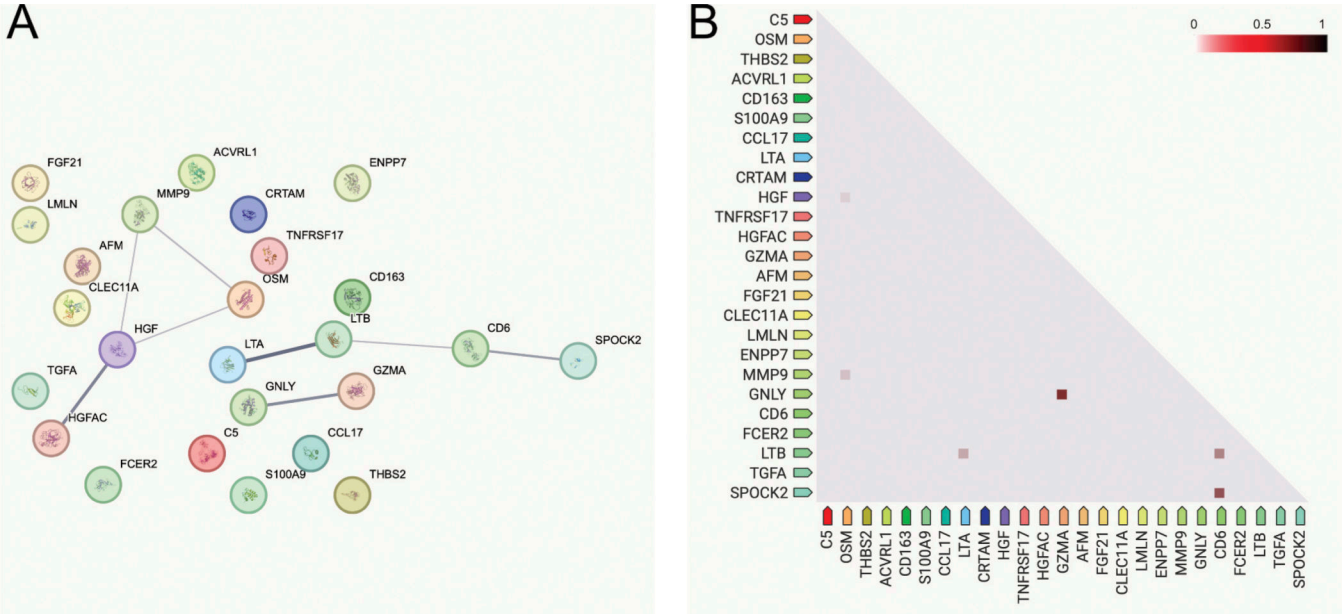

Figure 2.

Protein interaction and gene co-expression networks. Figure 2(A) presents a network map of protein-protein associations generated using the STRING database, incorporating data from experimental studies, co-expression, co-occurrence, and other interaction sources. Line thickness indicates the strength of supporting evidence. Figure 2(B) depicts a heat map of gene co-expression among genes corresponding to significantly associated proteins. Darker colors indicate greater co-expression.

Table 1.

Statistically significant associations of region with DNA methylation-predicted protein levels.

	Leukocyte-Adjusted Model			Sensitivity Adjusted Model w/o Leukocytes		
Protein	**Estimate (95% CI)**	***P*-value**	**FDR-Adjusted P-value**	**Estimate (95% CI)**	***P*-value**	**FDR-Adjusted P-value**
Afamin	−0.001 (−0.002, −0.001)	0.0002	0.005			
CCL17	0.01 (0.004, 0.01)	0.00001	0.001	0.01 (0.003, 0.01)	0.0001	0.002
CD163	0.01 (0.003, 0.01)	0.0004	0.005	0.01 (0.003, 0.01)	0.0002	0.004
CD6	−0.003 (−0.005, −0.001)	0.0002	0.005			
CLEC11A e1				−0.01 (−0.01, −0.004)	0.0003	0.005
CLEC11A e2				−0.01 (−0.01, −0.003)	0.0011	0.014
Complement C5a	−0.005 (−0.01, −0.001)	0.0058	0.04	−0.01 (−0.01, −0.003)	0.0001	0.002
CRTAM				0.004 (0.001, 0.01)	0.0029	0.022
ENPP7				0.01 (0.003, 0.02)	0.0061	0.042
FCER2	0.01 (0.003, 0.01)	0.0003	0.005	0.01 (0.003, 0.01)	0.0003	0.005
FGF21				0.004 (0.001, 0.01)	0.0034	0.024
Granulysin				0.01 (0.002, 0.01)	0.0025	0.022
GZMA	−0.002 (−0.003, −0.001)	0.0003	0.005			
HGF	0.003 (0.001, 0.01)	0.0067	0.04	0.003 (0.001, 0.01)	0.0017	0.019
HGFA	0.004 (0.002, 0.01)	0.0003	0.005			
HGF1	0.05 (0.02, 0.08)	0.003	0.03	0.04 (0.016, 0.07)	0.0029	0.022
Lymphotoxin abeta	0.003 (0.001, 0)	0.0031	0.03			
MMP-9	0.01 (0.003, 0.01)	0.0001	0.003			
OSM	0.004 (0.001, 0.01)	0.0065	0.04	0.01 (0.003, 0.01)	0.0001	0.002
S100-A9	0.002 (0.001, 0.003)	0.0045	0.04			
SKR3	0.01 (0.002, 0.01)	0.0008	0.01			
Testican-2	0.002 (0.001, 0.003)	0.0057	0.04	0.002 (0.001, 0.004)	0.0007	0.009
TGF.alpha	0.004 (0.001, 0.01)	0.0061	0.04	0.005 (0.002, 0.01)	0.0022	0.022
THBS2	0.003 (0.001, 0.005)	0.0062	0.04			
TNFRSF17				0.002 (0.001, 0.004)	0.00004	0.002

Model estimates are for Yakutian participants (Central Russia participants are the reference group).Model Adjustments: chronological age and sex. Leukocyte-adjusted models include additional adjustment for DNAm-estimated leukocyte proportions.

Epigenomics. Author manuscript; available in PMC 2026 January 16.

Table 2.

STRING protein pathway enrichment results.

Enrichment Term	Term ID	Observed Protein Count	Matching Proteins	FDR pvalue	Enrichment Database
PI3K-Akt Signaling	hsa04151	5	OSM, HGF, THBS2, TGFA, FGF21	0.01	KEGG
MET Receptor Activation	hsa6806942	2	HGF, HGFA	0.03	Reactome