IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Bioinformatics Analysis Identifies the Association Between Glucose Markers and Disease-Related Functions in Diabetes

Qi Chen

Shenzhen College of International Education, Huangganggongyuan St 1, Futian District, Shenzhen, Guangdong Province, 518000, China; CathQi06@gmail.com
Mentor: Dr. Xingjian Zhang

ABSTRACT: Diabetes, a chronic disease characterized by insufficient insulin production, affects approximately 422 million people globally. While researchers have achieved notable advancements in identifying new blood and tissue-based biomarkers related to diabetes, a more detailed and comprehensive investigation is needed. In this study, we conducted a systematic bioinformatics analysis using the Diabetes Complications Early Warning database to examine these associations. First, we identified the association of glucose-related markers with major markers related to blood lipid metabolism, renal function, and liver function. Second, we applied bioinformatics analysis to screen for major markers and diseases capable of categorizing subpopulations within the diabetic population and identified twelve blood markers and four disease markers. We discover that complete blood count (CBC) and blood chemistry (BC) biomarkers of different diabetic patients exhibit subtle changes in their performances. This may be attributed to their differential diabetic complications. Our study enhances the understanding of diabetes mechanisms and provides new avenues for developing precision medicine and therapeutic strategies.

KEYWORDS: Computational Biology and Bioinformatics; Other; Diabetes; Diabetic Subpopulations; Glucose-Related Markers.

Introduction

Diabetes is a metabolic disease characterized by the body's inability to produce sufficient insulin to regulate blood sugar levels. It may be caused by inadequate insulin production or inefficient insulin utilization.[1] Diabetes has become more prevalent in recent decades. In 2017, there were approximately 451 million adults with diabetes from around the world. By 2024, this figure is anticipated to increase to 693 million, comprising half the total population.[2] Diabetes can be categorized into type 1 and type 2. Type 1 diabetes, typically developed in childhood or adolescence, is caused by an autoimmune reaction. The autoimmune responses result in the pancreas producing negligible amounts of insulin or ceasing to produce it.[3] Conversely, type 2 diabetes tends to develop at an older age. However, there is a concerning trend of an increasing number of type 2 diabetic patients among teenagers and young adults.[4]

Regulating blood glucose levels is the primary approach in current diabetes treatments, such as insulin therapy[5] and lifestyle modifications.[6] By understanding how biomarkers affect glucose dynamics and the development of diabetes-related diseases, an integrated study of complete blood count (CBC) and blood chemistry (BC) markers can offer new perspectives on developing systemic treatment strategies. More specifically, investigating the connections between biomarkers and glucose levels, glucose dynamics, and patient age can enable us to learn about the effect of various biomarkers on different organs. Since diabetic patients are not just a homogenous population but potentially a composition of several subpopulations with distinct phenotypes, the categorization of subpopulations using blood markers and secondary diseases can help us identify specific subpopulations, improve models for predicting diabetes secondary diseases and potentially contribute to the development of more personalized treatment plans.

To systematically explore the relationships between blood markers and glucose dynamics, we conducted a bioinformatics analysis using the "Diabetes Complications Data Set" dataset. We used linear regressions and Students' t-tests to examine the link between glucose-related markers, blood lipid markers, kidney function markers, and liver function markers. Additionally, we conducted analyses to categorize subpopulations of diabetes from the perspectives of both biomarkers and secondary diseases using Uniform Manifold Approximation and Projection (UMAP) to enhance the understanding of the compositions of the diabetic patient population.

Methods

Database:

We used the "Diabetes Complications Data Set" from the Population Health Data Archive (PHDA) official website (https://www.ncmi.cn/phda/dataDetails.do?id=CSTR:A0006.11.A0005.201905.000282-V1.0). This comprehensive repository consists of a wide range of biomarkers from diabetic patients.

Processing Tools:

Python libraries, such as Pandas, NumPy, Matplotlib, Sklearn, Seaborn, Scipy, and Statsmodels, were used to analyze the data and provide statistical results. While processing data for each variable, patients with missing data points were removed to ensure accuracy.

DOI: 10.36838/v6i8.12

Statistical Analysis:

Data from a total of 200 diabetic patients were extracted; for variables with missing patient data, we removed them to ensure accuracy. Linear regressions and Students' t-tests were used to determine the r-value and p-value. *R square* ≥ 0.09; $P < 0.05$*; $P < 0.01$**; $P < 0.001$*** were considered statistically significant. The differences in the bar plots among multiple groups were examined using ANOVA with *post-hoc* analysis.

Data Visualization:

Scatter plots and bar charts were drawn to visualize the data. In scatter plots, individual data points were formed when variables were plotted on the x-axis and y-axis, making it easier to recognize trends, clusters, and outliers in the data. For bar charts with overlapping data, disparities between categories were highlighted. Under different circumstances, data were divided into various groups. The height of each bar represented the mean value of each data group. UMAP was utilized for visualizing multidimensional data.[7] Clusters of data in different colors can be easily spotted on the diagram, showing their similarities. This method was employed to discern the relationships and patterns of various indicators in diabetic patients.

Results

Significant markers that are sensitive to glucose levels and their association with glucose-related markers in diabetic patients:

We are interested in whether some major CBC and BC markers can reflect the abnormal glucose levels in diabetic patients. Before starting the investigation, we used three major glucose-related markers to reflect glucose levels from different perspectives. The glucose level (GLU) is measured to reflect the glucose concentration (mmol/L) in patients' blood after an overnight fast. The range of GLU in normal people ranges from 3.5 to 5.5 mmol/L.[8] We also included postprandial 2-hour glucose levels (GLU_2H). Generally, GLU_2H lower than 7.78 mmol/L is considered within the normal limit. GLU_2H between 7.78 mmol/L and 11.1 mmol/L is abnormal. GLU_2H greater than 11.1 mmol/L can be one of the leading indicators of diabetes symptoms.[9] We also developed a new marker glucose ratio (GLU_RATIO), calculated as GLU_2H divided by GLU to reflect the dynamics of patients' glucose levels after dietary intake. A high GLU_RATIO indicates a higher increase in patients' glucose levels and vice versa.

