Novel subgroups of adult-onset diabetes and their association with outcomes: A data-driven cluster analysis of six variables

Article *in* The Lancet Diabetes & Endocrinology · March 2018

DOI: 10.1016/S2213-8587(18)30051-2

CITATIONS
1,983

READS
10,119

24 authors, including:

Emma Ahlqvist
Lund University
194 PUBLICATIONS **9,086** CITATIONS
SEE PROFILE

Annemari Käräjämäki
Wellbeing services county of Ostrobothnia
48 PUBLICATIONS **5,225** CITATIONS
SEE PROFILE

Mats Martinell
Uppsala University
54 PUBLICATIONS **3,093** CITATIONS
SEE PROFILE

Mozhgan Dorkhan
Lund University
48 PUBLICATIONS **12,549** CITATIONS
SEE PROFILE

THE LANCET
Diabetes & Endocrinology

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Ahlqvist E, Storm P, Käräjämäki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018; published online Feb 28. http://dx.doi.org/10.1016/S2213-8587(18)30051-2.

Clustering of adult-onset diabetes into novel subgroups guides therapy and improves projection of outcome

Supplementary figures

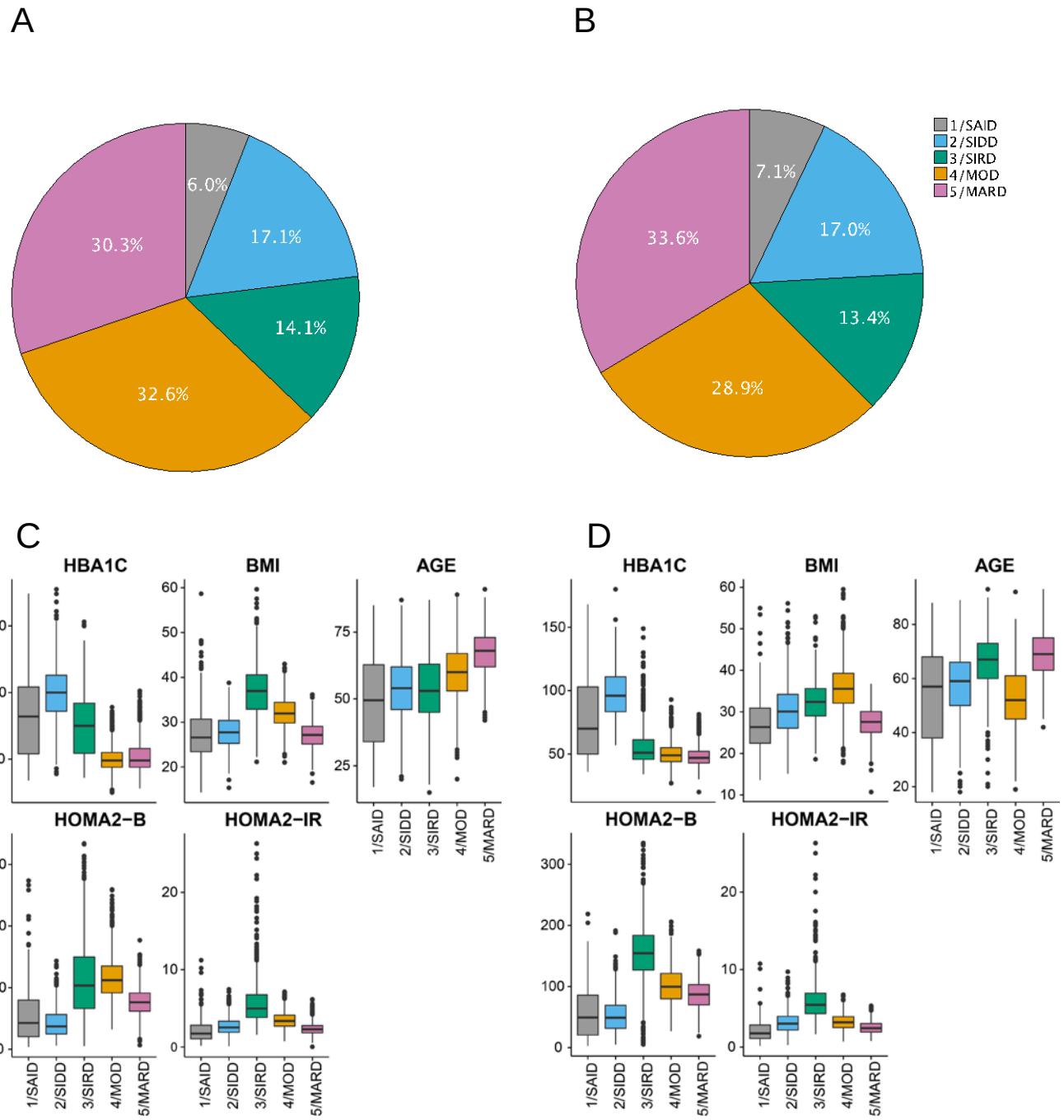

Figure S1. Results of sex-specific TwoStep clustering in ANDIS.

