Article

Liquid Biopsy in Type 2 Diabetes Mellitus Management: Building Specific Biosignatures via Machine Learning

Makrina Karaglani [1], Maria Panagopoulou [1], Christina Cheimonidi [1], Ioannis Tsamardinos [2], Efstratios Maltezos [3], Nikolaos Papanas [3], Dimitrios Papazoglou [3], George Mastorakos [4] and Ekaterini Chatzaki [1,5,*]

1 Laboratory of Pharmacology, Department of Medicine, Democritus University of Thrace, 68100 Alexandroupolis, Greece; mkaragla@med.duth.gr (M.K.); mpanagop@med.duth.gr (M.P.); chris_cheimonidi@imbb.forth.gr (C.C.)
2 JADBio Gnosis DA, Science and Technology Park of Crete, 71500 Heraklion, Greece; tsamard.it@gmail.com
3 Diabetes Centre, 2nd Department of Internal Medicine, Democritus University of Thrace, University Hospital of Alexandroupolis, 68100 Alexandroupolis, Greece; emaltez@med.duth.gr (E.M.); papanasnikos@yahoo.gr (N.P.); dpapazog@med.duth.gr (D.P.)
4 Endocrine Unit, 2nd Department of Obstetrics and Gynecology, National and Kapodistrian University of Athens, "Aretaieion" University Hospital, 11528 Athens, Greece; mastorakg@gmail.com
5 Institute of Agri-Food and Life Sciences, Hellenic Mediterranean University Research Centre, 71003 Heraklion, Greece
* Correspondence: achatzak@med.duth.gr; Tel.: +30-255-103-0533

Abstract: Background: The need for minimally invasive biomarkers for the early diagnosis of type 2 diabetes (T2DM) prior to the clinical onset and monitoring of β-pancreatic cell loss is emerging. Here, we focused on studying circulating cell-free DNA (ccfDNA) as a liquid biopsy biomaterial for accurate diagnosis/monitoring of T2DM. Methods: ccfDNA levels were directly quantified in sera from 96 T2DM patients and 71 healthy individuals via fluorometry, and then fragment DNA size profiling was performed by capillary electrophoresis. Following this, ccfDNA methylation levels of five β-cell-related genes were measured via qPCR. Data were analyzed by automated machine learning to build classifying predictive models. Results: ccfDNA levels were found to be similar between groups but indicative of apoptosis in T2DM. *INS* (Insulin), *IAPP* (Islet Amyloid Polypeptide-Amylin), *GCK* (Glucokinase), and *KCNJ11* (Potassium Inwardly Rectifying Channel Subfamily J member 11) levels differed significantly between groups. AutoML analysis delivered biosignatures including *GCK*, *IAPP* and *KCNJ11* methylation, with the highest ever reported discriminating performance of T2DM from healthy individuals (AUC 0.927). Conclusions: Our data unravel the value of ccfDNA as a minimally invasive biomaterial carrying important clinical information for T2DM. Upon prospective clinical evaluation, the built biosignature can be disruptive for T2DM clinical management.

Keywords: type 2 diabetes; circulating cell free DNA; DNA methylation; machine learning

1. Introduction

Just under half a billion people are living with diabetes worldwide, and the number is projected to increase by 25% in 2030 and 51% in 2045 [1]. Diabetes is a serious cause of blindness, kidney failure, stroke and amputations. Type 2 Diabetes mellitus (T2DM), the most common type of diabetes, is characterized by inadequate beta-pancreatic cell (β-cell) function, insulin insensitivity and chronic inflammation, all of which progressively lead to impaired glucose homeostasis [2]. In post-mortem specimens of T2DM patients, β-cell mass is reduced by 30–40% compared with specimens from non-diabetic subjects [3]. Increased β-cell apoptosis and reduced functional capacity of the remaining cells are important factors that contribute to the onset and the progression of the disease [4]. The inability to detect diabetes before the development of hyperglycemia limits our power for

diagnosis prior to clinical onset and for earlier interventions to preserve significant β-cell mass. Established biomarkers for monitoring T2DM progression such as HbA1c levels are unable to consistently and non-invasively detect/monitor the ongoing β-cell destruction in islets. Thus, as the number of people with T2DM worldwide keeps on rising, there is an emerging need for the development of minimally invasive biomarkers for the early diagnosis of diabetes as well as the development of monitoring potential, both of which could lead to better therapeutic decisions.

Given the significant obstacle of directly accessing the pancreas for biopsy, special attention has been recently given to minimally invasive biomarkers for monitoring or diagnosing the disease early. Towards this direction, epigenetic information detected in circulating cell-free DNA (ccfDNA) is currently tested as a clinically valuable biomarker of β-cell death in Type 1 Diabetes Mellitus (T1DM) by several groups [5–12]. These biomarkers rely on the principles that necrotic or apoptotic cells or even viable cells release amounts of their DNA into the bloodstream and/or other biofluids that can be easily detected [13] and that each cell type has a unique and stable methylome that controls gene expression [14].

In the field of biomarker discovery, it becomes increasingly apparent that a single biochemical or epi/genetic feature is unlikely to bear the sensitivity/specificity required to disrupt clinical practice, and more effort is put into describing feature patterns or models. Machine Learning (ML) is the application of artificial intelligence on data analysis to build trained models [15]. ML has penetrated biomarker discovery in many diseases [16–19] and in diabetes [20,21]. Nowadays, Automate ML (AutoML) tools have become available. They promise to democratize data analysis to non-experts, improve the replicability of the statistical analysis, and shield against common methodological analysis pitfalls such as overfitting [22]. AutoML had been used for the prediction and diagnosis of diseases such as Alzheimer's disease [23], lung and breast cancer [24–26], and suicide prediction amongst depressive patients [27].

In the present study, we focused on ccfDNA as a liquid biopsy biomaterial in T2DM, its characteristics and potential for clinical use. We first quantified ccfDNA levels of T2DM patients and healthy volunteers directly in serum, and we studied its fragment size distribution in order to evaluate its cellular release mechanism. We then evaluated the methylation profile of a panel of β-cell-related specific genes—*INS* (Insulin), *IAPP* (Islet Amyloid Polypeptide-Amylin), *GCK* (Glucokinase), *KCNJ11* (Potassium Inwardly Rectifying Channel Subfamily J member 11), and *ABCC8* (ATP Binding Cassette Subfamily C member 8)—via quantitative SYBR Green-based methylation-specific PCR. We selected this panel of β-cell-related specific genes based on existing literature proposing them as promising biomarkers for β-cell death (*INS*, *IAPP* and *GCK*) [28], as well as based on their implication for β-cell physiology (*INS*, *IAPP*, *GCK*, *KCNJ11* and *ABCC8*) [5–12,29] and potential response to antidiabetic therapy [30]. Most importantly, ad-hoc AutoML technology was applied on our ccfDNA experimental parameters in combination with patient demographical data (i.e., age, gender, BMI and smoking) to build accurate diagnostic/monitoring predictive biosignatures of clinical value for personalized diabetes management. Our study's workflow is presented in Figure 1.
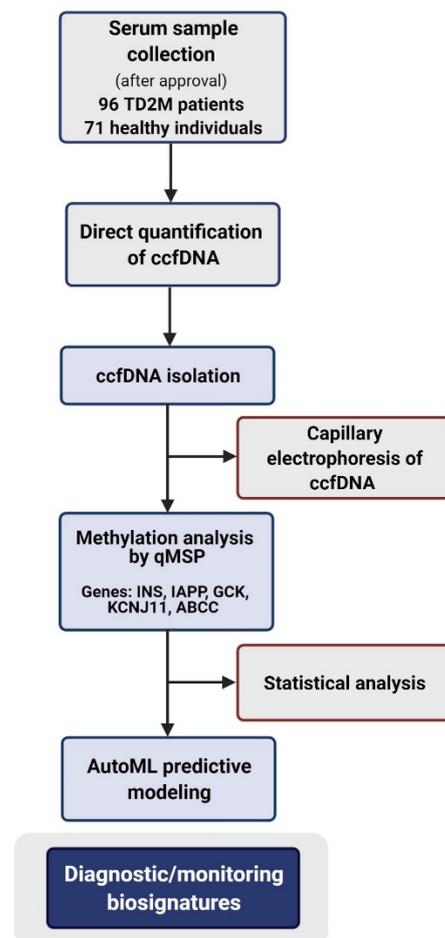
Figure 1. Workflow of our study. Abbreviations: TD2M: Type 2 diabetes mellitus, ccfDNA: circulating cell free DNA, qMSP: quantitative Methylation Specific PCR, AutoML: Automated Machine Learning.

