Send Orders for Reprints to reprints@benthamscience.net

Current Diabetes Reviews, XXXX, *XX,* 000-000

RESEARCH ARTICLE

Machine Learning and Augmented Intelligence Enables Prognosis of Type 2 Diabetes Prior to Clinical Manifestation

Jonathan RT Lakey[1,2,*], Krista Casazza[1], Waldemar Lernhardt[1], Eric J. Mathur[1], Ian Jenkins[1]

[1]*GATC Health, 2030 Main Street, Suite 660, Irvine, CA 92614, CA;* [2]*Departments of Surgery and Biomedical Engineering, University of California Irvine, Irvine, CA*

ARTICLE HISTORY

Received: October 03, 2023
Revised: December 07, 2023
Accepted: January 03, 2024

DOI:
10.2174/0115733998276990240117113408

Abstract: ***Background*:** The global incidence of type 2 diabetes (T2D) persists at epidemic proportions. Early diagnosis and/or preventive efforts are critical to attenuate the multi-systemic clinical manifestation and consequent healthcare burden. Despite enormous strides in the understanding of pathophysiology and on-going therapeutic development, effectiveness and access are persistent limitations. Among the greatest challenges, the extensive research efforts have not promulgated reliable predictive biomarkers for early detection and risk assessment. The emerging fields of multi-omics combined with machine learning (ML) and augmented intelligence (AI) have profoundly impacted the capacity for predictive, preventive, and personalized medicine.

***Objective*:** This paper explores the current challenges associated with the identification of predictive biomarkers for T2D and discusses potential actionable solutions for biomarker identification and validation.

***Methods*:** The articles included were collected from PubMed queries. The selected topics of inquiry represented a wide range of themes in diabetes biomarker prediction and prognosis.

***Results*:** The current criteria and cutoffs for T2D diagnosis are not optimal nor consider a myriad of contributing factors in terms of early detection. There is an opportunity to leverage AI and ML to significantly enhance the understanding of the underlying mechanisms of the disease and identify prognostic biomarkers. The innovative technologies being developed by GATC are expected to play a crucial role in this pursuit *via* algorithm training and validation, enabling comprehensive and in-depth analysis of complex biological systems.

***Conclusion*:** GATC is an emerging leader guiding the establishment of a systems approach towards research and predictive, personalized medicine. The integration of these technologies with clinical data can contribute to a more comprehensive understanding of T2D, paving the way for precision medicine approaches and improved patient outcomes.

Keywords: Type 2 diabetes (T2D), Incidence, Augmented Intelligence (AI), machine learning (ML), prediction, biomarkers.

1. INTRODUCTION

Type 2 diabetes (T2D) diagnosis is characterized by elevation in circulating blood glucose. The hyperglycemia is largely related to inadequate insulin response, often preceded by progressive insulin resistance. T2D affects an estimated 462 million people, or roughly 6% of the world's population. The consequential persistent hyperglycemia manifests in diabetes-associated complications[1]. The complications are apparent across systems, inducing alterations in protein synthesis and release. Mortality associated with T2D exceeded one million individuals in 2017, implicating it as the ninth leading cause of mortality[1]. In the US, approximately 25 million Americans ages 20 – 64 have T2D which corresponds to about 11% of this age group. Total health care and related costs are approximately $174 billion annually[2]. It has been well-established that the metabolic underpinnings that promote disease development occur well before the elevation in blood glucose and consequential uncompensated insulin response in T2D. While early identification is critical to prevent or delay T2D onset, many individuals with T2D have not been diagnosed[3]. The International Diabetes Federation (IDF) has reported approximately 50% of individuals who would meet the criteria for T2D do not know that they have the disease. As a result, treatment is delayed or absent,

*Address correspondence to this author at the Departments of Surgery and Biomedical Engineering, University of California Irvine, Irvine, CA; E-mail: jlakey@uci.edu

increasing risks for adverse health outcomes and substantial economic burden. Further, a substantial number of Americans with impaired glucose tolerance or prediabetes, who also incur significant healthcare costs[3]. The combined estimated US annual healthcare and public health outreach spending is in the in billions of dollars. Yet, a more comprehensive understanding of how to beneficially translate evidence-based practices from research towards enhanced dissemination and improvement of services continues to be limited. Collectively, the public health implications of T2D, undiagnosed T2D, and pre-T2D represent a significant national and international health challenge.

The differentiation among individuals with T2D, pre-T2D, or euglycemia resides in the implementation of well-accepted criteria. The 2019 ADA report states that the most widely used measures of fasting plasma glucose (FPG), 2-hr plasma glucose obtained from an oral glucose tolerance test (OGTT), and hemoglobin A1c ($HbA1_c$) are each generally appropriate for diagnosing T2D. However, classifications based on the glucose-based protocols and $HbA1_c$ values have been met with equivocal results [4,5]. A recent study by Tucker (2020), including over 7400 adults, reported that the $HbA1_c$ assay misdetected T2D nearly 67% of the time, and FPG was accurate in less than half of the cases, demonstrating that high specificity accompanied by low sensitivity may evoke unintended consequences in the context of T2D surveillance and preventive/intervention efforts[6]. Further, Khosla *et al*. (2021) reported that HbA_{1c} criteria inadequately characterized gluco-regulatory status among individuals of heterogeneous US residents of African descent (Africans residing in the US, African American, Caribbean) population[7]. Randomized controlled trials (RCT) are often fraught with heterogeneity which elicits significant limitations to clinical research, as well as confounds drug development efficacy studies in patients with T2D. Further, T2D and comorbidities are frequently different across racial/ethnic groups, potentially leading to over- or under-specification of biomarker relevance. This includes (but is not limited to) those with varying stages of the disease, different comorbidities like hypertension or cardiovascular disease, different ages, ethnic backgrounds, genetic variations and lifestyles, or those on different concurrent medications. As such, well-accepted criteria have well-established limitations.

Despite well-described limitations[8-10], glycated hemoglobin (HbA1c) maintains its position as one of the best indicators of glycemic control and an accurate proxy for estimating average serum glucose levels over the past three-month period across clinical and community settings. The clinical utility of the $HbA1_c$ biomarker resides in the minimal effect short, large fluctuations in serum glucose levels have on the measurement of glycated hemoglobin buildup over the lifetime of red blood cells. The consistent increase in glucose over time leads to an increase in the fraction of glycated hemoglobin. Unfortunately, elevated HbA1c levels indicate that cellular damage has already begun, combined with an increase in reactive oxygen species within red blood cells (RBCs). Free radicals and/or reactive oxygen species (ROS) can alter RBC membrane properties, inducing RBC aggregation, increased blood viscosity, impaired blood flow, and inflammation. ROS accumulation leads to the oxidation of Fe^{2+}-Hb to Fe^{3+}-Hb into the less stable ferryl Hb (Fe^{4+}-Hb)[11]. The instability of Fe^{4+} increases reactivity with several amino acids to restore the Fe^{3+}Hb oxidation state. This, in turn, induces cellular damage and leakage of Fe^{4+}-Hb into the vessel sub-endothelial matrix (endothelial permeability)[11], which further augments the activation of the pro-inflammatory cascade. The monocyte adhesion proteins lead to the accumulation of macrophages on blood vessel surfaces in response. The progressive build-up of these plaques links T2D to atherosclerosis and, ultimately, cardiovascular diseases. Although the chronic cascade described represents the typical progressive nature of T2D, the pathophysiology can be reversible[12,13]. If symptoms can be identified prior to damage to organ systems, glucose homeostasis can be reestablished. A novel platform that can be used to overcome barriers to T2D screening and prevention and promote equitable dissemination and implementation among underrepresented communities is highly warranted.

The clinical implementation of T2D intervention and prevention strategies resides in understanding heterogeneity in addition to the current understanding of traditional risk factors. The relative risk of T2D for each patient can be integrated into decision-making, *i.e.*, risk prediction models. In theory, risk prediction models can determine an individual's prognostic risk of T2D as well as the predictable outcome of the untreated disease or the response to treatment and management strategies for those already diagnosed. Machine learning (ML) and deep learning (DL) are augmented intelligence (AI) techniques that have been adopted to construct prediction models using multi-omic data, including genomics, transcriptomics, epigenomics, and proteomics. Leveraging the advanced technologies established through AI and analytical multi-omics methods, GATC has applied algorithms to "-omics" data from the population to establish efficient and accurate identification of biomarkers. The overarching goal is to optimize multiomic tools to help elucidate the pathophysiology of T2D and thus enable early prediction of predisposition and progression of T2D. Such scientific tools provide insightfulness to understanding due to the tools' ability to project causal associations between metabolite and protein markers, which can highlight pathways within T2D early detection. Integrating multiomic analyses will assist in identifying the causal biomarkers responsible for the early onset of T2D as well as differentiating among those with diabetes, prediabetes, or normal glucose metabolism[14].