We want to verify the data's quality by associating glucose levels and related markers with blood markers known to be affected by glucose and diabetes. First, we investigated the association between major glucose markers and glycated hemoglobin (HbA1c). HbA1c is a long-term blood test marker that measures the average glucose level in a person's bloodstream over the past two to three months.[10] It assesses the percentage of hemoglobin molecules bound to glucose, providing a more comprehensive view of long-term glucose control than regular blood glucose tests.[10] A notable positive correlation in the scatter plot and significant differences in the bar plot were detected when analyzing the relationship between HbA1c and GLU (Figure 1A, E). Additionally, we identified a significant positive correlation in the scatter plot and no significant difference in the bar plot when analyzing the relationship between HbA1c and GLU_2H (Figure 1B, F). A negative trend in the scatter plot and significant differences in the bar plot were found when analyzing the relationship between HbA1c and GLU_RATIO (Figure 1C, G). These results indicate that HbA1c is highly dependent on glucose and glucose dynamics. Furthermore, no significance was detected between HbA1c and the patient's age (Figure 1D, H).

Next, we analyzed the relationship between major glucose markers and glycated serum protein (GSP). GSP quantifies the amount of glucose attached to total serum proteins. It is a short-term biomarker of average blood glucose that lasts two to four weeks.[11] We found a positive trend in the scatter plot and significant differences in the bar plot when analyzing the relationship between GSP and GLU (Figure 1I, M). Additionally, the data suggest a positive trend in the scatter plot and no significant difference in the bar plot when analyzing the relationship between GSP and GLU_2H (Figure 1J, N). We also detected a negative trend in the scatter plot and no significant difference in the bar plot when analyzing the relationship between GSP and GLU_RATIO (Figure 1K, O). These results indicate that GSP depends on glucose, not glucose dynamics. Furthermore, no significance was detected between GSP and patients' age (Figure 1L, P).

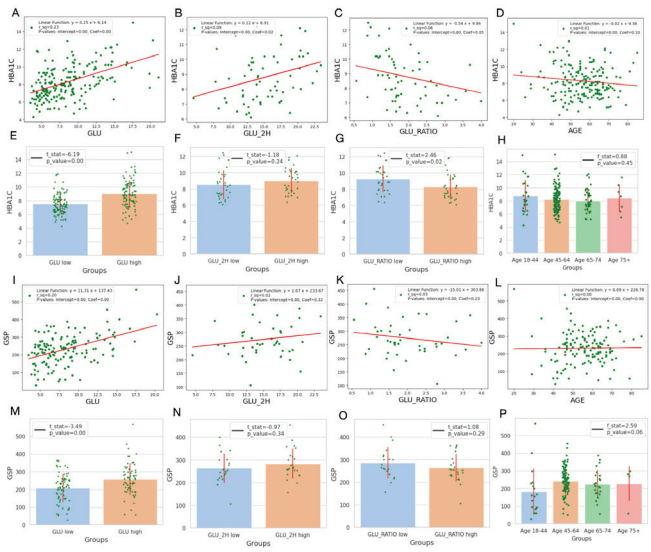

Figure 1: Major glucose-related markers and their association with GLU and GLU_2H levels in diabetic patients. (A-D) The plots show the linear correlation between GLU (A), GLU_2H (B), GLU_RATIO (C), AGE (D), and HbA1c. (E-H) Diabetic patients were categorized based on their GLU (E), GLU_2H (F), GLU_RATIO (G), or AGE (H). The bar charts represent the average HbA1c level of each group. (I-L) The plots show the linear correlation between GLU (I), GLU_2H (J), GLU_RATIO (K), AGE (L), and GSP. (M-P) Diabetic patients were categorized based on their GLU (M), GLU_2H (N), GLU_RATIO (O), or AGE (P). The bar charts represent the average GSP level of each group. Error bars are standard deviation. Related statistical analysis is indicated in the methods and materials.

Blood lipid markers and their association with glucose-related markers in diabetic patients:

Next, we are interested in the connection between major glucose and blood lipid markers, including triglyceride (TG) and high-density lipoprotein cholesterol (HDL_C). TG is a major constituent of body fat, derived from glycerol and three fatty acids.[12] Abnormal changes in TG can lead to arteriosclerosis, increasing the risk of stroke and heart disease.[13] Here, we detected a positive trend in the scatter plot and significant differences in the bar plot when analyzing the relationship between TG and GLU (Figure 2A, E). We also found a negative trend in the scatter plot and no significant difference in the bar plot when analyzing the relationship between TG and GLU_2H (Figure 2B, F). A significant negative correlation was detected in the scatter plot, and no significant difference was observed in the bar plot when analyzing the relationship between TG and GLU_RATIO (Figure 2C, G). These findings suggest that diabetic patients with higher GLU and GLU_RATIO tend to have higher TG. Moreover, we detected a negative trend in the scatter plot and no significant difference in the bar plot when analyzing the relationship between TG and age (Figure 2D, H).

We are then curious about the links between major glucose markers and high-density lipoprotein cholesterol (HDL_C). HDL_C is one of the five primary categories of lipoproteins. It is commonly linked to a decrease in the risk of cardiovascular diseases, such as atherosclerosis and myocardial infarction.[14,15] We detected a negative trend in the scatter plot and no significant difference in the bar plot when analyzing the correlation between HDL_C and GLU (Figure 2I, M). In addition, we show a positive trend in the scatter plot and no significant difference in the bar plot when examining the connection between HDL_C and GLU_2H, indicating HDL_C is independent of GLU_2H (Figure 2J, N). A significant positive correlation in the scatter plot and significant differences in the bar plot were also noted when analyzing the association between HDL_C and GLU_RATIO (Figure 2K, O). These results show that HDL_C is not influenced by GLU and GLU_2H while dependent on GLU_RATIO. Furthermore, no significance is seen between HDL_C and the age of the patients (Figure 2L, P).