2. Results

2.1. ccfDNA Levels in T2DM Patients and Healthy Volunteers

Total ccfDNA levels were directly quantified in sera from T2DM patients and healthy individuals (control group). Interestingly, ccfDNA levels did not differ between healthy individuals and diabetes patients (700(400–1590) ng/mL vs. 820(490–2430) ng/mL, respectively, p = 0.552) (Figure 2). Then, ccfDNA levels were compared within the group of diabetes patients in relation to different clinical characteristics. T2DM patients with more than 14 years of diabetes showed significantly higher ccfDNA levels than those with less years (p = 0.010). Other clinical parameters, such as gender, smoking, Body Mass Index (BMI), and insulin or oral antidiabetic treatment, showed no correlations to ccfDNA levels.

2.2. ccfDNA Fragment Size Analysis in T2DM Patients and Healthy Volunteers

DNA fragment size analysis was performed in isolated ccfDNA samples in order to reveal information about the cellular process of its release in the tissue of origin. It has been reported that samples containing DNA fragments of ~160 bp and multiples indicate release during apoptosis, whereas larger DNA fragments of ~2000–3000 bp are indicative of active release and above 10,000 bp are indicative of necrosis [31,32]. In the patients' group, peaks of ~160 bp (and multiples ×160 bp) were detected more often (47% of samples) than in healthy controls (21% of samples), indicating greater incidence of apoptosis (although the difference between percentages was not statistically significant). Peaks of larger DNA fragments of ~2000–3000 bp were observed in both groups. Notably, the ~2000 bp fragments appeared more often in T2DM patients with higher BMI, considered a continuous variable

($p = 0.037$). In Figure 3, representative electropherograms from two T2DM patients and one healthy individual are presented, showing a different pattern.
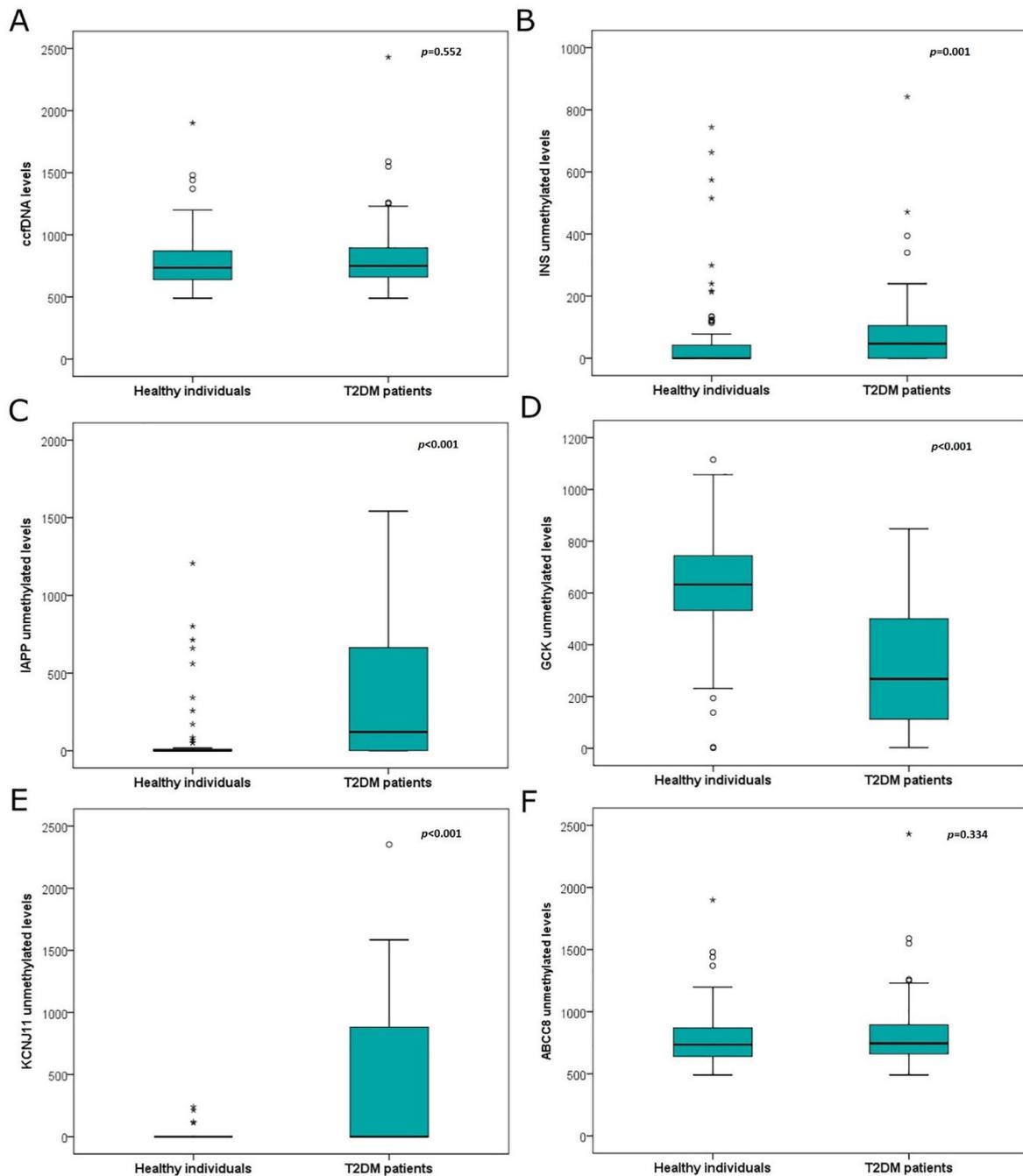

Figure 2. (**A**) Box plot of serum ccfDNA levels between healthy individuals ($n = 71$) and T2DM patients ($n = 96$) ($p = 0.552$). (**B**) Box plot of *INS* unmethylated alleles between healthy individuals and T2DM patients ($p = 0.001$). (**C**) Box plot of *IAPP* unmethylated alleles between healthy individuals and T2DM patients ($p < 0.001$). (**D**). Box plot of *GCK* unmethylated alleles between healthy individuals and T2DM patients ($p < 0.001$). (**E**) Box plot of *KCNJ11* unmethylated alleles between healthy individuals and T2DM patients ($p < 0.001$). (**F**) Box plot of *ABCC8* unmethylated alleles between healthy individuals and T2DM patients ($p = 0.534$). Small circles (°) correspond to "outlier" values and stars (*) to the "extreme" values of the dataset. Abbreviations: T2DM: Type 2 diabetes mellitus, ccfDNA: circulating cell free DNA.
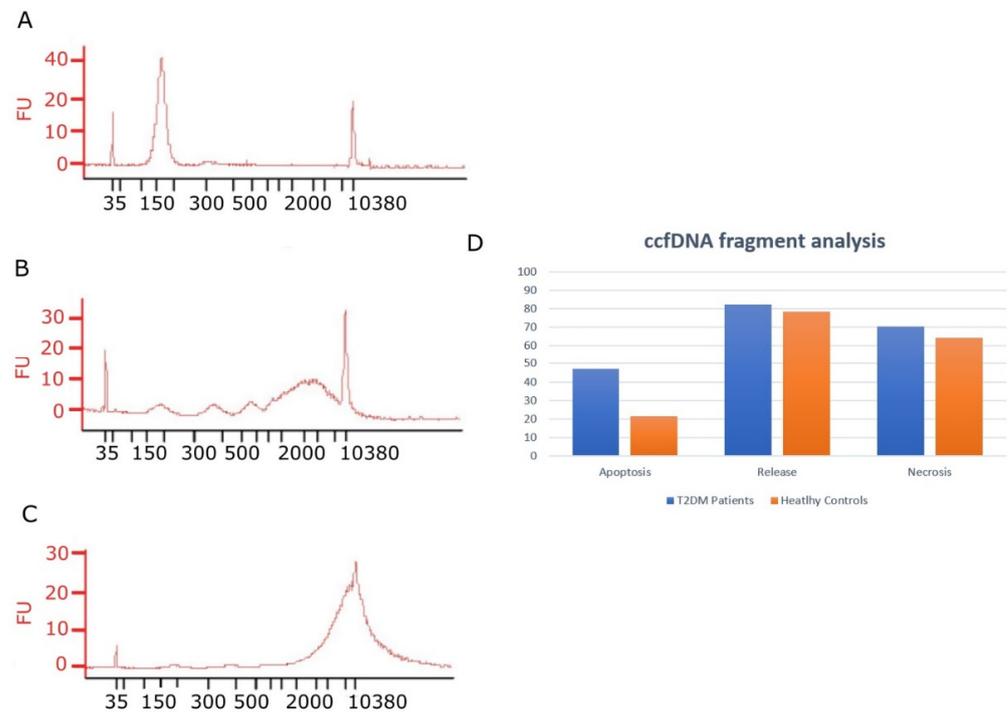
Figure 3. Representative capillary electropherograms showing DNA fragment size distribution in ccfDNA isolated from sera of two T2DM patients and one healthy control. (**A**) A T2DM patient ccfDNA sample showing a peak at ~160 bp indicative of apoptosis. (**B**) Another T2DM patient ccfDNA sample showing multiple peaks at ~160 bp, ~300 bp and ~500 bp indicative of apoptosis and an additional wide peak at ~2000–3000 bp indicative of active release. (**C**). A healthy volunteer ccfDNA sample with a wide peak at ~2000–3000 bp indicative of active release. Peaks at 35 bp and 10,380 bp in all electropherograms represent high and low ladders, respectively. (**D**) Distribution of ccfDNA fragment analysis in the patient's group and the healthy volunteer's group, respectively. Abbreviations: T2DM: Type 2 diabetes mellitus, ccfDNA: circulating cell free DNA, bp: base pairs.