The distribution of patients based on TwoStep cluster analysis in the ANDIS (All New Diabetes in Scania) cohort was similar in (A) men and (B) women. The different clusters in men (C) and women (D) had similar distribution of the variables used for clustering, i.e. of HAa1c (mmol/mol) at diagnosis; BMI (kg/m^2), age (years), HOMA2-B (%) and HOMA2-IR at registration, except for some differences in clusters 3 and 4 (women in cluster 4 were older and less obese compared to men in cluster 3).

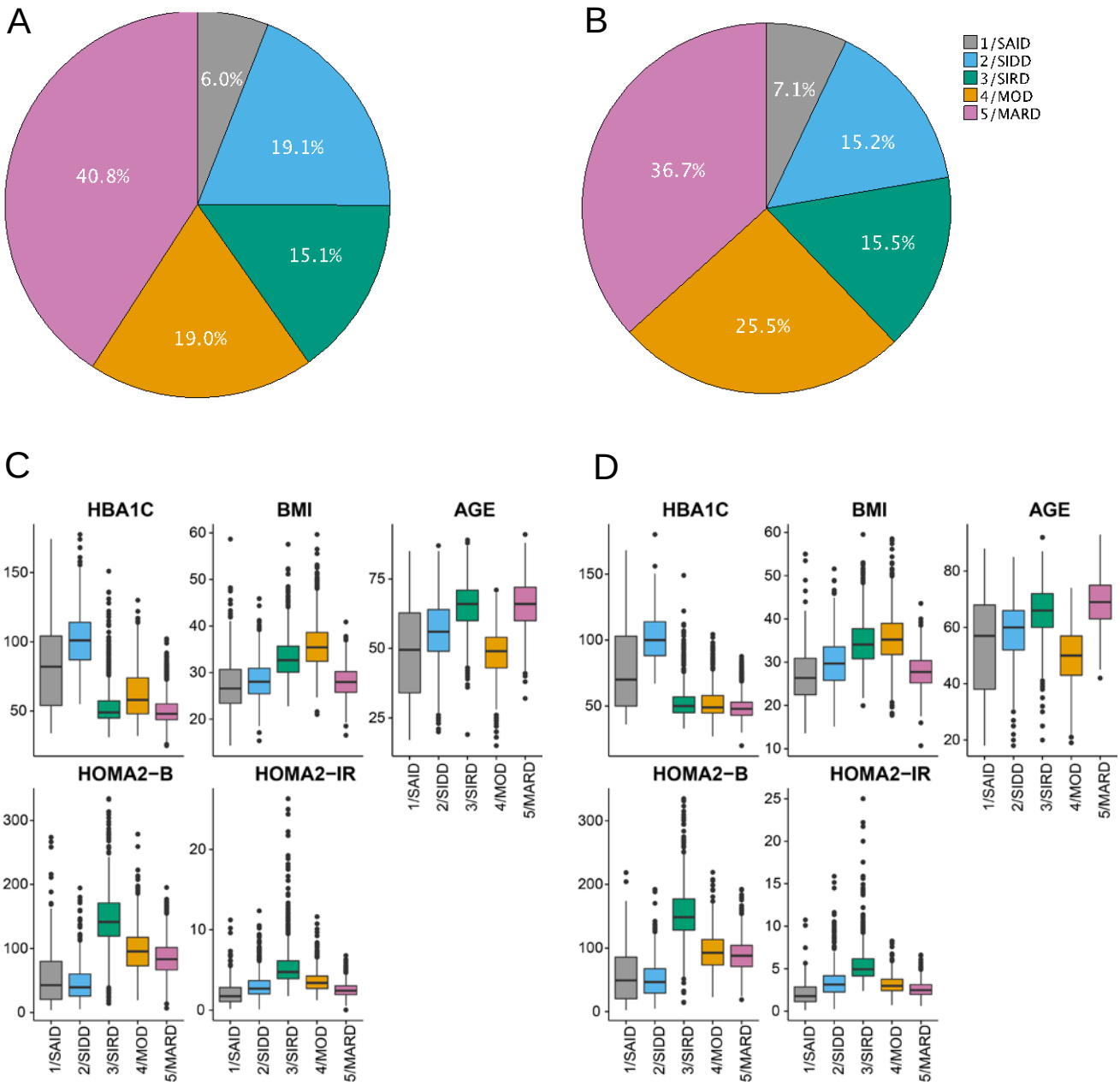

Figure S2. Results of sex-specific k-means clustering in ANDIS.

The distribution of patients based on k-means cluster analysis in ANDIS was similar in men (A) and women (B). The different clusters in men (C) and women (D) had similar distribution of the variables used for clustering, i.e. of Hba1c (mmol/mol) at diagnosis; BMI (kg/m^2), age (years), HOMA2-B (%) and HOMA2-IR at registration.