Given that T2D is not a single, uniform disease entity, the biological and mechanistic underpinnings of T2D may differ greatly across individuals. Heterogeneity makes it challenging to pinpoint biomarkers that apply universally across all cases. Diagnostic evaluation by OGTT, HbA1c, or FPG, although relatively cost-effective, does miss valuable understanding of etiopathologies, probable disease progression, or potential future risk (if glucose is within normal limits). Although an ideal biomarker necessitates consistent stability over time and across different physiological conditions, biomarker levels can be influenced by factors such as daily variations, diet, medications, and other transient factors, which can complicate their identification and interpretation. In addition, T2D often develops over a long period with a gradual progression from insulin resistance to overt hyper-

glycemia. To date, the identification of biomarker(s) that can accurately predict disease development and progression during this latency period is difficult due to the subtle and dynamic nature of these changes. Further, racial/ethnic differences in glucose handling have been documented and impact the accuracy of disease progression. The novel and patented predictive machine learning engines designed by GATC Health data scientists used input data and augmented intelligence (AI) tools to both identify causal biomarkers and calculate ratios and correlations between marker levels in patient samples, such as serum/plasma, saliva or urine, leading to the establishment of a gradient scale for assessment of a patient's risk for T2D. Importantly, circulating biomarkers are dynamic in nature, and changes may not reflect distinct patterns of physiological or pathological conditions. Accordingly, a modern approach to biomarker discovery involves integrating data from various "omics" sources (genomics, proteomics, metabolomics) along with clinical and lifestyle data. Analyzing and interpreting these complex datasets requires sophisticated computational techniques and bioinformatics expertise. Moreover, machine learning iterations that compare ratios of serum biomarkers with lipid intermediate correlations, combined with proteome data, provide a rich multi-omic dataset for the elucidation of T2D causal biomarkers.

The utility of AI continues to emerge, providing opportunities to improve or enhance the prognosis prior to the clinical manifestation of T2D along the health continuum. These AI applications have begun to be used for more targeted therapies across conditions, showing promise in improving screening and diagnosis early identification of risk for complications, thereby reducing morbidity and mortality, with the goal of improving quality of life, while decreasing healthcare costs. Continuously emerging yet compelling evidence suggests the high utility of AI in identifying potential prognostic biomarkers for the disease is highly relevant to T2D. Several studies focused on identifying associations between HbA1c and a number of biomarkers for T2D have been conducted, yet few have uncovered a potential predictive value related to protein and or lipid changes induced by T2D. Proteomics and lipidomics play a significant role in elucidating the key constituents of cellular membranes by providing unprecedented insight into T2D screening and biomarkers of this disease. Proteomics serves to classify all proteins that can participate in the biological processes of an organism as well as identify biological profiles of molecular products of transcription, whereas lipidomics allows for interrogation of, more specifically, molecular lipid species associated with T2DM risk factors. Progress in proteomics and lipidomics offers a new and innovative approach by which T2D prediction models can be utilized for early identification, differentiation and discrimination of pathways associated with T2D progression and has the potential to facilitate precise screening, diagnosis, prevention, management and treatment strategies.

While there are tradeoffs associated with the inclusion of multiple factors and homogeneity in populations, it is assumed these concerns can be overcome in the future, in AI models. As such, incorporating comprehensive datasets, including laboratory and genetic data, is essential to improve the accuracy of AI. In that regard, the inclusion of data only from low-risk or high-risk populations while improving accuracy within that specific group leads to potentially compromised generalizability. Conversely, data from non-selected populations in AI applications improve generalizability potentially as the "expense" of accuracy in predictive capacity across populations. The availability of novel data that can be incorporated is becoming increasingly more prevalent. As the data amasses, so does the capacity to improve the predictive ability of AI, yet usability is not sacrificed. Comprehensive datasets that include international datasets with diverse populations will undoubtedly lead to a vast improvement in the predictive ability while limiting population bias by race, ethnicity, geographic location, or other factors. The growing interest in leveraging large electronic medical record (EMR) data to develop prediction models using AI algorithms is increasingly gaining traction. Indeed, the incorporation of vast variables that have the prognostics capacity based on well-established influence on glucose homeostasis from very large cohorts of patients, allows the advanced AI technology platform to fill an evidence gap. Collectively, AI can identify and sample clinical factors that affect glucose levels; however, such applications are also limited to "patients," limiting prognostication potential.

Prediction and/or prognostication models capitalizing on the use of advanced statistical techniques integrating various modeling applications (*e.g.*, gradient boosting, random forest classification, recurrent neural net, and logistic regression) have rapidly emerged in therapeutic applications. Over the past five years, an AI diagnostic system that autonomously diagnoses patients with diabetic retinopathy using deep learning DL has been cleared by the Food and Drug Administration (FDA). The systems that have been granted FDA approval have demonstrated high sensitivities and specificities above 90% for the detection of referable diabetic retinopathy in clinical settings. However, in real-world settings, the performance of the algorithms has been less than optimal.

In the context of T2D, the capacity of AI to support a precision medicine paradigm necessitates multiple types of genetic, genomic, physiological, environmental biomarkers and behavioral data collection that are assembled and analyzed with methods leveraging augmented AI that can identify patterns without being specifically programmed to find them (*i.e.*, ML). Many biomarker studies suffer from limited sample sizes, making it challenging to detect subtle associations with disease outcomes. Additionally, findings from one study might not replicate consistently in other cohorts due to differences in study populations, methodologies, and data quality. Biomarker discovery is only the initial step. Validating the identified biomarkers in independent populations and demonstrating their clinical utility requires rigorous testing and long-term follow-up studies. This process can be time-consuming and resource-intensive. Genetic and environmental factors can vary across different ethnic populations, leading to differences in disease risk and biomarker expression. This variability complicates the identification of universally applicable biomarkers. The dynamic nature highlights the importance of the timing of biomarker measurements in relation to disease progression and treatment effects. Untangling the effects of these comorbidities on biomarker profiles re-

quires careful consideration and adjustment. As presented in Table **1**, recent investigations continue to be met with challenges of longitudinal assessment, appropriate methods of interrogating progression or non-generalizability. In addition, despite the vast interest in AI and ML applications, few studies have leveraged platforms outside of those already diagnosed with T2D (and frequently T2D with a co-morbidity). While Table **1** presents current research in the field, it is not meant to be an exhaustive list; rather, it provides an overview of the current landscape and highlights the need for comprehensive novel approaches. The present research highlights the lack of 1) a consistent methodological approach to the identification of prognostic biomarker(s), 2) gaps in longitudinal assessments, 3) inclusion of significant population-based differences, and 4) a comprehensive AI or ML-driven approach to identify prognostic biomarkers for T2D. Table **1** also serves as an indication of the persistent challenges in the identification of prognostic biomarkers, even with the availability of AI and ML.

Table 1. Recent publications reporting metabolic, inflammatory and cognitive predictive biomarkers for T2D.

Berezin AE, 2019[15]	Review Paper	Natriuretic peptide system alterations, stress-responsive cytokine, macrophage activation and high-sensitivity cardiac troponins are involved in the pathogenesis of T2D and complications.
Wigger L, *et al*, 2021[16]	Conducted a multi-omics analysis of pancreatic islets pancreatectomized living humans stratified by glucoregulatory capacity.	Islet gene expression was differentially regulated early in pre-T2D individuals as suggested by a progressive, pathogenesis remodeling of mature beta cells
Slieker RC *et al*. 2023[17]	Human pancreatic islets	Homocitrulline, isoleucine 2-aminoadipic acid, eight triacylglycerol species, and lowered sphingomyelin levels were associated with accelerated progression towards the need for insulin. fGDF15/MIC-1, IL-18Ra, CRELD1, NogoR, FAS, and ENPP7 were implicated as contributors to progression, while SMAC/DIABLO, SPOCK1 and HEMK2 appeared protective.
Moin ASM *et al*. 2022[18]	Included twenty-three subjects with T2D and matched non-diabetic control Caucasian subjects (ages 40–70 years).	Lower in T2D Adiponectin, Endocan and Mast/stem cell growth factor receptor-Kit (KIT). Cathepsin-D and Cadherin-E, and Kallikrein-4, Aminoacylase-1, Insulin-like growth factor-binding protein-4 (IGFBP4) and Reticulon-4 receptor (RTN4R) were higher in T2D. No changes in protein level expression were detected in T2D, suggesting protein synthesis is independent of glucose variability.
Zhang Y *et al*. 2021[19]	1932 patients with T2D and acute myocardial infarction (AMI) divided into tertiles according to their triglyceride glucose index (TyG) were included.	The TyG index was positively associated with MACCEs, suggesting that the TyG index is a valid marker for risk stratification and prognosis in patients with T2D and AMI.
Su WY *et al*. 2019[20]	The retrospective study enrolled 3524 patients with T2D 2009 until 2015.	FBG and the TyG represent useful predictive indices over HbA_{1c} and triglyceride for cardiovascular (CV) events and may offer an additional prognostic benefit in T2D
Cheema AK *et al*. 2020[21]	Secondary data analysis of metabolomic and proteomic profiling for investigation of the differential expression of the genes.	STAT3 and HIF, Interleukin 6 (IL6) were found to be predictive of T2D progression. Dysregulation of the coupled expression of (TNF, IL6, LEP, AGT, APOE, F2, SPP1, and INS) was also found to be predictive of T2D progression.
Alur V *et al*. 2023[22]	NexGen sequencing to identify the differentially expressed genes between T2D and healthy controls.	*APP, MYH9, TCTN2, USP7, SYNPO, GRB2, HSP90AB1, UBC, HSPA5*, and *SQSTM1*, might be linked with risk of T2D.
Resl M, *et al*, 2021[23]	Prospective study of 746 patients with T2D, being followed up for 60 months.	Prognostic performance of the biomarkers of interest (GDF-15, NT-proBNP, hs-TnT).
Bai Y *et al*. 2022[24]	A longitudinal cohort study of 516 without diabetes over 9 years, 51 developed T2D, and 92 pre-T2D.	Replacing FBG or OGTT or both with glycated albumin in T2D prediction models made no significant changes to the areas under the curve modeling
Schmidt MI, *et al*, 2019[25]	Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) cohort study of active or retired civil servants, 35-74y	ADA-criterion-based impaired fasting glucose (IFG) has the highest sensitivity but misclassifies ~50% of adults as having intermediate hyperglycemia and poorly predicts T2D.

