The Diabetome™ Knowledgebase : Over 300,000 curated diabetic patient records and analyses tools to gain new insights into risk factors of diabetes and characterize potential mechanisms of differential response to drugs

Background

Diabetes research today suffers from a number of challenges due to the fragmentation and complexity of the data. The lack of access to cohorts large enough to provide sufficient statistical power and the heterogeneity in the disease, hinders the study of disease complications. The chronic and multi-factorial nature of complications make the identification and validation of useful therapeutic targets very difficult. Samples are collected decades apart, often by disparate collaborators, under differing clinical and scientific conditions. This impairs researchers' ability to easily study diabetes and its complications with controlled cohorts over time, much less get reproducible results across multiple studies and labs.

To alleviate these challenges and help researchers develop new therapies, MedGenome Inc. in collaboration with Dr. Mohan's Diabetes Speciality Centre, has launched the Diabetome Knowledgebase. The Diabetome contains multiple data points collected over 25 years on over 300,000 Type 1 and Type 2 diabetes pa-tients. Information available in the database includes well-characterized clinical phenotypes, biochemical investigations, pharmaceutical prescriptions, genotype mapping, complications of diabetes, pedigree charts & basic statistical tools. By utilizing this integrated solution, researchers can stratify patients into sub-groups based on parameters such as rapid deteriorates, and differential therapeutic responders (positive and negative). This will allow researchers to further study the underlying mechanisms of the identified phenotypes by cor-relating with their clinical data , predict risk of diabetes, and find molecular features unique to the subgroups and help in identifying ideal treatment modalities for the sub-groups. Taken together, the Diabetome is a powerful tool to facilitate accurate drug target prediction and novel discoveries. In this white paper, we highlight the a) key features of the data-sets present , b) provide an overview of the output of the filters c) highlight examples of the types of data available for complications of the disease and show data on analysis and stratification of patients based on therapy.
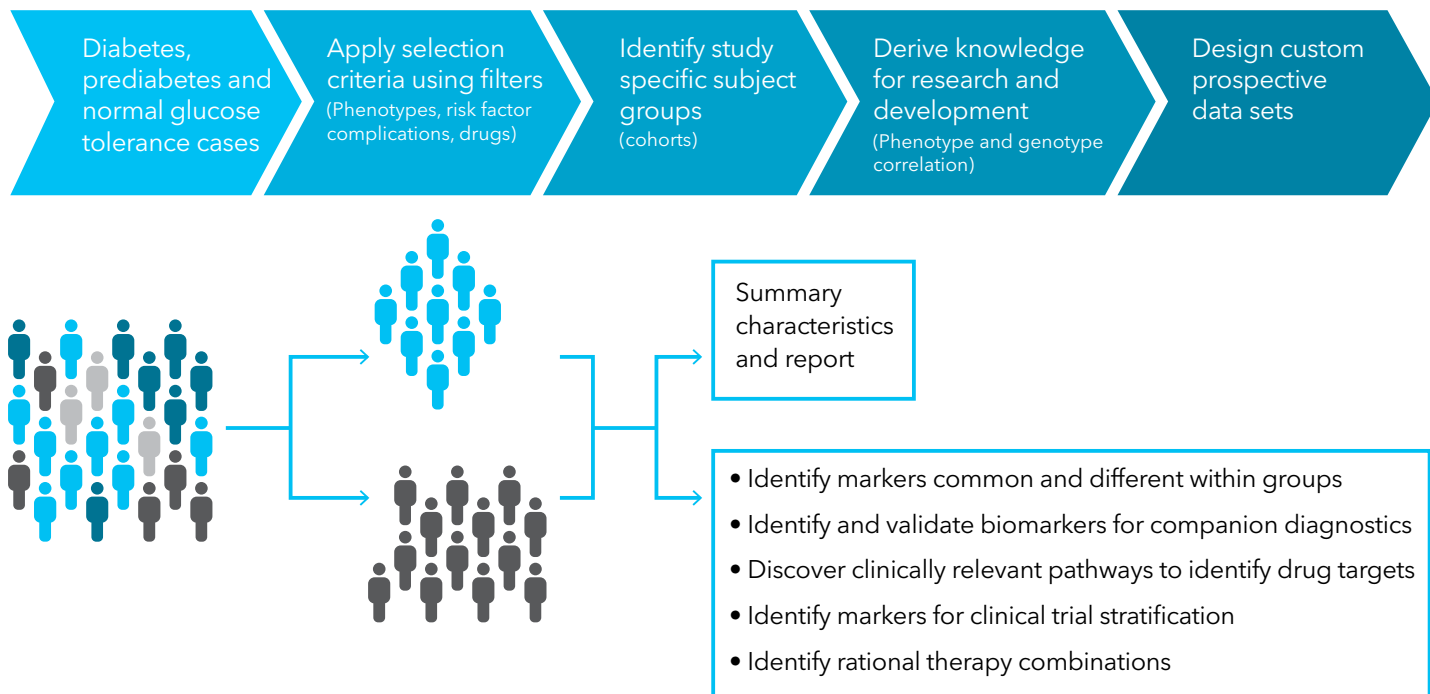

