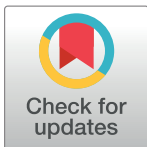


RESEARCH ARTICLE

Defining and characterizing the critical transition state prior to the type 2 diabetes disease

Bo Jin¹, Rui Liu^{2,3}, Shiyong Hao^{2,4}, Zhen Li^{2,5}, Chungqing Zhu¹, Xin Zhou⁶, Pei Chen⁷, Tianyun Fu¹, Zhongkai Hu², Qian Wu⁸, Wei Liu⁸, Daowei Liu⁸, Yunxian Yu⁹, Yan Zhang^{2,10}, Doff B. McElhinney^{2,4}, Yu-Ming Li⁶, Devore S Culver¹¹, Shaun T. Alfreds¹¹, Frank Stearns¹, Karl G. Sylvester², Eric Widen¹, Xuefeng B. Ling^{2,4,12*}

1 HBI Solutions Inc., Palo Alto, California, United States of America, **2** Stanford University, Stanford, California, United States of America, **3** School of Mathematics, South China University of Technology, Guangzhou, China, **4** Clinical and Translational Research Program, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford, California, United States of America, **5** School of Electrical Engineering, Southeast University, Nanjing, China, **6** Tianjin Key Laboratory of Cardiovascular Remodeling and Target Organ Injury, Pingjin Hospital Heart Center, Tianjin, China, **7** School of Computer Science and Engineering, South China University of Technology, Guangzhou, China, **8** China Electric Power Research Institute, Beijing, China, **9** School of Medicine, Zhejiang University, Hangzhou, China, **10** Department of Oncology, the First Hospital of Shijiazhuang, Shijiazhuang, Hebei, China, **11** HealthInfoNet, Portland, Maine, United States of America, **12** Health Care Big Data Center, School of Public Health, Zhejiang University, Hangzhou, China



OPEN ACCESS

Citation: Jin B, Liu R, Hao S, Li Z, Zhu C, Zhou X, et al. (2017) Defining and characterizing the critical transition state prior to the type 2 diabetes disease. PLoS ONE 12(7): e0180937. <https://doi.org/10.1371/journal.pone.0180937>

Editor: Cheng Hu, Shanghai Diabetes Institute, CHINA

Received: March 16, 2017

Accepted: June 24, 2017

Published: July 7, 2017

Copyright: © 2017 Jin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The work was performed under a business arrangement between HealthInfoNet (<http://www.hinonet.org>), the operators of the Maine Health Information Exchange and HBI Solutions, Inc. (HBI) located in California. The data is owned by the members of the HIN. HIN is responsible for security and access to its members' data and has established data service agreements (DSAs) restricting unnecessary exposure of information. HIN and its board (comprised from a cross section of its members) authorized the use of the de-identified data for this research. Access to the data used in the study

* These authors contributed equally to this work.

* bxling@stanford.edu

Abstract

Background

Type 2 diabetes mellitus (T2DM), with increased risk of serious long-term complications, currently represents 8.3% of the adult population. We hypothesized that a critical transition state prior to the new onset T2DM can be revealed through the longitudinal electronic medical record (EMR) analysis.

Method

We applied the transition-based network entropy methodology which previously identified a dynamic driver network (DDN) underlying the critical T2DM transition at the tissue molecular biological level. To profile pre-disease phenotypical changes that indicated a critical transition state, a cohort of 7,334 patients was assembled from the Maine State Health Information Exchange (HIE). These patients all had their first confirmative diagnosis of T2DM between January 1, 2013 and June 30, 2013. The cohort's EMRs from the 24 months preceding their date of first T2DM diagnosis were extracted.

Results

Analysis of these patients' pre-disease clinical history identified a dynamic driver network (DDN) and an associated critical transition state six months prior to their first confirmative T2DM state.

requires secure connection to HIN servers and should be requested directly to HIN. For any data request, researchers shall contact Phil Profenno (pprofenno@hinfonet.org) in HIN, and follow the HIN compliance requirements to submit an application. If the request is approved by HIN, the researchers will be granted access to the de-identified data via secure connections managed by HIN. HIN agrees to provide access to the de-identified data on a per request basis to interested researchers.

Funding: The authors received no public funding for this work. HBI Solutions, Inc. (HBI) is a private commercial company, and several authors are employed by HBI. HBI provided funding in the form of salaries to the authors employed by HBI: BJ, TF, CZ, FS and EW, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: We have the following interests: KGS, EW and XBL are co-founders and equity holders of HBI Solutions, Inc., which is currently developing predictive analytics solutions for healthcare organizations. BJ, TF, CZ, FS and EW are employed by HBI Solutions, Inc. From the Stanford University School of Medicine, Stanford, California, KGS and XBL conducted this research as part of a personal outside consulting arrangement with HBI Solutions, Inc. The research and research results are not, in any way, associated with Stanford University. There are no patents, further products in development or marketed products to declare. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

Conclusions

This 6-month window before the disease state provides an early warning of the impending T2DM, warranting an opportunity to apply proactive interventions to prevent or delay the new onset of T2DM.

Introduction

Type 2 diabetes mellitus (T2DM) is a metabolic disorder characterized by hyperglycemia caused by either or both insulin resistance or insufficient insulin production. People with T2DM are at high risk for several serious health complications including cardiovascular diseases, blindness, kidney failure, limb amputations, premature death, fractures, frailty, depression, and cognitive decline. Because of the shifting age structure of the global population and the trend of urbanization and lifestyle changes in developing countries, the prevalence of diabetes among adults, 422 million worldwide in 2016, nearly doubles from 1980 to 2016 [1]. The total estimated economic impact of those diagnosed with diabetes in the US in 2012 was \$245 billion, a cost that included \$176 billion in direct medical costs and \$69 billion in reduced productivity [2].

T2DM can often be prevented or delayed by maintaining a healthy life style, including a normal body weight, engaging in physical exercise smoking cessation, and consuming a healthful diet [3, 4]. A meta-analysis reported that 3-month lifestyle interventions decreased the risk for diabetes from the end of intervention up to 10 years later [5]. Additionally, a modifiable disease inflection point was hypothesized during the gradual progression to chronic diseases [6–11] after which the return to the pre-disease state is not possible. Therefore, it would be of high health priority to develop an early warning system for T2DM such that an opportunity window of greater than three months prior to the disease state shall warrant an effective intervention for the preventive care or the implementation of interventions to slow down the T2DM progression.

We hypothesized that complex diseases progress through three states, i.e., a normal state, a pre-disease state (or a critical transition state), and a disease state [12–15]. From population health point of view, it is essential to identify the pre-disease state to apply effective intervention, thus, qualitative deterioration can be managed. For many complex diseases, however, it is a difficult task to predict a pre-disease state because the state of the system may change little before the bifurcation point or the critical transition is reached, namely, there may be little difference between the normal and pre-disease states. A pre-disease state can be considered as a limit of the normal state but a disease state is different from the normal state. This is also why diagnoses based on traditional biomarkers or static measurements fail to distinguish a pre-disease state from a normal state. Additionally complexity of diseases, including T2DM, can involve thousands of genetic factors (e.g., SNPs and CNVs), epigenetic factors (e.g., methylation and acetylation), and different worldwide environment exposures. To resolve this problem, we piloted a novel model-free method to detect early warning signals of diseases [12–15]. Its theoretical foundation is based on dynamical driver network (DDN), which is also called as the driver (or leading) network of the disease because components in DDN drive the whole system from one state (e.g. normal state) to another (e.g. disease state). DDN can not only uncover significant signals among noises the emergence of the critical transitions for the commitment of disease progression, but can also help the formulation of discovery hypotheses

down to genetic and biochemical pathways. However, DDN itself cannot simply be recognized as a causal network of T2DM critical transition.

We applied through the application of the Transition Network Entropy (TNE) and DDN methods to T2DM disease transition analysis with high throughput genomic datasets. Specifically, based on the temporal-spatial gene expression data of T2DM, we identified tissue-specific DDNs corresponding to the critical transitions during T2DM development and progression. Our hypothesis of T2DM critical transition state at the tissue molecular level was validated during disease deterioration and characterized as responses to insulin resistance and serious inflammation. Based on the functional and network analysis on pathogenic molecular mechanism of T2DM, we showed that most of DDN genes, particularly the core ones, tended to locate at the biological pathways upstream of T2DM pathophysiology, which implied that DDN-identified genes act as the causal factors rather than the consequence to drive the downstream molecules to change their transcriptional activities. This finding validated our theoretical prediction of DDN as the driver network.

An improved understanding of disease at the system medicine level is anticipated to result from the big data analytics of the such statewide or nationwide electronic medical records (EMRs), particularly when combined with natural language processing and extraction of the unstructured clinical note information [16–22]. Statewide health information exchange (HIE) is a secure, standardized electronic system where providers can share important patient health information for treatment purposes. Before the HIE implementation, there have been lack of accurate and complete electronic medical data, inability to access real-time information, and missed or late information. Patient records were stored in several physician offices, different hospital storage systems, and independent laboratories. Patient records were available only when provider offices are open or certain staff are available. Primary care physician was unaware or alerted late to test results or hospital admission/discharge. With HIE, clinical medical records are now stored in one electronic location giving providers access to a complete patient view. Data can be accessible to providers in real-time. Once information is generated in native EMR, it's in the HIE within minutes. Event alerts and notifications can be sent from HIE to care manager through email. In addition, with the advance of machine learning and artificial intelligence, HIE leadership started to focus on not merely delivering information, but also on putting the vast amount of data they collect to work on big data analytics helping hospitals, accountable care organizations and physician practices put the power of near real-time data to work on improving patient care.

