Phenotype at diagnosis and disease progression over a 10-year period: a data driven approach in a large population with type 2 diabetes

Anand Thakarakkattil Narayanan Nair[1], Louise Donnelly[1] , Simona M Hapca[1], Sushrima Gan[1], Adem Y Dawed[1],RM Anjana[2], Viswanathan Mohan[2], Ewan R Pearson[1]
[1]School of Medicine, University of Dundee, Dundee, United Kingdom DD1 9SY, [2]Madras Diabetes Research Foundation, Chennai, India 600086

Introduction

Type 2 diabetes is a heterogeneous disease exemplified by the palette model of diabetes where each patient has a colour based upon the aetiological processes that result in their diabetes.[1]

Mapping individuals on this palette based upon multiple phenotypes at diagnosis may enable a greater understanding of who is at risk of progression or complications of diabetes, enabling targeted intervention to those at high risk.

Methodology

We used discriminative dimensionality reduction analysis (DDRTree)[2,3] on phenotypic data of 10572 individuals newly diagnosed with type 2 diabetes using comprehensive electronic medical records in Tayside and Fife, Scotland.

The phenotypic variables included age and sex residualized values of HbA1c, body mass index (BMI), HDL-cholesterol, total cholesterol (TC), triglycerides (TG), alanine aminotransferase (ALT) ,creatinine, systolic and diastolic blood pressure (BP). Prior to DDRTree analysis each variable was pre-processed by outlier detection and normalization. From the resulting tree structure, we grouped patients allocated to each tree branch and identified characteristics based on the phenotype expression on the tree. We evaluated time to insulin requirement, time to diabetic retinopathy ('R1' to 'R4' Scottish Retinopathy Grading Scheme) and time to major adverse cardiovascular event (Scottish Morbidity Record 01) separately using competing risk models (Fine and Gray model) with death as a competing event.

Results

The DDRTree reduced the multidimensional space of the phenotype data into a low dimensional space which can be represented in the form of a tree structure (*Fig1*). The tree structure was comprised of six branches from the principal tree, which were different in terms of baseline phenotypic characteristics (*Fig2*). The first two branches had high HDL, but one had high systolic (150.11±14.27 mm of Hg) and diastolic blood pressure (87±8.56mm of Hg). The next two branches had hypertension, high HbA1c and high total cholesterol, with the one them also having very high Triglycerides (3.6±1.36mmol/L) and HbA1c (8.88±2.02%). The fifth branch was mostly characterised by obesity (BMI 34±5.45 kg/m^2) and finally the sixth branch had the lowest systolic (123.89±10.87 mm of Hg) and diastolic (72.02±7.27 mm of Hg) blood pressure with lower total cholesterol (4.26±0.90 mmol/L) and HbA1c (*Fig3*). As the analysis was adjusted for age and sex these were similar across each group, with a mean age of 62.9 years, with 56.8% males.

