Identification and genetic validation of drug-comorbidity interactions in type 2 diabetes using data-driven disease trajectory analysis

Osamu Ichikawa[1,2], Benjamin S. Glicksberg[1,3], Noga Minsky[4], Hao-Chih Lee[1], Shuyu D. Li[5], Rong Chen[5,6], Li Li[1,5,6,#,*], Joel T. Dudley[1,5,#,*]

Affiliations:

1. Department of Genetics and Genomic Sciences, Institute for Next Generation Healthcare, Icahn School of Medicine at Mount Sinai, 770 Lexington Ave., New York, NY 10065.

2. Drug Research Division, Sumitomo Dainippon Pharma. Co. Ltd., 3-1-98 Kasugade-naka, Konohana-ku, Osaka, 554-0022, Japan.

3. Hasso-Plattner Institute for Digital Health, Icahn School of Medicine at Mount Sinai, 770 Lexington Ave 14th floor, New York, NY 10065

4. Department of Medicine, Diabetes Alliance, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place Box 1620 New York, NY 10029

5. Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology. Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place Box 1620 New York, NY 10029

6. Sema4, Stamford, CT, 06902

\# These authors contributed equally to this work

* Correspondence and requests for materials should be address to

Dr. Li Li (li.li@mssm.edu) or Dr. Joel Dudley (joel.dudley@mssm.edu)

Institute for Next Generation Healthcare

Icahn School of Medicine at Mount Sinai

770 Lexington Ave.

15th Floor

New York, NY 10065, USA

Abstract

Given various risk profiles, manifestations, comorbidities, and outcomes for individuals with Type 2 diabetes mellitus (T2D), a one-size-fits-all approach towards treatment and management is inadequate, and there is a clear need for personalized medications to reduce rates of complications and comorbidities. Here, we leveraged comprehensive Electronic Medical Records (EMR) to identify associations between T2D comorbidities and medications and evaluated their biological determinants using EMR-linked genetic information. We discovered clinically novel associations supported by the previous laboratory studies; e.g. 5-hydroxytryptamine 3 receptor antagonists, ondansetron and granisetron, are protective against cognitive disorders, and clopidogrel is protective against retinopathy. Furthermore, potentially novel associations were validated by genetic analysis; e.g. association between gabapentin and cognitive disorders was supported via variants of its target genes, *GRIN1* and *CACNA2D2*. These results of the current study open the door for optimizing treatment combinations to mitigate risks and for individualizing therapy based on T2D subtype risk profiles.

Type 2 diabetes mellitus (T2D) is a heterogeneous, multifactorial disease with a global diabetes prevalence that has more than doubled since 1980, rising from 4.3% to 9% in the adult population [1]. Associated with psychological and physical distress to both patients and caregivers, diabetes is a considerable determinant of disability and remains the leading cause of vision loss, kidney failure, and non-traumatic lower limb amputations. Moreover, diabetes more than doubles the risk of the myocardial infarction and cerebrovascular accidents [2,3]. According to the National Health and Nutrition Examination Survey (NHANES) the prevalence of diabetes in the United States is over 12% [4], and as such, diabetes is associated with a total direct estimated medical cost of $237 billion in 2017 [5,6]. At the time of diagnosis 86% of patients with T2D are affected by chronic comorbid conditions [7], and this figure rises over the natural course of disease, compounding the burden on patients, providers, and the economy. There is tremendous need to advance the way in which T2D is treated in order to further reduce rates of complications and co-morbidities.

Given various risk profiles, manifestations, comorbidities, and outcomes for individuals with T2D, a one-size-fits-all approach towards treatment and management is inadequate. There have been attempts to disentangle determinants of poor patient outcomes, as well as efforts to develop more personalized treatment recommendations [8]. One clear path forward is to model treatment recommendations based on disease profiles, since many conditions either co-manifest with or result from T2D. A previous study used Electronic Medical Records (EMR) to analyze patient characteristics and make treatments recommendations which improved diabetes control, measured by glycated hemoglobin A_{1C} (HbA1c), beyond the standard of care [9]. Another study sought to evaluate the molecular effects of various medications on diabetes comorbidities using a systems biology approach [10].

In our previous work, we characterized the heterogeneous patient landscape of T2D into three distinct subtypes using clinical and genomic data from an EMR-paired biobank [11]. We stratified patients using a topology-based, patient–patient network derived from demographic and clinical features such as disease diagnoses. We then layered on genetic information in the form of single nucleotide polymorphisms (SNPs) and characterized each subgroup according to enrichment of these features. For instance, subtype 1 was characterized by diabetic retinopathy and diabetic nephropathy; subtype 2 was enriched for lung cancer malignancy and cardiovascular diseases; and subtype 3 was associated most strongly with cardiovascular diseases and neurological diseases. Additionally, a recent study [12] discovered five replicable clusters of patients with T2D, representing different patient characteristics and risk of diabetic complications, based on a data-driven cluster analysis on six relevant variables. Taken together, the findings of those studies suggested that there might be subtypes of T2D that have different genetic and phenotypic characteristics, as well as risk factors for developing certain subsequent conditions.

As a data-driven analysis of a clinical population identified new T2D subtypes, we hypothesize that this approach can be applied to improve our knowledge regarding the association of medications with T2D related comorbidities and complications. In this study, we sought to push beyond our previous work to identify specific treatment strategies that are personalized according to T2D phenotypic patient profiles. Using the EMR of Mount Sinai Hospital, we investigated whether medications conferred a protective effect against, or the increased the risk of, subsequent comorbidities in all T2D patients (Fig. 1). Our results identified previously reported associations between medications and specific disease outcomes, and also uncovered potentially novel connections between medications and T2D comorbidities, some of which is supported by laboratory and animal data. We also evaluated the additive effects of medication pairs to identify potentially beneficial treatment regimens. In addition, we assessed

relationships between predicted protective medications and disease comorbidities through a genetic association analysis of T2D patients in the Mount Sinai Bio*Me* biobank cohort, a subset of EMR patients. The results of the current study open the door for optimizing treatment combinations to mitigate risks and for individualizing therapy based on T2D subtype risk profiles.

Results

Prevalence and risk estimation of comorbidities following T2D diagnosis

We began by assessing the prevalence and temporal ordering of the associations between T2D and other diseases (Fig. 1A). In particular, we sought to determine whether a given comorbidity tended to occur after diagnosis of T2D. We restricted our analysis to diseases with reported dates and only included the first reported encounter of a diagnosis, yielding 874,553 person–disease pairs, and analyzed their prevalence and temporal ordering at the patient level.

First, we calculated the prevalence of subsequent diseases based on the Kaplan–Meier product–limit estimator, in units of 1 year (Fig. 2A and Supplementary Table S1). In our cohort, the prevalence 10 years after T2D diagnosis was highest for 'Hypertension' (72% of cases), followed by 'Disorder of lipid metabolism' (63%). The macrovascular complication 'Coronary atherosclerosis and other heart disease' was had a prevalence of 46%, and the microvascular complication, 'Chronic kidney disease' was also quite common (34%). These results are consistent with previous studies demonstrating rates >67% for hypertension [7], 77% for 'Disorder of lipid metabolism' [13], 45–51% for 'Coronary atherosclerosis and other heart disease' [14,15], and 24–40% for 'Chronic kidney disease' [13,16], suggesting that our EMR data reflect a representative cohort of the US population. Ninety-six out of 219 diseases exhibited prevalence greater than 10% prevalence within 10 years after T2D diagnosis, and these were analyzed below. In addition, we analyzed all 13 neoplasms, because neoplasms are enriched in one of the T2D subtypes identified in our previous published study [17]. 109 T2D comorbidities were analyzed in total.

The increase in risk of each comorbidity after T2D diagnosis was quantified based on the timing of their occurrence in individual patients. Overall, out of the 109 T2D comorbidities analyzed, 93

7

diseases were significantly more likely to occur after T2D diagnosis (adjusted $p < 0.05$; Fig. 2B, Supplementary Table S1, and Supplementary Results).

Candidate medications associated with each T2D comorbidity

Fifteen medications specifically indicated for T2D (T2D medications) were associated with at least one T2D comorbidity, with 79 protective and 198 risk associations (Fig. 3), in the cross-sectional analysis with adaptive LASSO (Supplementary Results). For example, insulin was associated with protection against 11 comorbidities and risk for 33 comorbidities, and metformin was associated with protection against 12 comorbidities and risk for 14 comorbidities. Rosiglitazone had the largest number of risk associations, including circulatory diseases. This is consistent with the previously reported risk of cardiovascular disease and death, which led to rosiglitazone being withdrawn from European and other markets [18,19].

Nine T2D-related medications, which are not necessarily indicated for T2D, but are often used for treating or managing T2D co-morbidities (see Supplementary Methods), were associated with at least one disease, including three statins and five antihypertensive medications (three angiotensin II receptor [AT2R] blockers, two angiotensin-converting-enzyme [ACE] inhibitors, and niacin) (Fig. 3). These medications had 48 protective associations and 42 risk associations. Two high potency statins, atorvastatin and rosuvastatin, exhibited a wide variety of protective effects but few risks for T2D comorbidities, supporting that the use of these medications is a key factor in preventing and managing comorbidities related to lipid metabolism [20].