A major advantage in the application of EMR data mining stems from parallel analysis of thousands of demographical/clinical parameters and events, like the analytics of high dimensional genomics experiments, allowing the efficient detection of discriminant clinical patterns for sensitive measure of pathophysiological status and monitoring altered disease progression. Changes in clinical status is then based on the detection of perturbations in the frequencies and measures of specific clinical events and laboratory test results involved in several key disease-related or other specific clinical pathways [16–22]. Thus, EMR analytics can reveal crucial information that is closely related to the disease progression or outcomes of targeted therapeutics.

In line with our previous work of diabetes transition analysis with genomic datasets, we hypothesize that a critical transition state of T2DM disease progression can be identified through EMR mining. The detection of the T2DM transition state using HIE EMRs prior to definitive T2DM onset would provide an opportunity window for effective interventions to prevent or delay the T2DM development. To test this hypothesis, we set to apply our transition-based network entropy model [23] to identify the critical transition state and the leading network of longitudinal clinical histories to device early warning system of impending T2DM for population health management.

Result

Hypothesis testing of the EMR based critical transition state prior to T2DM

The T2DM initiation dynamics before sudden deterioration is usually too complicated to be fully expressed mathematically in high dimensional spaces. The drastic or a qualitative transition in a local system or network, from a normal (i.e. disease-free) state to a disease state, corresponds to a so-called bifurcation point in dynamical systems theory [24, 25]. If the system is approaching a bifurcation point, it will eventually be constrained to a one- or two-dimensional space (i.e., the center manifold), in which a dynamical system can be expressed in a very simple form. This is the theoretical basis for developing a general indicator that can detect the T2DM critical transition.

As shown in Fig 1, we collected the comprehensive historical longitudinal EMR datasets from all of Maine's HIE aggregated patients. A cohort of 7,334 patients with confirmatively diagnosis of T2DM between January 1, 2013 and June 30, 2013 was constructed (Fig 1A). We hypothesized three states during the T2DM progression can be characterized by clinical patterns in EMR: the normal state, the transition state, and the disease state (Fig 1B). To profile the critical transition state prior to T2DM, a feature correlation network with 33,403 features was analyzed using the network entropy algorithm (Fig 1C). The whole feature network was localized, and classified into three layers, which was calculated as the network entropy of the state change in a Markov process. Markov process and its ability to govern the dynamics of each node in the network was described in the following section. It was expected that there is a DDN with critical behaviors occurring when the system is in the transition state (Fig 1D).

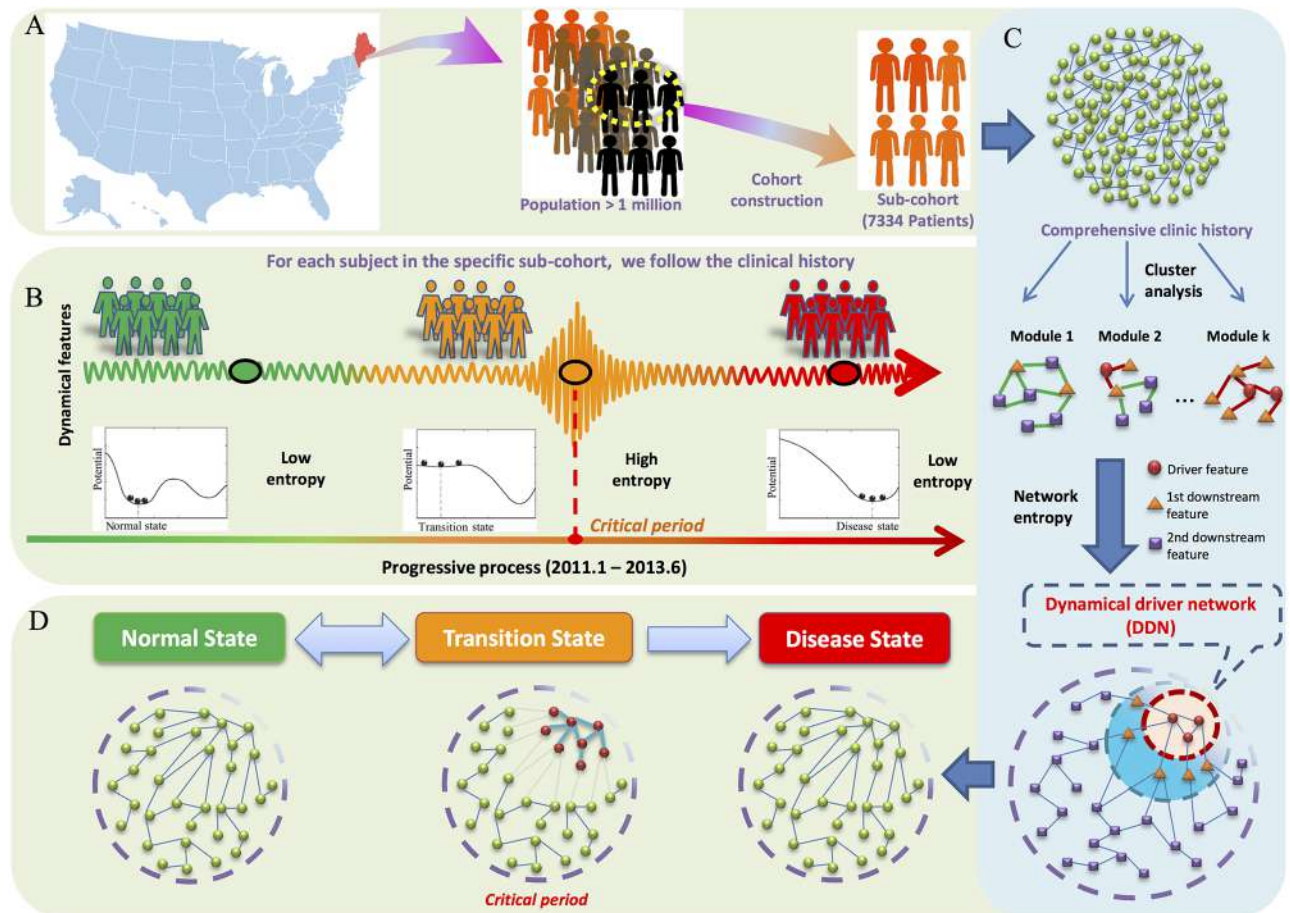
The critical transition detection prior to the first T2DM confirmative diagnosis

As shown in Fig 1B, by applying the TNE method to the cohort's clinical records, a DDN ($n = 257$ features) was identified to show the existence of the critical transition between a disease-free state and a T2DM disease state. Fig 2D presents the dynamical evolution in the whole feature network. For patients later developing T2DM, the DDN entropy becomes significantly high six months before the first confirmative diagnosis (Fig 2D), indicating a signal of the critical transition before the deterioration into T2DM. In contrast, there is no significant change in controls' feature network dynamical progression (S1 Fig). When the deterioration is impending, these selected features form a special DDN sub-network, and make the first move from the normal state toward the disease state during the transition. DDN's behavior validates our hypothesis that there is a transition state or pre-disease period between the normal and disease states. In the study population, the transition state occurred around 6 months before the T2DM confirmative state.

Fig 2A–2C show that the BMI, glucose, and A1C test results failed to identify the critical transition period for T2DM diagnosis on the same population. These results indicate that risk factors that are widely used by clinicians for T2DM diagnosis may not be able to serve as early-warning T2DM signals to alert patients prior to the confirmative diagnosis (i.e., prior to the time point 0 in Fig 2A–2C). It suggests that compared to the traditional diagnosis methods, the DDN-based approach is more effective in identifying the critical T2DM transition point.

DDN's functional characterization

DDN features were further characterized for their association and relatedness to T2DM clinical development. The 257 selected features, which constructed the DDN, classify into abnormal



We hypothesize there is a transition state between a normal and a disease state of diabetes.

Fig 1. The schematic of dynamic driver network (DDN) based model integrated to the Maine Health Information Exchange workflow to characterize the critical transition state prior to the type 2 diabetes mellitus (T2DM) disease. (A) Based on clinical records of 1.3 million people from Maine State, USA, we carried out a population study and extracted a sub-cohort with 7,334 patients with the first T2DM confirmative diagnosis during the study period. (B) The progression of T2DM can be divided into three stages, i.e., the normal state with relatively low entropy, the transition state right before the critical transition with relatively high entropy, the disease state with relatively low entropy. The sharp increase of entropy is expected to characterize the transition state before getting into the disease state. (C) With a transition-based network entropy, the features can be classified into three layers, and the DDN can be obtained. Based on the dynamical characteristics (such as comprehensive clinic history, or time-course information) the cluster analysis suffices to separate the network into a few functional modules. Further analysis via network entropy aggregates these modules and identifies the DDN. (D) Employing the transition-based network entropy method, we succeed in presenting the existence of a transition state (orange) between a normal state (green) and a disease state (red). The network structure of features can be divided into two parts, the DDN, and other downstream features. The DDN provides the indicative warning signals to the sudden deterioration of diabetes. The map in the figure was created by Adobe Photoshop CS6 (<https://helpx.adobe.com/x-productkb/policy-pricing/cs6-product-downloads.html>).