Figure 1: Overview of the workflow of the Diabetome: The schematic shows the workflow of the Diabetes knowledgebase : The Diabetome knowledgebase has curated data from over 300,000 cases and query filters to build cohorts for further analyses. These cohorts can be studied further to identify markers common and different within groups, identify and validate biomarkers for companion diagnostics, stratification of patients & identify relevant pathways for drug target discover and rational therapy combinations.

Summary of data curated in the Diabetome Knowledgebase

The Diabetome has over 300,000 patient records with clinical characteristics, biochemical lab results, secondary complications of diabetes, treatment history, and drug response data. In addition to the clinical information for the patients, another important aspect is that there is multi-generational diabetes familial history, family pedigree charts and information on consanguinous families available. Furthermore there is longitudinal data available for individual clinical profiles. Represented below is a summary of the types of data present in the Diabetome.
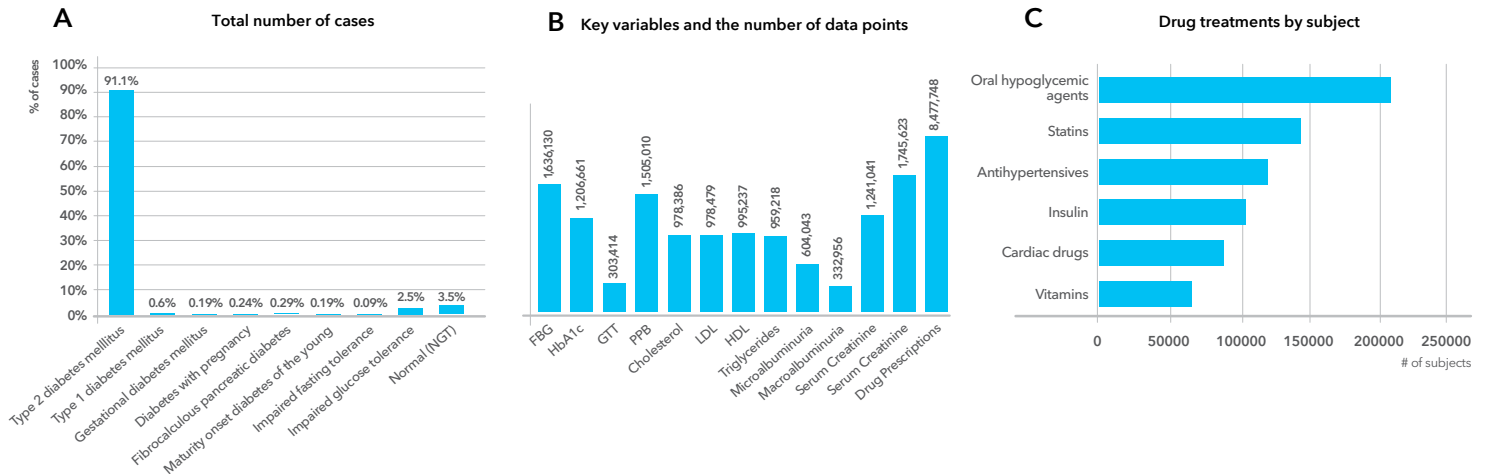

Figure 2: Summary of types of data present in the Diabetome: A) Shows a distribution of the datasets across the disease spectrum of Diabetes and the complications associated with Type I and 2 Diabetes. B) Shows the information available on the key clinical variables for Diabetes. C) Shows the data available for different classes of drugs for treatment of Diabetes.

Data on the secondary complications associated with Diabetes: microvascular and macrovascular complication case example

Microvascular complications include damage to eyes (retinopathy) leading to blindness, to kidneys (nephropathy) leading to renal failure and to nerves (neuropathy) leading to impotence. Diabetome contains over 140,000 diabetes patients with micro-vascular complications. Macrovascular complications include Cardiovascular Disease (CVD), Peripheral Vascular Disease (PVD) Table 1 shows : Number of cases with complications among Type 1, Type 2 and other diabetes groups. Arterial Disease (PAD). Diabetome contains over 15,000 diabetes patients with macrovascular complications.

