Using clustering techniques and polygenic risk scores to classify Type 2 diabetes patients

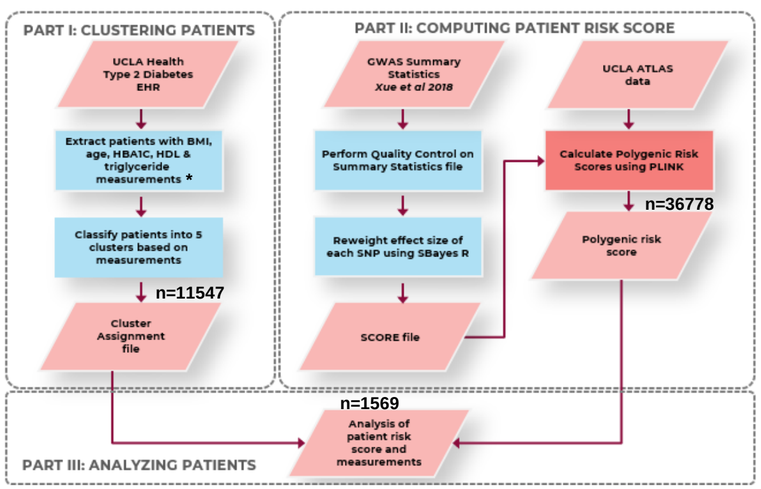
Aditi Kumar[1], Aditya Pimplaskar[2], Jeffrey N. Chiang[3,4]

Type 2 diabetes (T2D) accounts for over 90% of diabetes cases and is heterogeneous in terms of patient's symptoms, disease progression and response to treatment for the disease. Distinguishing T2D subtypes can be challenging. In contrast to clinical measurements, germline genetic data does not change and can be used to understand T2D heterogeneity. We used k-means clustering to classify UCLA patients into five clusters based on their onset age, BMI and triglyceride, HDL and %HBA1C levels. We extracted loci from an existing genome-wide association study (GWAS) and calculated a polygenic risk score (PRS) for each patient. On average patients of higher risk are more likely found in Cluster 5 but the highest risk patients were found in Cluster 4. Understanding the relationship between genetic risk and clinical measurements in T2D patients could improve our ability to understand and manage the disease.

Background

- Type 2 diabetes shows heterogeneity in symptoms, disease progression and response to disease treatment.
- Diagnosis of Type 2 diabetes is determined by eliminating other potential causes of high blood glucose.
- Germline genetic data does not change over time as it is unaffected by disease progression and treatment.
- Ahlqvist et al classified type 2 diabetes patients into 5 clusters based on 6 clinical measurements in Scandinavian patients (HOMA2-B, HOMA2-IR, onset age, HBA1C, BMI, fasting glucose).
- UCLA EHR for Type 2 diabetes patients includes BMI, onset age, fasting glucose, HBA1C, triglycerides, HDL measurements.
- We want to cluster patients based on clinical measurements and calculate genetic risk to determine relationship between clinical markers and genetic risk.

Methods



*Patients included in clustering had >= 1 measurement prior to diagnosis

Results

- Average patient in Cluster 5 had the lowest age of onset
- Average patient in Cluster 2 and average patient in Cluster 4 had the same genetic risk score but average patient in Cluster 4 had a lower age of onset
- Two clusters with overweight BMI and three clusters with obese BMI, one cluster with obese BMI, higher onset age, high triglyceride and low HDL level
- Highest proportion of patients in top decile were from Cluster 4

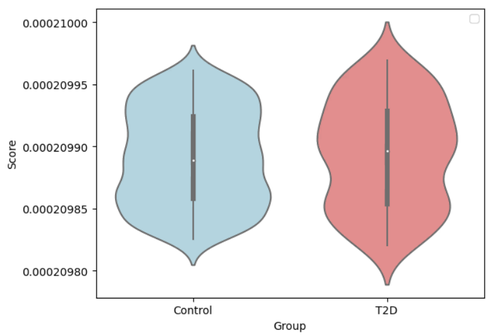


Fig. 1: Distribution of scores in control and T2D patients in the top decile of PRS

