Clustering on baseline clinical variables identifies subgroups of type 2 diabetes patients with different rate of progression over 18 months: a DIRECT study

Brorsson, Caroline Anna; Pedersen, Helle Krogh; Gudmundsdottir, Valborg; Mari, A.; Kurbasic, A.; Vinuela, A.; Fernandez, J.; Mahajan, A.; Gupta, Ramneek; Dermitzakis, E.; MacCarthy, M.; Franks, P.; Pearson, E.; Brunak, Søren

Published in:
Diabetologia

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
249 Clustering of diabetes into novel subgroups provides improved prediction of outcome
E. Ahlvist¹, A. Kärjäjärvi³, M. Martelini³, P. Storm¹, M. Dorkhan¹, P. Vikman¹, R.B. Prasad¹, D. Mansour Aly¹, N. Shaat¹, E. Lindholm¹, T. Tuomilehto⁴, A.H. Rosengren¹, L. Groop¹-²
¹Lund University Diabetes Centre, Lund University, Malmö, Sweden, ²Primary Health Care, Vaasa Central Hospital, Vaasa, Finland, ³Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden, ⁴Finnish Institute for Molecular Medicine, Helsinki University, ⁵Folkhälsoas 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 analyzed 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 SAID, 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; 15.3%) had the highest risk of kidney disease including CKD stage 3B (HR 3.30[2.67-4.08], p=3.6x10⁻⁷) and macroalbuminuria/end-stage renal disease (ESRD; HR 2.40[1.69-3.42], p=9.8x10⁻⁶). This was replicated in a cohort with longer follow-up (SDR, mean 11 years), where SIRD had five-fold increased risk of ESRD (HR 5.04[2.76-9.23], p=1.5x10⁻⁵). 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[1.18-1.56], p=2.7x10⁻⁵) and MARD (OR 1.41[1.28-1.55], p=1.1x10⁻⁵), but not with SIRD (OR 1.00[0.87-1.15], p=0.86). A variant in IGBP2BP2 (rs4402960) was associated with SIDD (OR 1.23[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 help identify people presenting with complications at younger age has uncertain significance, but has not been

250 Clustering on baseline clinical variables identifies subgroups of type 2 diabetes patients with different rate of progression over 18 months: a DIRECT study
C.A. Brorsson¹, H. Krogh Pedersen¹, V. Gudmundsdottir¹, A. Mari², A. Kurbasie², A. Viinuela³, J. Fernandez³, A. Mahajan³, R. Gupta¹, E. Demirtzakis⁴, M. MacCarthy⁵, P. Franks⁶, E. Pearson⁶, S. Brunak¹,², for the DIRECT Consortium;
¹Department of Bioinformatics, Technical University of Denmark, Lyngby, Denmark, ²Institute of Neuroscience - National Research Council, Padova, Italy, ³Lund University, Malmö, Sweden, ⁴University of Geneva, Geneva, Switzerland, ⁵University of Oxford, Oxford, ⁶University of Dundee, Dundee, UK, ⁷University of Copenhagen, Copenhagen, Denmark.

Background and aims: Type 2 diabetes is characterized by a heterogeneous presentation with varying degrees of insulin resistance and β-cell failure. Taking advantage of the detailed clinical information collected for type 2 diabetes patients included in the DIRECT study, our aim was to characterize inter-individual heterogeneity and identify subgroups with different presentations of diabetic phenotypes at baseline, and investigate the effect on the rate of progression during follow-up. Genotyping data was used to evaluate genetic differences between subgroups.

Materials and methods: 836 newly diagnosed individuals were enrolled in the DIRECT study. The clustering was based on 20 clinical variables from the baseline visit consisting of anthropometric, biochemical and glycaemic modeling variables. We clustered the individuals using an unsupervised, agglomerative clustering method. Diabetes progression was assessed using individual HbA1c slopes obtained from a conditional linear mixed-model using data at 0, 9 and 18 month adjusted for age, gender, body mass index, smoking status and baseline HbA1c. The subgroups showed differences in the rate of progression and in the rate of development of complications were followed using medical records and national registries. Risk of complications was analyzed using cox regression with the largest cluster as reference. Genetic loci known to be associated with T2D and related traits were analyzed 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 SAID, 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; 15.3%) had the highest risk of kidney disease including CKD stage 3B (HR 3.30[2.67-4.08], p=3.6x10⁻⁷) and macroalbuminuria/end-stage renal disease (ESRD; HR 2.40[1.69-3.42], p=9.8x10⁻⁶). This was replicated in a cohort with longer follow-up (SDR, mean 11 years), where SIRD had five-fold increased risk of ESRD (HR 5.04[2.76-9.23], p=1.5x10⁻⁵). 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[1.18-1.56], p=2.7x10⁻⁵) and MARD (OR 1.41[1.28-1.55], p=1.1x10⁻⁵), but not with SIRD (OR 1.00[0.87-1.15], p=0.86). A variant in IGBP2BP2 (rs4402960) was associated with SIDD (OR 1.23[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 help identify people presenting with complications at younger age has uncertain significance, but has not been