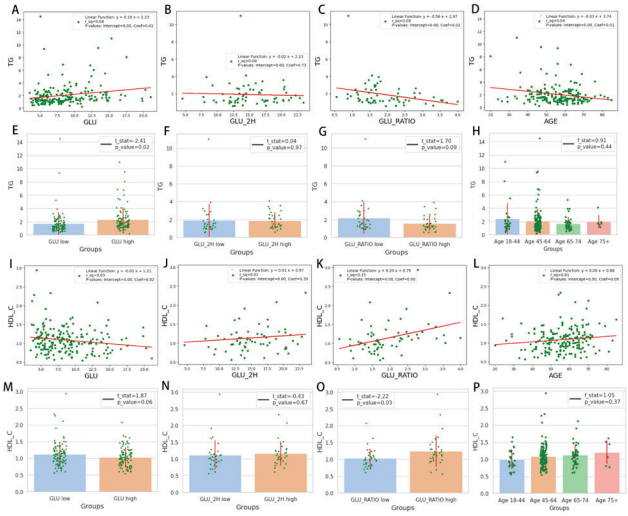

Figure 2: Blood lipid markers and their association with GLU and GLU_2H levels in diabetic patients. (A-D) The plots show the linear correlation between GLU (A), GLU_2H (B), GLU_RATIO (C), AGE (D), and TG. (E-H) Diabetic patients were categorized based on their GLU (E), GLU_2H (F), GLU_RATIO (G), or AGE (H). The bar charts represent the average TG level of each group. (I-L) The plots show the linear correlation between GLU (I), GLU_2H (J), GLU_RATIO (K), AGE (L), and HDL_C. (M-P) Diabetic patients were categorized based on their GLU (M), GLU_2H (N), GLU_RATIO (O), or AGE (P). The bar charts represent the average HDL_C level of each group. Error bars are standard deviation. Related statistical analysis is indicated in the methods and materials.

Renal function-related markers and their association with glucose-related markers in diabetic patients:

In addition, we looked at the connection between major glucose markers and renal function-related markers, such as blood urea (BU), serum creatinine (SCR), and serum uric acid (SUA). First, we investigated BU, which measures the total urea concentration in the blood. An abnormal BU can be seen as a sign of uremia, which can cause a reduction in the glomerular filtration rate.[16,17] Here, we identified a negative trend in the scatter plot and no significant differences in the bar plot when analyzing the relationship between BU and GLU (Figure 3A, E). We also detected a negative trend in the scatter plot and no significant difference in the bar plot when analyzing the relationship between BU and GLU_2H (Figure 3B, F).

Meanwhile, a significant negative correlation was identified in the scatter plot. However, no significant difference was noted in the bar plot while examining the link between BU and GLU_RATIO (Figure 3C, G). These data suggest that diabetic patients with elevated levels of GLU and GLU_RATIO appear to have higher BU. Moreover, no significance was detected between BU and patients' age (Figure 3D, H).

Next, we wanted to examine the relationship between major glucose markers and SCR. SCR is a byproduct that originates from the digestion of protein and the routine breakdown of muscle tissue.[18,19] An abnormal amount of SCR indicates impaired kidney function and chronic kidney disease.[20] A negative trend in the scatter plot and significant differences in the bar plot when analyzing the relationship between SCR and GLU were suggested (Figure 3I, M). We also detected a significant negative correlation in the scatter plot and significant differences in the bar plot when analyzing the relationship between

DOI: 10.36838/v6i8.12

SCR and GLU_2H (Figure 3J, N). Besides, we identified a negative trend in the scatter plot and no significant difference in the bar plot when analyzing the relationship between SCR and GLU_RATIO (Figure 3K, O). These findings show that SCR is dependent on GLU and GLU_2H. Furthermore, no correlation was found between SCR and the age of patients (Figure 3L, P).

We also investigated the association between major glucose markers and SUA. SUA quantifies the level of uric acid present in the blood. An abnormal level of SUA can cause hyperuricemia and gout.[21,22] Here, the data suggest a negative trend in the scatter plot and no significant difference in the bar plot when analyzing the relationship between SUA and GLU (Figure 3Q, U). We also detected a significant negative correlation in the scatter plot and significant differences in the bar plot when analyzing the relationship between SUA and GLU_2H, indicating SUA is highly dependent on GLU_2H (Figure 3R, V). We showed a negative trend in the scatter plot and no significant difference in the bar plot when analyzing the relationship between SUA and GLU_RATIO (Figure 3S, W). Moreover, no significance was detected between SUA and patients' age (Figure 3T, X).