2.3. Methylation Analysis of β-Cell-Specific Genes

Developed qMSP assays were validated in terms of analytical specificity, sensitivity, reproducibility, linearity and efficiency. The analytical specificity of all our developed assays was found to be 0.5% in the detection of methylated DNA molecules in a background of unmethylated DNA and 0.5% in the detection of unmethylated DNA molecules in a background of methylated DNA (specificity curves are presented in Supplementary Figure S1). The analytical sensitivity was found to be up to 0.1 ng for both methylated and unmethylated DNA molecules (sensitivity curves are shown in Supplementary Figure S1). The efficiency of all our developed assays ranged between 93 and 97%. Furthermore, high within—and between—sample reproducibility was also observed for all assays (CV ranging from 0.5% to 1.2%).

INS, *IAPP*, *GCK* and *KCNJ11* methylation differed significantly between T2DM patients and healthy individuals ($p = 0.001$, $p < 0.001$, $p < 0.001$ and $p < 0.001$, respectively), while *ABCC8* methylation did not differ between groups ($p = 0.534$) (Figure 2). Receiver operating characteristic (ROC) curve analysis showed that *GCK* methylation could provide high discrimination between T2DM patients and healthy individuals (AUC 0.848 (95% CI 0.787–0.910)) (Figure 4), while *IAPP* and *KCNJ11* methylation could offer lower discrimination between groups (AUC 0.727 (95% CI 0.649–0.805) and AUC 0.712 (95% CI 0.619–0.806), respectively) (Figure 4). *INS* methylation showed poor discrimination capacity between patients and controls (AUC 0.650 (95% CI 0.562–0.737)) (Figure 4).
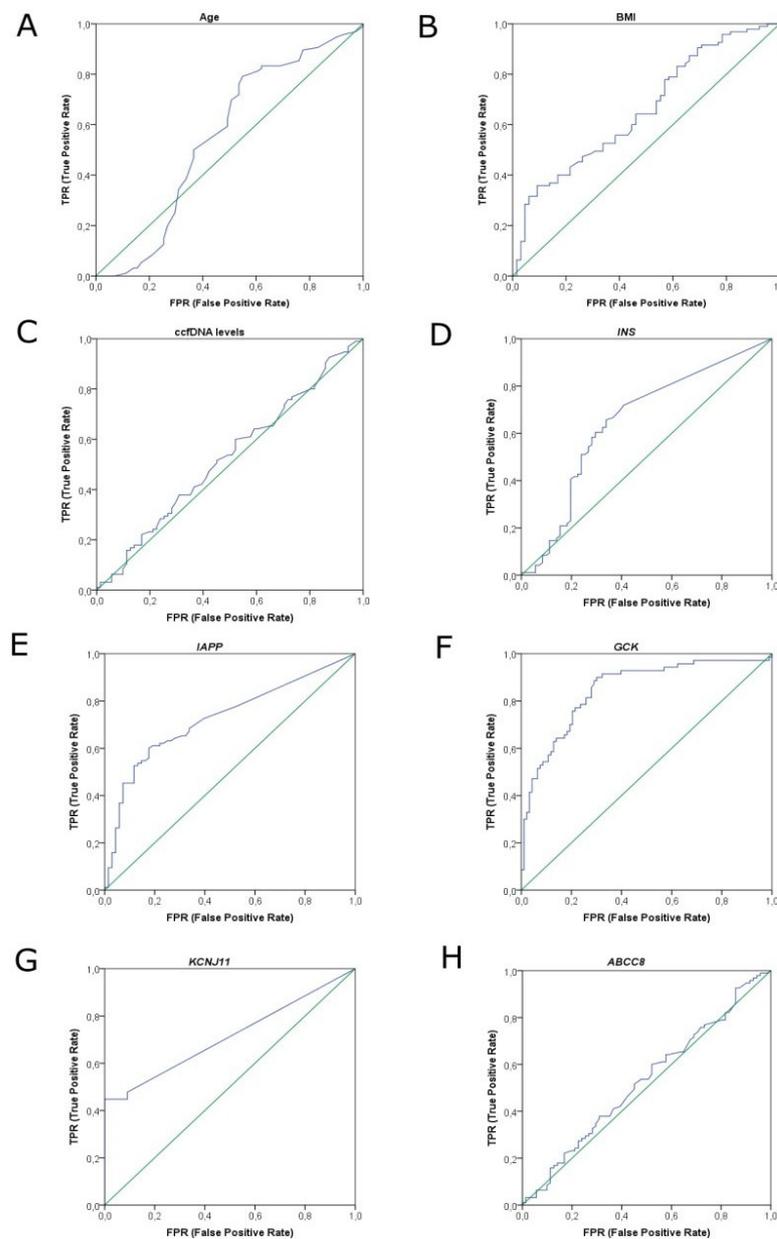
Figure 4. ROC curve analysis results. (**A**) ROC curve of age reaching an AUC of 0.566 (95% CI 0.468–0.664). (**B**) ROC curve of BMI reaching an AUC of 0.658 (95% CI 0.573–0.743). (**C**) ROC curve of ccfDNA levels reaching an AUC of 0.527 (95% CI 0.438–0.616). (**D**) ROC curve of *INS* gene reaching an AUC of 0.650 (95% CI 0.562–0.737). (**E**) ROC curve of *IAPP* gene reaching an AUC of 0.727 (95% CI 0.649–0.805). (**F**) ROC curve of *GCK* gene reaching an AUC of 0.848 (95% CI 0.787–0.910). (**G**) ROC curve of *KCNJ11* gene reaching an AUC of 0.712 (95% CI 0.619–0.806). (**H**) ROC curve of *ABCC8* gene reaching an AUC of 0.528 (95% CI 0.439–0.617). Abbreviations: ROC curve: receiver operating characteristic curve, BMI: Body Mass Index, AUC: area under the curve, CI: confidence interval, ccfDNA: circulating cell free DNA.

Methylation profiles of T2DM patients with complications were similar in all studied genes to those of patients without complications. Moreover, the duration of diabetes (less or more than 15 years of disease) and the levels of HbA1c (less or more than 8%) did not correlate with gene methylation. Analysis of methylation with respect to the other clinicopathological characteristics, such as gender, smoking, Body Mass Index (BMI), and insulin or oral antidiabetic treatment, did not reveal significant correlations.