T2D and T2D-related medications predicted to modulate the onset of various comorbidities

Among the candidate medications associated with each T2D comorbidity in the cross-sectional analysis, we assessed the timing effect of medications on disease onset by survival analysis with time-dependent covariates in propensity score–matched cohorts (Fig. 1B). Among T2D medications, indicated directly for diabetes treatment (see Supplementary Methods), and T2D-related medications, 64 out of 127 protective associations and 43 of 240 risk associations were also identified by the survival analysis (Fig. 3 and Supplementary Table S3). The analysis identified 30 medication-disease associations described previously, involving the medications commonly prescribed in patients with T2D, such as metformin, insulin, and statins. These associations include, for instance, atorvastatin and 'Pulmonary heart disease' (HR=0.62, adjusted p < 0.0001) [21], sitagliptin and 'Hepatitis' (HR=0.43, adjusted p=0.0060) [22], metformin and 'Nutritional deficiency' (HR=1.1, adjusted p=0.0023) [23], and nateglinide and 'Other endocrine disorders', which includes hypoglycemia (HR=3.7, adjusted p< 0.0001) [24] (Supplementary Table S3). Some associations were controversial due to inconsistent results in published studies. Among them, five medications had protective associations in our study, including metformin with 'Delirium, dementia, and amnestic and other cognitive disorders' (HR=0.74, adjusted p=0.00020) [25,26]; atorvastatin with 'Delirium, dementia, and amnestic and other cognitive' (HR=0.69, adjusted p=0.00041) [27,28]; rosuvastatin with 'Esophageal disorders' (HR=0.84, adjusted p=0.040) [29,30]; and sitagliptin with 'Acute and unspecified renal failure' (HR=0.78, adjusted p=0.027) [31] and 'Cardiac dysrhythmias' (HR=0.84, adjusted p=0.047) [32,33]. Lastly, we identified 40 protective associations which have not previously been described. These include insulin and 'Osteoporosis' (HR=0.72, adjusted p=0.0017), liraglutide and 'Urinary tract infections' (HR=0.46, adjusted p=0.048), and nateglinide and 'Abdominal pain' (HR=0.38, adjusted p=0.013). In addition, we identified 30 new risk associations, including exenatide and 'Glaucoma' (HR=1.6, adjusted p=0.046), metformin and 'Cataract' (HR=1.36, adjusted p < 0.0001),

and nateglinide and 'Nutritional deficiencies' (HR=1.52, adjusted p=0.023).

Non-T2D-related medications predicted to modulate the onset of well-known T2D complications

Two hundred thirty one medications, including non-T2D-related medications, exhibited protective effects for at least one of the 104 T2D comorbidities (Fig. 4 and Supplementary Table S4), and 270 medications increased the risk for at least one comorbidity (Fig. 4 and Supplementary Table S5). Table 1 summarizes seven comorbidities, including two major microvascular complications, 'Retinal detachments; defects; vascular occlusion; and retinopathy' and 'Chronic kidney disease'; two major macrovascular complications 'Coronary atherosclerosis and other heart disease' and 'Acute cerebrovascular disease'; and three comorbidities that were enriched in particular T2D subtypes in our previously published study [11], 'Neoplasm excl. benign', 'Delirium, dementia, and amnestic and other cognitive disorders', and 'Nutritional deficiencies'.

We identified medications whose use conferred protective effects against the development of specific comorbidities in T2D patients (Table 1A). Twelve medications were demonstrated to have protective associations which were previously described, including cholecalciferol and 'Coronary atherosclerosis and other heart disease' (HR=0.66, adjusted p=0.038) [34]; aspirin and 'Neoplasm excl. benign' (HR=0.57, adjusted p<0.0001) [35]; and nifedipine and 'Delirium, dementia, and amnestic and other cognitive disorders' (HR=0.54, adjusted p<0.0001) [36]. Intriguingly, our analysis confirms protective medication effects previously demonstrated only in non-clinical laboratory studies. Two 5-hydroxytryptamine 3 (5-HT$_3$) receptor antagonists (ondansetron and granisetron) and one alpha-adrenergic agonist (phenylephrine), exhibited protective effects against 'Delirium, dementia, and

amnestic and other cognitive disorders' (HR=0.71, adjusted p=0.0011; HR=0.44, adjusted p=0.045; HR=0.76, adjusted p=0.019; respectively) [37,38]. Bacitracin, azithromycin, and lidocaine protected against 'Neoplasm excl. benign' (HR=0.67, adjusted p=0.016; HR=0.72, adjusted p=0.00041; HR=0.80, adjusted p=0.029; respectively) [39-41]. Clopidogrel and hydroxyzine were protective against 'Retinal detachments; defects; vascular occlusion; and retinopathy' (HR=0.80, adjusted p=0.025; HR=0.52, adjusted p=0.017; respectively) [42,43].

We identified a list of potentially novel protective associations (Table 1A). In regard to microvascular complications, seven medications were significantly associated with 'Retinal detachments; defects; vascular occlusion; and retinopathy', and five medications were related to 'Chronic kidney disease'. In regard to macrovascular complications, six medications were associated with 'Coronary atherosclerosis and other heart disease', and four medications exhibited a protective effect on 'Acute cerebrovascular disease'. Four medications, including gabapentin (HR=0.74, adjusted p=0.0068), one of the first-line medications for the treatment of neuropathic pain in diabetic neuropathy, exhibited protective effects against 'Delirium, dementia, and amnestic and other cognitive disorders'. Three medications, vancomycin (HR=0.73, adjusted p=0.0049), betamethasone (HR=0.53, adjusted p=0.0060), and oxycodone (HR=0.79, adjusted p=0.0034), were associated with protection against 'Neoplasm excl. benign' and 11 medications were related to 'Nutritional deficiencies'. A complete list of all disease categories is provided in Supplementary Table S4.

We also discovered associations between medications and increased risk for comorbidities (Table 1B). We corroborated two previously reported risk medications, except for T2D medications, to 'Nutritional deficiencies', which were methotrexate (HR=1.6, adjusted p=0.013) [44] and ritonavir (HR=2.4, adjusted p<0.0001) [45]. In addition, we identified 9 potentially novel associations between medications and 'Retinal detachments; defects; vascular occlusion; and retinopathy' and 11 potentially

novel associations between medications and 'Chronic kidney disease' (Table 1B). Colchicine was associated with increased risk of 'Coronary atherosclerosis and other heart disease' (HR=1.7, adjusted p=0.00038), and phenytoin with increased risk of 'Acute cerebrovascular disease' (HR=2.1, adjusted p=0.0011). Four medications were associated with risk of 'Delirium, dementia, and amnestic and other cognitive disorders'. Six medications were associated with an increased risk of 'Neoplasms excl. benign'. Nine medications were associated with 'Nutritional deficiencies'. A complete list of all disease categories is available in Supplementary Table S5.

Validation of the predicted comorbidity-medication relationship by genetic association analysis

We evaluated the associations between the predicted medications and seven disease comorbidities of interest (Table 1A) within the Mount Sinai Bio*Me* biobank cohort (see Methods). Specifically, we analyzed the associations between each comorbidity and SNPs mapped to the genes targeted by medications predicted to be protective.

We found that 286 of 4,912 SNPs were significantly associated with comorbidities, with nominal p values less than 0.05 (Table 2 and Supplementary Table S9). We identified 58 SNPs, corresponding to 17 unique drug targets, that were significantly enriched in patients diagnosed with 'Retinal detachments; defects; vascular occlusion; and retinopathy'. These patients could, in theory, benefit from three corresponding medications: calcium, hydroxyzine, and ibuprofen, which demonstrated a protective effect. Twelve SNPs mapped to five genes targeted by metformin, naproxen, and ondansetron were associated with 'Chronic kidney disease'. Additionally, 45 SNPs mapped to 18 genes connected to six medications were associated with 'Coronary atherosclerosis and other heart disease'. Thirty-four SNPs

12

mapped to 11 genes targeted by five medications were associated with 'Acute cerebrovascular disease'. Twenty-five SNPs corresponding to six medications, specifically azithromycin, bacitracin, betamethasone, ketorolac, lidocaine, and oxycodone, were related to 'Neoplasms excluding benign'. Seventy-seven SNPs mapped to 15 genes related to four medications, gabapentin, nifedipine, cefazolin, and phenylephrine, were associated with 'Delirium, dementia, and amnestic and other cognitive disorders.' Sixteen SNPs linked to four medications were related to 'Nutritional deficiencies'. A complete list of each significant SNP-comorbidity association is provided in Supplementary Table S9.

In the SNP set–comorbidity combinations, we gathered SNPs based on gene-level annotations. There were nine associations between the gene level and comorbidities, all of which were consistently identified within the SNP–comorbidity combinations described above (Table 2).

These genetic associations between drug-target genes/SNPs and comorbidities in the T2D patient cohort strongly support our findings regarding the protective associations between medications and comorbidities and suggest mechanisms of actions for each medication.

Discussion

Because T2D is associated with various clinical manifestations and chronic comorbidities over time, the optimal therapeutic regimen evolves according to the changing disease profile of the individual patient. In this study, we developed the first data-driven deep phenotyping strategy for identifying the risk and benefits of medications based on the T2D comorbidity profiles. These T2D comorbidity profiles are based on our comprehensive large-scale EMR, which contains longitudinal clinical measurements, prescribed or administered medications, and disease diagnoses. Analyzing EMR enables us to explore longitudinal trajectory outcomes with more than 10 years of follow-up, increasing the likelihood of capturing long-term effects. We developed a systematic approach to estimate the association of medications, alone in combination, with future onset of comorbidities. Our approach could facilitate the precision of treatment regimens tailored to individual disease profiles.

Our analyses confirmed known associations between medications and diseases, and also uncovered potentially novel connections between medications and T2D comorbidities. Our approach began with evaluating the prevalence of comorbidities in T2D patients after the initial diagnosis of T2D. Next, statistical models were applied to assess the effects of medications on the risk of developing subsequent comorbidities. Lastly, medications and disease associations were assessed and explained by genetic variants using the T2D patient cohort in the genotyping biobank data from our Bio*Me* program [17].