Morgan-Benita J *et al*. 2022[26]	Implementations of genetic algorithms.	>100 metabolites identified with the five most accurate metabolites considered potential biomarkers for progression were ceramide, phosphatidylcholine, ganoderic acid C2, triglyceride and phosphatidylethanolamine
Banimfreg BH *et al*. 2022[27]	A case-control study with untargeted metabolomics to explore novel T2D biomarkers (n=92)	In the Emirati population, cortisol, glycocholic acid, bile acids, tryptoxine, and the tryptophan metabolite, 5-hydroxyindoleacetic acid were considered
Ungurianu A *et al*. 2022[28]	Review	Among patients with T2D, significant differences were identified in Inflammatory Predictive Biomarkers (AOPPs, AGEs, CRP, CRP/HDL, CRP/IL-6, IL-10/IL-6, IH1) compared to non-T2D controls. Glycemic control was strongly positively influenced (CRP/IL-10) and inversely (IL-10, IL-10/IL-1β ratio).
Kapłon-Cieślicka A *et al*. 2019[29]	284 T2D patients leptin, resistin, and TNF-α concentrations were measured	Higher resistin levels are associated with reduced survival in T2D, irrespective of TNF-α. Leptin was not a prognostic factor.
Garcia-Carretero R, *et al*, 2021[30]	Progression to overt T2D over the 12-year observation period. A total of 1576 hypertensive patients with out T2D.	A machine learning model included a homeostasis model assessment of insulin resistance, fibrinogen, and CRP. This model was more accurate than the logistic regression model, suggesting that inflammatory biomarkers and HOMA-IR have a strong prognostic value in predicting progression to T2D.
Ehtewish H *et al*. 2022[31]	Review	CRP, tau protein, brain-derived neurotrophic factor (BDNF), advanced glycation end products, glycosylated hemoglobin, and adipokines were identified as predictive biomarkers for Cognitive Effects

Indeed, significant strides have been made in identifying prognostic biomarkers for T2D, offering insights into disease mechanisms and potential avenues for personalized treatment. Studies have increasingly integrated data from the vast sources of "omics," including metabolomics, genomics, proteomics, lipidomics, and transcriptomics. Advanced computational techniques have been employed to analyze these large and significantly complex datasets. Computational modeling is exceedingly more efficient and effective than traditional statistics, in which the algorithm is trained to identify patterns and correlations within the data that may not otherwise have been apparent. Prognostic models have been developed using identified biomarkers to predict the risk of T2D onset, combining multiple biomarkers and clinical parameters to enhance prediction accuracy. However, optimization of network-based approaches to understanding the interactions between different molecular components and pathways involved in T2D to provide insights into the underlying mechanisms of the disease and potential therapeutic targets is still in its infancy. Validating identified biomarkers in independent cohorts is a critical step to confirm their reliability and relevance to enhance the credibility of the biomarkers and their potential clinical utility. As such, the discovery of prognostic biomarkers contributes to the development of personalized medicine strategies for T2D. The GATC base algorithm has integrated a multiomics platform to elucidate prognostic biomarkers for T2D. The novel GATC Health algorithms can be trained and optimized with muti-omic data derived from healthy and diseased patients, and a stepwise method for comparison of the ratios between causal biomarkers can afford a significant opportunity for altering the course of T2D.

T2D is well suited for proteomics and lipidomic analysis studies since it is a complex metabolic disease with significant contributions from proteins and lipid byproducts. Proteomic pathway analysis allows for the identification of altered protein expression patterns and provides insights into the signaling pathways involved in metabolic response. Most alterations occur at the protein level. Thus, proteomics allows the survey of simultaneous networks and an unbiased targeting of molecules. The networks and targeted molecules identified can then be applied to addiction, which has the potential for leveraging rigorous clinical application, accurate sample quality control, robust analytical statistics, and adding validity to the observed protein expression changes.

In the integration of algorithms of GATC, it is very likely that advanced bioinformatics will allow the simplification in the interpretation of mega-data sets and accelerate the identification of common underlying areas of physiological relevance. T2D exerts multi-system adaptive physiologic responses in protein synthesis and function. The resultant post-translational modifications comprise a crucial regulatory step in promoting proteomic variability. There are several different cellular activities and processes in determining protein function. However, the functional properties of the protein are fully determined by the collective attributes of the primary protein structure levels of expression and post-translational modification. In addition, the capacity to localize and target specific interactions with biological components such as receptors, ligands, cofactors, nucleic acids, gluconeogenic and lipidomic metabolites, as well as proteins used to assemble macromolecular structures, plays a major role in determining the structure and functionality of various proteins.

The rapid evolution of proteomics has highlighted the magnitude of changes in brain gene and protein expression induced by physiologic and psychological stressors. Further proteomics discovery allows for the inclusion of the synthesis of a variety of enzymes, including (but not limited to) gluco, lipo and protein kinases, endo/exo-nucleases, ligases, phosphatases, proteases, and transferases. By identifying

differentially expressed proteins and analyzing their functional annotations and interactions, we will gain a deeper understanding of the molecular mechanisms underlying the range of physiologic effects of each compound. Proteomic analysis facilitates the discovery of protein biomarkers associated with response and treatment outcomes. By comparing protein expression patterns in patients who progressed to T2D with healthy samples, potential biomarkers indicative of treatment response, drug resistance, or adverse effects will be identified. This, in turn, can be further validated and potentially used for patient stratification, monitoring treatment efficacy, or predicting therapeutic response.

Similarly, significant evidence implicate lipid imbalances as not only predictors of T2D but also causal mediators. Given the apparent links between lipid biosynthesis, metabolism and beta cell dysfunction lead to T2D. Since Minkowski discovered glucosuria, the role of lipids in T2D has been collectively understudied. Herein, we encourage employing lipidomics as an effective approach to identify the myriad of lipid metabolites in relation to pre-T2D and progression to T2D. For building the prognostication model, GATC has integrated ML to select algorithms that are well-suited for handling complex and high-dimensional data, as is often the case with proteomics datasets. The AI model is tested with the training set, while the hyperparameters are tuned by the validation set, and the overall performance is assessed by the test set. The algorithm learns patterns and relationships within the data that correlate with diabetes prognosis. The algorithm has demonstrated reliable performance. As new proteomic data becomes available, GATC will continue to refine and update the AI algorithm to improve its accuracy and adapt to changing patient populations and disease characteristics. In the long term, it is expected that the compounds proposed herein have the potential to be used for individual patients by identifying protein expression patterns and signatures that may influence physiologic response. By characterizing the proteomic profiles with T2D progression or maintenance of glucose regulation (controls), the research team can identify response-specific protein alterations that may guide treatment decisions and optimize future therapeutic strategies. The incorporation of the identified genomic, lipidomic and proteomic information has provided new insight into the pathophysiology of T2D with studies of the biomarkers and T2D suggesting roles for proteins in the progression from pre-T2D to T2D.

Lipidomic studies have also been shown to have the potential to elucidate T2D prognostication by identifying changes in the key constituents of cellular membranes. Lipids, as the major form of cellular energy storage, are integral mediators of structural and cell signaling. Fatty acids metabolism encompasses mobilization from adipose tissue triglyceride stores requiring the activity of TG lipases that generate fatty acids.

Eight major categories, encompassing over 80 major classes, 300 sub-classes, and thousands of lipid species, are represented by the lipidome. The various concentrations of components of the lipidome can be used to identify cellular physiology and pathophysiology. Fatty acids enter the bloodstream and are re-incorporated into triglyceride (TG) by hepatocytes. The lipidomic analysis allows an increased understanding of lipid metabolic pathways[32]. The well-described lipid metabolic pathway begins with the re-esterification of TG, which combines with apolipoprotein-B (APOB), resulting in the formation of very low-density lipoproteins (VLDL). VLDL enters the bloodstream under the regulation of microsomal TG transfer protein (MTTP) in conjunction with a neutral lipid core encapsulated by a phospholipid (PL) monolayer. The PL is enriched in phosphatidylcholine (PC) molecules, which contain polyunsaturated fatty acids (PUFA). Dysregulation of pathways in the synthesis of arachidonic acid (AA; 20:4n-6) and docosahexaenoic acid (DHA; 22:6n-3) alters surface recognition of the VLDL-TG particle and, in turn, FA enters the liver from the adipose tissue. The preferential oxidization of carbohydrates over fat is also dysregulated in T2D and uncouples hepatic TG synthesis from TG secretion, which is cytotoxic. A decline in the concentration of several lipid metabolites (*e.g.*, sphingolipids, TG, PL) is observed in the transition from pre-T2D to T2D. The metabolic byproducts of lipids represent the core components of cell membranes. Many of the metabolomes are linked to T2D progression.

The collective use of proteomic and lipidomic is a viable approach to determining the extent to which lipid biomarkers represent putative early-stage pathophysiological biomarkers for predicting and contributing to the progression of T2D. Untargeted lipidomics provide a possibility to analyze hundreds to thousands of individual lipid species simultaneously. Lipids are a major part of the metabolome. However, their hydrophobicity necessitates the use of different methods from that for aqueous components of the metabolome. To date, the dominant platform in lipidomics is mass spectrometry (MS), often enhanced by interfacing desorption ionization techniques of atmospheric pressure ionization and matrix-assisted laser desorption/ionization (MALDI). Enhancing MS with MALDI allows for conducting in thin-layer chromatography (TLC) plates, in which MALDI –MS imaging has become the more frequently adopted technology in lipidomics, which can provide spatial information to lipid classes in tissues.