2.4. AutoML Predictive Analysis

Our data were further analyzed by ML techniques in order to produce diagnostic/monitoring biosignatures of clinical value, combining the novel liquid biopsy-based methylation data emerged by our study and the clinical and demographical data of the study's groups. The JADBio automated machine learning (AutoML) platform employed for this analysis automatically performs and compares all standard, best practices and advanced ML techniques, and it produces upon feature selection the optimal best-performing model along with the most interpretable one.

In our AutoML analysis, the task was to predict T2DM versus health from the available ccfDNA parameters and the demographical patient data (age, gender, BMI, etc.). We first analyzed the whole dataset of the 96 T2DM patients on treatment and the 71 healthy individuals (control group). In this analysis, JADBio trained 3017 different machine learning pipelines (also called configurations), corresponding to different model types. Each one was employed many times during cross-validation (a repeated 10-fold CV without dropping), leading to fitting 90,510 model instances (https://app.jadbio.com/share/d59a0 8fb-e7ea-42e8-8eae-b225f512a38b, accessed on 13 January 2022). This classification analysis produced a best-performing five-feature biosignature via the Classification Random Forests algorithm that was able to discriminate between T2DM patients and healthy individuals with an AUC of 0.927 (95% CI 0.874–0.967) and an average precision of 0.951 (95% CI 0.914–0.980). Biosignature's features included *GCK*, *IAPP* and *KCNJ11* methylation as well as age and BMI, and their contribution in the model's performance defined as the percentage drop in predictive performance when the feature is removed from the model is shown in Figure 5C. The best-performing biosignature's performance is presented in Figure 5A,B. The most interpretable five-feature biosignature was also built via a ridge logistic regression algorithm reaching an AUC of 0.915 (95% CI 0.868–0.957) and an average precision of 0.941 (95% CI 0.901–0.975). This biosignature included as features *GCK*, *IAPP* and *KCNJ11* methylation, smoking status and BMI.

Most importantly, the size of the dataset allowed for further model automated validation. The whole dataset was split randomly into training and test sub-datasets by a 70/30 ratio via JADBio. In this analysis, JADBio trained 3017 different machine learning pipelines, corresponding to different model types and fitted 150,850 model instances (https://app. jadbio.com/share/42c8c603-06d4-47e7-8276-97d4fa970d6c, accessed on 13 January 2022). The training data from 66 T2DM patients and 51 healthy individuals (control group) led to a similar but not identical best-performing five-feature biosignature via the Classification Random Forests algorithm, which was able to discriminate between patients and healthy individuals with an AUC of 0.898 (95% CI 0.845–0.944) and an average precision of 0.937 (95% CI 0.893–0.968) (Figure 5D,E). The biosignature's features included *GCK*, *IAPP* and *KCNJ11* methylation as well as BMI and ccfDNA concentration, all but the last common to the original model from the whole dataset. Validating the model in the test sub-group data from 30 T2DM patients and 20 healthy individuals showed an AUC of 0.923 and an average precision of 0.945, verifying the model's performance stability. The best-performing biosignature is presented in Figure 5D–F. The most interpretable five-feature biosignature was also built via the Ridge Logistic Regression algorithm reaching an AUC of 0.879 (95% CI 0.826–0.927) and an average precision of 0.921 (95% CI 0.881–0.957). This biosignature's features included again *GCK*, *IAPP* and *KCNJ11* methylation, as well as BMI and ccfDNA concentration. In validation in the test dataset, model reached an increased AUC of 0.958 and an average precision of 0.972, again verifying no overfitting in the model construction. Supplementary Table S1 displays the algorithms and tuning hyper-parameter values that JADBio's AI decided to try in the analysis of the splitted training dataset.
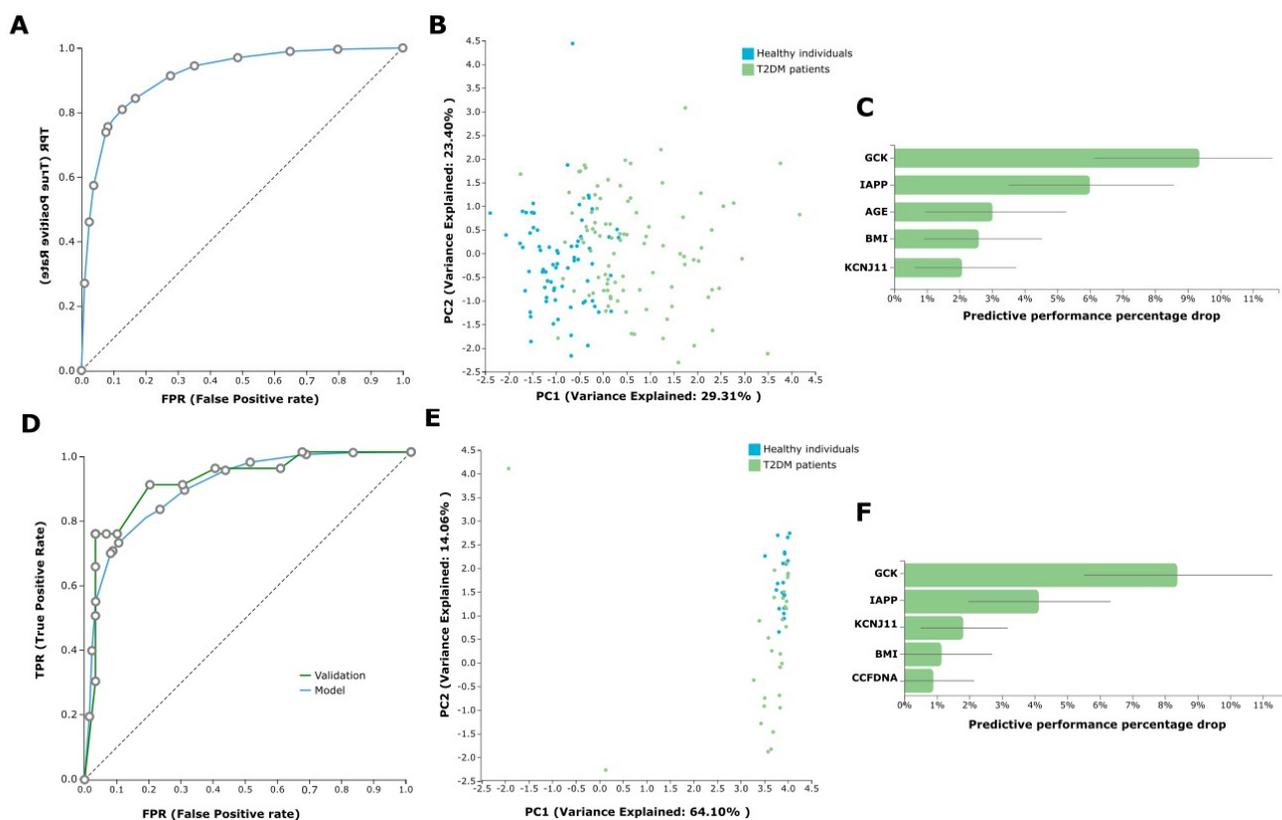
Figure 5. Predictive modelling results. (**A**) ROC curve of whole dataset reaching an AUC of 0.927 (95% CI 0.874–0.967). (**B**) Supervised Principal Component Analysis (PCA) plot of whole dataset depicting discrimination between T2DM patients and healthy individuals. (**C**) Feature importance plot of the features of the best-performing model for the whole dataset. Feature importance is defined as the percentage drop in predictive performance when the feature is removed from the model. (**D**) ROC curve of train sub-dataset (blue line) reaching an AUC of 0.898 (95% CI 0.845–0.944) and test sub-dataset (green line) showing an AUC of 0.923 for the best-performing model. (**E**) PCA plot of test sub-dataset depicting discrimination between T2DM patients and healthy individuals. (**F**) Feature Importance plot of the features of the best-performing model in the train/test 70/30 split sub-datasets. Abbreviations: ROC curve: receiver operating characteristic curve, AUC: area under the curve, CI: confidence interval, T2DM: type 2 diabetes mellitus.