Our analysis provides clinical evidence for associations previously demonstrated only in laboratory and animal studies. Two 5-hydroxytryptamine 3 (5-HT$_3$) receptor antagonists, ondansetron and granisetron, consistently showed protective effects to 'Delirium, dementia, and amnestic and other cognitive disorders'. Several studies on rats and nonhuman primates showed that ondansetron has positive effects on learning [37]. A clinical trial of ondansetron [46] revealed no significant therapeutic

14

benefit for Alzheimer's disease (AD), but that study was conducted on mild/moderate AD patients who received a short-term treatment (24 weeks), rather than pre-AD patients in which the condition might be prevented. This result, in conjunction with other two studies with 5-HT reuptake inhibitors [47,48], suggest that serotonergic strategies are not effective at improving cognition in patients with existing AD. Nevertheless, our analysis is the first to support a protective effect of 5-HT$_3$ antagonists against cognitive disorders, the first such evidence to be obtained in human patients. Thus, our EMR-based studies could reveal the long-term effects of prognostic factors for pre-dementia patients, particularly patients with T2D. To our knowledge, we are the first to describe a protective association in humans between clopidogrel, classified as a thienopyridine, and 'Retinal detachments; defects; vascular occlusion; and retinopathy'. A protective effect was established for ischemic retinopathy in a mouse model [43]. A clinical trial using a randomized, double-masked, placebo-controlled design showed that another thienopyridine, ticlopidine, significantly slows the progression of nonproliferative diabetic retinopathy in insulin-treated diabetic patients [49], and our result supports a protective class effect. Additionally, our analysis supported the previous studies, which have reported that some antibiotics are associated with decreased risk for neoplasms (See Supplementary Discussion).

Genetic analysis of the T2D patient cohort in our biobank provided potential insights into the protective associations between medication and disease comorbidities through an inferred underlying mechanism of action (Table 2). The potentially novel association between 'Delirium, dementia, and amnestic and other cognitive disorders' and the antiepileptic gabapentin, a first-line medications for neuropathic pain in diabetic neuropathy, is supported by the relationship between this disease and its target genes, *GRIN1* and *CACNA2D2* (Table 2). *GRIN1* encodes NMDA (N-methyl-D-aspartate) receptor 1, a target gene of the Alzheimer's disease drug, memantin. In addition, voltage-gated calcium channels, including CACNA2D2, are inhibited by gabapentin [50]; calcium channel blockade attenuates

amyloid-β-induced neuronal decline *in vitro* and is neuroprotective in animal models [51]. Further potentially novel associations were discussed in Supplementary Discussion.

Although some of the medication–comorbidity associations we identified have been supported by previous preclinical studies, the genetic association results further strengthen and guide our interpretation of the potential mechanisms of action. The association between 'Retinal detachments; defects; vascular occlusion; and retinopathy' and the histamine H1 blocker hydroxyzine is supported by two SNPs mapped to *HRH1*. The protective effect of an H1 blocker against retinopathy is consistent with a previous study showing that H1 blockage is effective in reversing leakage of albumin from the blood across the retinal barrier, a typical phenotype of diabetic retinopathy in diabetic rats [52]. We also describe an association between lidocaine and lower rates of 'Neoplasm excluding benign', and this observation is supported by studies which demonstrated that a clinical concentration of lidocaine inhibits epidermal growth factor receptor (EGFR) activity and suppresses EGF-induced in vitro cancer cell proliferation [53,54].

Though several of the associations demonstrated in our analysis have been previously reported by patient-level studies, our genetic association analysis sheds light on potential mechanisms of drug action. For example, the protective effect of metformin against 'Chronic kidney disease' is supported by the association with *PRKAB1*, the gene coding AMP-activated protein kinase (*AMPK*), which is activated by metformin. *AMPK* is an energy sensor that plays a pivotal role in cellular homoeostasis. Deficiency in *AMPK* activity promotes epithelial-to-mesenchymal transition and renal cell apoptosis, eventually contributing to chronic kidney disease (CKD) [55]. In addition, a meta-analysis of genome-wide association studies has identified a link between estimated glomerular filtration rate (eGFR) and another component of *AMPK*, encoded by *PRKAG2* [56]. Although, due to a risk of lactic acidosis, metformin is currently contraindicated in patients with severe CKD (eGFR < 30 ml/min/1.73 m^2), it has been

16

suggested that the benefits of its use may outweigh possible risks in these patients [57]. A novel *AMPK* activator, without the risk of lactic acidosis, could be a safer alternative to metformin in preventing progression of severe CKD.

In addition to identifying associations between medications and comorbidities, we analyzed whether medication pairing modified these associations (Supplementary Results and Supplementary Table S7). Some T2D medications were associated with nutritional deficiencies and this implies closer monitoring may be necessary with certain drugs. We observed novel associations between metformin and nateglinide with vitamin D deficiency, which is interesting because of data supporting that vitamin D supplementation may improve insulin sensitivity [58]. Other novel associations identified include glyburide and vitamin B-complex deficiencies and repaglinide and protein-calorie malnutrition.

In our current study, we build upon our previous work to identify specific treatment strategies according to individual T2D disease profiles and drug-disease associations. In our preceding publication, we characterized the heterogeneous patient landscape of T2D into three distinct subtypes using clinical data from an EMR and verified the disease association in a genetics biobank [11]. That work had suggested that there might be patterns of T2D with different genetic associated phenotypic characteristics, as well as risk factors for developing certain subsequent conditions. Combining this knowledge with our current observations regarding medication associations with comorbidities, a more personalized and precise approach toward medication regimens can be devised (Supplementary Discussion).

This study has its limitations. First, to define the patient cohort, we had to use ICD-9 codes instead of the well-validated eMERGE algorithm [59] because about 30% of the patients did not have lab test results for random/fasting glucose or HbA1c. To ensure high-quality phenotyping, however, we carefully ruled out type 1 diabetes mellitus, similarly to what is done by the eMERGE algorithm. Furthermore, the eMERGE algorithm was developed in 2012 and does not capture T2D medications

approved by FDA after 2012, such as some DPP-4 inhibitors, GLP-1 agonists, insulin analogs, and any SGLT-2 inhibitors. Second, the first diagnosis of T2D may not always have been documented at our hospital, so we cannot ensure that a given patient was not been diagnosed previously elsewhere. By the same reasoning, it is possible that some diagnoses and prescription of medications documented at different institutions may have been missed. One possible solution and future direction is to incorporate data from insurance claims. In addition, the definition of outcome comorbidities by CCS Diagnoses terms would reduce the resolution of disease category coded by ICD-9. However, we could collect sufficient sample size and reduce the noise of underdiagnoses, including assignment of a similar but different ICD-9 code, which was observed in diabetic neuropathy. Finally, the results of this observational study need to be interpreted with caution and will require confirmation in randomized clinical trials that stratify treatment groups based on the status of disease comorbidities and other clinical features.

Our study is the first to investigate a comprehensive EMR linked to a biobank with the aim of discovering the risk of T2D comorbidities and revealing treatment strategies tailored to patient risk profiles. Using systematic feature selection with a regularization technique and survival analysis with time dependent covariate in propensity score matched cohort, we identified a list of medications that may influence the risk spectrum of T2D comorbidities. We analyzed medications based on ingredient level irrespectively on drug form, and the approach enabled us to identify the ingredients that associate with comorbidities, which have not been reported in literature. Future studies would be followed up for further investigation. Finally, both previously observed and novel associations between medications and comorbidities were validated by genetic variants analysis using the T2D patient cohort in our biobank. With further evidence, if these recommendations can successfully be implemented in the clinic, it would undoubtedly reduce the economic burden by prioritizing prescription of the right medication at the right

time. Our medication–comorbidity model provides a real-world, data-driven approach for treatment

stratifications that consider future development of comorbidities, with the ultimate goal of guiding

precision medicine based on individual clinical phenotypes of patients with T2D.

Methods

Ethics statement

All data were fully anonymized. Ethical approval was provided by Program for the Protection of Human Subjects (PPHS) at Mount Sinai health system (HS#17-01058, IRB#17-02530). Waiver of informed consent was approved due to the purpose and scale of this study with retrospective and anonymous data only.

Study design

The objective of our study was to develop a data-driven approach to identify the risk and benefits of medications based on the T2D comorbidity profiles for guiding precision medicine tailored to individual clinical phenotypes. We used comprehensive Electronic Medical Records (EMR) system of Mount Sinai Hospital (MSH) to identify associations between T2D comorbidities and medications. We analyzed a T2D cohort of 40,144 individuals with well-annotated patient information. Please refer to Supplementary Methods section for patient cohort, clinical sources and term standardization, and definition of outcomes. We applied a systematic feature selection using a regularization technique and survival analysis method with time dependent covariates in a propensity score matched cohort. The relations between medications and comorbidities were validated by genetic variants analysis using the T2D patient cohort in our EMR-linked biobank.

Disease prevalence and disease pair temporal directionality

We calculated the disease prevalence based on the Kaplan–Meier product–limit estimator, with units of 1 year, and then we assessed disease-pair connectivity patterns for T2D comorbidities. Please refer to the Supplementary Methods section for detailed descriptions of the disease prevalence and disease pair temporal directionality analyses.