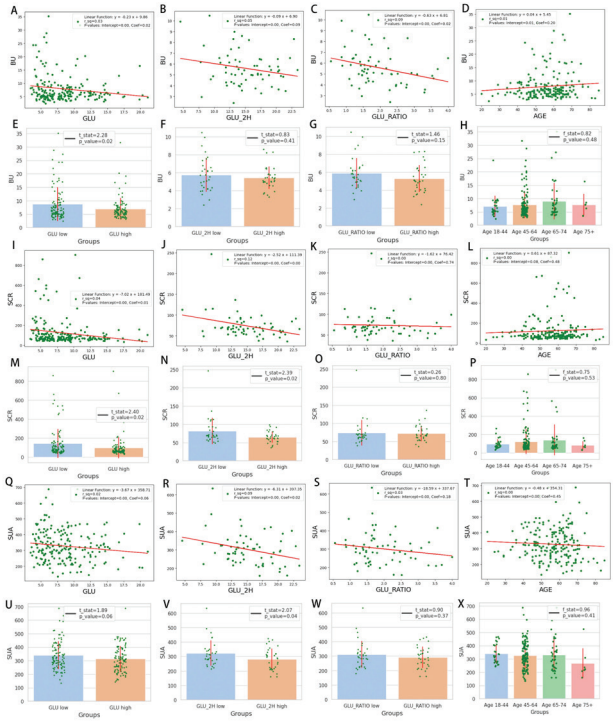

Figure 3: Renal function markers and their association with GLU and GLU_2H levels in diabetic patients. (A-D) The plots show the linear correlation between GLU (A), GLU_2H (B), GLU_RATIO (C), AGE (D), and BU. (E-H) Diabetic patients were categorized based on their GLU (E), GLU_2H (F), GLU_RATIO (G), or AGE (H). The bar charts represent the average BU level of each group. (I-L) The plots show the linear correlation between GLU (I), GLU_2H (J), GLU_RATIO (K), AGE (L), and SCR. (M-P) Diabetic patients were categorized based on their GLU (M), GLU_2H (N), GLU_RATIO (O), or AGE (P). The bar charts represent the average SCR level of each group. (Q-T) The plots show the linear correlation between GLU (Q), GLU_2H (R), GLU_RATIO (S), AGE (T), and SUA. (U-X) Diabetic patients were categorized based on their GLU (U), GLU_2H (V), GLU_RATIO (W), or AGE (X). The bar charts represent the average SUA level of each group. Error bars are standard deviation. Related statistical analysis is indicated in the methods and materials. The p values are plotted in the graphs.

Liver function-related markers and their association with glucose-related markers in diabetic patients:

Furthermore, we explore the connection between major glucose markers and liver function-related markers, such as aspartate aminotransferase (AST). AST is released into the blood when cells containing AST are damaged, commonly in the liver.[23] An abnormal amount of AST can indicate active liver disease.[24] Here, we identified a negative trend in the scatter plot and no significant difference in the bar plot when analyzing the relationship between AST and GLU (Figure 4A, E). A significant trend in the scatter plot and significant differences in the bar plot were found when analyzing the relationship between AST and GLU_2H (Figure 4B, F). We then detected a positive trend in the scatter plot and significant differences in the bar plot when analyzing the relationship between AST and GLU_RATIO (Figure 4C, G). The data suggest a positive trend in the scatter plot and significant differences in the bar plot when analyzing the relationship between AST and age (Figure 4D, H). This part detected no significance, indicating that the liver function-related marker AST is independent based on the given variables (GLU, GLU_2H, and GLU_RATIO).

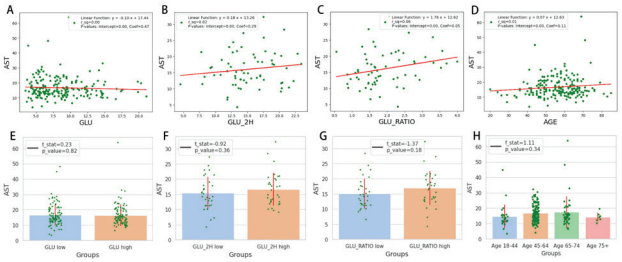

Figure 4: Liver function markers and their association with GLU and GLU_2H levels in diabetic patients. (A-D) The plots show the linear correlation between GLU (A), GLU_2H (B), GLU_RATIO (C), AGE (D), and AST. (E-H) Diabetic patients were categorized based on their GLU (E), GLU_2H (F), GLU_RATIO (G), or AGE (H). Error bars are standard deviation. Related statistical analysis is indicated in the methods and materials. The p values are plotted in the graphs.

Identification of subpopulations of diabetic patients using major blood markers:

In this section, we are curious about the ability of each biomarker or disease to identify subpopulations of diabetic patients. After cleaning the database by deleting patients and biomarkers with empty values for analysis optimization, we plotted a UMAP based on patients' CBC and BC test results. Then, we colored the UMAP according to the level of each biomarker or the presence of diseases in patients. The population was divided into three equal-sized subgroups based on the marker of interest. We investigated the role of 42 biomarkers in characterizing subpopulations of patients. Among these biomarkers, we identified 12 that can divide the patients into subpopulations. In previous sections, biomarkers related to renal functions and glucose metabolism, such as SCR, SUA, and GLU_RATIO, have already been introduced (Figure 5E, F, L).

We categorized the identified markers based on their associated organ functions. Among these biomarkers, we first investigated the ones related to individuals' cardiovascular well-being. This includes TG, total cholesterol (TC), HDL_C,

and low-density lipoprotein (LDL_C). TG and HDL_C have already been discussed in previous sections (Figure 5A, C). TC measures the total amount of cholesterol in the bloodstream, including HDL_C and LDL_C (Figure 5B). In contrast to HDL_C, LDL_C is commonly known as "bad" cholesterol since it aids in the transportation of cholesterol to peripherals (Figure 5D).[25]

We are then interested in biomarkers related to liver functions, such as DBILI, IBILI, TBILI, ALT, and AST. Bilirubin is generated by the degradation of red blood cells, metabolized by the liver, and eliminated from the body.[26] DBILI tests conjugated bilirubin, which is in a water-soluble form (Figure 5G).[26] In contrast, IBILI tests unconjugated bilirubin, a form of bilirubin excreted from the bile (Figure 5H).[27] Unconjugated bilirubin circulates the body while bound to a protein. Besides, TBILI is a test that measures the total bilirubin in the blood, including both conjugated and unconjugated bilirubin (Figure 5I). In addition, ALT is an enzyme essential for the metabolism of glucose and protein (Figure 5J).[28] AST has already been introduced previously (Figure 5K).

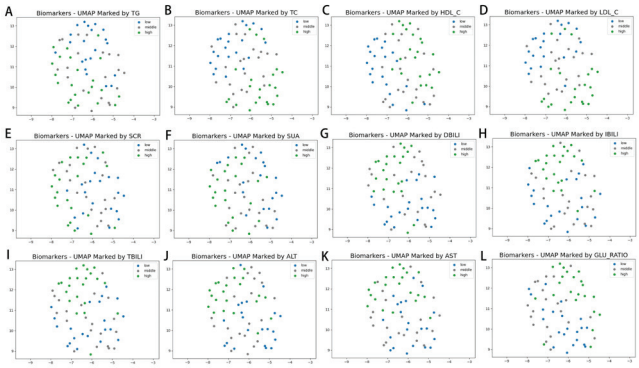

Figure 5: UMAP for diabetic patients is colored according to corresponding biomarkers. If a patient's corresponding biomarker is in the lowest one-third range, the data point will be marked blue; if a patient's corresponding biomarker is in the middle one-third range, the data point will be marked gray; if a patient's corresponding biomarker is in the highest one-third range, the data point will be marked orange. (A) UMAP is colored based on TG. (B) UMAP is colored based on TC. (C) UMAP is colored based on HDL_C. (D) UMAP is colored based on LDL_C. (E) UMAP is colored based on SCR. (F) UMAP is colored based on SUA. (G) UMAP is colored based on DBILI. (H) UMAP is colored based on IBILI. (I) UMAP is colored based on TBILI. (J) UMAP is colored based on ALT. (K) UMAP is colored based on AST. (L) UMAP is colored based on GLU_RATIO.