3. Discussion

ccfDNA is released into the blood, urine, and other biological fluids after cell apoptosis/necrosis or by active release from living cells [13]. Currently, ccfDNA-based biomarkers have emerged as promising minimally invasive options for the early diagnosis and monitoring of T1DM in several studies [5,8,11]. According to El Tarhouny et al., ccfDNA levels were significantly elevated in T2DM patients with or without complications compared to healthy individuals, indicating that ccfDNA can also be of significant clinical value in T2DM personalized management [33]. To expand knowledge on ccfDNA in T2DM with potentially valuable clinical applications, we investigated the levels, the fragment size distribution as well as the methylation profile of ccfDNA in a cohort of T2DM patients and healthy individuals with the ultimate goal of the production of diagnostic/monitoring biosignatures. To the best of our knowledge, this is the first study to evaluate such ccfDNA-based experimental parameters in a multi-level approach in diabetes.

By directly measuring ccfDNA in sera via fluorometry, our data revealed that ccfDNA levels were similar between T2DM patients and healthy individuals. This discrepancy with the data of El Tarhouny et al. [33] can possibly be attributed to different quantification methods and patient classification criteria. Thus, in our study, ccfDNA levels were quantified

via fluorometry, counting both nuclear and mitochondrial fractions of free/naked ccfDNA (unbounded DNA). On the other hand, the quantification of extracted ccfDNA by PCR for the *GAPDH* gene that was used by El Tarhouny et al. measures only the nuclear fraction of extracted ccfDNA and both the unbounded and bounded DNA in protein complexes.

Plasma ccfDNA is a popular specimen for liquid biopsy approaches, although in several previous studies on beta-cell death serum samples were used for ccfDNA isolation, so we decided to keep up with these approaches in order to have comparable results [5,8,9]. In any case, our pilot experiments not included in the manuscript showed no significant differences in the yield and quality of ccfDNA isolated between plasma and serum blood fractions. Serum volumes as low as those used here have not been highly reproducible in past assessments of similar assays; indeed, the cancer field where these assays are most highly advanced are in many cases moving toward sample volumes an order of magnitude larger than those used here. Our results from direct and indirect quantifications demonstrate that used sample volume was sufficient for the downstream analysis, as previously demonstrated by our group and others [9,25]. In addition, taking into consideration the short half-life of ccfDNA fragments and in order to avoid significant loss, isolation protocol was initiated within two hours of collection.

Several suggested cellular processes are responsible for ccfDNA release into biological fluids and the size of the DNA fragment content is informative for each one of them. Apoptosis is characterized by ~160 bp and multiple fragments, necrosis delivers fragments above ~10,000 bp, while active release delivers ~2000–3000 bp fragments [31,32]. We have recently shown that in breast cancer patient samples, DNA fragments of ~160 bp and multiples and those of ~10,000 bp appeared more often (in 60.0% and 53% of patients, respectively) than DNA fragments of ~2000 bp (in 37% of patients) [25]. In our previous study, using two cancer cell lines, we have shown that induction of apoptosis in vitro resulted in the release of these short ~160 bp and multiple fragments, confirming these data [34]. To reveal the mechanism(s) of ccfDNA release into the circulation in T2DM, fragment analysis by capillary electrophoresis showed mainly fragments of ~160 bp and multiples and that of ~2000 bp inT2DM patients, indicating greater incidence of apoptosis, although not reaching statistical significance. This could be attributed to the ongoing apoptosis of remaining functional insulin-producing β-cells in accordance with the fact that β-cell destruction is an important etiological factor in the development and progression of T2DM [35]. Moreover, in obese T2DM patients, peaks of ~2000 bp fragments were observed more often than in non-obese patients, possibly suggesting that obesity-induced DNA release from adipocytes is via active release.

In order to examine the tissue origin of ccfDNA in T2DM, we investigated the methylation status of five β-cell-specific genes based on previous literature findings that propose them as promising biomarkers for β-cell death (*INS*, *IAPP* and *GCK*), as well as their implication on β-cell physiology (*INS*, *IAPP*, *GCK*, *KCNJ11* and *ABCC8*). As methylation is a tissue-specific event, genes expressed exclusively in β-cells are unmethylated only in this cell type, and this status is expected to be detectable also in ccfDNA released by them. It has been previously shown that *INS*, *IAPP* and *GCK* are found to be unmethylated in serum ccfDNA of T1DM patients in relation to healthy individuals [5,6,8–11]. We have also previously studied the methylation levels of *KCNJ11* and *ABCC8* in another cohort of T2DM patients regarding their response to antidiabetic therapy, and we were interested in investigating further their use as biomarker of T2DM [30]. Here, we focused on T2DM and the methylation status of *INS*, *IAPP*, *GCK* and *KCNJ11*, and *ABCC8* is investigated for the first time in serum ccfDNA of T2DM patients. All but the *ABCC8* gene showed statistically significant differential methylation levels between patients and controls. Higher percentages of unmethylated alleles show the beta-pancreatic origin of ccfDNA in the blood of T2DM patients. It can be concluded that while on an apoptotic cell death pathway, β-cells release ccfDNA and enrich the physiological pool of free nucleic acids in the circulation.

To further exploit our ccfDNA experimental observations in order to build classifying biosignatures of higher diagnostic/monitoring performance, we performed classification

analysis by employing an innovative AutoML approach. We used JADBio, an AutoML platform that uses state-of-the-art statistical and machine learning methods [22]. Another innovative element of our approach was combining all ccfDNA experimental parameters (including levels and gene methylation) with clinical and demographical patient characteristics in one analysis executing feature selection and therefore eliminating redundancy. AutoML analysis in the dataset as a whole delivered a best-performing five-feature biosignature via the Classification Random Forests algorithm, including *GCK*, *IAPP* and *KCNJ11* methylation and age and BMI, showing very high performance in discriminating T2DM patients from healthy individuals (AUC 0.927). Similarly, the AutoML of split train/validation data led to a best-performing five-feature biosignature via the Classification Random Forests algorithm with the same but one new feature, i.e., *GCK*, *IAPP* and *KCNJ11* methylation and BMI and ccfDNA concentration, showing nearly as high discrimination capacity between groups (AUC 0.898). This was expected, as models built in smaller datasets demonstrate compromised performance. Most importantly, the validation of this biosignature in the independent validation subgroup showed an AUC of 0.923, verifying the model's performance stability and the absence of overfitting, which confirms what claimed to be achieved by the AutoML tool used. In both models, *GCK* and *IAPP* methylation are found to be the major features contributing to the performance, highlighting their value as biomarkers in diabetes.

To fully assess clinical relevance, we next plan to validate these biosignatures in independent, blinded validation cohorts of T2DM patients, obese and prediabetic individuals from a different clinical setting (cross-dataset analysis). We should note, however, that JADBio implements internal cross-validation with the bootstrap corrected cross-validation (BBC-CV) algorithm [36,37], shown to substitute external validation. The bootstrapping technique performs a correction to the estimation of out-of-sample performance of the final model. The correction (adjustment) is required because JADBio tries thousands of machine learning pipelines to identify the best one that produces the optimal, final model. The correction is conceptually similar to the Bonferroni adjustment required for multiple hypotheses testing due to performing multiple tests. Intuitively, the selection process, which selects the best out of numerous pipelines is bootstrapped. Internally, JADBio holds out a subset of the training data to test the generalization performance (out-of-sample) of the produced model, effectively simulating the presence of an external dataset. In fact, JADBio performs this training and test procedure numerous times to reduce the variance of the estimation. Specifically for small sample sizes, it employs a stratified, K-fold, repeated cross-validation protocol that exhibits small estimation variance. In addition, the final estimate is corrected for the fact that numerous machine learning pipelines have been tried (a form of the "winner's curse") using the recently developed bootstrap corrected cross-validation (BBC-CV) algorithm. This technique has been shown to produce conservative estimates of performance in massive evaluation experiments with general types of data [22]. In addition, the accurate estimation of AUC by JADBio in small sample settings has been tested in numerous studies. Indeed, we and others have previously shown in multimodal datasets that AUC estimations did not drop upon external validation, showing no over-estimations [25–27,38–40]. This body of work proves that JADBio estimates can be trusted and there is no need to have a separate hold-out dataset to statistically validate the results, a feature of particular importance for maximal extrapolation of precious biomedical datasets. In the case presented here, the model built in the dataset as a whole had marginally better performance that the one built in 70/30 train/test. Still, as we ourselves would like to promote trustworthiness of this ML approach, we plan to re-optimize and externally evaluate the biosignatures built in real-world prospective setting.