Fig1: Dimensionally reduced Phenotypic data

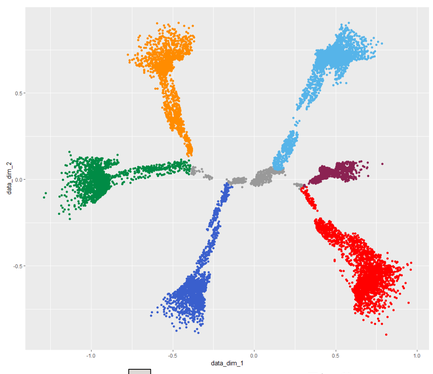


Fig3: Characterization of each tree ends

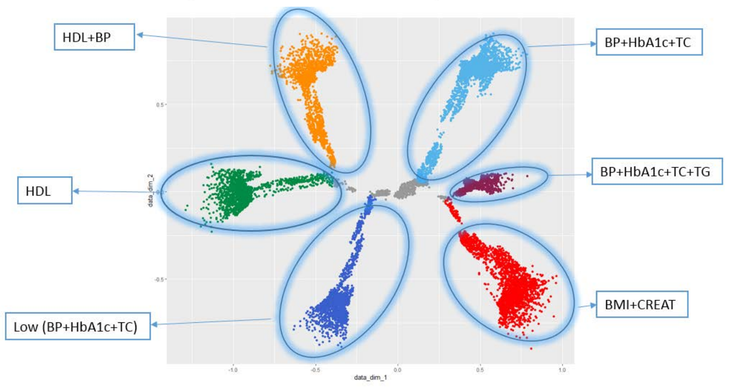
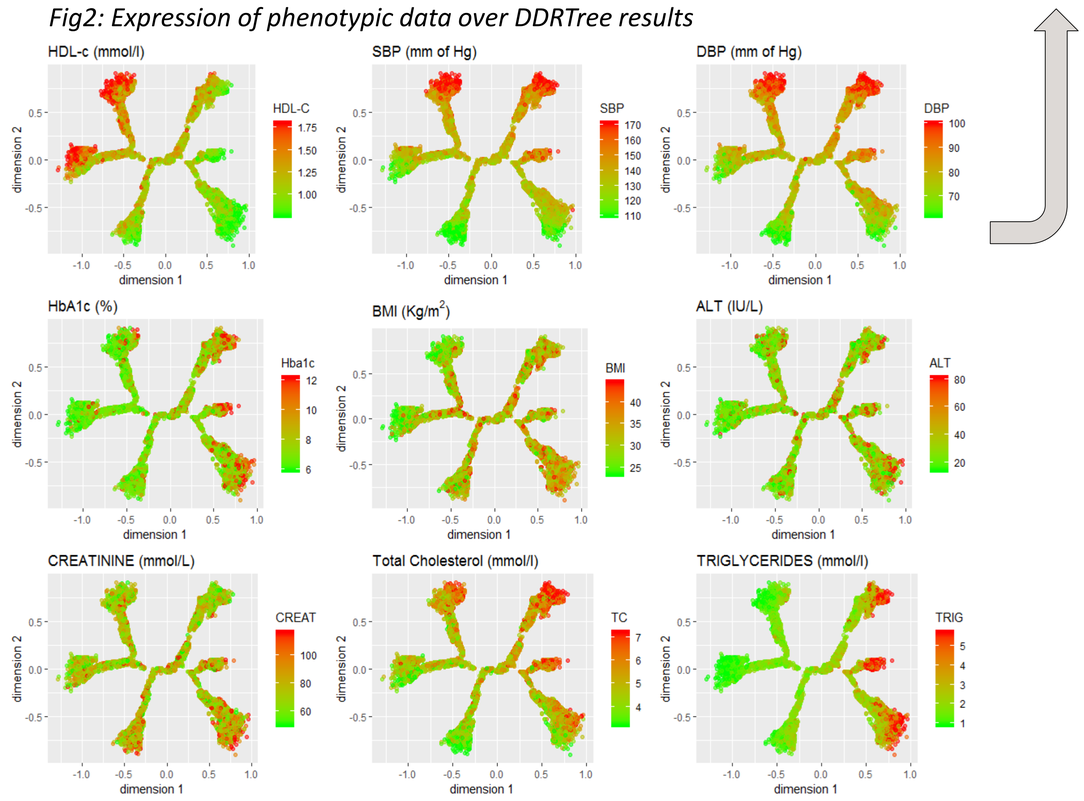


Fig2: Expression of phenotypic data over DDRTree results



Individuals from the branch with High HbA1c, Triglycerides and hypertension progressed fastest to insulin (Hazard Ratio (HR) 3.72, 95% CI 2.88-4.80 *vs* high HDL), whereas individuals from the branch with high cholesterol and highest blood pressure had higher risk for diabetic retinopathy (Hazard Ratio (HR) 1.34, 95% CI 1.27-1.62 *vs* high HDL) and major adverse cardiac event (Hazard Ratio (HR) 1.33, 95% CI 1.09-1.62 *vs* high HDL).

Discussion and conclusions

In this large data driven study on the phenotype of type 2 diabetes at diagnosis, adjusted for age and sex, individuals characterised by worse glycaemia at diagnosis progressed most rapidly to insulin; whilst individuals with elevated blood pressure and no protective elevation of HDL-cholesterol were more likely to develop diabetes related complications.

Future research that includes aetiology variables related to type 2 diabetes in the DDRTree analysis may provide more insights into heterogeneity of diabetes presentation and progression.

Acknowledgements

We would like to thank the INSPIRED research group and the Health Informatics Centre (HIC), Dundee for their continuous support.

This is a summary of independent research funded by National Institute for Health Research (NIHR)'s INSPIRED Programme. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

References

1.McCarthy MI. Painting a new picture of personalised medicine for diabetes. Diabetologia. 2017 May;60(5):793–799.

2. Qi Mao, Li Wang, Steve Goodison, and Yijun Sun. Dimensionality Reduction via Graph Structure Learning. The 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'15), 2015.

3. Cole Trapnell, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J. Lennon, Kenneth J. Livak, Tarjei S. Mikkelsen, and John L. Rinn. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature Biotechnology, 2014

FUNDED BY NIHR | National Institute for Health Research