Identification of subpopulations of diabetic patients using disease markers:

Next, we focus on exploring the ability of different diabetes complications to separate diabetic patients. Despite various biomarkers, our database also includes secondary diseases reported by patients. Instead of coloring the UMAP according to specific biomarkers, we colored it based on whether the diabetic patients have specific diseases. In Figure 6, blue dots (or 0 indicated in the legend) represent diabetic patients who do not report the corresponding disease. In contrast, orange dots (or 1 indicated in the legend) represent diabetic patients reporting the corresponding disease. We found out that 4 out of 32 diseases can successfully characterize diabetic patients into subpopulations, which include Carotid artery stenosis (CAS), Fatty liver disease (FLD), Biliary tract disease (BTD), and Nervous system disease (NSD). CAS is mainly caused by atherosclerotic plaque, which is part of atherosclerosis disease (Figure 6A).[29] FLD starts with an abnormal amount of triglycerides in the liver and is often linked to obesity and diabetes (Figure 6B).[30] BTD refers to diseases that affect the biliary tract, including the gallbladder, bile ducts, and their structures (Figure 6C).[31] NSD stands for disorders that impact the nervous system (Figure 6D).

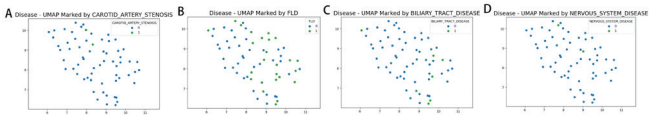

Figure 6: The UMAP for diabetic patients is colored according to the corresponding diseases. (A) UMAP is colored based on CAS. (B) UMAP is colored based on FLD. (C) UMAP is colored based on BTD. (D) UMAP is colored based on NSD.

Discussion

In previous diabetes investigations, researchers have made significant progress in identifying new blood and tissue-based biomarkers that can enable the identification, prevention, and therapy of diabetes.[32] Recently, we have been curious about the relationship between glucose-related biomarkers (GLU, GLU_2H, and GLU_RATIO) and different organ function and disease-related biomarkers, as well as how various biomarkers and diabetes complications can be used to identify diabetic patient subpopulations. Among the glucose-related biomarkers, GLU_RATIO is a newly derived marker aimed at determining the relative dynamics of glucose in patients. Our research uses the "Diabetes Complications Data Set" to conduct a detailed bioinformatic analysis of these biomarkers.

According to clinical data on diabetic patients, around 70% of type 2 diabetes patients are affected by diabetic dyslipidemia.[33] Diabetic dyslipidemia is a collection of anomalies related to lipoproteins. TG is a common kind of fat, and elevated TG can make the walls of arteries harder and thicker, resulting in an increased risk of cardiovascular diseases.[34] In our investigation, we identified that within diabetic patients, those with higher GLU also tend to have higher TG, which has been previously discussed (Figure 2E). The glucose levels of the diabetic population were already beyond the normal range. However, this population can be further divided into high and extra-high groups. This implies that absolute GLU within diabetic patients still matters and can be used as a marker to predict the major blood lipid marker TG. High TG levels potentially contribute to an increased likelihood of cardiovascular diseases in people with diabetes.

Diabetic nephropathy is a kidney disease resulting from diabetes. Diabetic kidney diseases may lead to disorders following persistent albuminuria and a gradual decrease in the rate of glomerular filtration. According to clinical data, about 30-40 percent of diabetic patients develop diabetic nephropathy.[35] For diabetic patients, a consistently high blood sugar level can lead to artery narrowing and blockage, causing harm to the kidneys.[36] Also, diabetes is often associated with prolonged hyperglycemia, leading to oxidative stress compounds building

DOI: 10.36838/v6i8.12

up. This can deteriorate bladder function.[37] These two cases will cause the kidneys not to work correctly, reducing BU production. Our study detected that diabetic patients with high GLU often have a lower BU (Figure 3E). This implies that diabetic individuals with higher GLU exhibit higher kidney stress and potential liver damage, which highlights that more blood sugar control and attention to the liver are needed for diabetic patients with high GLU.

In addition, SCR is a reliable biomarker for detecting late changes in kidney function, such as acute kidney injury.[38] Low SCR can be caused by liver disease, muscle disorders, or certain drugs.[39] We identified negative correlations between SCR and GLU or GLU_2H, indicating that diabetic patients with higher GLU or GLU_2H levels may require additional management to assess their kidney condition (Figure 3M, J, N). An abnormally elevated GLU_2H is considered hyperglycemia, which is related to a higher risk of diabetes.[40] This aligns with previous scientific research stating that low SCR is linked to a higher likelihood of developing type 2 diabetes mellitus.[41] Furthermore, we found a negative correlation between SUA and GLU_2H, which implies that diabetic patients with lower GLU_2H tend to have higher SUA (Figure 3V). Given that elevated SUA is detected in the initial stage of impaired glucose metabolism, it is reasonable to suggest that diabetic patients with low levels of GLU and GLU_2H can have high SUA.[42]

Even though there are 42 biomarkers and 32 diseases in our database, only 12 of the biomarkers and 4 of the diseases can influence patients and categorize the population into distinct subpopulations. As a result, some markers are more important than others because they can identify and separate subpopulations. When we examine an individual data point with an extreme value of one of these markers, we can infer that this patient belongs not only to a large diabetic population but also to a smaller population of people with specific biological characteristics. We identified a list of biomarkers that can categorize subpopulations of diabetic patients—TG, TC, HDL_C, LDL_C, SCR, SUA, DBILI, IBILI, TBILI, ALT, AST, and GLU_RATIO (Figure 5). For example, diabetic patients who belong to the elevated TC levels subpopulation may be more susceptible to developing cardiovascular ailments.[43] LDL_C is likely to contribute to the accumulation of plaque; high LDL_C can lead to higher risks of heart disease and stroke.[44,45] GLU_RATIO reflects the body's ability to regulate glucose before and after dietary intake. This evidence suggests that patients with different levels of the identified markers can exhibit distinct phenotypes and may be treated differently, given that they may belong to other subpopulations.