<https://doi.org/10.1371/journal.pone.0180937.g001>

lab test results, abnormal radiology test results, and medications. Medication features can be sub grouped by the associated primary diagnoses, procedures, or chronic conditions. Fig 3 summarizes the total 16 subgroups of features. Most features are correlated to the medication utilized in chronic diseases closely related to diabetes, such as cardiovascular diseases and metabolic disease. It indicates that the derived DDN is clinically reasonable for T2DM critical transition point identification, and thus highlights an opportunity to develop an early warning tool for clinical use based on the DDN method. Some features are related to abnormal lab test (n = 10 features) or the use of radiology tests (n = 3 features). These tests reflect the incubating risk factors, which may trigger the deterioration into T2DM.

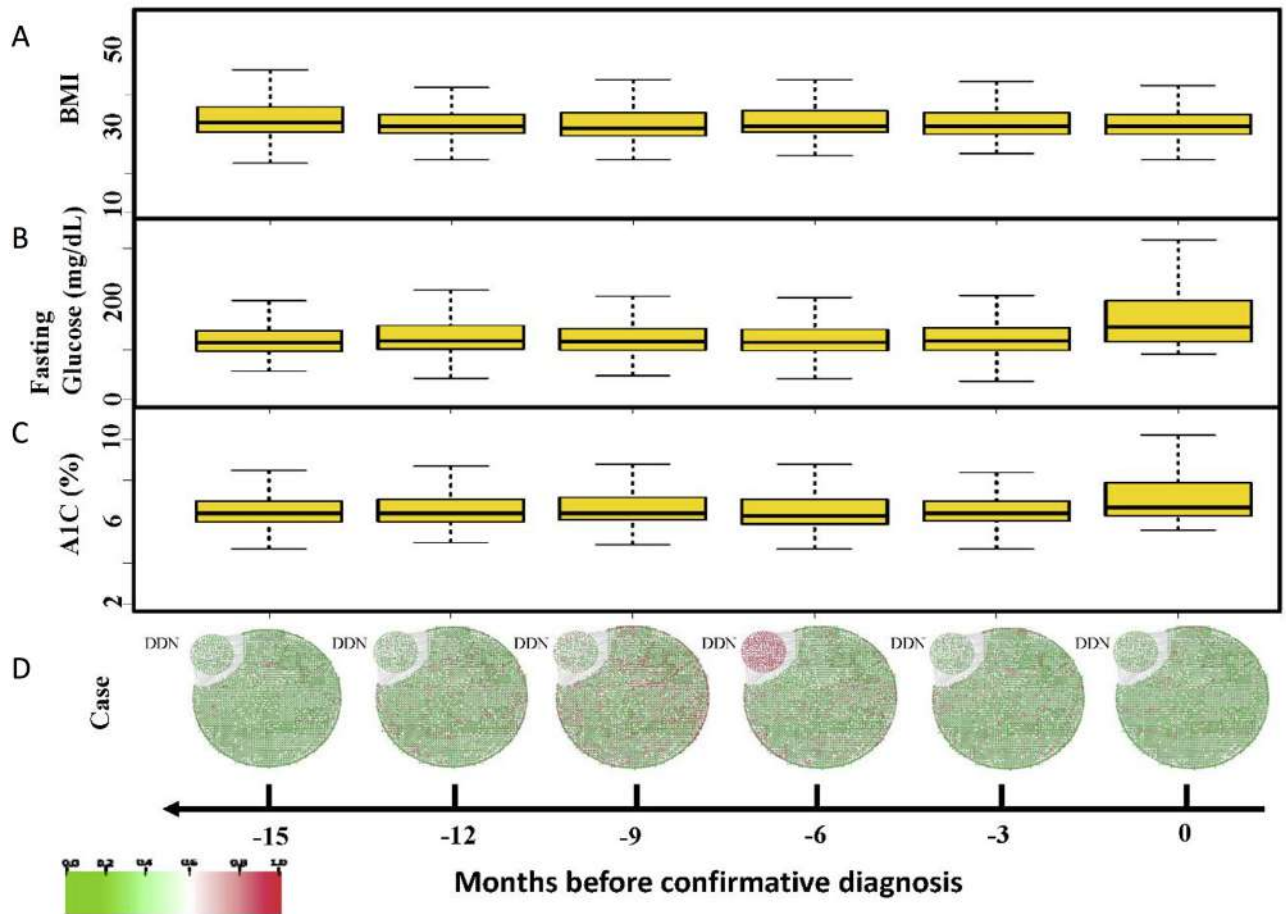


Fig 2. The comparison of dynamical evolution for the dynamic driver network (DDN) and the traditional biomarkers. (A) The trending of BMI illustrates no obvious increase tendency even when the patient is near the critical transition into type 2 diabetes mellitus (T2DM). (B) The glucose index remains a consistent value (less than 126 mg/dl) before the confirmative diagnosis (time point 0) with no indicative signal even when the patient is near the critical transition into T2DM. (C) The A1C index remains less than 6.5% before the confirmative diagnosis (time point 0) and does not change significantly even when the patient is near the critical transition into T2DM. (D) The evolution of DDN shows an early-warning signal can be detected 6 months before the diagnosis of T2DM.

<https://doi.org/10.1371/journal.pone.0180937.g002>

Clinical and resource utilization patterns during the transition to T2DM

The TNE-based DDN method profiles and differentiates T2DM's 3-state progression. We analyzed the changes of patients' clinical outcomes and resource utilization from the normal state to the disease state (Figs 4–7). It shows that among the patients diagnosed as T2DM in the disease period, 60% had essential hypertension, and 50% had abnormalities of lipid metabolism (Fig 4). The presence of these chronic conditions makes T2DM a costly chronic disease. As shown in Figs 4–7, the associated costs and resource utilization including clinical visits and laboratory utilization abruptly increase rapidly until reaching a peak at one month after the diagnosis was confirmed, then quickly decline in the following months.

Discussion

Consistent with our hypothesis, the results indicate that there is a critical transition that could be detected 6 months before the T2DM confirmative diagnosis. Unlike the detection of the

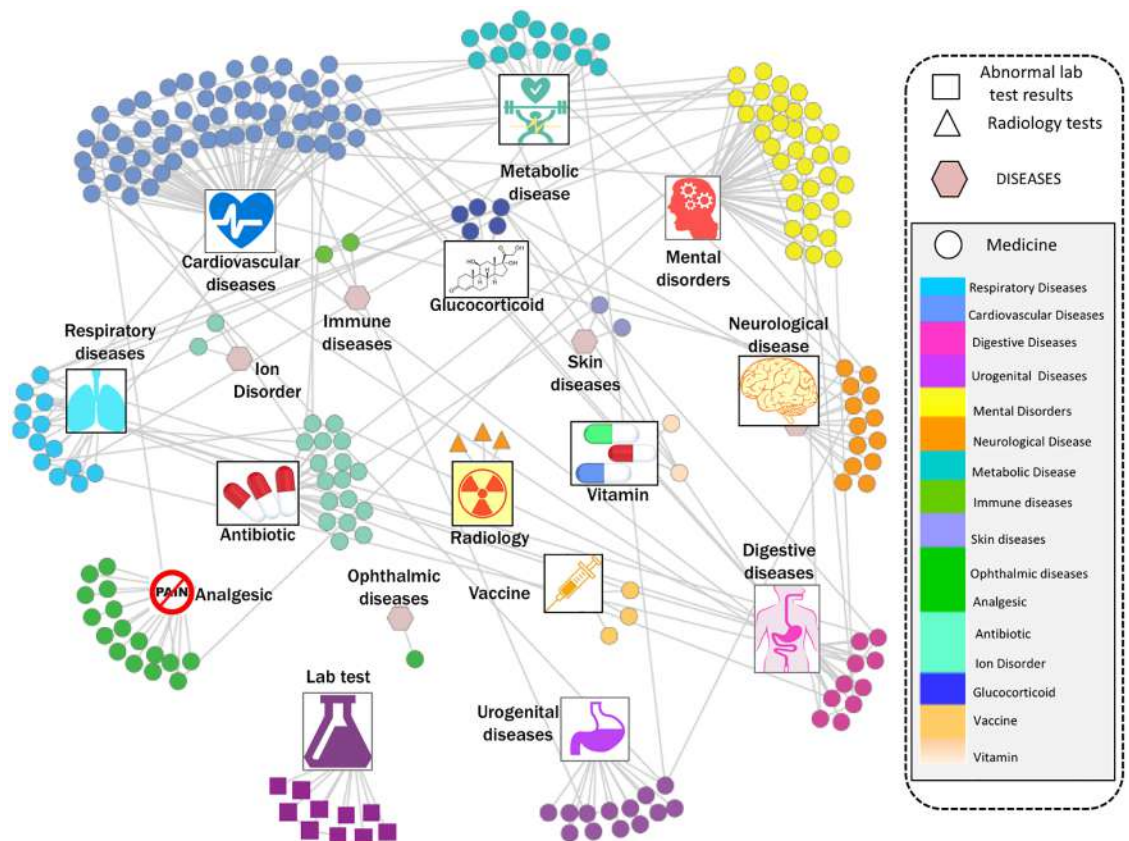


Fig 3. Modular structure of the dynamic driver network (DDN) features. The selected features of the DDN network can be classified into 16 subgroups of demographics, primary diagnosis, procedure, etc. Most features were found to be directly involving to the utilization of medicine to manage chronic diseases such as cardiovascular diseases and metabolic disease, which indicates the derived DDN is clinically reasonable for the critical transition identification of type 2 diabetes mellitus.

<https://doi.org/10.1371/journal.pone.0180937.g003>

disease state, the DDN enables the identification of an intermediate or transition state that generally has no clear metabolic abnormalities but with future trending of deterioration.