Feature selection

We considered many disease variables coded by Single-Level Clinical Classifications Software (CCS) diagnoses, medication variables mapped to RxNorm ingredient code, and lab/vital variables based on LOINC codes (Supplementary Methods). Accordingly, we adopted a feature selection method, specifically a penalized logistic regression with adaptive LASSO (Eq. 1), to identify variables of the highest relevance associated with ensuing comorbidities following T2D disease diagnosis. First, we pre-filtered disease and medication variables to only include those with more than 20 patients, and lab/vital variables measured for more than 50% of patients, yielding a total of 1,255 features including 246 diseases, 899 medications, 107 lab/vitals, and four demographic features. Adaptive LASSO is an extension of traditional LASSO [60] that uses coefficient-specific weights [61]. The adaptive LASSO estimator may achieve sparsity and selection consistency for the true model, i.e., correctly identifies the zero and nonzero parameters [62]. Let $\mathcal{L}_n(\beta; Y, X)$ be the negative log-likelihood parametrized by β for a sample of size *n*. The adaptive LASSO estimator is defined as:

$$\hat{\beta} = argmin_\beta \left\{ \mathcal{L}_n(\beta; Y, X) + \lambda_n \sum_{j=1}^d \widehat{\omega_j} |\beta_j| \right\} \tag{1}$$

21

where $\widehat{\omega_j} = |\widetilde{\beta_j}|^{-\gamma}$ is a coefficient specific weights vector, and λ_n is a regularization parameter. We set the positive constant γ as 1, according to Zou et al. [61], and obtained $\widetilde{\beta}$ by the maximum likelihood estimate of Ridge regression. The λ_n value for minimum AUC was chosen by 10-fold cross validation. We used the R package *glmnet* [63] for these penalized regressions.

Logistic regression model

In the cross-sectional analysis, we used the odds ratio (OR) from logistic regression to quantify the magnitude of the risk of comorbidity associated with prescription of each medication (i.e., increased risk or protective effect). Please see the Supplementary Methods section for a detailed description of logistic regression. Features that were significantly associated with each of the outcome comorbid diseases with nominal p values less than 0.05 were used for the following propensity score matching and survival analysis.

Survival analysis

We calculated hazard ratios using Cox proportional hazards models with propensity score matching of cases and controls. Propensity score matching is described in Supplementary Methods. Although many previous studies demonstrated that propensity score matching effectively suppresses confounder effects [64,65], we further adjusted the time-dependent effects after the baseline time point using time-dependent covariates [66]. We did so because our study focuses on the association between medications and the development of comorbid diseases, considering interventions after the baseline time point (Fig. 1B). For the time-dependent covariates, we used as many adaptive-LASSO-selected covariates as possible unless a given covariate did not violate the proportional hazards assumption.

Cox proportional hazard model with time-dependent covariates:

$$h(t) = h_0(t)\exp\left(\beta_x LASSO\_selected\_covariate_x(t)\right) \qquad (2)$$

where $h(t)$ is the expected hazard at time t, $h_0(t)$ is the baseline hazard, and

$LASSO\_selected\_covariate_x(t)$ is each of the time t-dependent binary variables selected by adaptive

LASSO, such as medications, disease history, lab values, and vital signs.

We confirmed the proportional hazards assumption by calculating Schoenfeld residuals

(Schoenfeld test p>0.05). We used the R packages *survival* and *survminer* for survival analysis. The p

values derived from the Cox models were adjusted for multiple testing using the Benjamini and

Hochberg false discovery rate method [67].

SNP/gene and disease association analyses

We assessed our predicted associations between medications and above seven well-known T2D

comorbidities [11] by performing a genetic association analysis. Specifically, we retrieved drug–gene

relationship information from DrugBank version 5.0.11 [68], and then obtained genotyping data for all

SNPs mapped to these genes in our cohort. We analyzed association a SNP–comorbidity basis as well as

a gene–comorbidity basis by using the SNP-set Kernel Association Test (SKAT) method [69]

(Supplementary Methods).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1   Collaboration, N. C. D. R. F. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* **387**, 1513-1530, doi:10.1016/S0140-6736(16)00618-8 (2016).

2   Emerging Risk Factors, C. *et al.* Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet* **375**, 2215-2222, doi:10.1016/S0140-6736(10)60484-9 (2010).

3   Chatterjee, S., Khunti, K. & Davies, M. J. Type 2 diabetes. *Lancet* **389**, 2239-2251, doi:10.1016/S0140-6736(17)30058-2 (2017).

4   Menke, A., Casagrande, S., Geiss, L. & Cowie, C. C. Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. *JAMA* **314**, 1021-1029, doi:10.1001/jama.2015.10029 (2015).

5   Seuring, T., Archangelidi, O. & Suhrcke, M. The Economic Costs of Type 2 Diabetes: A Global Systematic Review. *Pharmacoeconomics* **33**, 811-831, doi:10.1007/s40273-015-0268-9 (2015).

6   American Diabetes, A. Economic Costs of Diabetes in the U.S. in 2017. *Diabetes Care* **41**, 917-928, doi:10.2337/dci18-0007 (2018).

7   Suh, D. C., Choi, I. S., Plauschinat, C., Kwon, J. & Baron, M. Impact of comorbid conditions and race/ethnicity on glycemic control among the US population with type 2 diabetes, 1988-1994 to 1999-2004. *J Diabetes Complications* **24**, 382-391, doi:10.1016/j.jdiacomp.2009.07.001 (2010).

8   Kaul, N. & Ali, S. Genes, Genetics, and Environment in Type 2 Diabetes: Implication in Personalized Medicine. *DNA Cell Biol* **35**, 1-12, doi:10.1089/dna.2015.2883 (2016).

9   Bertsimas, D., Kallus, N., Weinstein, A. M. & Zhuo, Y. D. Personalized Diabetes Management Using Electronic Medical Records. *Diabetes Care* **40**, 210-217, doi:10.2337/dc16-0826 (2017).

10   Mayer, B. Using systems biology to evaluate targets and mechanism of action of drugs for diabetes comorbidities. *Diabetologia* **59**, 2503-2506, doi:10.1007/s00125-016-4032-2 (2016).

11   Li, L. *et al.* Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci Transl Med* **7**, 311ra174, doi:10.1126/scitranslmed.aaa9364 (2015).

12   Ahlqvist, E. *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol*, doi:10.1016/S2213-8587(18)30051-2 (2018).

13   Iglay, K. *et al.* Prevalence and co-prevalence of comorbidities among patients with type 2 diabetes mellitus. *Curr Med Res Opin* **32**, 1243-1252, doi:10.1185/03007995.2016.1168291 (2016).

14   Matsuura, T. *et al.* [Study of bacterial flora in the oral cavity and stomach of elderly patients receiving nasogastric tube feeding]. *Kansenshogaku Zasshi* **71**, 397-404 (1997).

15   Group, D. A. I. S. The prevalence of coronary heart disease in Type 2 diabetic patients in Italy: the DAI study. *Diabet Med* **21**, 738-745, doi:10.1111/j.1464-5491.2004.01230.x (2004).

16   Plantinga, L. C. *et al.* Prevalence of chronic kidney disease in US adults with undiagnosed diabetes or prediabetes. *Clin J Am Soc Nephrol* **5**, 673-682, doi:10.2215/CJN.07891109 (2010).

17   Li, L. *et al.* Disease risk factors identified through shared genetic architecture and electronic medical records. *Sci Transl Med* **6**, 234ra257, doi:10.1126/scitranslmed.3007191 (2014).

18   Nissen, S. E. & Wolski, K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* **356**, 2457-2471, doi:10.1056/NEJMoa072761 (2007).

19   Jain, R., Mullins, C. D., Lee, H. & Wong, W. Use of rosiglitazone and pioglitazone immediately after the cardiovascular risk warnings. *Res Social Adm Pharm* **8**, 47-59, doi:10.1016/j.sapharm.2010.12.003 (2012).

20   Mooradian, A. D. Dyslipidemia in type 2 diabetes mellitus. *Nat Clin Pract Endocrinol Metab* **5**, 150-159, doi:10.1038/ncpendmet1066 (2009).

21   Liu, H. F., Qi, X. W., Ma, L. L., Yao, D. K. & Wang, L. Atorvastatin improves endothelial progenitor cell function and reduces pulmonary hypertension in patients with chronic pulmonary heart disease. *Exp Clin Cardiol* **18**, e40-43 (2013).

22   Yanai, H. Dipeptidyl peptidase-4 inhibitor sitagliptin significantly reduced hepatitis C virus replication in a diabetic patient with chronic hepatitis C virus infection. *Hepatobiliary Pancreat Dis Int* **13**, 556 (2014).

23   Aroda, V. R. *et al.* Long-term Metformin Use and Vitamin B12 Deficiency in the Diabetes Prevention Program Outcomes Study. *J Clin Endocrinol Metab* **101**, 1754-1761, doi:10.1210/jc.2015-3754 (2016).

24   Group, N. S. *et al.* Effect of nateglinide on the incidence of diabetes and cardiovascular events. *N Engl J Med* **362**, 1463-1476, doi:10.1056/NEJMoa1001122 (2010).

25   Ng, T. P. *et al.* Long-term metformin usage and cognitive function among older adults with diabetes. *J Alzheimers Dis* **41**, 61-68, doi:10.3233/JAD-131901 (2014).

26   Kuan, Y. C., Huang, K. W., Lin, C. L., Hu, C. J. & Kao, C. H. Effects of metformin exposure on neurodegenerative diseases in elderly patients with type 2 diabetes mellitus. *Prog Neuropsychopharmacol Biol Psychiatry* **79**, 77-83, doi:10.1016/j.pnpbp.2017.06.002 (2017).

27   Bitzur, R. Remembering Statins: Do Statins Have Adverse Cognitive Effects? *Diabetes Care* **39 Suppl 2**, S253-259, doi:10.2337/dcS15-3022 (2016).

28   Geifman, N., Brinton, R. D., Kennedy, R. E., Schneider, L. S. & Butte, A. J. Evidence for benefit of statins to modify cognitive decline and risk in Alzheimer's disease. *Alzheimers Res Ther* **9**, 10, doi:10.1186/s13195-017-0237-y (2017).