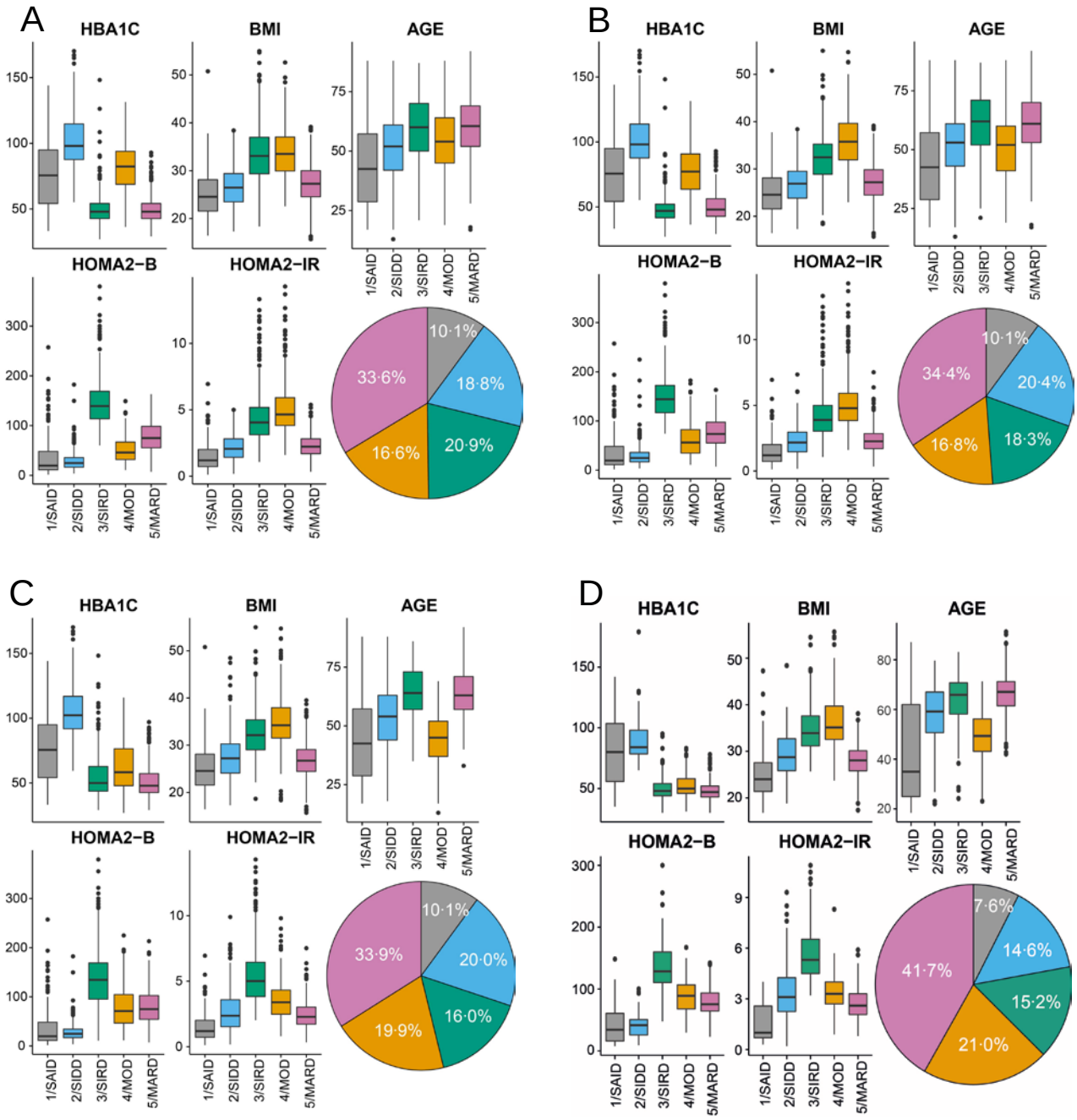

Figure S3. Replication of clustering in two independent Swedish cohorts, SDR and ANDIU.

Cluster distributions and characteristics in the Scania Diabetes Registry (SDR) using TwoStep (A), k-means (k=4) clustering (B) and cluster assignment based on ANDIS cluster coordinates (C). Cluster distributions and characteristics in ANDIU using k-means (k=4) clustering (D) and cluster assignment based on ANDIS cluster coordinates (E).