The DDN features discovered in this study may provide a repository of T2DM associative and/or disease driving clinical patterns underlying the progression of T2DM. This may help identify modifiable clinical features for proactive intervention for preventive care or to slow down the disease progression. Rather than using only a pre-defined set of clinical findings (such as IGT, IFG, A1C and BMI), we developed a mathematical approach of system medicine, focusing on the longitudinal evolution of each clinical feature and their relationships with each other during the procession of disease. The T2DM DDN network features (n = 257) can be divided into three subgroups: abnormal lab test results, abnormal radiology test results, and medications. Most of these features are highly correlated with T2DM or its complications, e.g. cardiovascular diseases and other metabolic diseases.

218/257 DDN features are medication prescriptions. Medication history can be used to gauge patient disease burden. Demographic characteristics (S2 Table) show that patients in transition to T2DM had significant chronic disease burden, including hypertension, heart problems and asthma, prior to the onset of T2DM, compared with controls. Such differences were reflected by medication prescriptions that comprised DDN. For example, Diltiazem, Metolazone, and Losartan can help to reduce risk of high blood pressure. Statin and Digoxin

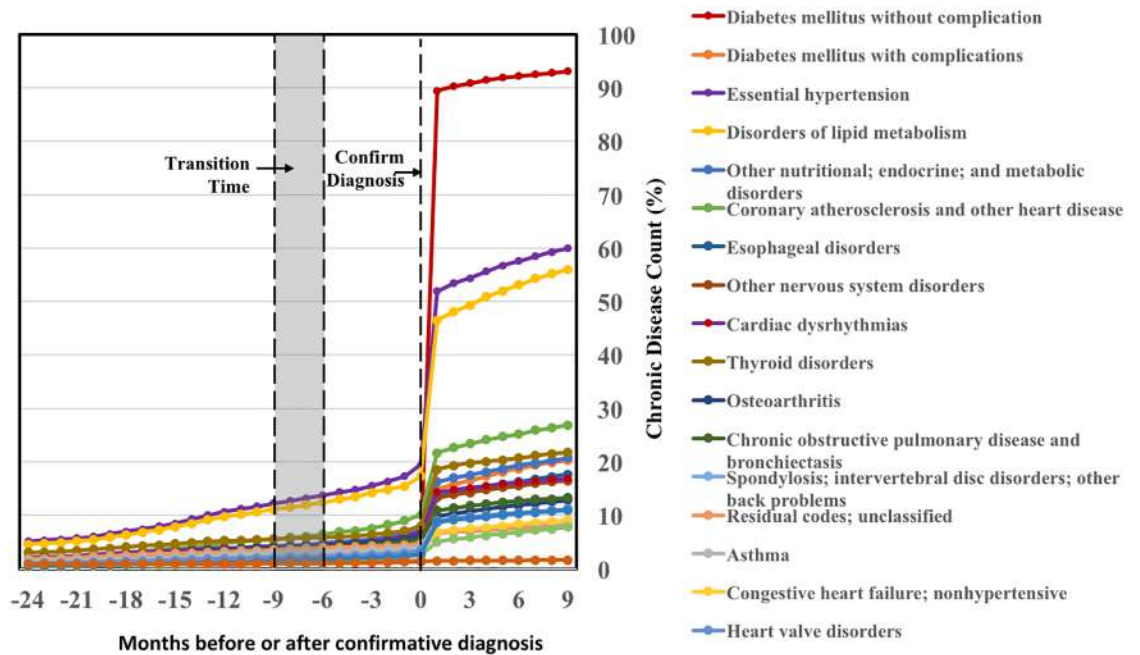


Fig 4. Trending of T2DM related chronic disease counts from 24 months prior to the diagnosis of T2DM to 6 months after the diagnosis of T2DM. The counts of diabetes mellitus without complication, the essential hypertension, and the disorders of lipid metabolism abruptly increase after the diagnosis was confirmed.

<https://doi.org/10.1371/journal.pone.0180937.g004>

can help with cholesterol lowering and heart diseases. These chronic diseases and T2DM are interrelated diseases, and medications to treat these diseases may have either positive or negative impact on critical transition from latent to T2DM onset. Cholesterol drugs, including Lovastatin, Atorvastatin, and Crestor that were selected as DDN features, help to maintain balance

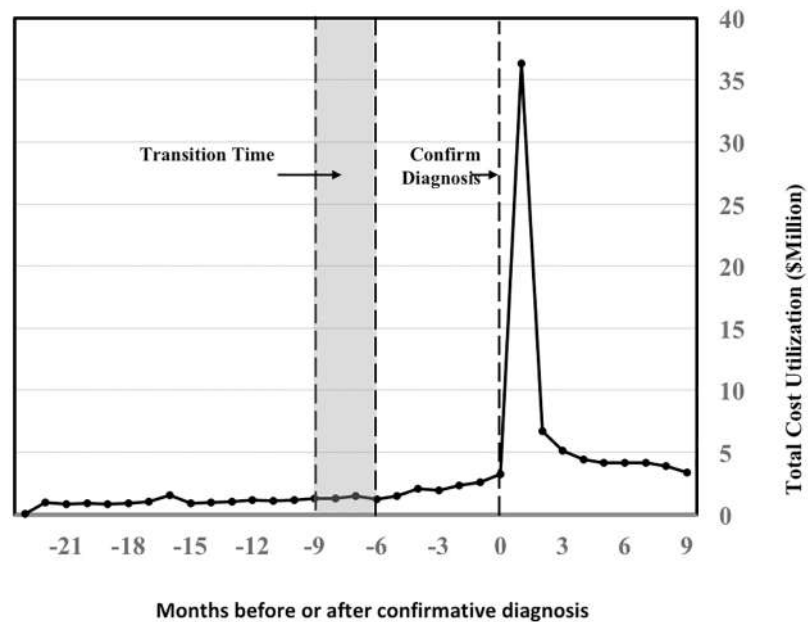


Fig 5. Trending of the total cost of utilization from 24 months prior to the diagnosis of type 2 diabetes mellitus (T2DM) to 6 months after the diagnosis of T2DM. The total cost reaches the peak rapidly after the confirmative diagnosis.

<https://doi.org/10.1371/journal.pone.0180937.g005>

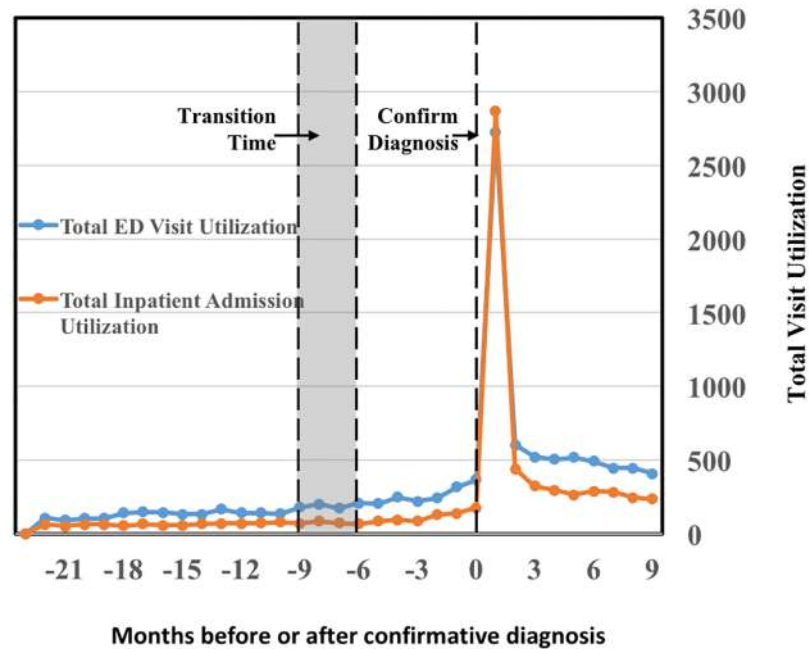


Fig 6. Trending of the total emergency department (ED) visit utilization and total inpatient admission utilization from 24 months prior to the diagnosis of type 2 diabetes mellitus (T2DM) to 6 months after the diagnosis of T2DM. The two curves both rise abruptly after confirmative diagnosis.

<https://doi.org/10.1371/journal.pone.0180937.g006>

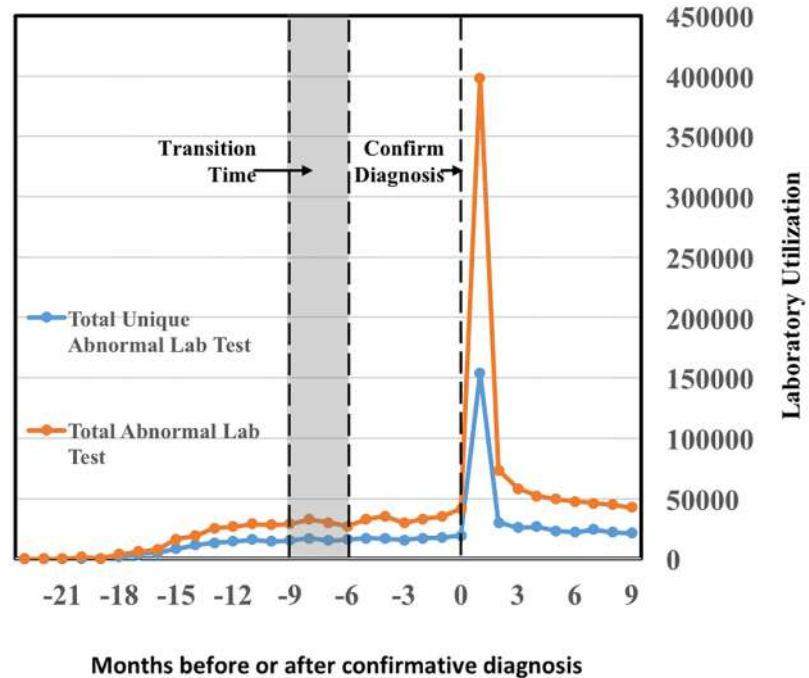


Fig 7. Trending of unique abnormal lab test volume and total abnormal lab test volume from 24 months prior to the diagnosis of type 2 diabetes mellitus (T2DM) to 6 months after the diagnosis of T2DM. Both volumes increase sharply after the confirmative diagnosis.