29   Beales, I. L., Dearman, L., Vardi, I. & Loke, Y. Reduced Risk of Barrett's Esophagus in Statin Users: Case-Control Study and Meta-Analysis. *Dig Dis Sci* **61**, 238-246, doi:10.1007/s10620-015-3869-4 (2016).

30   Smith, I., Schmidt, R., Halm, E. A. & Mansi, I. A. Do Statins Increase the Risk of Esophageal Conditions? Findings from Four Propensity Score-Matched Analyses. *Clin Drug Investig*, doi:10.1007/s40261-017-0589-x (2017).

31   Coppolino, G. *et al.* Exploring the effects of DPP-4 inhibitors on the kidney from the bench to clinical trials. *Pharmacol Res*, doi:10.1016/j.phrs.2017.12.001 (2017).

32   Chattipakorn, N., Apaijai, N. & Chattipakorn, S. C. Dipeptidyl peptidase-4 inhibitors and the ischemic heart: Additional benefits beyond glycemic control. *Int J Cardiol* **202**, 415-416, doi:10.1016/j.ijcard.2015.09.044 (2016).

33   Hung, Y. C., Lin, C. C., Huang, W. L., Chang, M. P. & Chen, C. C. Sitagliptin and risk of heart failure hospitalization in patients with type 2 diabetes on dialysis: A population-based cohort study. *Sci Rep* **6**, 30499, doi:10.1038/srep30499 (2016).

34   Aggarwal, R., Akhthar, T. & Jain, S. K. Coronary artery disease and its association with Vitamin D deficiency. *J Midlife Health* **7**, 56-60, doi:10.4103/0976-7800.185334 (2016).

35   Cao, Y. *et al.* Population-wide Impact of Long-term Use of Aspirin and the Risk for Cancer. *JAMA Oncol* **2**, 762-769, doi:10.1001/jamaoncol.2015.6396 (2016).

36   Lovell, M. A. *et al.* Calcium Channel Blockers, Progression to Dementia, and Effects on

        Amyloid Beta Peptide Production. *Oxid Med Cell Longev* **2015**, 787805,

        doi:10.1155/2015/787805 (2015).

37   Domeney, A. M. *et al.* The effect of ondansetron on cognitive performance in the marmoset.

        Pharmacol Biochem Behav **38**, 169-175 (1991).

38   Luo, F. *et al.* Phenylephrine enhances glutamate release in the medial prefrontal cortex through

        interaction with N-type Ca2+ channels and release machinery. *J Neurochem* **132**, 38-50,

        doi:10.1111/jnc.12941 (2015).

39   Li, S. *et al.* Bacitracin Inhibits the Migration of U87-MG Glioma Cells via Interferences of the

        Integrin Outside-in Signaling Pathway. *J Korean Neurosurg Soc* **59**, 106-116,

        doi:10.3340/jkns.2016.59.2.106 (2016).

40   Li, F. *et al.* Azithromycin effectively inhibits tumor angiogenesis by suppressing vascular

        endothelial growth factor receptor 2-mediated signaling pathways in lung cancer. *Oncol Lett* **14**,

        89-96, doi:10.3892/ol.2017.6103 (2017).

41   Zhang, L. *et al.* Lidocaine inhibits the proliferation of lung cancer by regulating the expression of

        GOLT1A. *Cell Prolif* **50**, doi:10.1111/cpr.12364 (2017).

42   Hollis, T. M. *et al.* Antihistamines reverse blood-ocular barrier breakdown in experimental

        diabetes. *J Diabet Complications* **2**, 47-49 (1988).

43   De La Cruz, P. *et al.* Effects of clopidogrel and ticlopidine on experimental diabetic ischemic

        retinopathy in rats. *Naunyn Schmiedebergs Arch Pharmacol* **367**, 204-210, doi:10.1007/s00210-

        002-0657-4 (2003).

44   Morgan, S. L., Baggott, J. E., Lee, J. Y. & Alarcon, G. S. Folic acid supplementation prevents

        deficient blood folate levels and hyperhomocysteinemia during longterm, low dose methotrexate

29

therapy for rheumatoid arthritis: implications for cardiovascular disease prevention. *J Rheumatol* **25**, 441-446 (1998).

45   Dave, J. A., Cohen, K., Micklesfield, L. K., Maartens, G. & Levitt, N. S. Antiretroviral Therapy, Especially Efavirenz, Is Associated with Low Bone Mineral Density in HIV-Infected South Africans. *PLoS One* **10**, e0144286, doi:10.1371/journal.pone.0144286 (2015).

46   Dysken, M., Kuskowski, M., Love, S. & Ondansetron Study, G. Ondansetron in the treatment of cognitive decline in Alzheimer dementia. *Am J Geriatr Psychiatry* **10**, 212-215 (2002).

47   Cutler, N. R. *et al.* Evaluation of zimeldine in Alzheimer's disease. Cognitive and biochemical measures. *Arch Neurol* **42**, 744-748 (1985).

48   Dehlin, O., Hedenrud, B., Jansson, P. & Norgard, J. A double-blind comparison of alaproclate and placebo in the treatment of patients with senile dementia. *Acta Psychiatr Scand* **71**, 190-196 (1985).

49   Ticlopidine treatment reduces the progression of nonproliferative diabetic retinopathy. The TIMAD Study Group. *Arch Ophthalmol* **108**, 1577-1583 (1990).

50   Maneuf, Y. P., Luo, Z. D. & Lee, K. alpha2delta and the mechanism of action of gabapentin in the treatment of pain. *Semin Cell Dev Biol* **17**, 565-570, doi:10.1016/j.semcdb.2006.09.003 (2006).

51   Nimmrich, V. & Eckert, A. Calcium channel blockers and dementia. *Br J Pharmacol* **169**, 1203-1210, doi:10.1111/bph.12240 (2013).

52   Hollis, T. M., Sill, H. W., Butler, C., Campos, M. J. & Gardner, T. W. Astemizole reduces blood-retinal barrier leakage in experimental diabetes. *J Diabetes Complications* **6**, 230-235 (1992).

53   Hirata, M. *et al.* Lidocaine inhibits tyrosine kinase activity of the epidermal growth factor receptor and suppresses proliferation of corneal epithelial cells. *Anesthesiology* **100**, 1206-1210 (2004).

54   Sakaguchi, M., Kuroda, Y. & Hirose, M. The antiproliferative effect of lidocaine on human tongue cancer cells with inhibition of the activity of epidermal growth factor receptor. *Anesth Analg* **102**, 1103-1107, doi:10.1213/01.ane.0000198330.84341.35 (2006).

55   Allouch, S. & Munusamy, S. AMP-activated Protein Kinase as a Drug Target in Chronic Kidney Disease. *Curr Drug Targets*, doi:10.2174/1389450118666170601130947 (2017).

56   Pattaro, C. *et al.* Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun* **7**, 10023, doi:10.1038/ncomms10023 (2016).

57   Heaf, J. Metformin in chronic kidney disease: time for a rethink. *Perit Dial Int* **34**, 353-357, doi:10.3747/pdi.2013.00344 (2014).

58   von Hurst, P. R., Stonehouse, W. & Coad, J. Vitamin D supplementation reduces insulin resistance in South Asian women living in New Zealand who are insulin resistant and vitamin D deficient - a randomised, placebo-controlled trial. *Br J Nutr* **103**, 549-555, doi:10.1017/S0007114509992017 (2010).

59   Pacheco, J. & Thompson, W. Northwestern University. Type 2 Diabetes Mellitus. PheKB; 2012 Available from: https://phekb.org/phenotype/18).

60   Tibshirani, R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **73**, 273-282, doi:DOI 10.1111/j.1467-9868.2011.00771.x (2011).

61   Zou, H. & Zhang, H. H. On the Adaptive Elastic-Net with a Diverging Number of Parameters. *Ann Stat* **37**, 1733-1751, doi:10.1214/08-AOS625 (2009).

62   Zou, H. & Li, R. One-step Sparse Estimates in Nonconcave Penalized Likelihood Models. *Ann Stat* **36**, 1509-1533, doi:10.1214/009053607000000802 (2008).

63   Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).

64   Schneeweiss, S. Learning from big health care data. *N Engl J Med* **370**, 2161-2163, doi:10.1056/NEJMp1401111 (2014).

65   Schermerhorn, M. L. *et al.* Long-Term Outcomes of Abdominal Aortic Aneurysm in the Medicare Population. *N Engl J Med* **373**, 328-338, doi:10.1056/NEJMoa1405778 (2015).

66   Fisher, L. D. & Lin, D. Y. Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health* **20**, 145-157, doi:10.1146/annurev.publhealth.20.1.145 (1999).

67   Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300 (1995).

68   Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* **34**, D668-672, doi:10.1093/nar/gkj067 (2006).

69   Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D. & Lin, X. Sequence kernel association tests for the combined effect of rare and common variants. *Am J Hum Genet* **92**, 841-853, doi:10.1016/j.ajhg.2013.04.015 (2013).

Acknowledgements

Author Contributions

Study supervision: L.L. and J.T.D. Conceived and designed the study: O.I. and L.L. Analyzed EMRs: O.I., B.G., and L.L. Performed machine learning, model development, survival and other statistical analysis: O.I. Figures and tables: O.I. and B.G. Analyzed genotyping data: O.I., L.L. Contributed Drug Bank and genetic variant analysis tools: B.G., S.D.L., and R.C. Contributed statistical insights: H-C. L. Contributed to literature review: O.I. and N.M. Contributed to clinical interpretation: L.L., O.I. and N.M. Wrote and edited the paper: O.I., B.G., N.M., L.L., S.D.L., and J.T.D.