A relatively small sample size is a limitation in this study. We [22] and others [40] argue that sample size is one of many important design elements contributing to the successful implementation in biomarker discovery. Machine learning, quickly penetrating the field, is there in order to overcome such limitations, aiding robust, optimized and maximal data extrapolation from small cohorts. As such, it is included in the recommended generic

steps for robust model building and evaluation. The novel biosignatures presented in our study, based on qMSP rather than whole-genome or digital PCR methodologies, have a great translatability to cost-effective assays, which can be implemented in any standard molecular diagnostic laboratory. They are, therefore, readily available to offer feasible liquid biopsy-based assays for predicting or diagnosing early diabetes as well as monitoring β-cell death upon prospective clinical validation.

Previous studies leveraged demographical, clinical and laboratory data, such as age, BMI, glycated hemoglobin (HbA1c), hypertension, smoking and glucose, and employed ML tools for predicting risk of diabetes with promising results [20,21,41]. The predictive abilities of the models built, however, (AUC ranging from 0.826 to 0.872) were inferior to those presented here (AUC reaching 0.927) incorporating for the first time parameters of a liquid biopsy biomaterial such as ccfDNA with demographical data.

Other approaches focusing on liquid biopsy biomarkers for the management of diabetes have examined methylation of *INS*, *IAPP* and *GCK* either as single parameters [5,8,9,12] or as combined parameters, i.e., *INS* and *CHTOP* in a duplex assay [42]. Previous reports showed that *GCK* methylation is a more suitable marker than *INS* methylation for the detection of β-cell death in T1DM and therefore can present an early diabetes biomarker [9]. Furthermore, the recent study of Arosemena et al. reported that none of the four groups (lean controls with normal glucose tolerance; overweight/obese with normal glucose tolerance; impaired glucose tolerance; and T2DM patients) showed statistically significant differences in *INS* methylation compared to the healthy controls [43]. These findings are in agreement with our observations in T2DM, where *INS* methylation showed a low capacity to discriminate T2DM, and *GCK* methylation revealed the best capacity, highlighting its potential clinical value as a minimally invasive biomarker for T2DM personalized management. Our model incorporates this biomarker with the maximal feature importance (*GCK*) in the model's performance, which, however, is dramatically increased by including other selected pancreatic gene methylation biomarkers and other parameters. These corroborated findings considered together, support the value of the herein model in the early detection of diabetes onset.

4. Materials and Methods

4.1. Study Groups and Serum Sampling

Our study's groups consisted of 96 T2DM patients on treatment and 71 healthy individuals (control group) of similar age without a history of diabetes. All samples were of Caucasian origin. T2DM was diagnosed according to the ADA guidelines [44]. All T2DM patients were on oral antidiabetic therapy of metformin, and 5 of them also received sulfonylureas treatment. Almost half of patients (*n* = 42) were also on insulin therapy. Demographic and clinical data of study groups are shown in Supplementary Table S2. Inclusion criteria of the study included age between 25 and 75 years old and ability to give informed consent. Exclusion criteria of all participants included the presence of a (another) chronic disease, underlying malignancies and systemic lupus erythematosus. Serum samples were obtained within 2 h of blood sampling through centrifugation at $3000\times g$ for 10 min. An additional high-speed centrifugation step at $14,000\times g$ for 10 min was performed to remove any cellular debris and contaminants, such as gDNA from damaged blood cells. Serum samples were stored at $-80\ °C$ until further use.

4.2. Direct Quantification of ccfDNA

The direct quantification of free unbounded/naked ccfDNA in 20 μL of serum was performed utilizing the Quant-iT dsDNA High-Sensitivity Assay Kit in Qubit 3.0 Fluorometer (Invitrogen, Karlsruhe, Germany) according to manufacturer specifications. A standard curve was generated using provided standards (0 and 10 ng/μL).

4.3. ccfDNA Isolation

All ccfDNA (naked, bound in nucleosomes or proteins, or internalized in vesicles) was extracted from 600 μL of serum using the QIAamp Blood Mini kit (Qiagen, Hilden, Germany) in a final elution volume of 30 μL, as specified by the manufacturer. ccfDNA isolation was performed manually in batches of 10 samples and no carrier DNA was used. Specifically, 20 μL Protease and 600 μL of each sample were added to a microcentrifuge tube, and then 200 μL Buffer AL was added to each sample. After lysis, samples were incubated for 10 min at 56 °C, and 200 μL 100% ethanol were added to each sample. Following this, the mixtures were applied to a QIAamp Mini spin column and two steps of wash and centrifuging at 6000× *g* were followed. Finally, 30 μL Buffer AE was added, and, through a new centrifugation step, serum ccfDNA was eluted.

The extracted ccfDNA was stored at −20 °C until further use. Then, 4 μL of extracted ccfDNA was subjected to quantitative PCR (qPCR) for the nuclear *GAPDH* gene to determine isolation efficiency in terms of ccfDNA quantity and quality following MIQE guidelines [45]. Primer sequences are shown in Supplementary Table S2.

4.4. Capillary Electrophoresis of Extracted ccfDNA

The High-Sensitivity DNA Kit on 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), an automated on-chip electrophoresis system, was used for the fragment size evaluation of 2 μL of extracted ccfDNA, following the manufacturer's instructions.

4.5. Methylation Analysis

A total of 20 μL of extracted ccfDNA was treated with sodium bisulfite (SB) using the EZ DNA Methylation-Gold kit (ZYMO Research Co., Irvine, CA, USA) in a final elution volume of 10 μL, following the manufacturer's procedure. In each reaction, CpGenome Human methylated and non-methylated DNA controls (Merck Millipore, Darmstadt, Germany) were included as negative and positive control samples, respectively. The SB-treated ccfDNA was stored at −80 °C until further use.

A methylation-independent PCR assay for the β-actin gene (*ACTB*) was used in order to verify sufficient quality and quantity of SB-treated ccfDNA. Methylation levels of β-cell-related genes *INS*, *IAPP*, *GCK*, *KCNJ11* and *ABCC8* were analyzed using quantitative SYBR Green-based methylation-specific PCR assays. For all genes, primers specific for methylated (m) and unmethylated (u) alleles were either newly designed using the MethPrimer [46] program or were based on bibliography with some modifications. Primer sequences are provided in Supplementary Table S3. To set up robust qMSP assays, extensive optimization was performed. Specificity and cross-reactivity of primers were evaluated using unconverted DNA, SB-treated methylated DNA and non-methylated DNA controls. Analytical specificity of qMSP assays was evaluated by using mixes of SB-converted methylated and non-methylated DNA standards (100%, 50%, 10%, 5%, 0.5%, 0%). Analytical sensitivity of assays was evaluated using serial dilutions of SB-treated methylated and non-methylated DNA controls in H_2O. The reproducibility (calculated as coefficients of variation, CVs), the efficiency and the linearity were also evaluated in order to complete the validation file of the established assays. All samples were run in duplicates. The percentage of methylation in a sample was estimated using following formula by Lu et al. [47]:

$$1 - \frac{1}{1 + 2^{(-\Delta Ct)}} \times 100\% \tag{1}$$

where

$$\Delta Ct = Ct_{unmeth} - Ct_{meth} \tag{2}$$

Following this, the percentage of methylation was multiplied by the concentration of total serum ccfDNA.

4.6. Statistics

Initially, Kolmogorov–Smirnov test was applied to check for normality in distribution of continuous data. Due to lack of normality in our data, non-parametric statistics were used for comparisons between groups (Kruskal–Wallis test or Mann–Whitney U test). The Spearman correlation coefficient was used as a measurement of correlation for continuous variables. The predictive power of the ccfDNA levels and β-cell-related genes methylation was tested using receiver operating characteristic (ROC) curve analysis and Area Under the Curve (AUC) metric. Continuous variables are expressed as median value (25th–75th percentile, interquartile range). Categorical variables are shown as absolute frequencies (percentage). In all tests performed, statistical significance was set at two-sided *p* value < 0.050. Standard statistical analysis was carried out with the SPSS version 21.0 statistical software package for Windows (IBM-SPSS Inc., New York, NY, USA).