The consistent elevation of blood glucose in T2D is often progressive in nature, frequently occurring secondary to the underlying metabolic changes. The amalgamation of hyperglycemia, hyperinsulinemia, and cellular changes previously described increases the risk for disease development and multi-organ damage. Thus, discovery-based lipidomics represents an effective strategy for early prediction and the identification of adverse outcomes pathways associated with T2D. GATC has identified metabolic biomarkers of subsequent T2D (unpublished data). Dyslipidemia is a hallmark of T2D, yet specific lipid molecules closely associated with the initiation and progression of diabetes remain unclear. The targeted lipidomics approach can be used to evaluate the complex lipid changes that occurred long before the diagnosis of T2D and to identify novel lipid markers for screening pre-T2D and T2D. Beyond the conventional risk lipid-based factors linked to T2D and co-morbidities (*i.e.*, plasma TG, total cholesterol (TC), small dense low-density lipoprotein cholesterol (LDL-C), and high-density lipoprotein cholesterol (HDL-C)), the particle size and mechanism by which the factors induce changes across the metabolome are less well-established. The utility beyond that of conventional factors is

warranted, such as large individual differences in other alternative omics and the complexity of the pathophysiological mechanisms of T2D across the disease progression timeline. Accordingly, an enhanced understanding of the lipid-mediated molecular and/or mechanistic pathways underlying T2D sequelae, especially for subtypes of pre-T2D, holds promise for developing innovative techniques to delay, reverse, or prevent gluco-dysregulation prior to clinical disease.

Bioinformatic tools that integrate proteomic and lipidomic data, as in the GATC algorithm, can help identify pathways responsible for metabolite dysregulation and provide novel targets for treatment. Investigating T2D using a system biology approach by integrating the data collection proposed herein with other omics data in the GATC base algorithm can highlight a comprehensive metabolic network used as the basis of causal inference. As such, the overarching objective of validating the GATC base algorithm is to augment molecular understanding, identify potential therapeutic targets, and improve the prevention and management of T2D and its sequelae. Collectively, the establishment of a validated biomarker panel presents tremendous clinical utility. The AI approach allows the focus to rest squarely on science. Understanding the physiology not only allows for early screening but expands the capacity to evaluate response to specific medications, including elucidation of the extent to which drug development pathways can be used to identify individuals who may most appropriately respond physiologically relative to those who may not respond biologically. In the context of T2D, improvement of prevention and risk stratification, in addition to maximization of the effectiveness of interventions and novel insights into the etiology, diagnosis, and prognosis of T2D are highly warranted.

GATC data scientists have the capacity to interrogate data, clean and preprocess the collected data to handle missing values, outliers, and inconsistencies, identify the most informative and relevant features by applying techniques, such as statistical tests, correlation analysis, or dimensionality reduction methods to select or extract the features that contribute the most to the prediction task. GATC scientists also have the capacity to divide the dataset into training, validation, and test sets. The training set is used to train the algorithm, the validation set is used to fine-tune model parameters and hyperparameters, and the test set is used to evaluate the final performance of the trained model. The algorithm learns the underlying patterns and relationships between the features and the target variable based on the provided examples. This involves optimizing the parameters of the model to minimize the prediction error. The hyperparameters of the chosen algorithm are fine-tuned using the validation set. Hyperparameters are configuration settings that are not learned from the data but influence the performance of the model. Techniques such as grid search or random search can be employed to explore different combinations of hyperparameters and select the ones that yield the best performance. The model will be evaluated by applying the trained algorithm to the test set and calculating performance metrics.

To address the lack of predictive biomarkers, with enough discriminative power across the heterogeneity of the disease, GATC is leveraging a machine learning artificial intelligence multi-omics platform but requires further validation (Fig. **1**).

GATC will use the identified and validated biomarkers of T2D development prior to the onset of the disease. Subsequently, machine learning analysis augmented by proteomics and lipidomics will be used in the GATC base algorithm that can predict both the development of persistent concentration of the biomarkers with normoglycemia and T2D prior to the threshold elevation in $HbA1_c$. We believe evaluation of these promising predictive in human cohorts could aid in the development of prognostics and therapeutics. Bioinformatic tools that integrate proteomic and lipidomic data, as in the GATC algorithm, enhance opportunities for pathway identification that may underlie the dysregulation of metabolic parameters and elucidate novel targets for treatment.

The platform of GATC serves to deepen molecular understanding, help identify potential therapeutic targets, and improve the prevention and management of T2D sequelae. Prevention and risk mitigation efforts for pre-T2D and T2D and complications reside in the capacity to maximize the beneficial effects of prevention and intervention efforts. Whereas the specifics of the process may vary depending on the data collected, the base GATC algorithm will have enzyme-linked immunosorbent assay (ELISA), proteomic, and lipidomic datasets integrated, including demographic data, medical history, lifestyle factors, and laboratory test results. GATC data scientists will clean and preprocess the collected data to handle missing values, outliers, and inconsistencies and identify the most informative and relevant features by applying techniques, such as statistical tests, correlation analysis, or dimensionality reduction methods to select or extract the features that contribute the most to the prediction task. Next, the GATC scientists will divide the dataset into training, validation, and test sets. The training set is used to train the algorithm, the validation set is used to fine-tune model parameters and hyperparameters, and the test set is used to evaluate the final performance of the trained model.

The applicability and utility of promising biomarkers in clinical practice require demonstration of clinical utility and novel biomarkers to outperform currently available biomarkers. While the theoretical basis for multi-marker panels suggests increased prognostication relative to a single biomarker, such as $HbA1_c$, an absolute prognostic value is still insufficient for clinical application. Rather than the single biomarker approach, the development of biomarker panels has been proposed to have greater potential for risk prediction. Unfortunately, the high correlation between the putative biomarkers associated with disease states is still lacking; thus, the predictive value of current biomarker panels is limited. A large panel of candidate biomarkers is widely known to play a role in inflammation, endothelial dysfunction, and vascular selection based on *in vitro* and *in vivo* studies[33]. However, the utility of such a panel is significantly limited by the current inclusion of the well-described, yet population-limited biomarkers found associated with relevant outcomes in large longitudinal studies, even when considering confounders and/or additional risk factors. Twelve putative causal T2D biomarkers, thought to be involved in early-stage progression of T2D, have been identified by the GATC team. Stepwise analysis of these biomarkers with modern AI methods can afford a significant

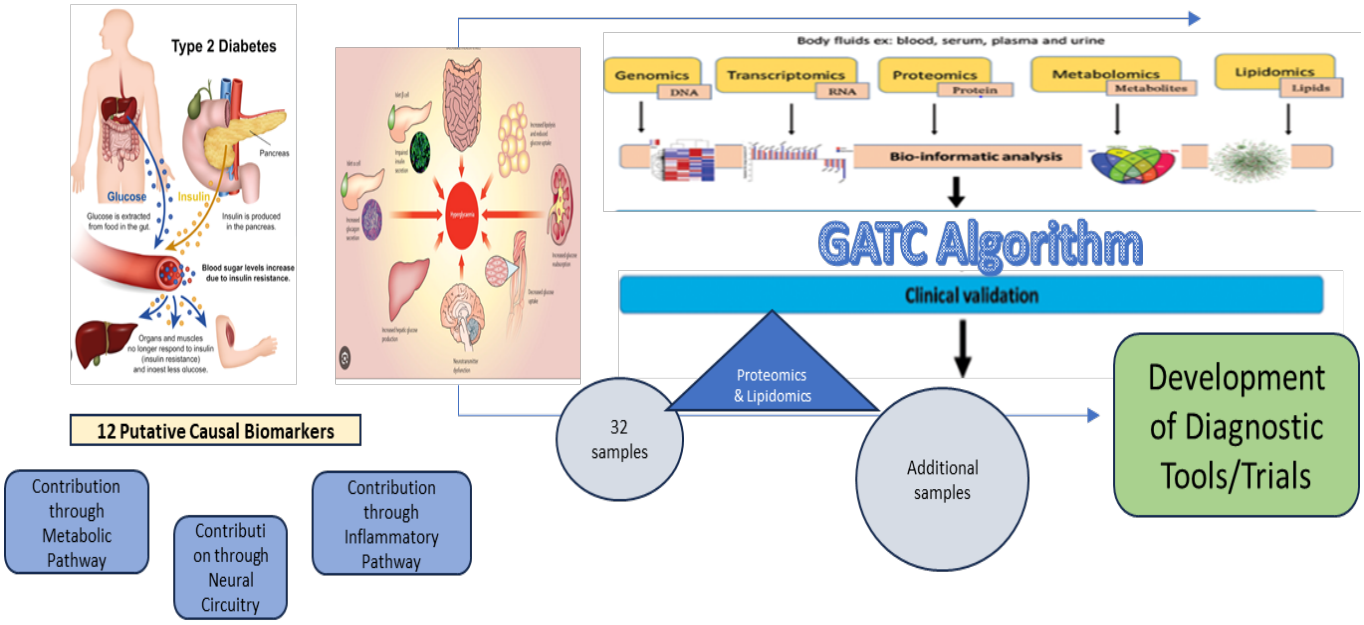

Fig. (1). Leveraging multi-omics to enhance the AI algorithm of GATC for validation of putative causal biomarkers of T2D. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

opportunity for altering the clinical progression of T2D. The putative T2D causal biomarkers include glycated fructosamine albumin (FA); proinsulin; misfolded proinsulin including dimerization and trimerization; hemoglobin A1c; C-reactive protein; Interleukin 18; Interleukin-1 receptor antagonist; lipoxin A4; resolvin E1; resolvinD1; protectinD1 and maresin1. Unpublished results at GATC Health suggest comparison of the ratios of these serum biomarkers, in addition to understanding the relationships between lipid intermediates and the proteome, may provide accurate early prediction (1 to 5 years) of predisposition of an individual to T2D much earlier than elevated blood glucose and HBA1c. However, the uniqueness and novelty proposed herein reside in the longitudinal application of the banked serum samples obtained from 32 individuals diagnosed in the prediabetic state and samples from the same individuals following the onset of T2D symptoms. Stepwise analysis of 12 putative causal biomarkers with modern AI methods can afford a significant opportunity for altering the clinical course and progression of T2D. Biomarkers include glycated fructosamine; proinsulin; misfolded proinsulin; hemoglobin $A1_c$; C-reactive protein; Interleukin 18; Interleukin-1 receptor antagonist; lipoxin A4; resolvin E1; resolvinD1; protectinD1, and maresin1.