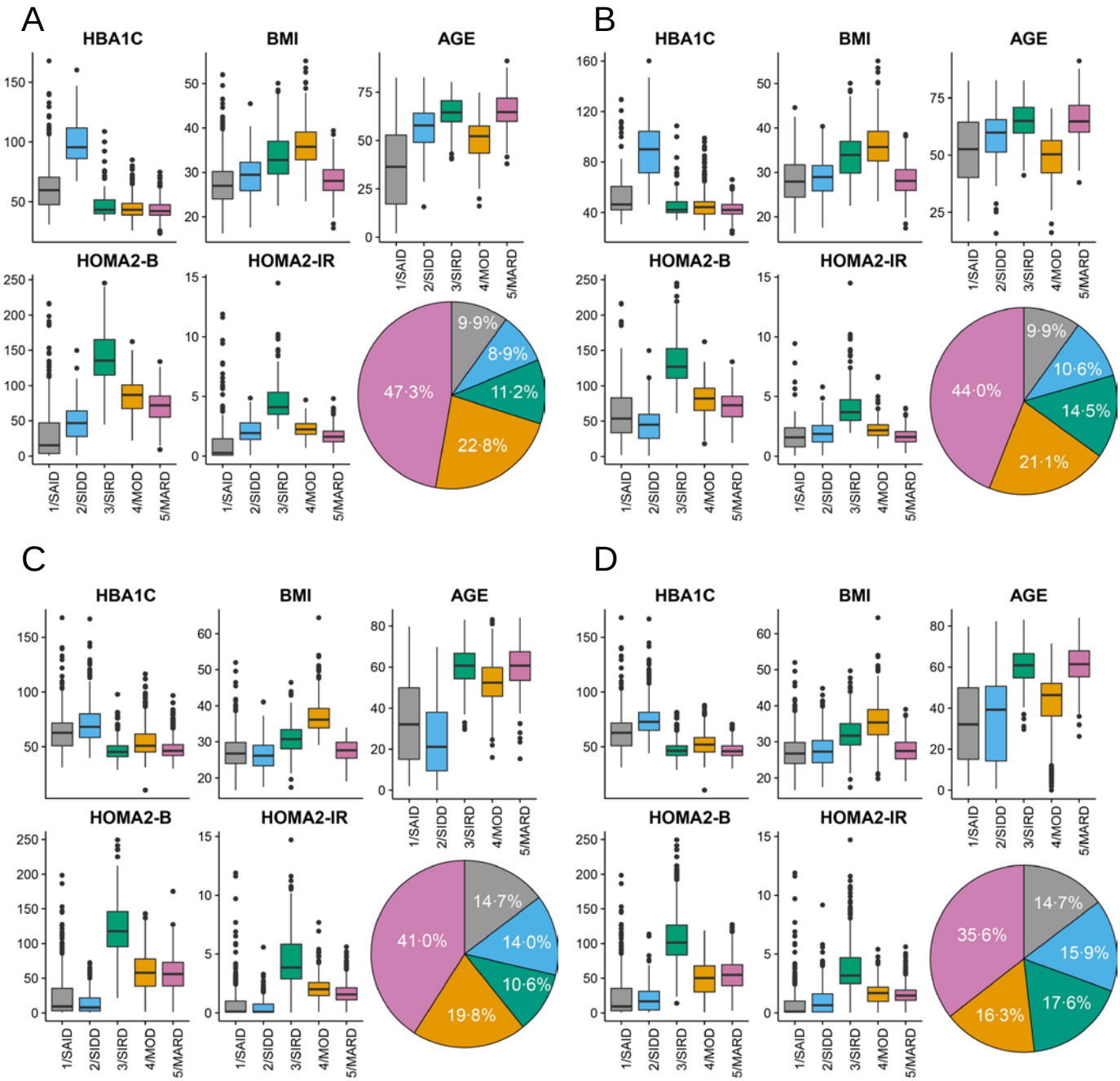

Figure S4. Clustering in the Finnish DIREVA cohort comparing patients with newly diagnosed diabetes and longer duration.

Patient distribution in DIREVA in newly diagnosed (diabetes duration at sampling less than 2 years) using de novo k-means clustering (A), and cluster assignment based on ANDIS cluster coordinates (B). Patient distribution in patients with longer duration at sampling (mean 10·15±10·34 years) using de novo k-means clustering (C) and cluster assignment based on ANDIS cluster coordinates (D).

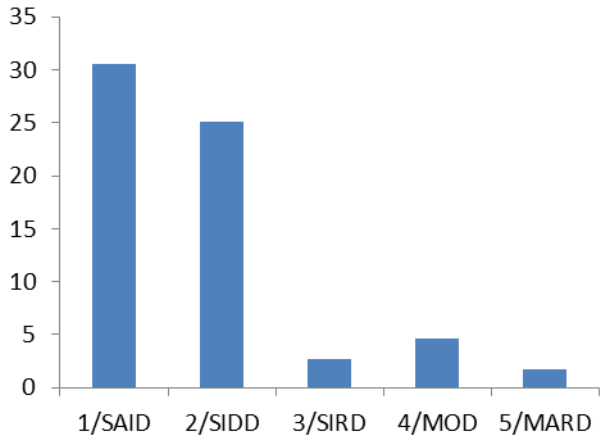

Figure S5. Prevalence of ketoacidosis at diagnosis in ANDIS.