Competing Interests

J.T.D. has received consulting fees or honoraria from Janssen Pharmaceuticals, GlaxoSmithKline, AstraZeneca, and Hoffman-La Roche; is a scientific advisor to LAM Therapeutics; and holds equity in NuMedii Inc., Ayasdi Inc., and Ontomics, Inc. O.I. is an employee of Sumitomo Dainippon Pharma Co., Ltd. All other authors declare no competing interests.

Figure Legends

Fig. 1. Schematic work flow of our study design and methods. (A) Workflow of current study. (B) Diagram of survival analysis with time dependent covariates after propensity score matching. The baseline time point, T_0, was defined as the first prescription day of the targeted medication or 30 days after the initial T2D diagnosis date, whichever was later (see Supplementary Methods).

Fig. 2. Prevalence and increased risk after T2D. (A) The 20 most prevalent diseases in our T2D cohort. The color coding of diseases is shown in the right of the plot. (B) The 20 diseases with the greatest increase in risk after T2D diagnosis (adjusted binomial p values < 0.05, using the Bonferroni method). The values on the edge and line thickness indicate relative risks. A complete list of prevalence and increased risk is provided in Supplementary Table S1.

Fig. 3. Heatmap of associations between T2D disease comorbidity and T2D /T2D- related medication. Rectangle border color represents odds ratio (OR) by the cross-sectional analysis using logistic regression (nominal p values < 0.05), and rectangle fill color represents hazard ratio (HR) by the survival analysis using time-dependent covariate Cox regression after propensity score matching (adjusted p values < 0.05, using the Benjamini-Hochberg method). The protective associations (OR < 1.0 or HR < 1.0) are displayed in blue, and the risk associations (OR > 1.0 or HR > 1.0) are displayed in red. The drug indications are colored gray.

Fig. 4. Network of T2D disease comorbidities and medications. This network includes all kinds of medications, including non-T2D medications, and associations with adjusted p values < 0.01, using the

Benjamini-Hochberg method. For visualization purposes, we showed the diseases categorized into six

Multi-Level CCS codes (Neoplasms, Nutritional, Mental, Sense, Circulatory, and Genitourinary), except

for residual diseases beginning with "other" (e.g., "Other circulatory disease"). Blue ellipses represent

medications, and squares represent disease comorbidities with the fill color corresponding to the disease

categories by Multi-Level CCS codes. The protective associations (HR < 1.0) are displayed by blue

lines, and the risk associations (HR > 1.0) are displayed by red lines. Line thickness represents the

magnitude of HR. A complete list of all associations (adjusted p values < 0.05) is available in

Supplementary Tables S4 and S5.

Tables

Table 1. Protective or risk medications for the selected comorbidities.

Table 1 (A)

Comorbidity	Medication	HR cox	95% CI	Adjusted P value	Reference (PMID)
Retinal detachments; defects; vascular occlusion; and retinopathy	Psyllium	0.47	0.26-0.83	3.35E-02	
	Hydroxyzine	0.52	0.33-0.81	1.66E-02	
	Aluminum (IN[a])	0.52	0.32-0.85	3.23E-02	
	Naproxen	0.70	0.54-0.91	2.74E-02	
	Lidocaine	0.71	0.59-0.87	4.89E-03	
	Nystatin	0.72	0.57-0.92	2.67E-02	
	Calcium	0.73	0.57-0.92	3.20E-02	
	Ibuprofen	0.73	0.6-0.88	7.66E-03	
	Clopidogrel	0.80	0.68-0.94	2.54E-02	
Chronic kidney disease	Levetiracetam	0.53	0.37-0.76	3.95E-03	
	Metformin[b]	0.63	0.63-0.63	0.00E+00	22641582, 21325870
	Sodium phosphate	0.66	0.51-0.86	1.11E-02	
	Ondansetron	0.68	0.6-0.78	1.86E-06	
	Naproxen	0.69	0.54-0.9	2.38E-02	
	Acetaminophen	0.72	0.64-0.81	1.13E-06	
	Hydrocortisone	0.76	0.65-0.9	8.91E-03	28415636
Coronary atherosclerosis and other heart disease	Atovaquone	0.22	0.1-0.52	3.76E-03	
	Pseudoephedrine	0.34	0.21-0.55	1.90E-04	
	Timolol	0.48	0.28-0.81	2.33E-02	15819591
	Diclofenac	0.60	0.42-0.86	2.28E-02	
	Potassium phosphate	0.60	0.45-0.81	5.52E-03	
	Sulfamethoxazole	0.62	0.48-0.81	3.95E-03	

	Iron	0.65	0.51-0.82	2.93E-03	18294483
	Cholecalciferol	0.66	0.48-0.91	3.80E-02	27499590
	Hydromorphone	0.74	0.6-0.92	2.55E-02	
	Metformin[b]	0.79	0.69-0.89	1.72E-03	28476766
Acute cerebrovascular disease	Succinylcholine	0.53	0.39-0.71	4.95E-04	
	Repaglinide[b]	0.54	0.37-0.8	1.02E-02	
	Ondansetron	0.59	0.47-0.75	1.84E-04	
	Bacitracin	0.61	0.42-0.89	3.60E-02	
	Sitagliptin[b]	0.65	0.46-0.92	4.67E-02	26687195
Neoplasms excl. Benign	Betamethasone	0.53	0.37-0.77	6.01E-03	
	Aspirin	0.57	0.57-0.57	0.00E+00	26940135
	Ketorolac	0.65	0.53-0.79	4.78E-04	26071482
	Bacitracin	0.67	0.51-0.87	1.56E-02	
	Azithromycin	0.72	0.62-0.84	4.11E-04	
	Vancomycin	0.73	0.61-0.88	4.88E-03	
	Hydralazine	0.74	0.61-0.89	8.55E-03	28277033
	Oxycodone	0.79	0.7-0.9	3.39E-03	
	Amoxicillin	0.80	0.67-0.95	3.30E-02	28529929
	Lidocaine	0.80	0.68-0.95	2.94E-02	
Delirium, dementia, and amnestic and other cognitive disorders	Granisetron	0.44	0.22-0.85	4.48E-02	
	Nifedipine	0.54	0.42-0.7	8.29E-05	26221415
	Hydrocortisone	0.65	0.53-0.81	1.64E-03	
	Cefazolin	0.68	0.58-0.8	3.42E-05	
	Atorvastatin	0.69	0.58-0.82	4.10E-04	[c] 28212683, 27440840
	Ondansetron	0.71	0.59-0.84	1.12E-03	
	Hydromorphone	0.72	0.58-0.89	1.23E-02	
	Metformin[b]	0.74	0.65-0.85	1.95E-04	[c] 24577463, 28583443
	Gabapentin	0.74	0.62-0.89	6.80E-03	
	Phenylephrine	0.76	0.64-0.92	1.87E-02	
Nutritional deficiencies	Psyllium	0.59	0.44-0.8	3.68E-03	

37

Bupivacaine	0.60	0.5-0.71	1.67E-07	
Gemfibrozil	0.67	0.48-0.93	4.59E-02	
Acetaminophen	0.67	0.61-0.74	6.28E-13	
Chlorhexidine	0.67	0.5-0.91	3.09E-02	
Aspirin	0.69	0.63-0.75	5.98E-14	
Atovaquone	0.69	0.52-0.93	4.21E-02	
Dextromethorphan	0.71	0.56-0.9	2.07E-02	
Tiotropium	0.74	0.59-0.92	2.84E-02	
Allopurinol	0.79	0.67-0.93	1.66E-02	
Clopidogrel	0.79	0.71-0.88	1.98E-04	

Table 1 (B)

Comorbidity	Medication	HR cox	95% CI	Adjusted P value	Reference (PMID)
Retinal detachments; defects; vascular occlusion; and retinopathy	Aprepitant	4.81	2.4-9.65	1.84E-04	
	HYLAN G-F 20	4.37	2.29-8.32	1.44E-04	
	Prednisolone	2.99	2.14-4.18	1.70E-08	
	Desloratadine	2.18	1.15-4.11	4.72E-02	
	Timolol	1.76	1.3-2.38	2.28E-03	
	Repaglinide[b]	1.47	1.18-1.82	3.86E-03	
	Insulin[b]	1.36	1.17-1.58	7.48E-04	
	Lisinopril	1.36	1.2-1.54	3.89E-05	
	Glipizide	1.20	1.04-1.38	3.66E-02	
Chronic kidney disease	Indigotindisulfonate	2.65	1.55-4.55	3.26E-03	
	Omega-3 Acid Ethyl Esters (USP)	2.48	1.41-4.36	9.23E-03	
	Calcitriol	1.85	1.38-2.49	6.93E-04	
	Pyridoxine	1.64	1.15-2.34	2.43E-02	
	Nifedipine	1.52	1.3-1.78	9.49E-06	