In addition, we identified CAS, FLD, BTD, and NSD as the secondary diseases that make diabetes patients exhibit different phenotypes. These four diseases affect the blood profile of diabetic patients, leading to the characterization of subpopulations (Figure 6). CAS is caused by the accumulation of plaque inside the arteries. Since diabetic patients often experience hyperglycemia and high cholesterol levels, these risk factors increase the probability of plaque building up in the carotid arteries.[46,47] Therefore, a specific subpopulation of diabetic patients faces elevated susceptibility to CAS, which is characterized in our plot (Figure 6A). Besides, since insulin resistance is linked to an elevated amount of free fatty acids, it is reasonable to reflect that this imbalance exacerbates the advancement of FLD in diabetic patients (Figure 6B).[48] BTD refers to diseases affecting the biliary tract. Diabetes affects both the production and release of bile acids, leading to excess cholesterol in the bile. As the level of cholesterol increases, gallstones are more likely to form.[49] Consequently, certain diabetic patients would have a higher risk of developing BTD and are characterized differently from blood tests (Figure 6C). Furthermore, diabetic patients commonly experience hyperglycemia, referred to as high blood glucose levels and increased amounts of lipids. Prolonged hyperglycemia can lead to nerve damage, causing NSD (Figure 6D).[50] These results suggest that diabetic patients with these identified secondary diseases belong to subpopulations that exhibit distinct phenotypes compared to the general population and may require different treatments. More importantly, the indications and presentations of the UMAP provide us with a template to evaluate the risk of new patients who have not exhibited symptoms of the identified secondary disease but have fallen into the corresponding secondary disease subpopulation. This could suggest that extra attention and precautions may be needed to manage their risk of developing related diseases.

Although we utilized a large amount of data from the "Diabetes Complications Data Set," there are still some study limitations that can be further explored. Firstly, most of the reported associations are correlations, not causations. Future experiments that control one of the glucose markers, potentially through dietary interventions and evaluation of the associated markers, can help further identify the underlying mechanisms and causality. Secondly, given the limited number of patients and the diversity of reported secondary diseases, our investigations have not covered all potential diabetic complications, such as diabetic dermopathy, diabetic osteoporosis, and cataracts.[51–53] If a more extensive investigation of diabetic complications can be conducted, even including some rare diseases, this might aid in developing more personalized and targeted medical treatments and make our understanding of the challenges faced by diabetic patients more comprehensive.

■ Conclusion

In this study, we conducted a detailed bioinformatics analysis using the "Diabetes Complications Data Set". We revealed intricate correlations between glucose-related markers and different function-related biomarkers, including blood lipids, renal function, and liver function. We also identified twelve biomarkers and four diseases that are important in categorizing diabetic patients into subpopulations. These discoveries enhance our understanding of the nuanced bodily mechanisms of diabetic patients compared to people without diabetes. The results highlight that different diabetic patients exhibit subtle differences in their biomarker performances, which are potentially influenced by differential diabetic complications. This study potentially encourages the development of targeted treatments for diabetic subpopulations, considering their differential CBC and BC biomarker performances. Overall,

our findings provide new insights into how markers affected by diabetes could be altered and the development of targeted interventions and medical treatments for diabetes.

Acknowledgments

I am thankful to my mentor for guiding me through the project and providing detailed feedback, and I am also grateful to my parents for their encouragement and support.