Ketoacidosis at diagnosis was most common in clusters 1/SAID (124 of 406; 30·5%) and 2/SIDD (259 of 1033;25·1%), but only occurred in 23 of 842 SIRD patients, 56 of 1215 of MOD patients and 38 of 2273 MARD patients.

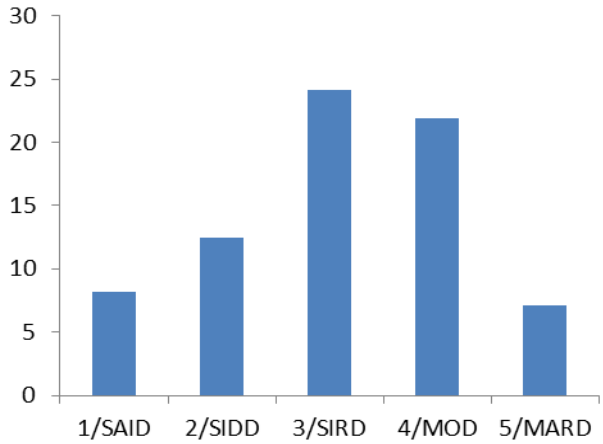

Figure S6. Prevalence of non-alcoholic fatty liver disease (NAFLD) in ANDIS estimated from ALT measurements.

A total of 11999 patients had at least one P-ALT measurement in the database of the Clinical Chemistry unit. Of them, 3739 had at least two readings exceeding the upper reference values for the assays used (>1·1-1·2 µkat/L for men and >0·7-0·85 µkat/L for women). Cluster 3/SIRD had the highest prevalence of NAFLD (302 of 1556; 24·1%), defined as two pathological ALT measurements and BMI>28 (OR 3·96[3·27-4·78], $p=5{\cdot}8x10^{-46}$ compared to MARD (226 of 3415; 7·1%) and OR 1·56[1·24-1·95], $p=1{\cdot}4x10^{-4}$ compared to MOD (381 of 2120; 21.9%) after adjustment for sex and age).

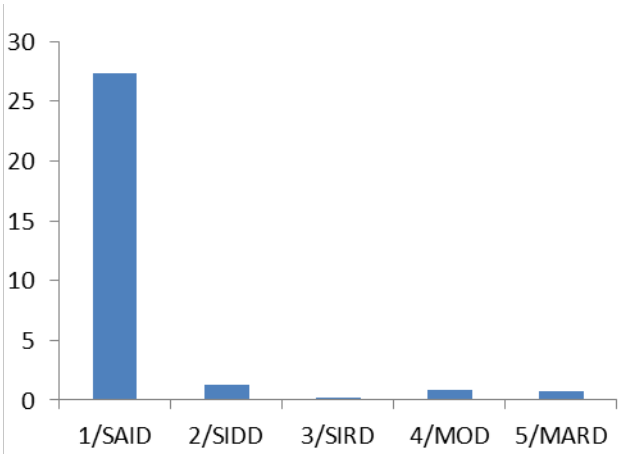

Figure S7. Prevalence of auto-antibodies directed against zinc transporter 8A (ZnT8A).

ZnT8A auto-antibody positivity was mainly observed in patients from cluster 1/SAID (79 of 289 positive; 27·3%). OR 52·74[26·89-103·45], $p=8{\cdot}6x10^{-31}$ compared to MARD (10 of 1412 positive). Importantly only 9 of 685 (1·3%) SIDD patients were positive.

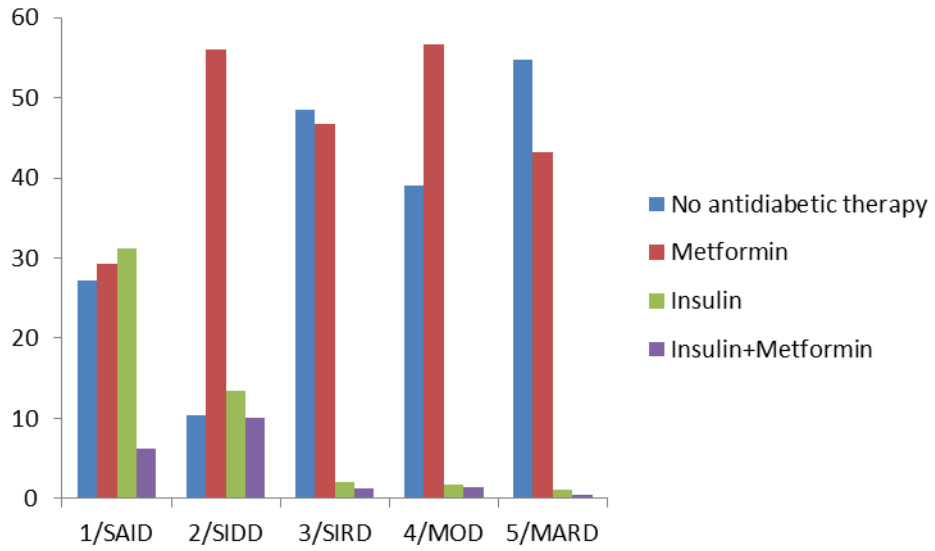

Figure S8. Antidiabetic medication (%) at time of registration in ANDIS.