	Pseudoephedrine	1.43	1.08-1.91	4.21E-02	
	Prochlorperazine	1.41	1.09-1.82	3.27E-02	
	Insulin[b]	1.40	1.25-1.58	1.05E-06	
	Labetalol	1.36	1.19-1.55	1.48E-04	
	Ferrous sulfate	1.30	1.14-1.48	1.15E-03	
	Carvedilol	1.18	1.04-1.34	3.24E-02	
Coronary atherosclerosis and other heart disease	Colchicine	1.71	1.33-2.18	3.76E-04	
	Insulin[b]	1.19	1.03-1.37	4.89E-02	[c] 28958751, 22686416
Acute cerebrovascular disease	Phenytoin	2.08	1.44-3.01	1.13E-03	
Neoplasms excl. Benign	Cosyntropin	2.62	1.55-4.43	2.91E-03	
	Trimethobenzamide	2.10	1.27-3.48	1.77E-02	
	Cetylpyridinium	1.93	1.14-3.28	4.47E-02	
	Nadolol	1.91	1.38-2.64	1.19E-03	
	Tenofovir	1.85	1.3-2.62	4.61E-03	
	Sodium sulfate	1.46	1.09-1.94	3.39E-02	
Delirium, dementia, and amnestic and other cognitive disorders	Iohexol	2.25	1.32-3.83	1.41E-02	
	Bupropion	1.59	1.22-2.08	4.32E-03	
	Repaglinide[b]	1.37	1.11-1.69	1.49E-02	
	Heparin	1.33	1.11-1.59	1.06E-02	
Nutritional deficiencies	HYLAN G-F 20	2.51	1.32-4.74	2.00E-02	
	Ritonavir	2.36	1.86-2.99	3.35E-10	26633015
	Thiopental	2.11	1.19-3.76	3.54E-02	
	Tretinoin	1.95	1.19-3.19	2.81E-02	
	Methotrexate	1.60	1.18-2.18	1.32E-02	9517760
	Nateglinide[b]	1.52	1.13-2.05	2.29E-02	
	Dobutamine	1.52	1.08-2.14	4.82E-02	
	Fexofenadine	1.47	1.2-1.79	1.67E-03	
	Ramipril	1.26	1.09-1.45	9.73E-03	
	Homatropine	1.26	1.07-1.48	2.49E-02	
	Glyburide[b]	1.24	1.06-1.44	2.33E-02	1070515

Repaglinide[b]	1.22	1.06-1.4	2.60E-02	
Metformin[b]	1.14	1.06-1.22	2.31E-03	22179958

(A) Protective medications for the selected comorbidities (adjusted p < 0.05, using Benjamini-Hochberg method). All comorbidities are listed in Supplementary Tables S4. (B) Risk medications for the selected comorbidities (adjusted p < 0.05, using Benjamini-Hochberg method). All comorbidities are available in Supplementary Tables S5.

[a] Ingredient.

[b] Medications for T2D.

[c] Multiple published studies showed inconsistent results.

Table 2. Summary of associations between T2D comorbidities and SNPs/genes related to predicted protective medications (p<0.05).

Disease	Medication	SNP-level association	Gene-level association		
		Gene (# of significant SNPs; P < 0.05)	Gene; P < 0.05	P value	# of marker used in SKAT
Retinal detachments; defects; vascular occlusion; and retinopathy	Calcium	ATP2C1(1), CACNA1C(23), CALM1(1), CAST(2), CP(2), S100A2(1), SPTBN1(2), TFRC(2)	CP	6.23E-04	28
	Hydroxyzine	HRH1(2)	-	-	-
	Ibuprofen	BCL2(5), PLAT(2), PPARA(3), PPARG(3), PTGS1(1), PTGS2(1), SLC15A1(6), THBD(1)	PTGS2	2.27E-02	4
Chronic kidney disease	Metformin	PRKAB1(3)	PRKAB1	1.26E-02	8
	Naproxen	PTGS1(1)	-	-	-
	Ondansetron	HTR3A(2), HTR4(4), OPRM1(2)	OPRM1	2.89E-02	98

	Cholecalciferol	VDR(3)	-	-	-
Coronary atherosclerosis and other heart disease	Diclofenac	ALOX5(1), KCNQ2(1), KCNQ3(4), SCN4A(1)	-	-	-
	Hydromorphone	OPRD1(2), OPRK1(3), OPRM1(2)	-	-	-
	Iron	CP(1), EGLN1(1), TF(5)	-	-	-
	Pseudoephedrine	ADRA1A(7), ADRB1(1), ATF6(1), GRIN3B(1), NFATC1(2), SLC6A2(6), SLC6A3(3)	-	-	-
	Timolol	ADRB1(1)	-	-	-
Acute cerebrovascular disease	Bacitracin	IDE(3)	-	-	-
	Ondansetron	HTR4(9), OPRM1(4)	-	-	-
	Repaglinide	ABCC8(2), PPARG(4)	-	-	-
	Sitagliptin	DPP4(3)	-	-	-
	Succinylcholine	CHRM1(1), CHRM2(3), CHRM3(3), CHRNA10(1), CHRNA7(1)	CHRM2	8.36E-03	46
Neoplasms excl. benign	Azithromycin	PADI4(2)	-	-	-
	Bacitracin	IDE(4)	-	-	-
	Betamethasone	NR3C1(7)	-	-	-
	Ketorolac	PTGS1(1)	-	-	-
	Lidocaine	EGFR(5), SCN4A(3), SCN5A(1), SCN9A(5)	SCN4A	1.08E-02	48
	Oxycodone	OPRK1(2), OPRM1(5)	-	-	-
Delirium, dementia, and amnestic and other cognitive disorders	Cefazolin	PON1(10)	-	-	-
	Gabapentin	ADORA1(3), CACNA1B(2), CACNA2D1(5), CACNA2D2(2), GABBR2(3), GRIN1(3), GRIN2A(6), GRIN2B(8)	CACNA2D2	3.21E-03	26
	Gabapentin	ADORA1(3), CACNA1B(2), CACNA2D1(5), CACNA2D2(2),	GRIN1	4.40E-03	6

		GABBR2(3), GRIN1(3), GRIN2A(6), GRIN2B(8)			
	Nifedipine	CACNA1C(13), CACNA1D(15), CACNA2D1(5), CACNB2(3), CALM1(1)	-	-	-
	Phenylephrine	ADRA1A(3), ADRA1B(2), ADRA1D(1)	ADRA1A	3.43E-03	73
Nutritional deficiencies	Bupivacaine	SCN10A(5)	-	-	-
	Dextromethorphan	CHRNA7(1), NCF2(1), OPRK1(2), OPRM1(2)	-	-	-
	Gemfibrozil	PPARA(2)	-	-	-
	Tiotropium	CHRM1(1), CHRM3(1), CHRM5(1)	CHRM1	1.25E-02	4

Reference SNP ID number, position, OR, 95% confidence interval, and p-values for SNP-level associations are provided in Supplementary Table S9.

Supporting information captions

Supplementary Methods.

Supplementary Results.

Supplementary Discussion.

Fig. S1. Types of selected features for each disease comorbidities. Types of selected features for each

disease comorbidities. Medication, disease, lab values/vital signs, demographics were colored by

blue, green, red, and purple, respectively.

Fig. S2. Exclusion of the first 30days after T2D diagnosis. (A) Diagnosis distribution. (B)

Prescription distribution. The order of the dates of T2D and its comorbidities was important for

this study, but it is not biologically meaningful to assign an exact date of diagnosis for either

T2D or its comorbidities. This is especially T2D true because, from a biological point of view,

T2D is subject to considerable uncertainty in regard to the date of onset. As a result, a very sharp

distinction of T2D comorbidities diagnoses immediately after diagnosis of T2D is not likely to

be meaningful.

Table S1. Prevalence and increased risk after T2D.

Table S2. All features selected by adaptive LASSO and odds ratio estimated by logistic regression

(nominal p < 0.05).

Table S3. All associations between T2D comorbidity and T2D /T2D-related medications (adjusted

p < 0.05, using the Benjamini-Hochberg method). References to the previous studies are also

included. [a] Multiple published studies showed inconsistent results.

Table S4. All protective medications identified by the survival analysis (adjusted p < 0.05, using the Benjamini-Hochberg method).

Table S5. All risk medications identified by the survival analysis (adjusted p < 0.05, using the Benjamini-Hochberg method).

Table S6. Balance of covariates before and after propensity score matching in the survival analysis.

Table S7. Candidate medications for adjunctive/combination therapy with T2D medications. (A) Significant additive effects in each subgroup of T2D patients taking a T2D medication in Table 1B. All medications in Table 1A were tested for each comorbidity, except for 'Nutritional deficiencies' (see Supplementary Methods). (B) Significant additive effects in each subgroup of T2D patients taking a T2D medication in Table 1A. All medications in Table 1A were tested across each comorbidity in Table 1. (C) High-resolution association between nutritional deficiency and high-risk T2D medications. Significant associations between diseases coded by ICD-9, and T2D medications (p < 0.05) clarified the type of nutritional deficiency. [a] Ingredient.

Table S8. Balance of covariates before and after propensity score matching in the additive effects analysis.

