Circulating osteocalcin as a bone-derived hormone is inversely correlated with body fat in patients with type 1 diabetes

Introduction

Type 1 diabetes is a chronic autoimmune disease that requires careful monitoring and management, particularly in patients undergoing insulin therapy. Advanced analysis of clinical data through data mining techniques offers a promising opportunity to better understand the physiological and behavioral characteristics of patients, with the goal of identifying hidden patterns, homogeneous subgroups, or correlations that may support personalized care.

In this study, we aim to apply unsupervised clustering methods to a dataset of patients with type 1 diabetes, with the objective of identifying clinically meaningful subgroups. For model construction, we selected key variables of anthropometric, metabolic, hormonal, and functional nature, collected from a reference clinical study. These include:

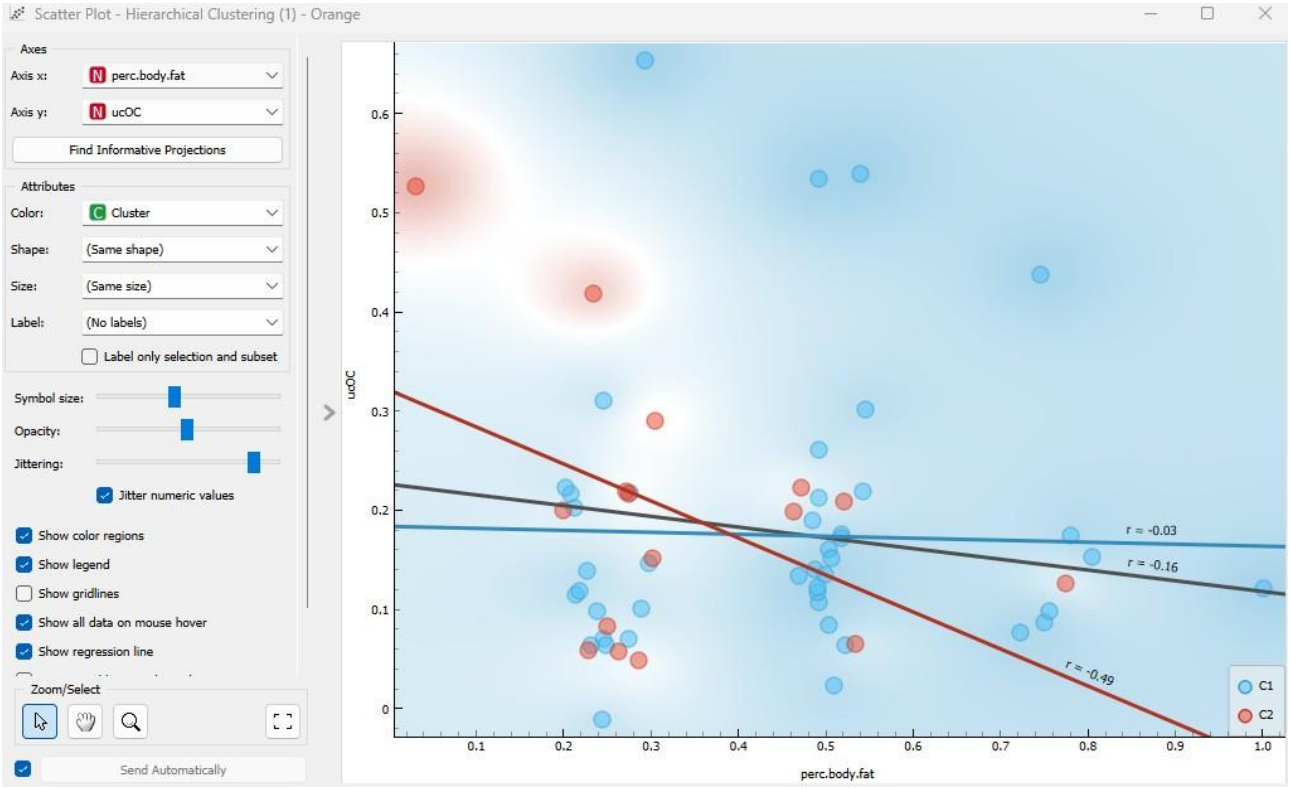
- Age and duration of diabetes, to assess the disease's progression over time;
- Body composition indicators such as body mass index (BMI), body fat percentage, and weight in kilograms;
- Insulin treatment parameters, including TDD (total daily insulin dose), basal dose, bolus dose, and type of insulin regimen (MDI/CSII);
- Metabolic markers such as HbA1c (glycated hemoglobin), eGFR (estimated glomerular filtration rate), and adiponectin;
- Muscle strength and mobility indicators, including grip strength, knee extension strength, and gait speed;
- Hormonal and bone health assessments, such as free testosterone (for male subjects), ucOC, and OC (osteocalcins);
- Skeletal muscle mass index (SMI) as an indicator of sarcopenia;
- Finally, patient sex and insulin regimen type were encoded in binary format for analysis.

By using these parameters, the goal is to explore the presence of clinically interpretable clusters within the studied population, providing new insights for patient classification, monitoring disease progression, and optimizing therapeutic protocols.

Analysis

For the purpose of creating the clusters, the following features were selected and used as active variables in the clustering process: body fat percentage, total osteocalcin (OC), undercarboxylated osteocalcin (ucOC), body mass index (BMI), sex, and age. All other available variables in the dataset were excluded from the clustering model and instead treated as meta-variables, to be used for later interpretation and validation of the identified clusters. This selection was driven by the study's focus on exploring the relationship between osteocalcin levels and body fat, while accounting for key demographic and physiological factors.