Figure S8 shows the frequency of use of at least metformin, at least insulin or both at registration in ANDIS (at the time of measurement of plasma glucose and C-peptide) stratified by cluster. More patients in cluster 2/SIDD had been prescribed insulin and/or metformin than in clusters 3-5, reflecting the higher HbA1c at diagnosis.

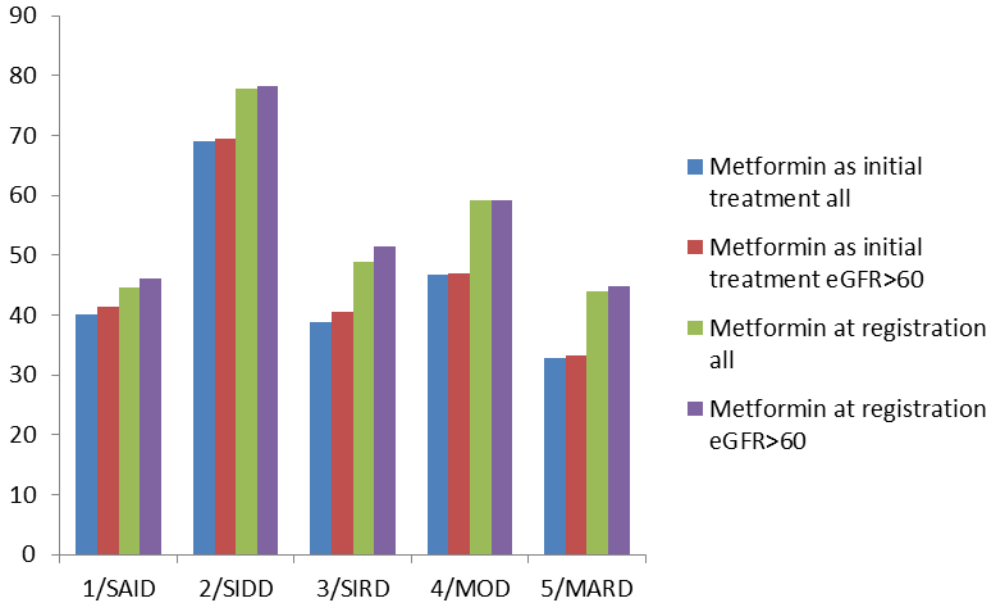

Figure S9. Metformin treatment (%) by cluster in ANDIS.

The percentage of patients in each cluster prescribed metformin as their first treatment after diagnosis (initial treatment), and percent of renally sufficient (eGFR>60 mL/min/1·73m^2) patients on metformin at registration. This shows that the difference between clusters is not a result of discontinuation of metformin due to adverse effects or contra-indication of metformin in patients with kidney disease.

