

## ABSTRACT

TEIXEIRA, Carolynne Martins, Federal University of Viçosa, March 2025. **Interaction between omics biomarkers, dietary intake, and methyl-donor nutrients in the incidence of obesity, BMI trajectory, and MASLD: analysis of the longitudinal CUME and FLiO studies.** Advisors: Helen Hermana Miranda Hermsdorff, Marian de los Ángeles Zulet Alzórriz. Co-advisor: Josefina Bressan

Obesity is a chronic and multifactorial condition associated with an increased prevalence of metabolic dysfunction-associated steatotic liver disease (MASLD). The development of these diseases results from the interaction between genetic factors, insulin resistance, inadequate dietary habits, and physical inactivity, which compromise metabolic homeostasis and promote fat accumulation in both adipocytes and hepatocytes. Despite the various therapeutic strategies available, many approaches to obesity treatment have shown limited efficacy. In the context of MASLD, caloric restriction for weight loss induction remains the primary recommendation to prevent disease progression. However, evidence indicates that individuals responses to the same dietary intervention vary significantly. Genetic and epigenetic factors, modulated by the intake of methyl-donor nutrients (MetD), may contribute to this variability. In this scenario, the identification of omics biomarkers and the potential modulatory effect of diet emerge as promising strategies for the treatment of obesity and MASLD, in addition to representing an alternative to more invasive or costly interventions. Biomarkers play a fundamental role in elucidating the cellular mechanisms underlying these diseases, as well as in prediction, prognosis, and the development of personalized therapies. Thus, the present study aims to identify omics biomarkers associated with the incidence of obesity, BMI trajectory, and MASLD, in addition to evaluating the role of diet as an effect modifier in the interactions between biomarkers and the aforementioned outcomes. To achieve this objective, two prospective studies will be analyzed: the FLiO Study, a randomized controlled clinical trial comparing the effect of two hypocaloric diets in individuals with obesity and MASLD diagnosed by ultrasonography; and the CUME study, a prospective cohort investigating the influence of the Brazilian dietary pattern on the occurrence of non-communicable chronic diseases. Dietary intake will be assessed using a previously validated questionnaire, and MetD intake will be estimated based on the United States Department of Agriculture (USDA) food composition table. Obesity and MASLD will be diagnosed based on BMI ( $\geq 30$  kg/m<sup>2</sup>) and the presence of at least 5% hepatic fat,

respectively. Outcomes, as well as anthropometric and biochemical data, will be collected at baseline and throughout follow-up. Oral mucosa cell collection will be performed for genetic analysis of single nucleotide polymorphisms (SNPs) using q-PCR. Metabolomics analysis will be conducted using ultra-high-performance liquid chromatography coupled with mass spectrometry (UHPLC-MS) to identify metabolites in urine samples. Serum analyses will follow previously established protocols. Normality of variables will be assessed using the Shapiro-Wilk test. For comparisons of continuous and categorical variables, Student's t-test or Mann-Whitney and ANOVA or Kruskal-Wallis tests will be applied, as appropriate. Categorical variables will be analyzed using Pearson's chi-square test, with results expressed as absolute and relative frequencies. Linear, logistic, or Poisson regression models will be employed according to the statistical assumptions of each analysis. The area under the receiver operating characteristic (ROC) curve will be used to evaluate the diagnostic capacity of the analyzed omics biomarkers. The Cox proportional hazards analysis will estimate the independent association between exposure variables and the incidence of obesity, expressed through hazard ratios (HR) and respective 95% confidence intervals (CI95%). Latent Class Growth Modeling (LCGM) will be applied to examine weight trajectory and the degree of steatosis, allowing for the classification of individuals into distinct groups according to their individual response patterns. The MetD intake score will be calculated based on the daily intake of these nutrients, categorized as low (value 0) or high (value 1) according to the 75th percentile of the sample. The Genetic Risk Score (GRS) will be estimated under the assumption that each SNP contributes independently and equally to the investigated outcomes. The Hardy-Weinberg equilibrium will be tested for each gene ( $p > 0.05$ ). Identification of confounding variables will be conducted using a Directed Acyclic Graph (DAG), developed on the DAGitty platform (<http://www.dagitty.net/>). The presence of interactions will be tested using an interaction term in the adjusted models. All statistical analyses will be performed using STATA software, version 13.0, with a significance level of 5%.