References

1. Kharroubi, A. T.; Darwish, H. M. Diabetes Mellitus: The Epidemic of the Century. *World J. Diabetes* **2015**, *6* (6), 850.
2. Cho, N. H.; Shaw, J. E.; Karuranga, S.; Huang, Y.; da Rocha Fernandes, J. D.; Ohlrogge, A. W.; Malanda, B. IDF Diabetes Atlas: Global Estimates of Diabetes Prevalence for 2017 and Projections for 2045. *Diabetes Res. Clin. Pract.* **2018**, *138*, 271–281.
3. Steele, C.; Hagopian, W. A.; Gitelman, S.; Masharani, U.; Cavaghan, M.; Rother, K. I.; Donaldson, D.; Harlan, D. M.; Bluestone, J.; Herold, K. C. Insulin Secretion in Type 1 Diabetes. *Diabetes* **2004**, *53* (2), 426–433.
4. Wilmot, E. G.; Davies, M. J.; Yates, T.; Benhalima, K.; Lawrence, I. G.; Khunti, K. Type 2 Diabetes in Younger Adults: The Emerging UK Epidemic. *Postgrad. Med. J.* **2010**, *86* (1022), 711–718.
5. Treatment Strategies against Diabetes: Success so Far and Challenges Ahead. *Eur. J. Pharmacol.* **2019**, *862*, 172625.
6. Lieb, K.; von der Osten-Sacken, J.; Stoffers-Winterling, J.; Reiss, N.; Barth, J. Conflicts of Interest and Spin in Reviews of Psychological Therapies: A Systematic Review. *BMJ Open* **2016**, *6* (4), e010606.
7. Demidova, L. A.; Stepanov, M. A. *Approach to the Analysis of the Multidimensional Time Series Based on the UMAP Algorithm in the Problems of the Complex Systems Proactive Maintenance*. https://ieeexplore.ieee.org/abstract/document/9211008/ (accessed 2024-02-28).
8. Güemes, M.; Rahman, S. A.; Hussain, K. What Is a Normal Blood Glucose? *Arch. Dis. Child.* **2016**, *101* (6), 569–574.
9. Piché, M. E.; Lemieux, S.; Pérusse, L.; Weisnagel, S. J. High Normal 2-Hour Plasma Glucose Is Associated with Insulin Sensitivity and Secretion That May Predispose to Type 2 Diabetes. *Diabetologia* **2005**, *48* (4), 732–740.
10. Weykamp, C. HbA1c: A Review of Analytical and Clinical Aspects. *Ann. Lab. Med.* **2013**, *33* (6), 393–400.
11. Little, R. R.; Goldstein, D. E. Long-Term Glucose Monitoring With Glycated Proteins. *Lab. Med.* **1992**, *23* (8), 533–538.
12. *Triglyceride*. https://doi.org/10.1016/j.jff.2021.104724.
13. Varbo, A.; Nordestgaard, B. G. Remnant Cholesterol and Triglyceride-Rich Lipoproteins in Atherosclerosis Progression and Cardiovascular Disease. *Arterioscler. Thromb. Vasc. Biol.* **2016**. https://doi.org/10.1161/ATVBAHA.116.308305.
14. Silbernagel, G.; Schöttker, B.; Appelbaum, S.; Scharnagl, H.; Kleber, M. E.; Grammer, T. B.; Ritsch, A.; Mons, U.; Holleczek, B.; Goliasch, G.; Niessner, A.; Boehm, B. O.; Schnabel, R. B.; Brenner, H.; Blankenberg, S.; Landmesser, U.; März, W. High-Density Lipoprotein Cholesterol, Coronary Artery Disease, and Cardiovascular Mortality. *Eur. Heart J.* **2013**, *34* (46), 3563–3571.
15. Abbott, R. D.; Wilson, P. W.; Kannel, W. B.; Castelli, W. P. High Density Lipoprotein Cholesterol, Total Cholesterol Screening, and Myocardial Infarction. The Framingham Study. *Arteriosclerosis: An Official Journal of the American Heart Association, Inc.* **1988**. https://doi.org/10.1161/01.ATV.8.3.207.
16. A History of Uremia Research. *J. Ren. Nutr.* **2017**, *27* (6), 449–452.
17. Zemaitis, M. R.; Foris, L. A.; Katta, S.; Bashir, K. Uremia. In *StatPearls [Internet]*; StatPearls Publishing, 2023.
18. Patel, S. S.; Molnar, M. Z.; Tayek, J. A.; Ix, J. H.; Noori, N.; Benner, D.; Heymsfield, S.; Kopple, J. D.; Kovesdy, C. P.; Kalantar-Zadeh, K. Serum Creatinine as a Marker of Muscle Mass in Chronic Kidney Disease: Results of a Cross-Sectional Study and Review of Literature. *J. Cachexia Sarcopenia Muscle* **2013**, *4* (1), 19.
19. Hosten, A. O. BUN and Creatinine. In *Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition*; Butterworths, 1990.
20. Jafar, T. H.; Schmid, C. H.; Levey, A. S. Serum Creatinine as Marker of Kidney Function in South Asians: A Study of Reduced GFR in Adults in Pakistan. *J. Am. Soc. Nephrol.* **2005**, *16* (5), 1413.
21. Jin, M.; Yang, F.; Yang, I.; Yin, Y.; Luo, J. J.; Wang, H.; Yang, X.-F. Uric Acid, Hyperuricemia and Vascular Diseases. *Front. Biosci. 17*, 656.
22. Logan, J. A.; Morrison, E.; McGill, P. E. Serum Uric Acid in Acute Gout. *Ann. Rheum. Dis.* **1997**, *56* (11), 696–697.
23. McGill, M. R. The Past and Present of Serum Aminotransferases and the Future of Liver Injury Biomarkers. *EXCLI J.* **2016**, *15*, 817.
24. Aspartate Aminotransferase Isoenzymes. *Clin. Biochem.* **1990**, *23* (4), 311–319.
25. Das, P.; Ingole, N. Lipoproteins and Their Effects on the Cardiovascular System. *Cureus* **2023**, *15* (11). https://doi.org/10.7759/cureus.48865.
26. Kalakonda, A.; Jenkins, B. A.; John, S. Physiology, Bilirubin. **2017**.
27. Fevery, J.; Van Damme, B.; Michiels, R.; De Groote, J.; Heirwegh, K. P. M. Bilirubin Conjugates in Bile of Man and Rat in the Normal State and in Liver Disease. *J. Clin. Invest.* **1972**, *51* (9), 2482–2492.
28. Yang, R.-Z.; Park, S.; Reagan, W. J.; Goldstein, R.; Zhong, S.; Lawton, M.; Rajamohan, F.; Qian, K.; Liu, L.; Gong, D.-W. Alanine Aminotransferase Isoenzymes: Molecular Cloning and Quantitative Analysis of Tissue Expression in Rats and Serum Elevation in Liver Toxicity. *Hepatology* **2009**, *49* (2), 598.
29. Diagnosis and Treatment of Carotid Artery Stenosis. *Journal of Neurology & Stroke* **2017**, *7* (3), 62–65.
30. Cohen, J. C.; Horton, J. D.; Hobbs, H. H. Human Fatty Liver Disease: Old Questions and New Insights. *Science* **2011**. https://doi.org/10.1126/science.1204265.
31. *Biliary tract disease - Latest research and news*. Springer Nature. https://www.nature.com/subjects/biliary-tract-disease (accessed 2024-02-27).
32. Lyons, T. J.; Basu, A. BIOMARKERS IN DIABETES: HEMOGLOBIN A1c, VASCULAR AND TISSUE MARKERS. *Transl. Res.* **2012**, *159* (4), 303.
33. Jialal, I.; Singh, G. Management of Diabetic Dyslipidemia: An Update. *World J. Diabetes* **2019**, *10* (5), 280–290.
34. Zhang, B.-H.; Yin, F.; Qiao, Y.-N.; Guo, S.-D. Triglyceride and Triglyceride-Rich Lipoproteins in Atherosclerosis. *Frontiers in Molecular Biosciences* **2022**, *9*. https://doi.org/10.3389/fmolb.2022.909151.
35. Gross, J. L.; de Azevedo, M. J.; Silveiro, S. P.; Canani, L. H.; Caramori, M. L.; Zelmanovitz, T. Diabetic Nephropathy: Diagnosis, Prevention, and Treatment. *Diabetes Care* **2005**, *28* (1), 164–176.
36. The Prevalence of Renal Artery Stenosis among Patients with Diabetes Mellitus. *Eur. J. Intern. Med.* **2012**, *23* (7), 639–642.
37. Daneshgari, F.; Liu, G.; Birder, L.; Hanna-Mitchell, A. T.; Chacko, S. Diabetic Bladder Dysfunction: Current Translational Knowledge. *J. Urol.* **2009**, *182* (6 0), S18.
38. Gounden, V.; Bhatt, H.; Jialal, I. Renal Function Tests. In *StatPearls [Internet]*; StatPearls Publishing, 2023.
39. Ostermann, M.; Kashani, K.; Forni, L. G. The Two Sides of Creatinine: Both as Bad as Each Other? *J. Thorac. Dis.* **2016**, *8* (7), E628.