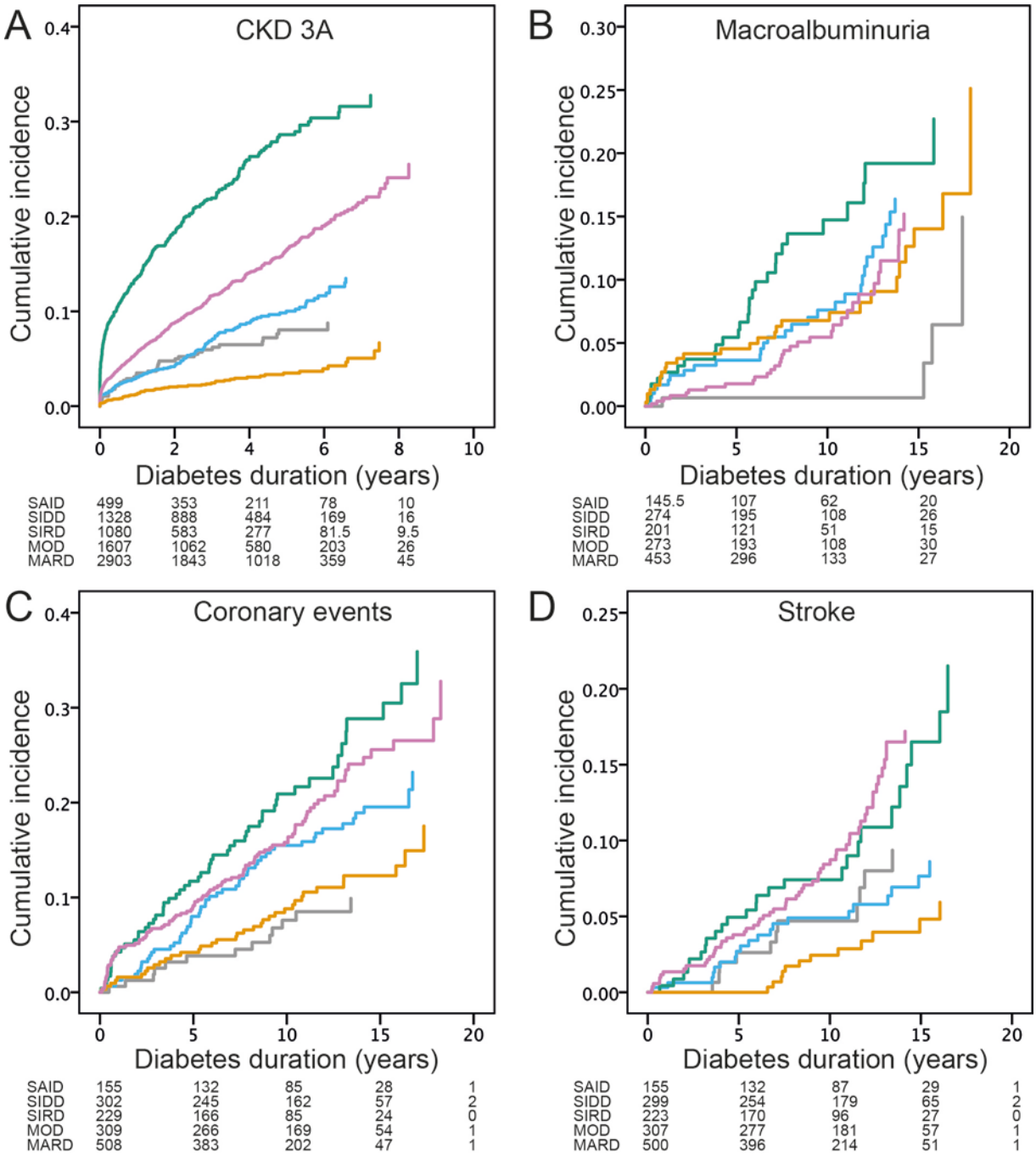

Figure S10. Risk of diabetic complications by cluster.

Cox regressions of diabetic complications for (A) CKD stage 3A (eGFR <60 ml/min) in ANDIS and macroalbuminuria (B) in SDR, coronary events (C) and stroke (D) in SDR. Tables show number of individuals at risk. For statistics see Table S7 and S11.

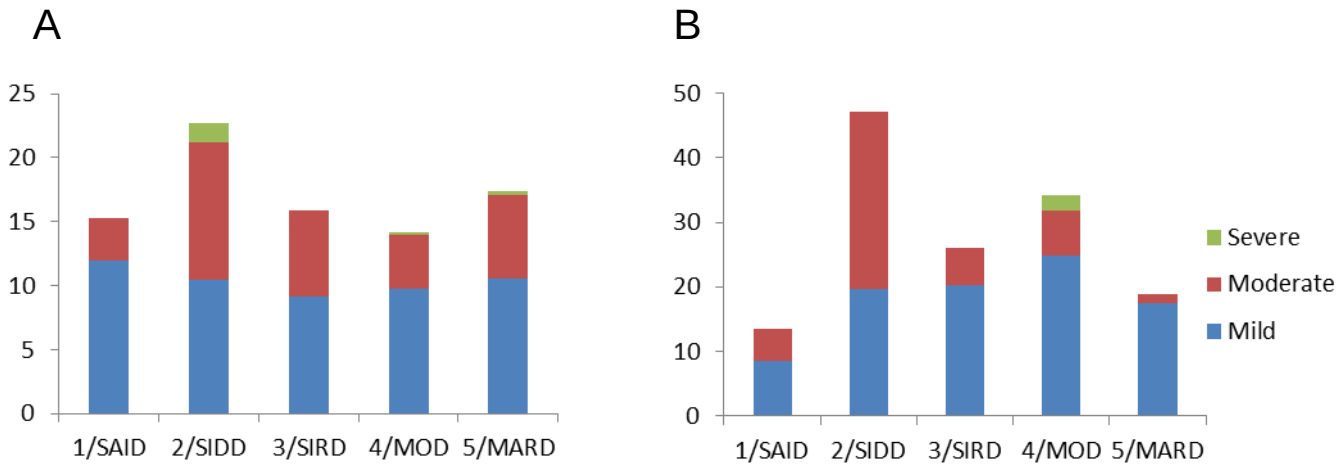

Figure S11. Diabetic retinopathy (DR) at diagnosis.

Prevalence of different stages of diabetic retinopathy in (A) ANDIS and (B) ANDIU. In ANDIS DR risk was significantly higher in SIDD than in the reference cluster MARD (OR 1·6[1·3-1·9], $p=9·7x10^{-7}$). In ANDIU DR risk was elevated in both SIDD (OR 4·6[3·0-7·0], $p=4·1x10^{-13}$) and MOD compared to MARD (OR 2·2[1·4-3·3], $p=2·8x10^{-4}$).