**Table 1:** *Number of cases with complications among Type 1, Type 2 and other diabetes groups.*

Complications	Among T2DM	Among T1DM	Others	Total
Diabetic retinopathy	74,074	353	1,022	75,449
Diabetic nephropathy	64,363	260	1,772	66,395
Diabetic neuropathy	73,467	58	1,308	74,833

Table 2: *Number of cases with complications among Type 2 and other diabetes groups.*

Complications	Among T2DM	Others	Total
Cardio Vascular Disease (CVD)	8,314	194	8,508
Peripheral Vascular Disease (PVD)	6,184	316	6,500

Visualizing longitudinal clinical data:

The Diabetome Knowledgebase has tools that can identify trajectories and temporal patterns in longitudinal clinical data (e.g. fasting glucose or HbA1c) in diabetes patients to better understand the disease and its progression (Figure 3A). Using this feature one can visualize patients, and then select the specific subsets of patients and gain deeper insights into differences in clinical parameters among the groups. In addition, advanced features of the Diabetome can allow to stratify responders and non-responders to specific drugs (Figure 3B) and help in identifying the molecular basis for differences in responders versus non-responders based on other clinical data present in the database for each of the sub-groups.
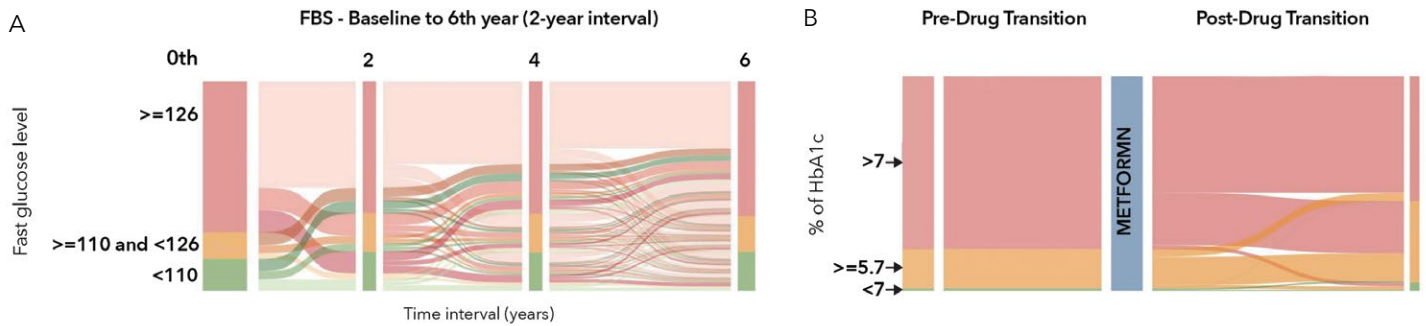
Figure 3A: Shows changes occurring in the levels of fasting gluscose in patients over a period of 6 years: The different colors represent different levels of fasting blood glucose levels in patients . The groups of patients with different levels of HbA1c can be selected and analyzed further for other biochemical and clinical parameters present in the database, and infer underlying causes for differences among the different groups. B) The Sankey chart shows the response pattern to Metformin treatment measured based on the levels of HbA1c levels. The responders and the non-responders are divided into groups and can be selected for further analysis of the underlying factors that lead to these differences in response.

Conclusions

MedGenome Inc. has developed the Diabetome™ Knowledgebase, a database of >300,000 diabetes phenotypes with longitudinal clinical, biochemical, therapeutic data and family details. In addition to the curated datasets, the solution has integrated analytic and visualization tools that can enable researchers to select specific diabetic cohorts, understand the biology, risk prediction of diabetes and stratify patients into responders and non-responders for drugs of interest and understand the molecular basis for the differences in response by utilizing the biochemical data captured in the database. This in turn can facilitate studying the underlying mechanisms of disease via correlation to other phenotypic information present in the database and a) enable early-stage drug discovery, b) assist in clinical decision making and c) understand post-market drug surveillance.

MEDGENOME INC.

348 Hatch Drive,
Foster City,
CA 94404, USA

(888) 440-0954

research@medgenome.com

www.medgenome.com

References

1. International Diabetes Federation. IDF Diabetes Atlas. 8th edn. Brussels, Belgium: International Diabetes Federation. 2017 (global prevalence)

2. Ranjit Unnikrishnan, Ranjit Mohan Anjana and Viswanathan Mohan. Diabetes mellitus and its complications in India. Nature Reviews Endocrinology, 12 357-370, 2016. (statistics section)