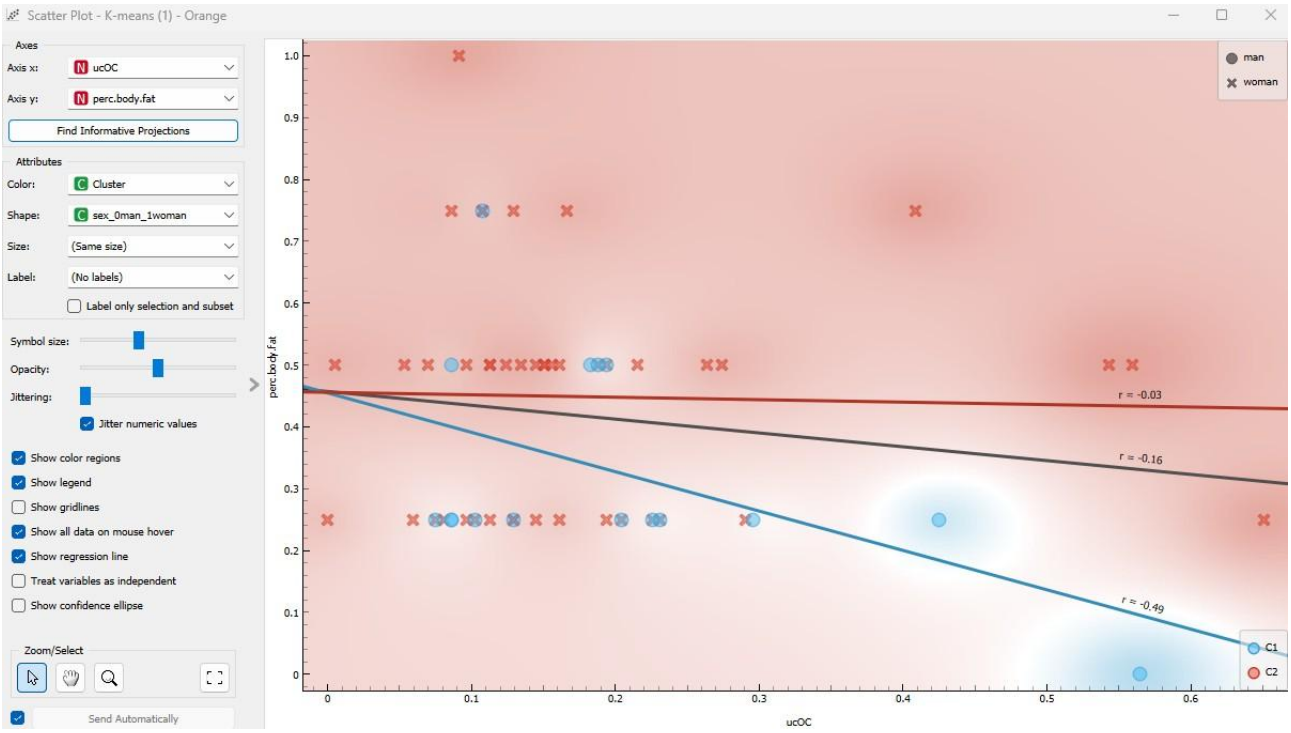
Preliminary observations showed that, when analyzing the entire sample of 67 patients as a single group, the inverse correlation between total osteocalcin (OC), undercarboxylated osteocalcin (ucOC), and body fat percentage (% body fat) was highly variable—especially among individuals with normal weight. This variability seems partly due to physiological sex-based differences affecting body composition and bone-hormonal metabolism. When the analysis was restricted to males with a BMI between 25 and 35, the correlation between OC and % body fat reached as low as -0.487, while in a comparable female subgroup it was only -0.094. These findings justify stratifying patients by sex in the unsupervised clustering to avoid masking clinically relevant patterns.



Hierarchical Clustering – Scatter Plot

To improve the internal consistency of the clusters and reduce the influence of extreme values, a BMI filter was applied, leading to the exclusion of 7 out of 67 patients (approximately 10%), all of whom had BMI values far outside the considered range (25 < BMI < 35). The analysis was therefore conducted on 60 patients, divided into two main clusters: Cluster **C1**, composed of female subjects, included 44 individuals (about 63% of the total), while Cluster **C2**, representing the male component, comprised the remaining 16 patients (27%). This stratification allowed for a segmentation more consistent with the patients' physiological characteristics and helped reveal specific patterns within more homogeneous subgroups.

Interestingly, applying k-Means clustering instead of hierarchical clustering yielded virtually identical results. The same subgroup divisions were observed, with minimal variation in cluster composition. This consistency between methods reinforces the reliability of the identified patient segmentation and suggests that the observed patterns are robust across different unsupervised clustering approaches.



K-means – Scatter Plot

Also in this case, the correlation within the two clusters—automatically separated by k-Means, which ultimately mirrored the sex-based division—showed clear differences between men and women. The inverse correlation between osteocalcin levels and body fat percentage remained strong in the male cluster, reaching approximately -0.49, while it was considerably weaker in the female cluster, around -0.03. These nearly identical values to those obtained with hierarchical clustering further confirm the sex-specific nature of the association and reinforce the validity of the segmentation approach.