1.1. Hemoglobin $A1_c$ ($HbA1_c$)

It is reported as a percentage that represents the relative amount of glycohemoglobin- glucose attached to hemoglobin. The $HbA1_c$ test also provides an algorithmic method of calculating the average level of blood sugar of patients over the past 121 days (*i.e.*, the average lifespan of a red blood cell). The $HbA1_c$ test is routinely used by T2D patients to determine whether they need to adjust their diet and/or medications[34]. This test can also be used to diagnose T2D in new patients but represents the average past blood glucose[35]. Predicate HbA1c tests are in the market and are relatively expensive. HbA1c also has important limits as a prognostic tool. While progression to T2D in patients with elevated $HbA1_c$ is often likely, $HbA1_c$ may not accurately reflect the severity of the organ damage, including the kidney and heart and when $HbA1_c$ reaches the threshold of 6.5%, tissue damage has already ensued. Indeed, a value of less than 48mmol/mol (6.5%) does not exclude diabetes or the myriad of risks associated with poor glucose control. The heterogeneity of clinical presentation and manifestation across tissues has important prognostic implications. A patient without renal or cardiovascular damage and a patient with renal or cardiovascular damage can have an identical HbA1c, but their prognosis is dramatically different. Notwithstanding, $HbA1_c$ represents one important biomarker in a comprehensive suite of predictive candidates[36].

1.2. Misfolded Proinsulin

It is of particular interest in this proposal because it can be an indicator of impaired beta cell function and insulin processing[37]. Endoplasmic reticular (ER) stress signaling may offer the potential to identify new drug targets to treat diabetes. ER, stress plays a significant role in diabetes pathogenesis and contributes to insulin resistance in the context of misfolded proinsulin. Misfolded proinsulin refers to the precursor form of insulin that has not properly folded into its functional structure. Proinsulin is synthesized in the beta cells of the pancreas and undergoes a series of enzymatic processing steps to convert it into mature insulin, which is then released into the bloodstream. In healthy individuals, proinsulin is efficiently converted into mature insulin, and only minimal amounts of unfolded proinsulin are present. However, in individuals with beta cell dysfunction or insulin resistance, the conversion process may be compromised, leading to increased levels of unfolded proinsulin, misfolded, dimerized or trimerized proinsulin, which in some cases is not able to cross into the extracellular matrix. Elevated levels

of unfolded proinsulin have been observed in individuals with impaired glucose tolerance, prediabetes, and early-stage T2D[38]. The constant demand for insulin prohormone in diabetics results in improper tertiary protein folding of the proinsulin during passage through the endoplasmic reticulum. Critical disulfide bonds are not properly formed, resulting in reduced insulin production. Thus, the accumulation of unfolded proinsulin is a consequence of beta cell dysfunction, as the beta cells are unable to properly process and secrete insulin. Measuring unfolded proinsulin levels can provide insights into beta cell health and dysfunction[39].

1.3. Proinsulin

It refers to the properly folded and processed precursor form of insulin. Proinsulin is synthesized in the beta cells of the pancreas and undergoes specific enzymatic cleavage to generate mature insulin, which is then released into the bloodstream. Correctly folded proinsulin is relevant as it reflects the normal functioning of beta cells and the efficient processing of proinsulin into mature insulin. In individuals with healthy beta cell function, most proinsulin is converted into mature insulin, and only minimal amounts of correctly folded proinsulin are present. However, in individuals with impaired beta cell function or insulin resistance, the conversion process may be disrupted, leading to an imbalance between correctly folded proinsulin and mature insulin[12]. Analysis of data from patients as well as tissue samples, indicates that processing of proinsulin is frequently reduced, and at an early stage in type 1 diabetes[40,41]. A recent study implicated cytokines as modulators in interactions between proinsulin and type 1 and T2D[42]. Alterations in the ratio of correctly folded proinsulin to mature insulin have been observed in various stages of T2D[40]. Increased levels of correctly folded proinsulin relative to mature insulin have been associated with impaired glucose tolerance and prediabetes. The adaptive mechanisms as T2D progresses, additive result in the β-cell damage that cannot be restored, leading to chronic metabolic dysfunction, altering glucose metabolism, and leading to a decline in nutrient-regulated secretory functions. Metabolic dysfunction also impairs proinsulin processing and, as a consequence, therefore, there is a deficit in mature insulin-containing secretory granules. This indicates a disruption in the processing of proinsulin, potentially reflecting beta cell dysfunction[43]. Measurement of the ratio of correctly folded proinsulin to mature insulin has been investigated as a potential marker for identifying risk for developing T2D or monitoring progression[44].

1.4. Follistatin

It is a protein that is naturally produced in the body and is expressed in almost all tissues belonging to the TGF-beta superfamily. Follistatin binds TGF-β family members, serving to neutralize the proteins. Follistatin is also essential for the growth and development of muscle fibers and is involved in the development of muscle fiber hypertrophy. As a binding protein for activins and myostatin[45], regulating various biological processes, follistatin is involved in the modulation of inflammation and has gained attention as a potential biomarker due to its involvement in metabolic regulation and insulin sensitivity. Evidence suggests that follistatin has multiple auto- and paracrine functions in various tissues[46]. Follistatin is thought to influence glucose metabolism by modulating the action of insulin and signaling pathways, such that higher levels of follistatin are associated with insulin resistance, impaired glucose tolerance, and increased risk of developing T2D. Serum follistatin has been suggested as a predictive biomarker in gestational diabetes [47,48]. Conversely, lower levels of follistatin have been correlated with improved insulin sensitivity and better glucose control. Thus, measuring follistatin levels serves as a biomarker for identifying individuals at risk of developing diabetes or monitoring disease progression[45].

1.5. C-Reactive Protein (CRP)

While it is important to consider CRP as a useful marker of low-grade inflammation, it is not specific to T2D. However, chronic low-grade inflammation plays a crucial role in the development and progression of T2D. Increased levels of CRP have been observed in individuals with obesity, insulin resistance, and T2D, with an inextricable link between impaired insulin sensitivity and inflammation. Longitudinal studies have shown that elevated CRP levels in apparently healthy individuals are associated with an increased risk of developing T2D later in life [49], suggesting CRP is a potential prognostic biomarker among individuals who are at a higher risk of developing the disease and may benefit from preventive interventions. Further, in the context of the concomitant cardiovascular complications in T2D, CRP has been established as a strong predictor of cardiovascular events in individuals with T2D[50]. Monitoring CRP levels can also facilitate evaluation of treatment response/self-management in individuals with T2D. Monitoring changes in CRP levels can provide valuable insights into the effectiveness of therapeutic interventions in managing T2D and associated complications.

1.6. Interleukin 18 (IL-18)

It is a pro-inflammatory cytokine. The pathogenesis of various inflammatory diseases, including T2D, has been linked to IL-18[51]. While the role of IL-18 is still being elucidated, evidence suggests its involvement in the progression of T2D. IL-18 has been shown to promote inflammation in both adipose tissue and skeletal muscle, leading to the disruption of insulin signaling pathways and subsequent insulin resistance. Additionally, IL-18 has been implicated in the impairment of pancreatic beta-cell function, which further contributes to the development and progression of T2D. IL-18 levels are elevated in obese individuals, particularly those with visceral adiposity. Adipose tissue produces and releases IL-18, and its increased secretion in obesity contributes to a state of chronic inflammation, promoting insulin resistance and impairing glucose metabolism. Elevated IL-18 levels represent an independent contributor to the risk of cardiovascular events [52].

1.7. Interleukin 1 (IL-1Ra)

It is a naturally occurring protein that competitively inhibits the actions of interleukin 1 (IL-1), a pro-inflammatory cytokine. IL-1 plays a role in promoting inflammation and impairing insulin signaling. IL-1Ra acts as an endogenous inhibitor of IL-1, counteracting its effects and reducing inflammation. By modulating the inflammatory response, IL-

1Ra may help improve insulin sensitivity and mitigate the progression of insulin resistance in T2D. Whereas IL-1 can exert detrimental effects on beta-cells, leading to impaired insulin secretion and increased beta-cell apoptosis, IL-1Ra counteracts these effects by blocking IL-1R and protecting beta-cells from the damaging effects of IL-1. By preserving beta-cell function and survival, IL-1Ra may help maintain adequate insulin production and contribute to glycemic control. Several clinical trials have investigated the therapeutic potential of IL-1Ra[53,54].