<https://doi.org/10.1371/journal.pone.0180937.g007>

between HDL and LDL cholesterol, which contributes to T2DM preventive care as low levels of HDL cholesterol are associated with elevated risk of T2DM [26]. Persistent medication history associated with elevated risk of T2DM illustrates the activation of the potential underlying clinical pathways responsible for the transition from latent to T2DM state. DDN medication features such as antibiotics (Levofloxacin, Fluoroquinolone, Azithromycin, and Nystatin), β blockers (Sotalol and Metoprolol), diuretics (Metolazone and Hydrochlorothiazide), and antidepressants (Amitriptyline, Citalopram, and Fluoxetine), can actually increase the risk of T2DM [27–30]. Both the positive and negative correlated medications in regard to T2DM disease transition collectively suggest that our findings suggest the clinical care utilities of the DDN features.

Like medication prescription, abnormal laboratory test results that were selected by DDN learning process also reflect clinical status of our cohort. Abnormal results of metabolic panel, cardiac panel, and urine culture are consistent with the cohort subjects' comorbidities (S2 Table), including increased risk of having metabolic syndrome and obesity, heart problem, kidney diseases, and infection.

The cost and resource utilization among T2DM new onset cases primarily occurred and peaked shortly after the disease was confirmed. Such findings indicate that identification of transition state prior to T2DM may allow the timely initiation of interventions to reduce clinical resource utilization.

The DDN method is a generalized methodology for mining the dynamical information from the longitudinal clinical EMR patterns and identifying the transition state. This method was previously applied to genomics based diabetic disease progression analysis to identify the states of normal, critical transition, and new disease onset [13–15]. In this study, we applied this method to profile the EMR clinical patterns and identify the critical transition state of T2DM. The work constructs the first step towards the development of an early warning system of T2DM for population health.

Compared with other early warning methods to identify the critical transitions, the DDN method has several advantages. First, it is a more efficient method. Only the local information from a local network was involved in the calculation, which largely reduced the complexity. Second, by ignoring the nodes where the TNE changes slightly the noise in the calculation was essentially reduced, and the bias in identifying the critical transitions was avoided. Third, the conventional correlation-based methods analyzed the linear relationships between the nodes only, while the DDN method can depict the nonlinear relationships among the nodes. Fourth, the DDN features obtained by this method provide an approach to the underlying mechanisms during the progression of the disease, which would be very helpful in the development of the clinical treatment of the diseases.

T2DM is a progressive disease that is associated with a tremendous economic burden to both the U.S. health care system and the whole society. Reducing T2DM incidence and its complications through intensive lifestyle and pharmacologic interventions will result in better patient outcomes [31–33]. According to two clinical trials, a 3-month intensive lifestyle intervention in high-risk patients significantly reduced those diagnosed with diabetes during a 3- to 10-year period [34, 35]. The effectiveness of interventions, however, is largely dependent on the timing of their application relative to disease transitions. Our study shows that clinical profile of patients have detectable change prior to confirmative diagnosis of T2DM, compared with normal patients. Such change could happen during or even before the insulin resistance state when glucose level is still in a normal range. The clinical profile change, defined as critical transition state in this study, can be identified by our DDN algorithm. Identification of critical transition state of T2DM improves awareness of both providers and patients, and thus helps decision making on diet and life style interventions before the commitment of the T2DM

progression. It warrants the timely interventions targeted to the high-risk patients prior to deterioration, which can device effective preventive care, improve the quality of disease management, and reduce the overall clinical resource utilization.

In our previous work, we demonstrated it is efficient method to detect both T1DM and T2DM using high dimensional genomics datasets [13–15]. This study integrated statewide HIE EMR longitudinal patterns and applied the TNE-based DDN method to uncover the disease transition state prior to the new onset T2DM. The DDN analytics with HIE real time EMR datasets can be generalized to apply to other complex chronic diseases. Our approach, to identify complex disease critical transition state as an early warning system and case-find high-risk patients statewide or nationwide in real time, can promise to translate the medicine of all complex chronic diseases.

We notice there are some limitations of this study. First, the average number of patient records prior to diagnosis of T2DM was 8.4, while 14% of patients had no patient record before diagnosis. These patients might have been diagnosed with T2DM somewhere out of HIE network, or have developed T2DM but without awareness. For those patients without EMR records of medication, laboratory test results, or chronic disease history, they may be bucketed into the baseline subgroup without commitment to T2DM progression. Second, data sources of this study are EMR datasets routinely collected by Maine HIE. Study population therefore are limited to those who had visited any of the care facilities participating HIE network. Adjustments would be needed when our statistical learning and the associated healthcare management protocols need to be transferred to populations in other geographical and/or cultural areas outside HIE network. Geographical, environmental, and racial disparity analyses may uncover additional T2DM critical transition features for T2DM population health.

Methods

Data availability

The work was performed under a business arrangement between HealthInfoNet (<http://www.hinfonyet.org>), the operators of the Maine Health Information Exchange and HBI Solutions, Inc. (HBI) located in California. By business arrangement we mean HBI is a contracted vendor to HealthInfoNet (HIN), and HBI is under contract to deploy its proprietary applications and risk models on the HIN data for use by HIN members. HIN is a steward of the data on behalf of its members which includes health systems, hospitals, medical groups and federally qualified health centers. The data is owned by the HIN members, not HIN. HIN is responsible for security and access to its members' data and has established data service agreements (DSAs) restricting unnecessary exposure of information. HIN and its board (comprised from a cross section of its members) authorized the use of the de-identified data for this research, as the published research helps promote the value of the HIE and value to Maine residents.

HBI receives revenue for providing this service, which is performed remotely. HBI does not own or have access to the data outside of providing services to HIN. HIN manages and controls the data within its technology infrastructure. The research was conducted on HIN technology infrastructure, and the researchers accessed the de-identified data via secure remote methods. All data analysis and modeling for this manuscript was performed on HIN servers and data was accessed via secure connections controlled by HIN.

Access to the data used in the study requires secure connection to HIN servers and should be requested directly to HIN. Researchers may contact Phil Prefenno at pprofenno@hinfonyet.org, (207) 541–4115 to request data. Data will be available upon request to all interested researchers. HIN agrees to provide access to the de-identified data on a per request basis to interested

researchers. Future researchers will access the data through exact the same process as the authors of the manuscript.

Ethics statement

The studies were performed in accordance with the Declaration of Helsinki and relevant local guidelines and regulations. This work was done under a business arrangement between HealthInfoNet (HIN), operator of the Maine HIE (a nonprofit organization to benefit the population health of the Maine State residents), and HBI Solutions, Inc (HBI). The data use is governed by a business agreement (BAA) between HIN and HBI. No PHI was released during the analysis. Instead, HBI implemented a product that was the foundation for the agreement and then reported their results from applying this model to the products/services that HIN now provides. Because this study analyzed de-identified data, the Stanford University Institutional Review Board considered it exempt (October 16, 2014).

Data warehouse

The data warehouse used in this study consists of all the aggregated patient histories found in Maine's HIE. Incorporated data elements from the HIE EMRs included patient encounter (visit/admission) history, patient demographics, abnormal laboratory tests and results, radiographic procedures, medication prescriptions, diagnosis and procedures, which were coded according to the *International Classification of Diseases, 9th Revision, Clinical Modification* (ICD-9-CM), and AHRQ Chronic Condition Indicator (CCI) [25]. Census data from the U.S. Department of Commerce Census Bureau were integrated into our data warehouse. We categorized patients by socioeconomic status utilizing residence zip codes as an approximation to the average household mean and median family income and average level of education. The details of data extraction, management and storage are described in [S1 Text](#).

Maine HIE patient clinical histories were organized as hospital episode level relational database tables. We processed the database at the patient level based on the Enterprise Master Patient Index (EMPI) for population analysis within the HIE's 36 inpatient and 400+ ambulatory facilities. Since the time point of first confirmatory diagnosis of T2DM varies for each patient, all cohort patients' first confirmatory diagnoses of T2DM were aligned as time zero. A longitudinal analysis was performed monthly up to 24 months prior to time zero. Multiple patient-level database tables were transformed, integrated, and pivoted to construct a high dimensional table (33,403 clinical variables available for each patient) to support the machine learning process. To reduce the table dimension, i.e. the number of variables for each patient, we exploited the data variance for each data category (primary diagnosis/procedure, secondary diagnosis/procedure, laboratory result, radiology result and outpatient prescription) as the simplest criterion [24], which essentially projects the data points along the dimensions of maximum variances. This dimension reduction exercise ended with a set of 257 clinical longitudinal features from the preceding 24 months before time zero ([S1 Table](#)).

