Clustering of diabetes into novel subgroups provides improved prediction of outcome


1 Lund University Diabetes Centre, Lund University, Malmö, Sweden, 2 Primary Health Care, Vaasa Central Hospital, Vaasa, Finland, 3 Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden, 4 Finnish Institute for Molecular Medicine, Helsinki University, 5 Folkhälsan Research Center, Helsinki, Finland.

Background and aims: The current classification of diabetes into two main forms (T1D and T2D) has been useful in delineating T1D as an insulin-deficient form requiring insulin therapy but less useful for dissecting the heterogeneity of T2D. A refined classification could provide a powerful tool to identify those at greatest risk of complications already at diagnosis and tailor individualized treatment.

Materials and methods: We performed unsupervised clustering (k-means and hierarchical) based on six variables (age, BMI, HbA1c, HOMA2-B, HOMA2-IR and GAD auto-antibodies) in ANDIS, a Swedish cohort of 15,000 newly diagnosed diabetes patients. Disease progression, treatment and development of complications were followed using medical records and national registries. Risk of complications was calculated using Cox regression with the largest cluster as reference. Genetic loci known to be associated with T2D and related traits were analysed using MLE comparing each cluster to a non-diabetic cohort from the same geographical region.

Results: ANDIS patients clustered into one GADA-positive cluster (referred to as SAD), Severe Autoimmune Diabetes) and four GADA-negative clusters. This was replicated in three independent cohorts from Sweden and Finland. Cluster 2 (SIDD, Severe Insulin Deficient Diabetes; 17.5% of patients) was characterized by early onset, insulin deficiency and high HbA1c. During a mean follow-up of 4 years SIDD had higher HbA1c, was more likely to be prescribed insulin and develop diabetic retinopathy compared to other GADA-negative clusters. Cluster 3 (SIRD, Severe Insulin Resistant Diabetes; 15.5%) had the highest risk of kidney disease including CKD stage 3B (HR 3.30 [2.67-4.08], p=3.6x10^-15) and macroalbuminuria/end-stage renal disease (ESRD; HR 2.40 [1.69-3.42], p=9.8x10^-15). This was replicated in a cohort with longer follow-up (HR 5.04 [2.76-9.23], p=1.5x10^-14). Cluster 4 (MOD, Mild Obese Diabetes; 21.6%) and cluster 5 (MARD, Mild Age-Related Diabetes; 39.1%) showed only modest metabolic derangements. Using p<0.01 as cut-off, no genetic variant was associated with all clusters. Strikingly, the T2D-associated variant, rs7903146, in the TCF7L2 gene was associated with SIDD (OR 1.51 [1.33-1.71], p=2.8x10^-10), MOD (OR 1.38 [1.15-1.65], p=5.7x10^-7) and MARD (OR 1.41 [1.28-1.55], p=1.1x10^-10), but not with SIRD (OR 1.00 [0.87-1.15], p=0.86). A variant in IGF2BP2 (rs4402960) was associated with SIDD (OR 1.25 [1.08-1.40], p=2.0x10^-4) and MARD (OR 1.22 [1.11-1.33], p=2.1x10^-5), but not with SIRD (OR 1.01 [0.88-1.16], p=0.53) or MOD (OR 1.04 [0.92-1.18], p=0.31).

Conclusion: This study can be a first step towards a more precise, clinically useful, classification of diabetes representing an important step towards precision medicine in diabetes.

Supported by: Swedish Research Council, ERC, Novo Nordisk Foundation, ALF.

Disclosure: E. Ahlvist: None.