1.8. Glycated Fructosamine Albumin (FA)

Glycated Fructosamine Albumin (FA) in circulation includes all the stable ketoamines in circulation. FA is produced through the non-enzymatic glycation of circulating serum proteins (albumins, globulins, and other minority proteins). In T2D, the concentration of circulating FA increases due to the increased glycation products due to impaired glucose regulation in the blood. Elevated FA levels have also been implicated as causal in the incidence of vascular complications associated with T2D, and high FA levels indicate a more aggressive disease progression. As such, FA reflects glucose control. Unlike $HbA1_c$, FA reflects glucose dysregulation over a two-to-three-week period[55].

1.9. Lipid Mediators

They are long-chain polyunsaturated fatty acids that play a well-established role in modulating the inflammatory cascade as well as contributing to peripheral glucose uptake and insulin response. More recent mechanistic investigations have revealed glucose regulatory involvement is at least in part related to specialized pro-resolving lipid mediators (SPMs), resolvins, maresins, and protectins.

1.10. Lipoxin A4 (LXA4)

It is a SPM made from arachidonic acid (AA). As a part of the AA pathway, LXA4 plays a role in the resolution of inflammation and the regulation of immune responses. LXA4 has been shown to have anti-inflammatory properties, inhibiting the production of pro-inflammatory molecules and promoting the resolution of inflammation, plausibly mitigating processes associated with insulin resistance and beta-cell dysfunction. LXA4 has been shown to preserve beta-cell mass and function[56], thereby delaying progressive pancreatic decline and/or delaying the need for exogenous insulin.

1.11. Resolvin E1 (RvE1) and Resolvin D1 (RvD1):

They are SPMs derived from eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), respectively. Protectin D1 (PD1) and Maresin 1 (MaR1) are SPMs derived primarily from DHA. The SPMs are involved in the resolution of inflammation and the regulation of immune responses, with potent anti-inflammatory actions, counteracting the inflammatory processes associated with insulin resistance and beta-cell dysfunction. RvE1 has shown potential in enhancing insulin sensitivity in various target tissues, such as adipose tissue and skeletal muscle [57]. RvE1 and RvD1 have also been shown to have beneficial effects on lipid metabolism by reducing triglyceride levels, improving high-density lipoprotein (HDL) cholesterol function, and promoting lipid clearance. These lipid-modulating properties of RvE1 may have implications for managing cardiovascular risk factors associated with T2D. PD1 has been found to modulate lipid metabolism by reducing triglyceride levels and promoting lipid clearance[58].

Augmenting the utility of the base algorithm of GATC with the addition of these 12 identified putative causal biomarkers for T2D allows the algorithm to capture a broader range of biological signals, leading to a more accurate and comprehensive predictive model, improving the ability of the algorithm to identify individuals at risk of developing the disease. As different individuals may have varying levels of risk for developing T2D based on their biomarker profiles, incorporating multiple biomarkers, the algorithm can stratify individuals into different risk categories, enabling targeted interventions and personalized treatment strategies. Further, the use of these evidenced-based multiple biomarker strategies provides valuable insights into the underlying biological mechanisms. Enhancing the GATC algorithm with proteomic and lipidomic data allows greater capacity to analyze how these biomarkers interact and contribute to disease progression can enhance our understanding of the complex pathways involved in the condition. Incorporating these 12 identified putative causal biomarkers increases the likelihood of the validity and reproducibility of the algorithm. It becomes easier to replicate the findings in independent datasets, which is essential for building trust in the predictive model. For monitoring the progression or treatment response in individuals, a panel of biomarkers can provide a more comprehensive assessment of how the disease is changing over time or how it responds to interventions increasing the chances early detection, providing an opportunity for timely intervention to prevent or delay the onset of the disease with adaptability to changes in our understanding of the disease. As new research identifies additional relevant biomarkers or refines the list of causal factors, the model can be updated without significant restructuring. Indeed, a robust algorithm based on 12 biomarkers can serve as a valuable clinical decision support tool, however, it is important to note that while using a panel of biomarkers has significant advantages, the quality of the biomarkers, the size and diversity of the training dataset, proper validation, and integration into clinical practice are crucial factors that determine the success of such an algorithm. GATC leverages collaborations between biomarker researchers, data scientists, and healthcare providers to realize the full potential of predictive models based on these biomarkers. The algorithm learns the underlying patterns and relationships between the features and the target variable (T2D) based on the provided examples. This involves optimizing the parameters of the model to minimize the prediction error. The hyperparameters of the chosen algorithm(s)are fine-tuned using the validation set. Hyperparameters are configuration settings that are not learned from the data but influence the model's performance. Techniques such as grid search or random search can be employed to explore different combinations of hyperparameters and select the ones that yield the best performance. These metrics provide insights into the predictive ability and generalization of the algorithm to unseen data. If the model performance is unsatisfactory, iterate on the previous steps by adjusting the feature selection, trying different algorithms, or modifying

hyperparameters until a satisfactory performance is achieved. Once the model is deemed satisfactory, it can be deployed in real-world scenarios for diabetes prediction. Ongoing monitoring and validation may be necessary to ensure that the performance of the model remains reliable over time and to address any potential biases or limitations that may arise.

A large panel of candidate biomarkers is widely known to play a role in inflammation, endothelial dysfunction, vascular selected based on *in vitro* and *in vivo* studies [58,33,59]. The uniqueness and novelty proposed herein resides in the longitudinal application of the banked serum samples obtained from 32 individuals diagnosed in the prediabetic state and samples from the same individuals following the onset of T2D symptoms. Stepwise analysis of 12 putative causal biomarkers with modern AI methods can afford a significant opportunity for altering the clinical course and progression of T2D. ELISA is used to measure and quantify analyte levels. Serum samples were banked from 32 healthy, aged-matched non-T2D controls. While the theoretical basis for multi-marker panels suggested increased prognostication relative to a single biomarker such as HbA1c, absolute prognostic value is still insufficient for clinical application.

The identical serum samples used in the ELISA study will be used to measure protein levels in the 96 serum samples to elucidate changes in protein expression and modifications in the progression of pre-to T2D state, compared to healthy control individuals. The analysis will involve a shotgun approach that integrates measurements of 16 classes of lipids, with subsequent calculation of maresin, protectin, and lipoxin. By identifying differentially expressed proteins and analyzing their functional annotations and interactions, we will gain a deeper understanding of the molecular mechanisms underlying the range of physiologic effects of each compound. Proteomic analysis will also be used to facilitate the discovery of protein biomarkers associated with response and treatment outcomes. By comparing protein expression patterns in T2D progressed and healthy samples, potential biomarkers indicative of treatment response, drug resistance, or adverse effects will be identified, which in turn can be further validated and potentially used for patient stratification, monitoring treatment efficacy, or predicting therapeutic response. In the long term, it is expected that the compounds proposed herein will serve as a foundation for the potential for personalized medicine by identifying patient-specific protein expression patterns and signatures that may influence drug response. By characterizing the proteomic profiles with T2D progression or maintenance of glucose regulation (controls), the research team can identify response-specific protein alterations that may guide treatment decisions and optimize future therapeutic strategies.

CONCLUSION

Using AI for identifying predictive biomarkers for type 2 diabetes offers several potential benefits that can significantly impact the understanding, diagnosis, and management of the disease. The GATC algorithm can efficiently analyze large and complex datasets, including genetic, genomic, and clinical data, enabling our interdisciplinary team to identify patterns and correlations that may be challenging for traditional methods to uncover, as well as identify subtle patterns and relationships in data that may serve as early indicators for proactive intervention, potentially preventing the development or progression of the disease. The potential for the development of personalized treatment plans by considering individual variations in genetics, lifestyle, and other factors can lead to more targeted and effective interventions, optimizing patient outcomes *via* the integration of information from various sources. The GATC algorithm excels at identifying complex patterns and interactions within these factors, providing a deeper understanding of the underlying mechanisms of the disease.

AI can aid in the identification of potential drug targets and the development of new therapies. By understanding the molecular pathways involved in T2D, we enhance our capacity to leverage specific biomarkers for drug discovery, potentially leading to more effective and targeted treatments. Further, AI-powered tools, such as the GATC algorithm, can enable continuous monitoring of patients, allowing for real-time adjustments to treatment plans based on dynamic changes in biomarkers and other health parameters. The comprehensive approach presented herein can accelerate progress in understanding T2D and holds great promise in revolutionizing our approach to the disease, from early detection to personalized treatment and ongoing management.