Study design

The overall study design was illustrated in [Fig 1](#). With high throughput genomics datasets, we previously discovered the critical transition point or the early detection of diabetes through the application of the TNE and DDN. In this study, we tested our hypothesis that high dimensional clinical patterns in EMR can be used to define a pre-disease critical transition state prior to definitive T2DM diagnosis. The hypothesis was tested with newly diagnosed patients'

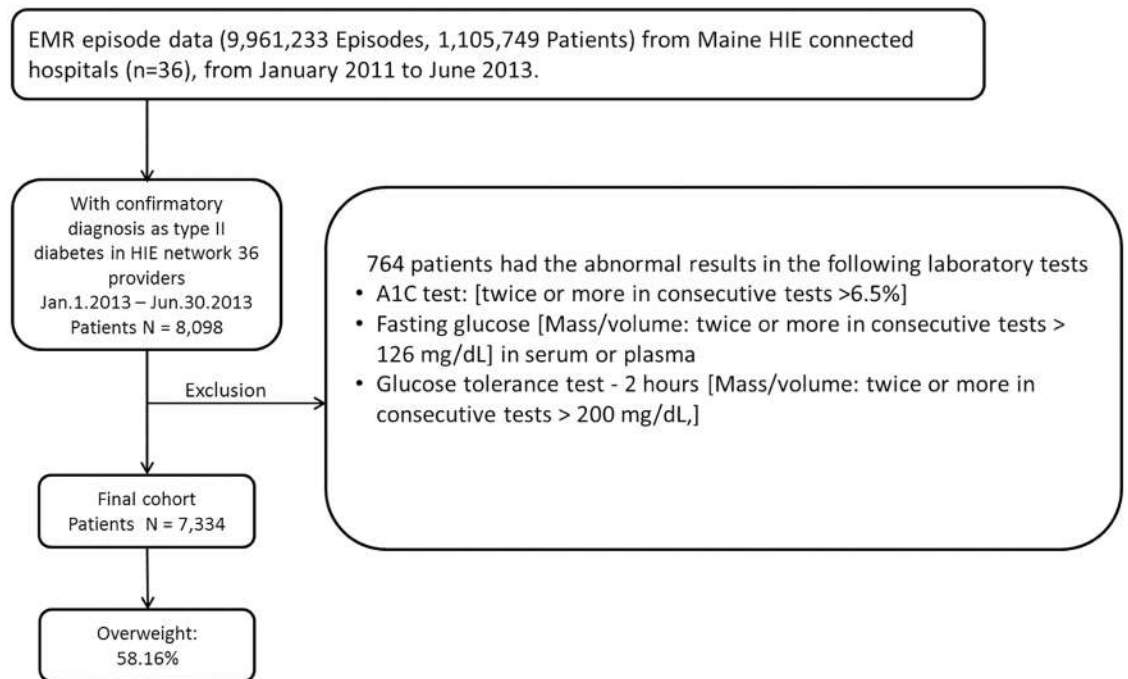


Fig 8. The flow chart of the cohort construction. Based on EMR episode data, a cohort of 8,098 patients was screened out with confirmatory diagnosis as T2DM. 764 patients who had the obviously abnormal results before the confirmative diagnosis in laboratory tests (positive twice) such as fasting glucose test, glucose tolerance test, A1C test, etc., were excluded for analysis, since these subjects might already suffer from T2DM before the confirmative diagnosis. A final cohort of 7,334 patients was constructed. EMR: electronic medical record; HIE: Health information exchange.

<https://doi.org/10.1371/journal.pone.0180937.g008>

clinical information that was extracted from EMR administrated by the HIE in Maine. An early-warning DDN and the associated clinical parameters were recognized to define the T2DM pre-disease transition state prior to the emerging confirmative diagnosis.

Cohort construction

Based on the EMR data from 1.3 million people residing in Maine, the ‘case cohort’ considered of those who were alive and diagnosed with T2DM between January 1, 2013 and June 30, 2013. Patients with discharge ICD-9-CM diagnosis codes of 25000, 25002, 25010, 25020, 25022, 25030, 25032, 25040, 25042, 25050, 25052, 25060, 25062, 25070, 25072, 25080, 25082, 25090, and 25092 were considered as having the first confirmatory T2DM diagnosis. Patients who had repeated abnormal glucose laboratory test results of A1C, fasting plasma glucose (FPG), or oral glucose tolerance test (OGTT) before the first confirmatory of T2DM diagnosis were excluded from the case cohort, as they might had developed T2DM before T2DM diagnosis was made by their physicians. For example, the patients who had more than twice the positive results (>6.5%) in A1C test were excluded. The final case cohort included 7,334 patients (Fig 8). Patients without any T2DM-related ICD-9-CM diagnosis codes from June, 2011 to December, 2013 were considered as control. 10,145 age and gender matched controls were selected to construct the control cohort. The demographic data of the case and control cohorts is shown in S2 Table. Compared with control cohort subjects, patients in case cohort exhibited complex comorbidities, making them ‘heavy users’ of healthcare resource with higher costs ($p < 0.001$, Mann—Whitney U test).

Markov process

Generally, it is difficult to analyze a large network directly because of the complexity and time cost. We used a local network structure instead of the whole network to represent its dynamical properties. Specifically, each node has a local network, i.e., the local network centered on node i consists with itself and its m linked first-order neighbor nodes i_1, i_2, \dots, i_m whose local transition state is $X^i(t) = (x_i(t), x_{i_1}(t), \dots, x_{i_m}(t))$ at time t .

The transition state coefficient x_i of the Markov process was defined as:

$$x_i(t) = \begin{cases} 1 & \text{if } |\Delta z_i(t)| > d_i \\ 0 & \text{if } |\Delta z_i(t)| \leq d_i \end{cases} \tag{1}$$

Where $\Delta z_i(t) = z_i(t) - z_i(t-1)$ is the changing of z_i at time t , d_i is the threshold for discretization.

When $|\Delta z_i(t)|$ is sufficiently large ($> d_i$), the transition state coefficient $x_i = 1$, otherwise $x_i = 0$.

Based on this definition, $X(t) = (x_1(t), \dots, x_n(t))$ is the transition state coefficient for the network at time t .

In a local network centered on node i with m linked first-order neighbors i_1, i_2, \dots, i_m , we first need to determine the threshold parameters $d = \{d_i, d_{i_1}, \dots, d_{i_m}\}$, which measure whether the state $z_i(t)$ has a large change or a state transition from its former state $z_i(t-1)$. Note that a system is in a stable state during its normal state, whereas a system is sensitive to perturbation in its transition state. Thus, d_i should be set to distinguish the “small changes” in normal state from the “large changes” in transition state.

For this local network, therefore, we select $d = \{d_i, d_{i_1}, \dots, d_{i_m}\}$ when the system is in a normal state such that for each node k , $p(|z_k(t_0)| > d_k) = \alpha$, if time point t_0 is in a normal state.

Obviously, for such threshold d , it holds that

$$p(|z_k(t_0)| > d_i, |z_{i_1}(t_0)| > d_{i_1}, \dots, |z_{i_m}(t_0)| > d_{i_m}) \leq \alpha \tag{2}$$

With such a distribution, features with large deviations clearly correspond to high probabilities.

The following description and derivation are all based on a local node, i.e. i . To simplify the notation, we omit i and denote $X^i(t)$ as $X(t)$, while we also denote the transition state simply as a state.

Given the current state at time t is $X(t)$, at the next time point $t+1$ there are a total of 2^{m+1} possible transition states, each of which is a stochastic event that is denoted, respectively, as $\{A_u\}_{u=1,2,\dots,2^{m+1}}$, where

$$A_u = \{x_i = \gamma_0, x_{i_1} = \gamma_1, \dots, x_{i_m} = \gamma_m\} \tag{3}$$

where $\gamma_l \in \{0,1\}$, for $l \in \{0,1,2,\dots,m\}$. The next state depends on the current state and its transition functions. Obviously, for this local subnetwork, the discrete stochastic process

$$\{X(t+i)\}_{i=0,1,\dots} = \{X(t), X(t+1), \dots, X(t+i), \dots\} \tag{4}$$

With $X(t+i) = A_u$, $u \in \{1,2,\dots,2^{m+1}\}$, is a stochastic Markov process during a period or phase of the system. This stochastic process is defined or given by a Markov matrix $P = (p_{u,v})$, which describes the transition rates from state u to state v as follows

$$p_{u,v}(t) = \Pr(X(t+1)) = A_v | X(t) = A_u \tag{5}$$

where $u, v \in \{1,2,\dots,2^{m+1}\}$, and $\sum_v p_{u,v}(t) = 1$.

Transition-based network entropy (TNE)

The TNE is a type of Shannon entropy [36] that is conditional on the previous state of the local dynamical network in a Markov process. In the feature space network, each node represents a feature, and each edge represents a regulatory relation between two features. For a local structure centered on node i and its m linked first-order neighbor nodes i_1, i_2, \dots, i_m , we already know that its state transition process is a stochastic Markov process. Within a period or phase, we assume that there is no change in the transition matrix, i.e., the transition probabilities $p_{u,v}(t)$, between any two possible states A_u and A_v . Thus, the process $\{X(t)\}_{t \in [t_1, t_2]}$ is a stationary stochastic Markov process during a specific period $[t_1, t_2]$, i.e., the normal stage or the transition stage.