4.7. AutoML Predictive Modelling with JADBio

The innovative AutoML technology JADBio, version 1.2.8 (https://www.jadbio.com/, accessed on 13 January 2022) [22], was applied to produce diagnostic/monitoring biosignatures/classifiers based on the ccfDNA parameters, methylation data and demographical information. Given a 2D matrix of data, JADBio automatically preprocesses data (Mean Imputation, Mode Imputation, Constant Removal, Standardization), performs feature selection by employing LASSO or Statistical Equivalent Signatures (SES) algorithms, tries several algorithms (Classification Random Forests, Support Vector Machines (SVM), Ridge Logistic Regression and Classification Decision Trees) and thousands of algorithmic configurations, and then selects the best-performing model, estimates the out-of-sample model's performance after bootstrap correction and cross-validation, and provides several visualizations.

The training set is used to train multiple machine learning pipelines that include the steps of preprocessing, imputation, feature selection, and modeling. Internally and automatically, JADBio holds out a subset of the training data to test the generalization performance (out-of-sample) of the produced model, effectively simulating the presence of an external dataset. In fact, JADBio performs this training and test procedure numerous times to reduce the variance of the estimation. Specifically, for small sample sizes, it employees a stratified, K-fold, repeated cross-validation protocol BBC-CV algorithm [37] that exhibits small estimation variance and removes the estimation bias due to the fact that it was selected among many. The features selected in the winning pipeline are the ones included in the returned model. In addition, our dataset was randomly automatically split into train and validation sub-datasets by a 70/30 ratio via JADBio.

For our AutoML classification analysis, we used extensive model tuning effort, we chose the AUC metric for optimization of performance and we set classifier maximum size to five features. The predictive power of the biosignature was assessed using AUC and average precision (also known as area under the precision–recall curve) metrics.

5. Conclusions

The identification of markers that can predict future onset of type 2 diabetes is of great interest to the field; the ability to measure such markers consistently and uninvasively is critical. Our data unravel the value of ccfDNA as a liquid-biopsy biomaterial carrying important clinical information for minimally invasive T2DM diagnosis and monitoring. Overall, by adopting ad-hoc AutoML technology in our study, a highly potent predictive biosignature based on ccfDNA methylation parameters in combination with demographical data has emerged, readily available to be translated into a cost-effective laboratory test. Upon prospective clinical evaluation, our biosignature could aid the early diagnosis and monitoring of T2DM, meeting the need for a minimally invasive advancement in the direction of personalized diabetes management.

6. Patents

The contents of this manuscript were included in an EPO patent filing. Makrina Karaglani, Ioannis Tsamardinos and Ekaterini Chatzaki act as inventors.

References

1. Saeedi, P.; Petersohn, I.; Salpea, P.; Malanda, B.; Karuranga, S.; Unwin, N.; Colagiuri, S.; Guariguata, L.; Motala, A.A.; Ogurtsova, K.; et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res. Clin. Pract.* **2019**, *157*, 1–43. [CrossRef] [PubMed]
2. Kharroubi, A.T.; Darwish, H.M. Diabetes mellitus: The epidemic of the century. *World J. Diabetes* **2015**, *6*, 850–867. [CrossRef] [PubMed]
3. Rahier, J.; Guiot, Y.; Goebbels, R.M.; Sempoux, C.; Henquin, J.C. Pancreatic beta-cell mass in European subjects with type 2 diabetes. *Diabetes Obes. Metab.* **2008**, *10* (Suppl. S4), 32–42. [CrossRef] [PubMed]
4. Butler, A.E.; Janson, J.; Bonner-Weir, S.; Ritzel, R.; Rizza, R.A.; Butler, P.C. Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes* **2003**, *52*, 102–110. [CrossRef]
5. Akirav, E.M.; Lebastchi, J.; Galvan, E.M.; Henegariu, O.; Akirav, M.; Ablamunits, V.; Lizardi, P.M.; Herold, K.C. Detection of beta cell death in diabetes using differentially methylated circulating DNA. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 19018–19023. [CrossRef]
6. Husseiny, M.I.; Kaye, A.; Zebadua, E.; Kandeel, F.; Ferreri, K. Tissue-specific methylation of human insulin gene and PCR assay for monitoring beta cell death. *PLoS ONE* **2014**, *9*, e94591. [CrossRef]

7. Herold, K.C.; Usmani-Brown, S.; Ghazi, T.; Lebastchi, J.; Beam, C.A.; Bellin, M.D.; Ledizet, M.; Sosenko, J.M.; Krischer, J.P.; Palmer, J.P.; et al. beta cell death and dysfunction during type 1 diabetes development in at-risk individuals. *J. Clin. Investig.* **2015**, *125*, 1163–1173. [CrossRef]

8. Olsen, J.A.; Kenna, L.A.; Spelios, M.G.; Hessner, M.J.; Akirav, E.M. Circulating Differentially Methylated Amylin DNA as a Biomarker of beta-Cell Loss in Type 1 Diabetes. *PLoS ONE* **2016**, *11*, e0152662. [CrossRef]

9. Sklenarova, J.; Petruzelkova, L.; Kolouskova, S.; Lebl, J.; Sumnik, Z.; Cinek, O. Glucokinase Gene May Be a More Suitable Target Than the Insulin Gene for Detection of beta Cell Death. *Endocrinology* **2017**, *158*, 2058–2065. [CrossRef]

10. Lehmann-Werman, R.; Neiman, D.; Zemmour, H.; Moss, J.; Magenheim, J.; Vaknin-Dembinsky, A.; Rubertsson, S.; Nellgard, B.; Blennow, K.; Zetterberg, H.; et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E1826–E1834. [CrossRef]

11. Usmani-Brown, S.; Lebastchi, J.; Steck, A.K.; Beam, C.; Herold, K.C.; Ledizet, M. Analysis of beta-cell death in type 1 diabetes by droplet digital PCR. *Endocrinology* **2014**, *155*, 3694–3698. [CrossRef] [PubMed]

12. Fisher, M.M.; Perez Chumbiauca, C.N.; Mather, K.J.; Mirmira, R.G.; Tersey, S.A. Detection of islet beta-cell death in vivo by multiplex PCR analysis of differentially methylated DNA. *Endocrinology* **2013**, *154*, 3476–3481. [CrossRef] [PubMed]

13. Gahan, P.B. Biology of circulating nucleic acids and possible roles in diagnosis and treatment in diabetes and cancer. *Infect. Disord. Drug Targets* **2012**, *12*, 360–370. [CrossRef] [PubMed]

14. Bergman, Y.; Cedar, H. DNA methylation dynamics in health and disease. *Nat. Struct Mol. Biol.* **2013**, *20*, 274–281. [CrossRef] [PubMed]

15. Davenport, T.; Kalakota, R. The potential for artificial intelligence in healthcare. *Future Healthc. J.* **2019**, *6*, 94–98. [CrossRef] [PubMed]

16. Myszczynska, M.A.; Ojamies, P.N.; Lacoste, A.M.B.; Neil, D.; Saffari, A.; Mead, R.; Hautbergue, G.M.; Holbrook, J.D.; Ferraiuolo, L. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat. Rev. Neurol.* **2020**, *16*, 440–456. [CrossRef]

17. Schaefer, J.; Lehne, M.; Schepers, J.; Prasser, F.; Thun, S. The use of machine learning in rare diseases: A scoping review. *Orphanet J. Rare Dis.* **2020**, *15*, 145. [CrossRef]

18. Kourou, K.; Exarchos, T.P.; Exarchos, K.P.; Karamouzis, M.V.; Fotiadis, D.I. Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 8–17. [CrossRef]

19. Chatzaki, E.; Tsamardinos, I. Somatic copy number aberrations detected in circulating tumor DNA can hold diagnostic value for early detection of hepatocellular carcinoma. *EBioMedicine* **2020**, *57*, 102851. [CrossRef]

20. Zhang, L.; Wang, Y.; Niu, M.; Wang, C.; Wang, Z. Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: The Henan Rural Cohort Study. *Sci. Rep.* **2020**, *10*, 4406. [CrossRef]