LIST OF ABBREVIATIONS

ADA	=	American Diabetes Association
AI	=	Augmented Intelligence
APOB	=	Apolipoprotein B
AUC-ROC	=	Area Under the Receiver Operating Characteristic Curve
CRP	=	C-Reactive Protein
DHA	=	Docosahexaenoic Acid
DL	=	Deep Learning
ELISA	=	Enzyme-linked Immunosorbent Assay
FA	=	Glycated Fructosamine Albumin
FBG	=	Fasting Blood Glucose
FDA	=	Food and Drug Administration
HbA1c	=	Hemoglobin A1c (Glycated Hemoglobin)
HDL	=	High-Density Lipoprotein Cholesterol
HOMA-IR	=	Homeostatic Model Assessment for Insulin Resistance
IDF	=	International Diabetes Federation
IFG	=	Impaired Fasting Glucose
IL	=	Interleukin
LC	=	Liquid Chromatography
LDL	=	Low-Density Lipoprotein Cholesterol
MALDI	=	Matrix-Assisted Laser Desorption/ Ionization
ML	=	Machine Learning

MTTP	=	Microsomal TG Transfer Protein
NP	=	Natriuretic Peptide
PL	=	Phospholipid
PUFA	=	Polyunsaturated Fatty Acids
RCT	=	Randomized Control Trial
SPM	=	Specialized Pro-Resolving Lipid Mediators
T2D	=	Type 2 Diabetes
TLC	=	Thin-layer Chromatography
TNF	=	Tumor Necrosis Factor
TG	=	Triglyceride
TyG	=	Triglyceride Glucose Index

AUTHORS' INFORMATION

JRTL is a scientific advisor and Chair of the Scientific Advisory Board at GATC. IJ is the CSO of GATC and has a patent pending for the algorithm. EM is a scientific innovation consultant. WL serves as a chief advisor for GATC. KC serves as a scientific consultant for GATC.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

Financial support for this review paper was provided by GATC Health, Irvine, CA.

CONFLICTS OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

The authors would like to thank Michael Alexander from the Department of Surgery and the University of California Irvine for his assistance.

REFERENCES

[1] Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2018; 14(2): 88-98. http://dx.doi.org/10.1038/nrendo.2017.151 PMID: 29219149

[2] Evans M, Chandramouli AS, Faurby M, Matthiessen KS, Mogensen PB, Verma S. Healthcare costs and hospitalizations in US patients with type 2 diabetes and cardiovascular disease: A retrospective database study (OFFSET). Diabetes Obes Metab 2022; 24(7): 1300-9. http://dx.doi.org/10.1111/dom.14703 PMID: 35504854

[3] Ogurtsova K, Guariguata L, Barengo NC, *et al.* IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021. Diabetes Res Clin Pract 2022; 183: 109118. http://dx.doi.org/10.1016/j.diabres.2021.109118 PMID: 34883189

[4] Diabetes statistics. Available from: https://www.niddk.nih.gov/health-information/health-statistics/diabetes-statistics

[5] 2. Classification and diagnosis of diabetes: *Standards of medical care in diabetes—2019*. Diabetes Care 2019; 42(Suppl. 1): S13-28. http://dx.doi.org/10.2337/dc19-S002 PMID: 30559228

[6] Tucker LA. Limited agreement between classifications of diabetes and prediabetes resulting from the OGTT, Hemoglobin A1c, and fasting glucose tests in 7412 U.S. J Clin Med 2020; 9(7): 2207. http://dx.doi.org/10.3390/jcm9072207

[7] Khosla L, Bhat S, Fullington LA, Horlyck-Romanovsky MF. HbA $_{1c}$ performance in african descent populations in the united states with normal glucose tolerance, prediabetes, or diabetes: A scoping review. Prev Chronic Dis 2021; 18: 200365. http://dx.doi.org/10.5888/pcd18.200365 PMID: 33705304

[8] Gomez-Peralta F, Choudhary P, Cosson E, Irace C, Rami-Merhar B, Seibold A. Understanding the clinical implications of differences between glucose management indicator and glycated haemoglobin. Diabetes Obes Metab 2022; 24(4): 599-608. http://dx.doi.org/10.1111/dom.14638 PMID: 34984825

[9] Weykamp C. HbA1c: A review of analytical and clinical aspects. Ann Lab Med 2013; 33(6): 393-400. http://dx.doi.org/10.3343/alm.2013.33.6.393 PMID: 24205486

[10] Vigersky RA. Going beyond HbA1c to understand the benefits of advanced diabetes therapies. J Diabetes 2019; 11(1): 23-31. http://dx.doi.org/10.1111/1753-0407.12846 PMID: 30151979

[11] Chintagari NR, Jana S, Alayash AI. Oxidized ferric and ferryl forms of hemoglobin trigger mitochondrial dysfunction and injury in alveolar type I cells. Am J Respir Cell Mol Biol 2016; 55(2): 288-98. http://dx.doi.org/10.1165/rcmb.2015-0197OC PMID: 26974230

[12] Perugini RA, Malkani S. Remission of type 2 diabetes mellitus following bariatric surgery: Review of mechanisms and presentation of the concept of 'reversibility'. Curr Opin Endocrinol Diabetes Obes 2011; 18(2): 119-28. http://dx.doi.org/10.1097/MED.0b013e3283446c1f PMID: 21522001

[13] Mottalib A, Sakr M, Shehabeldin M, Hamdy O. Diabetes Remission after nonsurgical intensive lifestyle intervention in obese patients with type 2 diabetes. J Diabetes Res 2015; 2015: 1-4. http://dx.doi.org/10.1155/2015/468704 PMID: 26114120

[14] Azarova I, Polonikov A, Klyosova E. Molecular genetics of abnormal redox homeostasis in type 2 diabetes mellitus. Int J Mol Sci 2023; 24(5): 4738. http://dx.doi.org/10.3390/ijms24054738 PMID: 36902173

[15] Berezin AE. Prognostication of clinical outcomes in diabetes mellitus: Emerging role of cardiac biomarkers. Diabetes Metab Syndr 2019; 13(2): 995-1003. http://dx.doi.org/10.1016/j.dsx.2019.01.018 PMID: 31336558

[16] Wigger L, Barovic M, Brunner AD, *et al.* Multi-omics profiling of living human pancreatic islet donors reveals heterogeneous beta cell trajectories towards type 2 diabetes. Nat Metab 2021; 3(7): 1017-31. http://dx.doi.org/10.1038/s42255-021-00420-9 PMID: 34183850

[17] Slieker RC, Donnelly LA, Akalestou E, *et al.* Identification of biomarkers for glycaemic deterioration in type 2 diabetes. Nat Commun 2023; 14(1): 2533. http://dx.doi.org/10.1038/s41467-023-38148-7 PMID: 37137910

[18] Moin ASM, Sathyapalan T, Atkin SL, Butler AE. Diagnostic and prognostic protein biomarkers of β-cell function in type 2 diabetes and their modulation with glucose normalization. Metabolites 2022; 12(3): 196. http://dx.doi.org/10.3390/metabo12030196 PMID: 35323639

[19] Zhang Y, Ding X, Hua B, *et al.* Predictive effect of triglyceride-glucose index on clinical events in patients with type 2 diabetes mellitus and acute myocardial infarction: Results from an observational cohort study in China. Cardiovasc Diabetol 2021; 20(1): 43. http://dx.doi.org/10.1186/s12933-021-01236-3 PMID: 33573649

[20] Su WY, Chen SC, Huang YT, *et al.* Comparison of the effects of fasting glucose, hemoglobin A_{1c}, and triglyceride–glucose index on cardiovascular events in type 2 diabetes mellitus. Nutrients 2019; 11(11): 2838. http://dx.doi.org/10.3390/nu11112838 PMID: 31752391

[21] Cheema AK, Kaur P, Fadel A, Younes N, Zirie M, Rizk NM. Integrated datasets of proteomic and metabolomic biomarkers to predict its impacts on comorbidities of type 2 diabetes mellitus. Diabetes Metab Syndr Obes 2020; 13: 2409-31. http://dx.doi.org/10.2147/DMSO.S244432 PMID: 32753925

[22] Alur V, Raju V, Vastrad B, Vastrad C, Kavatagimath S, Kottarshetti S. Bioinformatics analysis of next generation sequenc-

ing data identifies molecular biomarkers associated with type 2 diabetes mellitus. Clin Med Insights Endocrinol Diabetes 2023; 16 http://dx.doi.org/10.1177/11795514231155635 PMID: 36844983

[23] Resl M, Vila G, Heinzl M, *et al.* Changes in the prognostic values of modern cardiovascular biomarkers in relation to duration of diabetes mellitus. J Diabetes Complications 2021; 35(9): 107990. http://dx.doi.org/10.1016/j.jdiacomp.2021.107990 PMID: 34294516

[24] Bai Y, Fang Y, Ming J, *et al.* Serum glycated albumin as good biomarker for predicting type 2 diabetes: A retrospective cohort study of china national diabetes and metabolic disorders survey. Diabetes Metab Res Rev 2022; 38(1): e3477. http://dx.doi.org/10.1002/dmrr.3477 PMID: 34041844

[25] Schmidt MI, Bracco PA, Yudkin JS, *et al.* Intermediate hyperglycaemia to predict progression to type 2 diabetes (ELSA-Brasil): An occupational cohort study in Brazil. Lancet Diabetes Endocrinol 2019; 7(4): 267-77. http://dx.doi.org/10.1016/S2213-8587(19)30058-0 PMID: 30803929

[26] Morgan-Benita J, Sánchez-Reyna AG, Espino-Salinas CH, *et al.* Metabolomic selection in the progression of type 2 diabetes mellitus: A genetic algorithm approach. Diagnostics 2022; 12(11): 2803. http://dx.doi.org/10.3390/diagnostics12112803 PMID: 36428864

[27] Banimfreg BH, Shamayleh A, Alshraideh H, Semreen MH, Soares NC. Untargeted approach to investigating the metabolomics profile of type 2 diabetes emiratis. J Proteomics 2022; 269: 104718. http://dx.doi.org/10.1016/j.jprot.2022.104718 PMID: 36100153