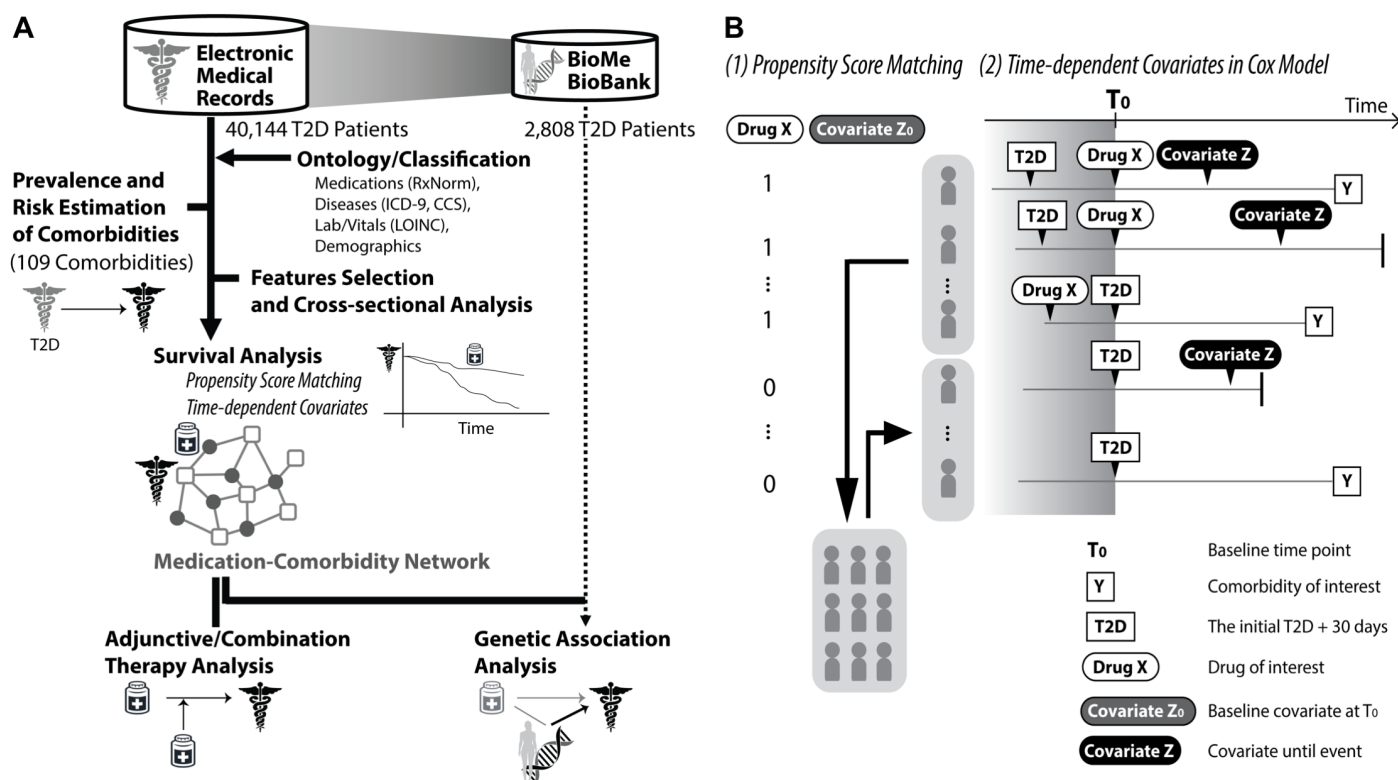**Table S9. All association between T2D comorbidities and SNPs related to predicted protective medications (p < 0.05).**
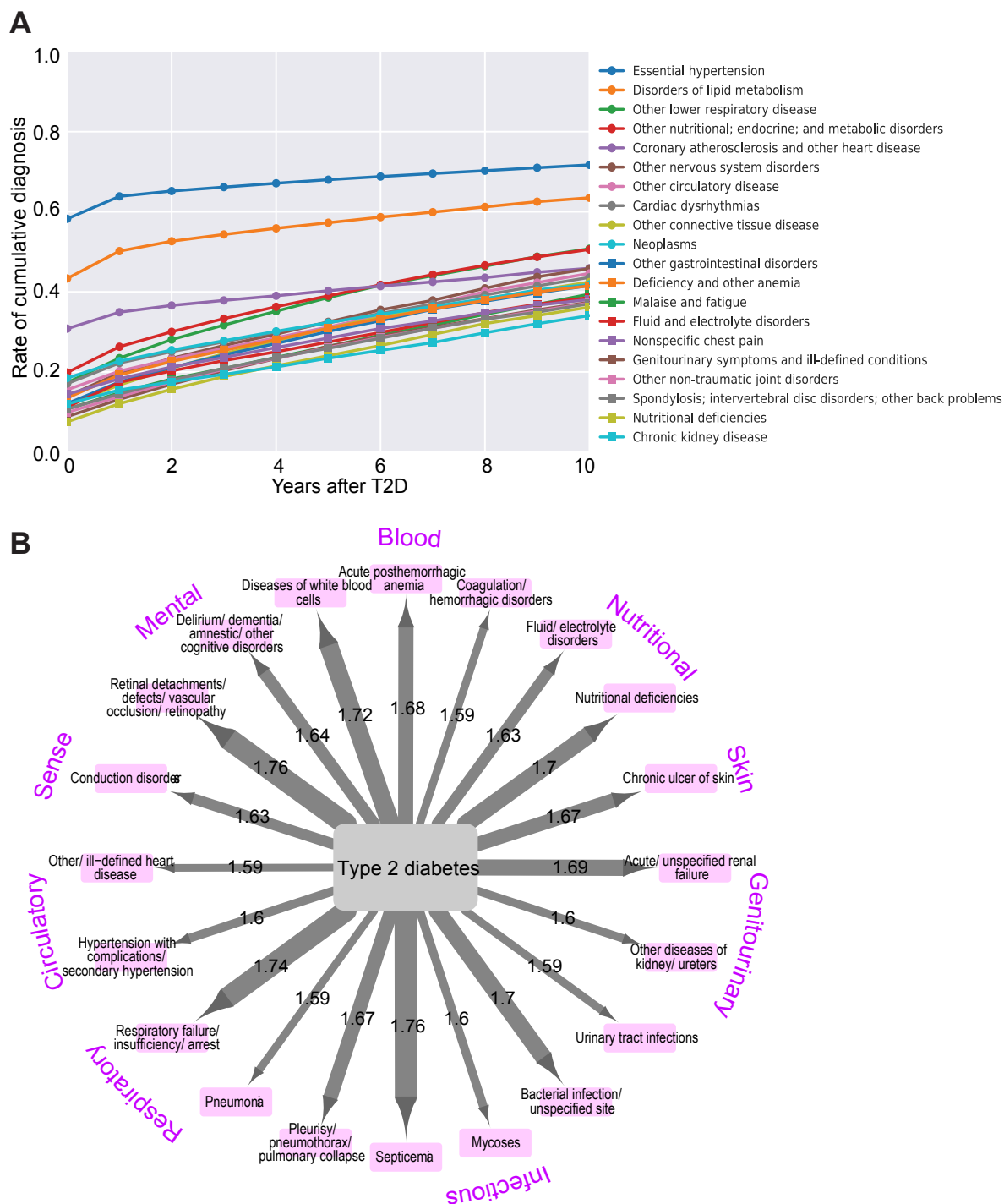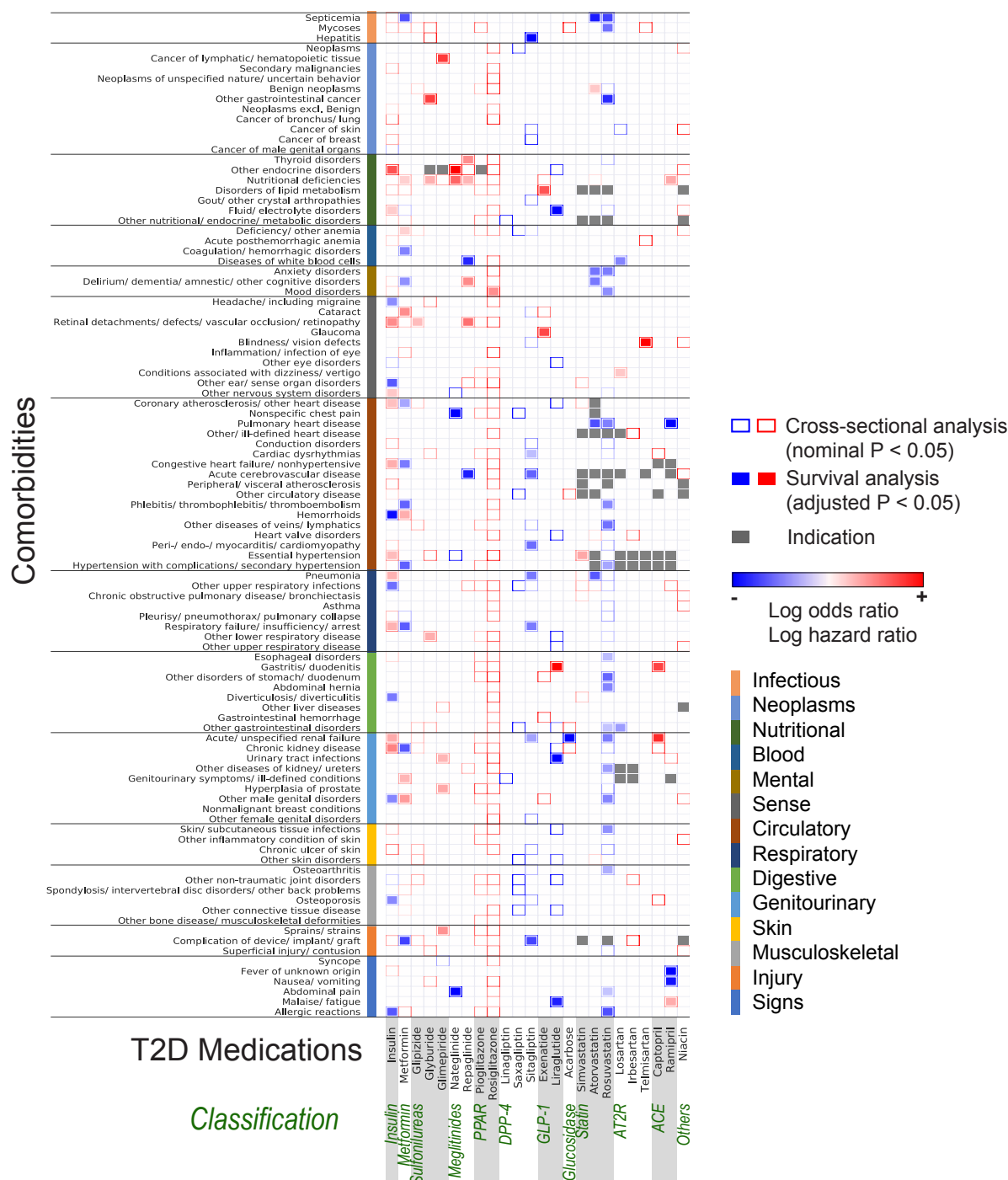
Fig. 1

Fig. 2

Fig. 3

Fig. 4