DOI: 10.36838/v6i8.12

40.Klein, R.; Klein, B. E. K.; Moss, S. E.; Cruickshanks, K. J. Relationship of Hyperglycemia to the Long-Term Incidence and Progression of Diabetic Retinopathy. *Arch. Intern. Med.* **1994**, *154* (19), 2169–2178.
41.Qin, P.; Lou, Y.; Cao, L.; Shi, J.; Tian, G.; Liu, D.; Zhou, Q.; Guo, C.; Li, Q.; Zhao, Y.; Liu, F.; Wu, X.; Qie, R.; Han, M.; Huang, S.; Zhao, P.; Wang, C.; Ma, J.; Peng, X.; Xu, S.; Chen, H.; Zhao, D.; Zhang, M.; Hu, D.; Hu, F. Dose-Response Associations between Serum Creatinine and Type 2 Diabetes Mellitus Risk: A Chinese Cohort Study and Meta-Analysis of Cohort Studies. *J. Diabetes* **2020**, *12* (8), 594–604.
42.Rafiullah, M.; Siddiqui, K. Serum Uric Acid and Metabolic Markers in Diabetes. *Biomarkers in Diabetes* **2022**, 1–21.
43.Jeong, S.-M.; Choi, S.; Kim, K.; Kim, S. M.; Lee, G.; Park, S. Y.; Kim, Y.-Y.; Son, J. S.; Yun, J.-M.; Park, S. M. Effect of Change in Total Cholesterol Levels on Cardiovascular Disease Among Young Adults. *J. Am. Heart Assoc.* **2018**, *7* (12). https://doi.org/10.1161/JAHA.118.008819.
44.Al-shoaibi, A. A. A.; Li, Y.; Song, Z.; Chiang, C.; Hirakawa, Y.; Saif-Ur-Rahman, K. M.; Shimoda, M.; Nakano, Y.; Matsunaga, M.; Aoyama, A.; Tamakoshi, K.; Ota, A.; Yatsuya, H. Association of Low-Density Lipoprotein Cholesterol with Risk of Coronary Heart Disease and Stroke among Middle-Aged Japanese Workers: An Analysis Using Inverse Probability Weighting. *J. Appl. Toxicol.* **2023**, *30* (5), 455–466.
45.Rafieian-Kopaei, M.; Setorki, M.; Doudi, M.; Baradaran, A.; Nasri, H. Atherosclerosis: Process, Indicators, Risk Factors and New Hopes. *Int. J. Prev. Med.* **2014**, *5* (8), 927.
46.Tanaka, M.; Masuda, S.; Matsuo, Y.; Sasaki, Y.; Yamakage, H.; Muranaka, K.; Wada, H.; Hasegawa, K.; Tsukahara, T.; Shimatsu, A.; Satoh-Asahara, N. Hyperglycemia and Inflammatory Property of Circulating Monocytes Are Associated with Inflammatory Property of Carotid Plaques in Patients Undergoing Carotid Endarterectomy. *J. Appl. Toxicol.* **2016**, *23* (10), 1212–1221.
47.Goldstein, J. L.; Brown, M. S. A Century of Cholesterol and Coronaries: From Plaques to Genes to Statins. *Cell* **2015**, *161* (1), 161–172.
48.Boden, G. Role of Fatty Acids in the Pathogenesis of Insulin Resistance and NIDDM. *Diabetes* **1997**, *46* (1), 3–10.
49.Staels, B.; Fonseca, V. A. Bile Acids and Metabolic Regulation: Mechanisms and Clinical Responses to Bile Acid Sequestration. *Diabetes Care* **2009**, *32 Suppl 2* (Suppl 2). https://doi.org/10.2337/dc09-S355.
50.Diabetes and the Peripheral Nerve. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2009**, *1792* (10), 931–940.
51.McCASH, S.; Emanuel, P. O. Defining Diabetic Dermopathy. *J. Dermatol.* **2011**, *38* (10), 988–992.
52.Inzerillo, A. M.; Epstein, S. Osteoporosis and Diabetes Mellitus. *Rev. Endocr. Metab. Disord.* **2004**, *5* (3), 261–268.
53.Javadi, M.-A.; Zarei-Ghanavati, S. Cataracts in Diabetic Patients: A Review Article. *J. Ophthalmic Vis. Res.* **2008**, *3* (1), 52.

Author

Qi Chen, an 18-year-old girl, studies at Shenzhen College of International Education and will graduate in 2025. When she is taking care of her grandmother – a diabetic patient with coronary heart disease - in the hospital, her interest in diabetes emerged. She is passionate about using data analysis to solve practical problems.