[28] Ungurianu A, Zanfirescu A, Grădinaru D, Ionescu-Tîrgovişte C, Dănciulescu Miulescu R, Margină D. Interleukins and redox impairment in type 2 diabetes mellitus: Mini-review and pilot study. Curr Med Res Opin 2022; 38(4): 511-22. http://dx.doi.org/10.1080/03007995.2022.2033049 PMID: 35067142

[29] Kapłon-Cieślicka A, Tymińska A, Rosiak M, *et al.* Resistin is a prognostic factor for death in type 2 diabetes. Diabetes Metab Res Rev 2019; 35(2): e3098. http://dx.doi.org/10.1002/dmrr.3098 PMID: 30447052

[30] Garcia-Carretero R, Vigil-Medina L, Barquero-Perez O. The use of machine learning techniques to determine the predictive value of inflammatory biomarkers in the development of type 2 diabetes mellitus. Metab Syndr Relat Disord 2021; 19(4): 240-8. http://dx.doi.org/10.1089/met.2020.0139 PMID: 33596118

[31] Ehtewish H, Arredouani A, El-Agnaf O. Diagnostic, prognostic, and mechanistic biomarkers of diabetes mellitus-associated cognitive decline. Int J Mol Sci 2022; 23(11): 6144. http://dx.doi.org/10.3390/ijms23116144 PMID: 35682821

[32] Khan SR, Mohan H, Liu Y, *et al.* The discovery of novel predictive biomarkers and early-stage pathophysiology for the transition from gestational diabetes to type 2 diabetes. Diabetologia 2019; 62(4): 687-703. http://dx.doi.org/10.1007/s00125-018-4800-2 PMID: 30645667

[33] Khanijou V, Zafari N, Coughlan MT, MacIsaac RJ, Ekinci EI. Review of potential biomarkers of inflammation and kidney injury in diabetic kidney disease. Diabetes Metab Res Rev 2022; 38(6): e3556. http://dx.doi.org/10.1002/dmrr.3556 PMID: 35708187

[34] Vijayakumar P, Nelson RG, Hanson RL, Knowler WC, Sinha M. HbA1c and the prediction of type 2 diabetes in children and adults. Diabetes Care 2017; 40(1): 16-21. http://dx.doi.org/10.2337/dc16-1358 PMID: 27810987

[35] Emery KA, Robins J, Salyer J, Thurby-Hay L, Djira G. Type 2 diabetes self-management variables and predictors. Clin Nurs Res 2022; 31(7): 1250-62. http://dx.doi.org/10.1177/10547738211067322 PMID: 34961341

[36] Li Z, Han D, Qi T, *et al.* Hemoglobin A1c in type 2 diabetes mellitus patients with preserved ejection fraction is an independent predictor of left ventricular myocardial deformation and tissue abnormalities. BMC Cardiovasc Disord 2023; 23(1): 49. http://dx.doi.org/10.1186/s12872-023-03082-5 PMID: 36698087

[37] Ajoolabady A, Liu S, Klionsky DJ, *et al.* ER stress in obesity pathogenesis and management. Trends Pharmacol Sci 2022; 43(2): 97-109. http://dx.doi.org/10.1016/j.tips.2021.11.011 PMID: 34893351

[38] Sun J, Cui J, He Q, Chen Z, Arvan P, Liu M. Proinsulin misfolding and endoplasmic reticulum stress during the development and progression of diabetes. Mol Aspects Med 2015; 42: 105-18. http://dx.doi.org/10.1016/j.mam.2015.01.001 PMID: 25579745

[39] Arunagiri A, Haataja L, Cunningham CN, *et al.* Misfolded proinsulin in the endoplasmic reticulum during development of beta cell failure in diabetes. Ann N Y Acad Sci 2018; 1418(1): 5-19. http://dx.doi.org/10.1111/nyas.13531 PMID: 29377149

[40] Rodriguez-Calvo T, Chen YC, Verchere CB, *et al.* Altered β-cell prohormone processing and secretion in type 1 diabetes. Diabetes 2021; 70(5): 1038-50. http://dx.doi.org/10.2337/dbi20-0034 PMID: 33947721

[41] Sokooti S, Dam WA, Szili-Torok T, *et al.* Fasting proinsulin independently predicts incident type 2 diabetes in the general population. J Pers Med 2022; 12(7): 1131. http://dx.doi.org/10.3390/jpm12071131 PMID: 35887628

[42] Tran DT, Pottekat A, Lee K, *et al.* Inflammatory cytokines rewire the proinsulin interaction network in human islets. J Clin Endocrinol Metab 2022; 107(11): 3100-10. http://dx.doi.org/10.1210/clinem/dgac493 PMID: 36017587

[43] Rohli KE, Boyer CK, Blom SE, Stephens SB. Nutrient regulation of pancreatic islet β-cell secretory capacity and insulin production. Biomolecules 2022; 12(2): 335. http://dx.doi.org/10.3390/biom12020335 PMID: 35204835

[44] Then C, Gar C, Thorand B, *et al.* Proinsulin to insulin ratio is associated with incident type 2 diabetes but not with vascular complications in the KORA F4/FF4 study. BMJ Open Diabetes Res Care 2020; 8(1): e001425. http://dx.doi.org/10.1136/bmjdrc-2020-001425 PMID: 32423965

[45] Wu C, Borné Y, Gao R, *et al.* Elevated circulating follistatin associates with an increased risk of type 2 diabetes. Nat Commun 2021; 12(1): 6486. http://dx.doi.org/10.1038/s41467-021-26536-w PMID: 34759311

[46] Hedger MP, Winnall WR, Phillips DJ, de Kretser DM. The regulation and functions of activin and follistatin in inflammation and immunity. Vitam Horm 2011; 85: 255-97. http://dx.doi.org/10.1016/B978-0-12-385961-7.00013-5 PMID: 21353885

[47] Lorenzo-Almorós A, Hang T, Peiró C, *et al.* Predictive and diagnostic biomarkers for gestational diabetes and its associated metabolic and cardiovascular diseases. Cardiovasc Diabetol 2019; 18(1): 140. http://dx.doi.org/10.1186/s12933-019-0935-9 PMID: 31666083

[48] Näf S, Escote X, Ballesteros M, *et al.* Serum activin A and follistatin levels in gestational diabetes and the association of the Activin A-Follistatin system with anthropometric parameters in offspring. PLoS One 2014; 9(4): e92175. http://dx.doi.org/10.1371/journal.pone.0092175 PMID: 24763182

[49] Calle MC, Fernandez ML. Inflammation and type 2 diabetes. Diabetes Metab 2012; 38(3): 183-91. http://dx.doi.org/10.1016/j.diabet.2011.11.006 PMID: 22252015

[50] Sharif S, Van der Graaf Y, Cramer MJ, *et al.* Low-grade inflammation as a risk factor for cardiovascular events and all-cause mortality in patients with type 2 diabetes. Cardiovasc Diabetol 2021; 20(1): 220. http://dx.doi.org/10.1186/s12933-021-01409-0 PMID: 34753497

[51] Banerjee M, Saxena M. Interleukin-1 (IL-1) family of cytokines: Role in type 2 diabetes. Clin Chim Acta 2012; 413(15-16): 1163-70. http://dx.doi.org/10.1016/j.cca.2012.03.021 PMID: 22521751

[52] Suri S, Mitra P, Abhilasha A, *et al.* Role of interleukin-2 and interleukin-18 in newly diagnosed type 2 diabetes mellitus. J Basic Clin Physiol Pharmacol 2022; 33(2): 185-90. http://dx.doi.org/10.1515/jbcpp-2020-0272 PMID: 33711216

[53] Akash MSH, Shen Q, Rehman K, Chen S. Interleukin-1 receptor antagonist: A new therapy for type 2 diabetes mellitus. J Pharm Sci 2012; 101(5): 1647-58. http://dx.doi.org/10.1002/jps.23057 PMID: 22271340

[54] Volarevic V, Al-Qahtani A, Arsenijevic N, Pajovic S, Lukic ML. Interleukin-1 receptor antagonist (IL-1Ra) and IL-1Ra producing mesenchymal stem cells as modulators of diabetogenesis. Autoimmunity 2010; 43(4): 255-63. http://dx.doi.org/10.3109/08916930903305641 PMID: 19845478

[55] Das UN. Arachidonic acid and lipoxin A4 as possible endogenous anti-diabetic molecules. Prostaglandins Leukot Essent Fatty Acids 2013; 88(3): 201-10. http://dx.doi.org/10.1016/j.plefa.2012.11.009 PMID: 23295193

[56] Repossi G, Das UN, Eynard AR. Molecular basis of the beneficial actions of resveratrol. Arch Med Res 2020; 51(2): 105-14. http://dx.doi.org/10.1016/j.arcmed.2020.01.010 PMID: 32111491

[57] Freire MO, Dalli J, Serhan CN, Van Dyke TE. Neutrophil resolvin E1 receptor expression and function in type 2 diabetes. J Immunol 2017; 198(2): 718-28. http://dx.doi.org/10.4049/jimmunol.1601543 PMID: 27994073

[58] Aydin Ozgur B, Coskunpinar E, Bilgic Gazioglu S, *et al.* Effects of complement regulators and chemokine receptors in type 2 diabetes. Immunol Invest 2021; 50(5): 478-91. http://dx.doi.org/10.1080/08820139.2020.1778022 PMID: 32611246

[59] Yang Q, Vijayakumar A, Kahn BB. Metabolites as regulators of insulin sensitivity and metabolism. Nat Rev Mol Cell Biol 2018; 19(10): 654-72. http://dx.doi.org/10.1038/s41580-018-0044-8 PMID: 30104701