Thus, there is a stationary distribution $\pi = (\pi_1, \dots, \pi_{2^{m+1}})$ that satisfies $\sum_v \pi_v p_{u,v} = \pi_u$. Therefore, we can define the transition-based network entropy, i.e., the TNE, as

$$H_i(t) = H(\chi) = - \sum_{u,v} \pi_v p_{u,v} \log p_{u,v} \tag{6}$$

where the subscript index i indicates the center node i of this local network, while χ represents the transition process $X(t), X(t + 1), \dots, X(t + T), \dots$ of the local network. The entropy is equivalent to the so-called entropy rate

$$H(\chi) =: \lim_{T \rightarrow \infty} \frac{1}{T} H(X(t), X(t + 1), \dots, X(t + T)) = \lim_{T \rightarrow \infty} H(X(t + T) | X(t + T - 1), H(t + T - 2), \dots, X(t)) \tag{7}$$

when the limit exists. The TNE is the conditional entropy and describes the average transition entropy, depending on the state transition, i.e. $H_i(t) = H(X(t) | X(t - 1)) = H(X(t), X(t - 1) | H(X(t - 1)))$. We also note that $X(t)$ or $(Z(t) - Z(t - 1))$ are variation variables. Clearly, in a normal state (or a disease state), a system recovers from a small perturbation quickly because of high resilience, i.e., $X(t)$ and $X(t-1)$ are almost independent. Thus, we have $H_i(t) \approx H(X(t))$ due to, $H(X(t), X(t - 1)) \approx H(X(t)) + H(X(t - 1)) > 0$ which results in a high TNE. By contrast, the system has difficulty recovering from a small perturbation in a transition state because of low resilience, i.e., $X(t)$ and $X(t-1)$ are strongly correlated, which implies that $H_i(t)$ rapidly approaches the minimum, $H_i(t) \approx 0$ due to $H(X(t), X(t - 1)) \approx H(X(t - 1))$.

Dynamic driver network (DDN)

The TNE method described above was employed to identify the DDN. Considering the following equations with noise perturbations near the equilibrium \bar{Z} :

$$Z(t + 1) = f(Z(t); P) \tag{8}$$

Where $Z(t) = (z_1(t), \dots, z_n(t))$ is an n -dimensional state vector or variables at time t that represent features in the system, while $P = (p_1, \dots, p_s)$ is a parameter vector or driving factors that represent slowly changing factors. Mappings $f: R^n \times R^s \rightarrow R^n$ are generally nonlinear functions. For such dynamical evolution function, there is a bifurcation if the following conditions hold:

1. \bar{Z} is a fixed point of system such that $\bar{Z} = f(\bar{Z}; P)$.
2. There is a value P_c such that one or a pair of the eigenvalues of the Jacobian matrix $\frac{\partial f(Z; P_c)}{\partial Z} |_{Z=\bar{Z}}$ is equal to 1 in the modulus.
3. When $P \neq P_c$ the eigenvalues are not always equal to 1 in the modulus.

The above three conditions with other transverse conditions imply that the system undergoes a phase change at \bar{Z} or a codimension-one bifurcation when P reaches the threshold P_c . For system near \bar{Z} and before P reaches P_c , we assume that the system is at a stable fixed point \bar{Z} so all the eigenvalues are within $(0, 1)$ in the modulus. The parameter value P_c at which the state shift of the system occurs, is known as a bifurcation parameter value or a critical transition value.

The generic properties in dynamics of Eq (8) was derived based on a consideration of the linearized equations for Eq (8) and the noise perturbations near \bar{Z} . Specifically, introducing new variables $Y(t) = (Y_1(t), \dots, Y_n(t))$ and a transformation matrix S , i.e., $Y(t) = S^{-1}(Z(t) - \bar{Z})$, we have

$$Y(t + 1) = \Lambda(P)Y(t) + \zeta(t) \tag{9}$$

where $\Lambda(P)$ is the diagnosis matrix of $\frac{\partial f(Z; P_c)}{\partial Z} \Big|_{Z=\bar{Z}}$, $\zeta(t) = (\zeta_1(t), \dots, \zeta_n(t))$ are small Gaussian noises with zero means.

We denote σ_i as the small standard deviation of ζ_i for all k . Without any loss of generality, the diagnosis matrix $\Lambda = (\lambda_1, \dots, \lambda_n)$ for each λ_i is between 0 and 1. (In fact, three typical cases arise during the diagonalization process [15], but we only illustrate the diagonal case with different real eigenvalues for simplicity. The derivations of the other two cases are similar. Among all eigenvalues of Λ , the largest one (in the modulus), say λ_1 , approaches 1 in the modulus when parameter $P \rightarrow P_c$. The eigenvalue λ_1 characterizes the system's rate of change around a fixed point and is known as the dominant eigenvalue. The normal state corresponds to a period of $|\lambda_1| < 1$, whereas the transition stage corresponds to the period with $\lambda_1 \rightarrow 1$. Without generality loss, we assume that the first variable y_1 in Y corresponds to λ_1 , namely $(y_1, 0, \dots, 0)$ is the eigenvector of λ_1 . Close to a fixed point, we have shown that there is a dominant group or a dynamical driver network (DDN) that satisfies the following conditions when the system approaches a critical transition point [15].

We now summarize the critical properties of the stochastically perturbed linear system as described in Eq (9). Denote z_i as the value of feature i . When P approaches the bifurcation point, the following results hold:

1. For any feature i in the DDN, the fluctuation of z_i increase sharply;
2. If both i and j are in the DDN, then the correlation between z_i and z_j sharply increases;
3. If i is in the DDN but j is not, then the correlation between z_i and z_j decrease sharply;
4. For two features i nor j which do not belong to the DDN, the correlation between z_i and z_j show no significant changes.

This result describes the generic critical properties of features in DDN. The DDN is the sub-network that makes the first move from one state toward another state at the critical transition point, so we refer to the DDN as the driving network in this critical transition.

In contrast to the original variables Z analysis [15], this work focuses on the variation equation for Eq (8) with variation variables ΔZ . We note that

$$z_i = s_{ij}y_1 + \dots + s_{in}y_n + \bar{z}_i \tag{10}$$

Let the variation variables be $\Delta Z = Z(t) - Z(t-1)$, then Eq (10) is transformed into

$$\Delta z_i = s_{ij}\Delta y_1 + \dots + s_{in}\Delta y_n \tag{11}$$

Where

$$\Delta Y = Y(t) - Y(t - 1) \tag{12}$$

We refer to $\Delta z_i(t)$ and $\Delta y_i(t)$ as the variation variables for $z_i(t)$ and $y_i(t)$, respectively. From Eqs (9) and (12), it clearly holds that

$$\Delta Y(t + 1) = \Delta Y(t) + \zeta(t) \tag{13}$$

where $\zeta(t) = \varsigma(t) - \varsigma(t - 1)$ are Gaussian noises with zero means and covariances $k_{ij} = Cov(\xi_i, \xi_j)$. It is clear that the standard deviation of $\xi_i(t)$ is $\sqrt{2}\sigma_i$, where σ_i is the standard deviation of ζ for all t . The variable Δy_i corresponds to the dominant eigenvalue λ_1 .

For a stochastically perturbed linear system of Eq (8):

$$Z(t + 1) = A(P)Z(t) + \varepsilon(t) \tag{14}$$

where $\varepsilon(t)$ is the Gaussian noise and P is a parameter vector that controls the Jacobian matrix A . We denote the variation variable as $\Delta z_i(t) = z_i(t) - z_i(t-1)$ and assume T to be sufficiently large.

1. When P is not in the vicinity of a critical transition point or a bifurcation point, it holds that for any i and j including $i = j$, $\Delta z_i(t+T)$ is statistically independent of $\Delta z_j(t)$, where $i, j = 1, 2, \dots, n$.
2. When P approaches to a critical transition point, the following holds.
 - a. If both i and j are in the dominant group, or DDN members, then there is a strong correlation between $\Delta z_i(t+T)$ and $\Delta z_j(t)$;
 - b. If neither i nor j in the dominant group, then $\Delta z_i(t+T)$ is statistically independent of $\Delta z_j(t)$. Note that this also holds for $i = j$.

We point out that for a nonlinear case at a fixed point, the dynamical behavior has the same trend described.

We combine the TNEs for all nodes and define the average network entropy for the whole network with n nodes as the average TNE as follows:

$$H(t) = \frac{1}{n} \sum_{i=1}^n H_i(t) \tag{15}$$

Suppose that there are control samples and case samples, then we define the comparative entropy as:

$$E(t) = \frac{H_{control}(t)}{H_{case}(t)} \tag{16}$$

where $H_{control}(t)$ is the TNE based on control samples in the form of Eq (15), and $H_{case}(t)$ is the TNE based on case samples in the form of Eq (15).

Supporting information

S1 Fig. The dynamical progression of feature networks of case and control population. (A). The dynamical progression of feature networks of case population; (B). The dynamical progression of feature networks of control population. (TIFF)

S1 Text. Data extraction, management and storage.
(DOCX)

S1 Table. A list of DDN features.
(DOCX)

S2 Table. The demographic characteristics of the case and control cohort.
(DOCX)

Acknowledgments

We thank and express our gratitude to the hospitals, medical practices, physicians and nurses participating in Maine's HIE. We also thank the biostatistics colleagues at the Department of Health Research and Policy for critical discussions.

Author Contributions

Conceptualization: KGS EW XBL.

Data curation: CZ TF QW DSC STA.

Formal analysis: BJ RL SH ZL XZ PC ZH DL YZ YML.

Investigation: CZ QW WL FS DSC STA.

Methodology: BJ RL ZL YY DBM.

Project administration: EW XBL.

Software: BJ RL PC.

Supervision: XBL.

Visualization: RL PC.

Writing – original draft: BJ RL SH ZL XZ PC YY YML DSC STA FS KGS EW XBL.