**Keywords:** biomarkers, GRS, metabolomics, obesity, steatosis

## RESUMO

TEIXEIRA, Carolynne Martins, Universidade Federal de Viçosa, Março, 2025. **Interação entre biomarcadores ômicos, ingestão alimentar e nutrientes doadores de metil na incidência de obesidade, trajetória do IMC e MASLD: análise dos estudos longitudinais CUME e FLiO.** Orientadora: Helen Hermana Miranda Hermsdorff, Marian de los Ángeles Zulet Alzórriz. Coorientadora: Josefina Bressan

A obesidade é uma condição crônica e multifatorial associada ao aumento da prevalência da doença hepática esteatótica associada à disfunção metabólica (MASLD). O desenvolvimento dessas doenças decorre da interação entre fatores genéticos, resistência à insulina, hábitos alimentares inadequados e sedentarismo, os quais comprometem a homeostase metabólica e favorecem o acúmulo de gordura tanto nos adipócitos quanto nos hepatócitos. Apesar das diversas estratégias terapêuticas disponíveis, muitas abordagens para o tratamento da obesidade demonstraram eficácia limitada. No contexto da MASLD, a restrição calórica para a indução da perda ponderal permanece a principal recomendação para evitar a progressão da doença. No entanto, evidências indicam que a resposta dos indivíduos a uma mesma intervenção dietética varia significativamente. Fatores genéticos e epigenéticos, modulados pela ingestão de nutrientes doadores de metil (MetD), podem contribuir para essa variabilidade. Nesse cenário, a identificação de biomarcadores ômicos e o potencial efeito modulador da dieta emergem como estratégias promissoras para o tratamento da obesidade e da MASLD, além de representar uma alternativa às intervenções mais invasivas ou dispendiosas. Os biomarcadores desempenham um papel fundamental na elucidação de mecanismos celulares subjacentes a essas doenças, bem como na predição, prognóstico e desenvolvimento de terapias personalizadas. Diante desse contexto, o presente estudo tem como objetivo identificar biomarcadores ômicos associados à incidência de obesidade, à trajetória do IMC e à MASLD, além de avaliar o papel da dieta como modificador de efeito nas interações entre biomarcadores e os desfechos mencionados. Para atingir esse objetivo, serão analisados dois estudos prospectivos: o FLiO *Study*, um ensaio clínico randomizado e controlado que compara o efeito de duas dietas hipocalóricas em indivíduos com obesidade e MASLD diagnosticada por ultrassonografia; e o estudo CUME, uma coorte prospectiva que investiga a influência do padrão alimentar brasileiro na ocorrência de doenças crônicas não transmissíveis. O consumo alimentar será avaliado por meio de um questionário previamente validado, e a ingestão de MetD será estimada a partir da tabela

do Departamento de Agricultura dos Estados Unidos (USDA). A obesidade e a MASLD serão diagnosticadas com base no IMC ( $\geq 30$  kg/m<sup>2</sup>) e na presença de pelo menos 5% de gordura hepática, respectivamente. Os desfechos, bem como os dados antropométricos e bioquímicos, serão coletados na linha de base e ao longo do acompanhamento. A coleta de células da mucosa oral será realizada para análise genética de polimorfismos de nucleotídeo único (SNPs) por meio de q-PCR. A metabolômica será conduzida utilizando cromatografia líquida de ultra-alta eficiência acoplada à espectrometria de massas (UHPLC-MS) para a identificação de metabólitos em amostras de urina. As análises séricas seguirão protocolos previamente estabelecidos. A normalidade das variáveis será avaliada pelo teste de Shapiro-Wilk. Para comparação de variáveis contínuas e categóricas, serão aplicados o teste t de Student ou Mann-Whitney e ANOVA ou Kruskal-Wallis, conforme a adequação. As variáveis categóricas serão analisadas pelo teste do qui-quadrado de Pearson, com resultados expressos em frequências absolutas e relativas. Modelos de regressão linear, logística ou de Poisson serão empregados conforme os pressupostos estatísticos de cada análise. A área sob a curva ROC (*Receiver Operating Characteristic*) será utilizada para avaliar a capacidade diagnóstica dos biomarcadores ômicos. A análise de riscos proporcionais de Cox estimará a associação independente entre as variáveis de exposição e a incidência de obesidade, expressa por meio das razões de risco (HR) e respectivos intervalos de confiança de 95% (IC95%). A Modelagem de Crescimento de Classe Latente (LCGM, do inglês *Latent Class Growth Modeling*) será aplicada para examinar a trajetória do peso e o grau de esteatose, permitindo a classificação dos indivíduos em grupos distintos de acordo com seus padrões de resposta individual. O escore de ingestão de MetD será calculado com base na ingestão diária desses nutrientes, sendo categorizado como baixo (valor 0) ou alto (valor 1) conforme o 75º percentil da amostra. O escore de risco genético (*Genetic Risk Score*, GRS) será estimado sob a suposição de que cada SNP contribui de maneira independente e equivalente para os desfechos investigados. O equilíbrio de Hardy-Weinberg será testado para cada gene ( $p > 0,05$ ). A identificação de variáveis de confusão será conduzida por meio de um Gráfico Acíclico Direcionado (DAG), elaborado na plataforma DAGitty (<http://www.dagitty.net/>). A presença de interações será testada por meio de um termo de interação nos modelos ajustados. Todas as análises estatísticas serão realizadas no software STATA, versão 13.0, adotando-se um nível de significância de 5%.

**Palavras-chave:** biomarcadores, GRS, metabolômica, obesidade, esteatose