21. Muhammad, L.J.; Algehyne, E.A.; Usman, S.S. Predictive Supervised Machine Learning Models for Diabetes Mellitus. *SN Comput. Sci.* **2020**, *1*, 240. [CrossRef] [PubMed]

22. Tsamardinos, I.; Charonyktakis, P.; Lakiotaki, K.; Borboudakis, G.; Zenklusen, J.C.; Juhl, H.; Chatzaki, E.; Lagani, V. Just Add Data: Automated Predictive Modeling and BioSignature Discovery. *bioRxiv* **2020**. [CrossRef]

23. Karaglani, M.; Gourlia, K.; Tsamardinos, I.; Chatzaki, E. Accurate Blood-Based Diagnostic Biosignatures for Alzheimer's Disease via Automated Machine Learning. *J. Clin. Med.* **2020**, *9*, 3016. [CrossRef] [PubMed]

24. Markaki, M.; Tsamardinos, I.; Langhammer, A.; Lagani, V.; Hveem, K.; Røe, O.D. A Validated Clinical Risk Prediction Model for Lung Cancer in Smokers of All Ages and Exposure Types: A HUNT Study. *EBioMedicine* **2018**, *31*, 36–46. [CrossRef] [PubMed]

25. Panagopoulou, M.; Karaglani, M.; Balgkouranidou, I.; Biziota, E.; Koukaki, T.; Karamitrousis, E.; Nena, E.; Tsamardinos, I.; Kolios, G.; Lianidou, E.; et al. Circulating cell-free DNA in breast cancer: Size profiling, levels, and methylation patterns lead to prognostic and predictive classifiers. *Oncogene* **2019**, *38*, 3387–3401. [CrossRef] [PubMed]

26. Panagopoulou, M.; Karaglani, M.; Manolopoulos, V.G.; Iliopoulos, I.; Tsamardinos, I.; Chatzaki, E. Deciphering the Methylation Landscape in Breast Cancer: Diagnostic and Prognostic Biosignatures through Automated Machine Learning. *Cancers* **2021**, *13*, 1677. [CrossRef]

27. Adamou, M.; Antoniou, G.; Greasidou, E.; Lagani, V.; Charonyktakis, P.; Tsamardinos, I.; Doyle, M. Toward Automatic Risk Assessment to Support Suicide Prevention. *Crisis* **2019**, *40*, 249–256. [CrossRef]

28. Liu, Y.; Tan, Q.; Liu, F. Differentially methylated circulating DNA: A novel biomarker to monitor beta cell death. *J. Diabetes Its Complicat.* **2018**, *32*, 349–353. [CrossRef]

29. McTaggart, J.S.; Clark, R.H.; Ashcroft, F.M. The role of the KATP channel in glucose homeostasis in health and disease: More than meets the islet. *J. Physiol.* **2010**, *588*, 3201–3209. [CrossRef]

30. Karaglani, M.; Ragia, G.; Panagopoulou, M.; Balgkouranidou, I.; Nena, E.; Kolios, G.; Papanas, N.; Manolopoulos, V.G.; Chatzaki, E. Search for Pharmacoepigenetic Correlations in Type 2 Diabetes Under Sulfonylurea Treatment. *Exp. Clin. Endocrinol. Diabetes* **2019**, *127*, 226–233. [CrossRef] [PubMed]

31. Aucamp, J.; Bronkhorst, A.J.; Badenhorst, C.P.S.; Pretorius, P.J. The diverse origins of circulating cell-free DNA in the human body: A critical re-evaluation of the literature. *Biol. Rev. Camb. Philos Soc.* **2018**, *93*, 1649–1683. [CrossRef] [PubMed]

32. Laktionov, P.P.; Tamkovich, S.N.; Rykova, E.Y.; Bryzgunova, O.E.; Starikov, A.V.; Kuznetsova, N.P.; Sumarokov, S.V.; Kolomiets, S.A.; Sevostianova, N.V.; Vlassov, V.V. Extracellular circulating nucleic acids in human plasma in health and disease. *Nucleosides Nucleotides Nucleic Acids* **2004**, *23*, 879–883. [CrossRef]

33. El Tarhouny, S.A.; Hadhoud, K.M.; Ebrahem, M.M.; Al Azizi, N.M. Assessment of cell-free DNA with microvascular complication of type II diabetes mellitus, using PCR and ELISA. *Nucleosides Nucleotides Nucleic Acids* **2010**, *29*, 228–236. [CrossRef] [PubMed]

34. Panagopoulou, M.; Karaglani, M.; Balgkouranidou, I.; Pantazi, C.; Kolios, G.; Kakolyris, S.; Chatzaki, E. Circulating cell-free DNA release in vitro: Kinetics, size profiling, and cancer-related gene methylation. *J. Cell. Physiol.* **2019**, *234*, 14079. [CrossRef]

35. Donath, M.Y.; Ehses, J.A.; Maedler, K.; Schumann, D.M.; Ellingsgaard, H.; Eppler, E.; Reinecke, M. Mechanisms of β-Cell Death in Type 2 Diabetes. *Diabetes* **2005**, *54*, S108. [CrossRef]

36. Tsamardinos, I.; Greasidou, E.; Borboudakis, G. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. *Mach. Learn.* **2018**, *107*, 1895–1922. [CrossRef] [PubMed]

37. Borboudakis, G.; Tsamardinos, I. Extending greedy feature selection algorithms to multiple solutions. *Data Min. Knowl. Discov.* **2021**, *35*, 1393–1434. [CrossRef] [PubMed]

38. Deutsch, L.; Stres, B. The Importance of Objective Stool Classification in Fecal 1H-NMR Metabolomics: Exponential Increase in Stool Crosslinking Is Mirrored in Systemic Inflammation and Associated to Fecal Acetate and Methionine. *Metabolites* **2021**, *11*, 172. [CrossRef]

39. Nagy, Á.; Ligeti, B.; Szebeni, J.; Pongor, S.; Gyrffy, B. COVIDOUTCOME-estimating COVID severity based on mutation signatures in the SARS-CoV-2 genome. *Database* **2021**, *2021*. [CrossRef]

40. Glaab, E.; Rauschenberger, A.; Banzi, R.; Gerardi, C.; Garcia, P.; Demotes, J. Biomarker discovery studies for patient stratification using machine learning analysis of omics data: A scoping review. *BMJ Open* **2021**, *11*, e053674. [CrossRef]

41. Lai, H.; Huang, H.; Keshavjee, K.; Guergachi, A.; Gao, X. Predictive models for diabetes mellitus using machine learning techniques. *BMC Endocr. Disord.* **2019**, *19*, 101. [CrossRef]

42. Syed, F.; Tersey, S.A.; Turatsinze, J.V.; Felton, J.L.; Kang, N.J.; Nelson, J.B.; Sims, E.K.; Defrance, M.; Bizet, M.; Fuks, F.; et al. Circulating unmethylated CHTOP and INS DNA fragments provide evidence of possible islet cell death in youth with obesity and diabetes. *Clin. Epigenetics* **2020**, *12*, 116. [CrossRef]

43. Arosemena, M.; Meah, F.A.; Mather, K.J.; Tersey, S.A.; Mirmira, R.G. Cell-Free DNA Fragments as Biomarkers of Islet β-Cell Death in Obesity and Type 2 Diabetes. *Int. J. Mol. Sci.* **2021**, *22*, 2151. [CrossRef]

44. American Diabetes Association. 2. Classification and Diagnosis of Diabetes. *Diabetes Care* **2015**, *38*, S8. [CrossRef]

45. Bustin, S.A.; Benes, V.; Garson, J.A.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M.W.; Shipley, G.L.; et al. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **2009**, *55*, 611–622. [CrossRef]

46. Li, L.C.; Dahiya, R. MethPrimer: Designing primers for methylation PCRs. *Bioinformatics* **2002**, *18*, 1427–1431. [CrossRef]

47. Lu, L.; Katsaros, D.; de la Longrais, I.A.; Sochirca, O.; Yu, H. Hypermethylation of let-7a-3 in epithelial ovarian cancer is associated with low insulin-like growth factor-II expression and favorable prognosis. *Cancer Res.* **2007**, *67*, 10117–10122. [CrossRef]