Writing – review & editing: ZL SH CZ TF ZH QW WL DL YZ DBM XBL.

References

1. Organization WH. Global Report on Diabetes. Geneva. 2016.
2. American Diabetes A. Economic costs of diabetes in the U.S. in 2012. *Diabetes care*. 2013; 36(4):1033–46. <https://doi.org/10.2337/dc12-2625> PMID: 23468086;
3. Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J. Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA: the journal of the American Medical Association*. 2007; 298(22):2654–64. <https://doi.org/10.1001/jama.298.22.2654> PMID: 18073361.
4. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. *BMJ*. 2016; 354:i3857. <https://doi.org/10.1136/bmj.i3857> PMID: 27510511;
5. Schellenberg ES, Dryden DM, Vandermeer B, Ha C, Korownyk C. Lifestyle interventions for patients with and at risk for type 2 diabetes: a systematic review and meta-analysis. *Ann Intern Med*. 2013; 159(8):543–51. <https://doi.org/10.7326/0003-4819-159-8-201310150-00007> PMID: 24126648.
6. He D, Liu ZP, Honda M, Kaneko S, Chen L. Coexpression network analysis in chronic hepatitis B and hepatic lesions reveals distinct patterns of disease progression to hepatocellular carcinoma. *Journal of molecular cell biology*. 2012; 4(3):140–52. <https://doi.org/10.1093/jmcb/mjs011> PMID: 22467683.

7. Hirata Y, Bruchoovsky N, Aihara K. Development of a mathematical model that predicts the outcome of hormone therapy for prostate cancer. *Journal of theoretical biology*. 2010; 264(2):517–27. <https://doi.org/10.1016/j.jtbi.2010.02.027> PMID: 20176032.
8. Litt B, Esteller R, Echaz J, D'Alessandro M, Shor R, Henry T, et al. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron*. 2001; 30(1):51–64. PMID: 11343644.
9. McSharry PE, Smith LA, Tarassenko L. Prediction of epileptic seizures: are nonlinear methods relevant? *Nature medicine*. 2003; 9(3):241–2; author reply 2. <https://doi.org/10.1038/nm0303-241> PMID: 12612550.
10. Paek SH, Chung HT, Jeong SS, Park CK, Kim CY, Kim JE, et al. Hearing preservation after gamma knife stereotactic radiosurgery of vestibular schwannoma. *Cancer*. 2005; 104(3):580–90. <https://doi.org/10.1002/cncr.21190> PMID: 15952200.
11. Venegas JG, Winkler T, Musch G, Vidal Melo MF, Layfield D, Tgavalekos N, et al. Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature*. 2005; 434(7034):777–82. <https://doi.org/10.1038/nature03490> PMID: 15772676.
12. Tan Z, Liu R, Zheng L, Hao S, Fu C, Li Z, et al. Cerebrospinal fluid protein dynamic driver network: At the crossroads of brain tumorigenesis. *Methods*. 2015; 83:36–43. Epub 2015/05/20. <https://doi.org/10.1016/j.ymeth.2015.05.004> PMID: 25982164.
13. Li M, Zeng T, Liu R, Chen L. Detecting tissue-specific early warning signals for complex diseases based on dynamical network biomarkers: study of type 2 diabetes by cross-tissue analysis. *Briefings in bioinformatics*. 2014; 15(2):229–43. <https://doi.org/10.1093/bib/bbt027> PMID: 23620135.
14. Liu X, Liu R, Zhao XM, Chen L. Detecting early-warning signals of type 1 diabetes and its leading biomolecular networks by dynamical network biomarkers. *BMC medical genomics*. 2013; 6 Suppl 2:S8. Epub 2013/07/17. PMID: 23819540
15. Chen L, Liu R, Liu ZP, Li M, Aihara K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Scientific reports*. 2012; 2:342. Epub 2012/03/31. <https://doi.org/10.1038/srep00342> PMID: 22461973;
16. Zheng L, Wang Y, Hao S, Shin AY, Jin B, Ngo AD, et al. Web-based Real-Time Case Finding for the Population Health Management of Patients With Diabetes Mellitus: A Prospective Validation of the Natural Language Processing-Based Algorithm With Statewide Electronic Medical Records. *JMIR Med Inform*. 2016; 4(4):e37. <https://doi.org/10.2196/medinform.6328> PMID: 27836816.
17. Wang Y, Luo J, Hao S, Xu H, Shin AY, Jin B, et al. NLP based congestive heart failure case finding: A prospective analysis on statewide electronic medical records. *Int J Med Inform*. 2015; 84(12):1039–47. <https://doi.org/10.1016/j.ijmedinf.2015.06.007> PMID: 26254876.
18. Jin B, Zhao Y, Hao S, Shin AY, Wang Y, Zhu C, et al. Prospective stratification of patients at risk for emergency department revisit: resource utilization and population management strategy implications. *BMC Emerg Med*. 2016; 16:10. <https://doi.org/10.1186/s12873-016-0074-5> PMID: 26842066;
19. Hu Z, Jin B, Shin AY, Zhu C, Zhao Y, Hao S, et al. Real-time web-based assessment of total population risk of future emergency department utilization: statewide prospective active case finding study. *Interactive journal of medical research*. 2015; 4(1):e2. <https://doi.org/10.2196/ijmr.4022> PMID: 25586600;
20. Hu Z, Hao S, Jin B, Shin AY, Zhu C, Huang M, et al. Online Prediction of Health Care Utilization in the Next Six Months Based on Electronic Health Record Information: A Cohort and Validation Study. *J Med Internet Res*. 2015; 17(9):e219. <https://doi.org/10.2196/jmir.4976> PMID: 26395541.
21. Hao S, Wang Y, Jin B, Shin AY, Zhu C, Huang M, et al. Development, Validation and Deployment of a Real Time 30 Day Hospital Readmission Risk Assessment Tool in the Maine Healthcare Information Exchange. *PloS one*. 2015; 10(10):e0140271. <https://doi.org/10.1371/journal.pone.0140271> PMID: 26448562;
22. Hao S, Jin B, Shin AY, Zhao Y, Zhu C, Li Z, et al. Risk prediction of emergency department revisit 30 days post discharge: a prospective study. *PloS one*. 2014; 9(11):e112944. <https://doi.org/10.1371/journal.pone.0112944> PMID: 25393305;
23. Liu R, Li M, Liu ZP, Wu J, Chen L, Aihara K. Identifying critical transitions and their leading biomolecular networks in complex diseases. *Scientific reports*. 2012; 2:813. <https://doi.org/10.1038/srep00813> PMID: 23230504;
24. Gilmore R. *Catastrophe theory for scientists and engineers*. New York: Dover Publications; 1993. xvii, 666 p. p.
25. Murray JD. *Mathematical Biology*. New York: Springer; 1993.
26. Haase CL, Tybjaerg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. HDL Cholesterol and Risk of Type 2 Diabetes: A Mendelian Randomization Study. *Diabetes*. 2015; 64(9):3328–33. <https://doi.org/10.2337/db14-1603> PMID: 25972569.

27. Bangalore S, Parkar S, Grossman E, Messerli FH. A meta-analysis of 94,492 patients with hypertension treated with beta blockers to determine the risk of new-onset diabetes mellitus. *The American journal of cardiology*. 2007; 100(8):1254–62. <https://doi.org/10.1016/j.amjcard.2007.05.057> PMID: 17920367.
28. Shen L, Shah BR, Reyes EM, Thomas L, Wojdyla D, Diem P, et al. Role of diuretics, beta blockers, and statins in increasing the risk of diabetes in patients with impaired glucose tolerance: reanalysis of data from the NAVIGATOR study. *BMJ*. 2013; 347:f6745. <https://doi.org/10.1136/bmj.f6745> PMID: 24322398;
29. Mikkelsen KH, Knop FK, Frost M, Hallas J, Pottegard A. Use of Antibiotics and Risk of Type 2 Diabetes: A Population-Based Case-Control Study. *The Journal of clinical endocrinology and metabolism*. 2015; 100(10):3633–40. <https://doi.org/10.1210/jc.2015-2696> PMID: 26312581;
30. Telfer SJ. Fluoroquinolone antibiotics and type 2 diabetes mellitus. *Medical hypotheses*. 2014; 83(3):263–9. <https://doi.org/10.1016/j.mehy.2014.05.013> PMID: 24947193.
31. Eriksson KF, Lindgarde F. Prevention of type 2 (non-insulin-dependent) diabetes mellitus by diet and physical exercise. The 6-year Malmo feasibility study. *Diabetologia*. 1991; 34(12):891–8. PMID: 1778354.
32. Pan XR, Li GW, Hu YH, Wang JX, Yang WY, An ZX, et al. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. *Diabetes care*. 1997; 20(4):537–44. PMID: 9096977.
33. Inzucchi SE. Clinical practice. Diagnosis of diabetes. *N Engl J Med*. 2012 6:542–50. <https://doi.org/10.1056/NEJMcp1103643> PMID: 22873534
34. Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *The New England journal of medicine*. 2001; 344(18):1343–50. <https://doi.org/10.1056/NEJM200105033441801> PMID: 11333990.
35. Knowler WC B-CE, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM; Diabetes Prevention Program Research Group. REDUCTION IN THE INCIDENCE OF TYPE 2 DIABETES WITH LIFESTYLE INTERVENTION OR METFORMIN. 2002; 6:393–403. <https://doi.org/10.1056/NEJMoa012512> PMID: 11832527
36. Shannon CE. A Mathematical Theory of Communication. *Bell System Technical Journal*. 1948; 27(3):379–423.