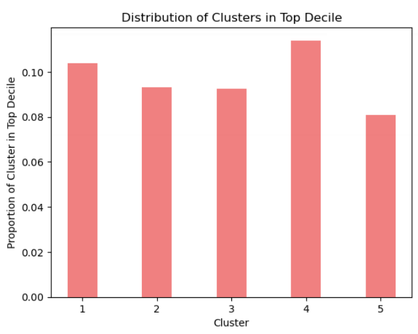


Fig. 2: Proportion of T2D patients from each cluster in the top decile of PRS

	White	Other/Unknown	Latino	Asian	Black	American Indian/Alaska Native	Pacific Islander	Age 0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	Female	Male
Control	45.283459	24.896485	20.586221	6.579808	2.202609	0.308210	0.143208	0.000127	0.000694	0.015202	0.046031	0.078187	0.113012	0.152709	0.187215	0.203364	0.203459	57.650758	42.349242
T2D	40.599108	20.777565	25.812620	7.584449	4.652645	0.446144	0.127470	0.000000	0.000000	0.014659	0.025494	0.064372	0.154238	0.272785	0.291268	0.176546	0.000637	45.188018	54.811982

Table 1: Patient Demographics (Control n=7889, T2D n=1569)

	BMI	CHOLHDL (mg/dL)	HGBA1C (%)	TRIGLY (mg/dL)	Onset Age (years)	SCORE
0	32.633029	48.726141	6.749710	169.658506	47.052532	0.000060
1	27.943495	57.061204	6.233378	118.118729	78.425629	0.000050
2	29.300466	52.880444	6.442238	141.213710	68.011743	0.000068
3	30.625879	50.372028	6.647762	153.608625	58.132567	0.000050
4	31.550354	47.432323	6.742727	179.212121	30.476602	0.000075

Table 2: Average patient in each cluster

	BMI	CHOLHDL (mg/dL)	HGBA1C (%)	TRIGLY (mg/dL)	Onset Age (years)	SCORE
0	32.070000	46.712500	7.245000	192.537500	47.565982	0.00021
1	27.665333	59.566667	6.221333	131.110000	78.161918	0.00021
2	28.774306	52.285714	6.458571	137.857143	67.327537	0.00021
3	31.818929	52.000000	6.596190	150.376190	59.393085	0.00021
4	31.347000	53.960000	7.532000	141.760000	28.072603	0.00021

Table 3: Average patient in each cluster in the top decile of PRS

Discussion

- On average high triglyceride level, low HDL level and high PRS could be associated with lower age of onset type 2 diabetes
- GWAS based on 62892 cases and 596424 controls of European ancestry
- White people constitute a higher proportion of patients in the control group (45.3%) than in the T2D group (40.6%)
- Only Type 2 diabetes patients with sequencing sample collected and at least 1 clinical measurement post diagnosis were included
- Time series measurements were averaged
- Future directions: incorporating time series measurements and patient medications

Acknowledgements

We would like to thank Callahan Clark for her feedback on this research

1 BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
2 Bioinformatics Interdepartmental PhD Program, UCLA
3 Department of Neurosurgery, David Geffen School of Medicine, UCLA
4 Department of Computational Medicine, David Geffen School of Medicine, UCLA

References

Ahlqvist, Emma, et al. Novel Subgroups of Adult-Onset Diabetes and Their Association With Outcomes: A Data-Driven Cluster Analysis of Six Variables. *The Lancet.* https://doi.org/10.1016/S2213-8587(18)30051-2.

Kullo, I.J., Lewis, C.M., Inouye, M. *et al.* Polygenic scores in biomedical research. *Nat Rev Genet* **23**, 524–532 (2022). https://doi.org/10.1038/s41576-022-00470-z

Udler, M.S. Type 2 Diabetes: Multiple Genes, Multiple Diseases. *Curr Diab Rep* **19**, 55 (2019). https://doi.org/10.1007/s11892-019-1169-7

Udler, Miriam S., et al. Type 2 Diabetes Genetic Loci Informed by Multi-Trait Associations Point to Disease Mechanisms and Subtypes: A Soft Clustering Analysis. *PLOS Medicine.* https://doi.org/10.1371/journal.pmed.1002654

Xue A, et al. (2018). Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nature Communications. 2018 (accepted).