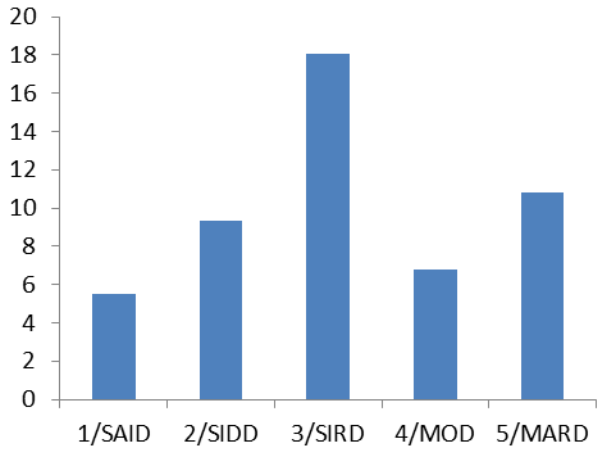

Figure S12. Prevalence of chronic kidney disease in DIREVA.

CKD (eGFR<60 ml/min) in DIREVA by clusters assigned based on ANDIS cluster coordinates. SIRD had increased risk of CKD (OR 2·02[1·38-2·96], $p=3·4x10^{-4}$) after adjustment for age, sex and duration of diabetes. Median follow-up time was 8·53 years (IQR 4·88-14·63).

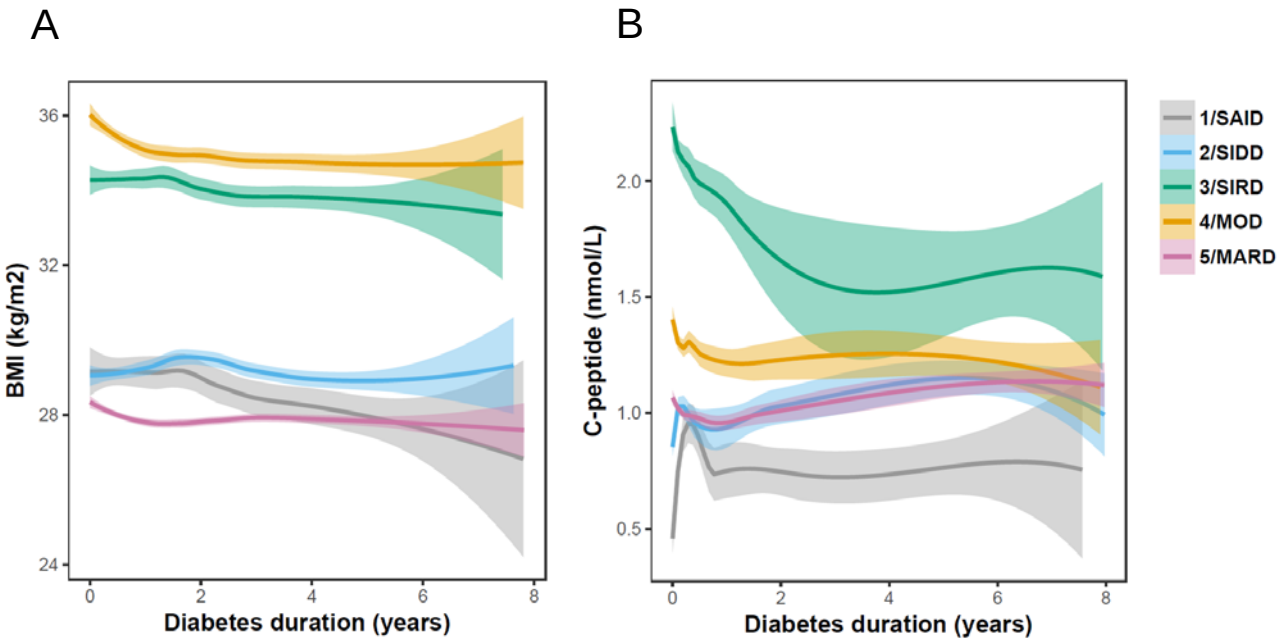

Figure S13. Change in cluster variables over time in ANDIS.

Figure S13 shows change in BMI (A) and C-peptide (B) during follow-up by cluster.

Supplementary tables

Table S1. Patient characteristics in ANDIS using the traditional classification.*

	T1D	LADA	T2D
N	204	723	12112
Frequency in total cohort, %	1·5	5·3	88·3
Men, %	64·7	54·1	59·7
Hba1c at diagnosis, mmol/l	107·88(27·06)	73·65(28·26)	62·32(24·12)
BMI, kg/m^2	22·02(3·50)	28·77(6·32)	30·93(5·72)
Age at diagnosis, years	34·03(13·84)	54·86(12·67)	60·93(12·25)
HOMA2-B	23·50(20·94)	65·46(46·88)	91·95(48·19)
HOMA2-IR	0·66(0·35)	2·64(2·08)	3·41(2·55)
Insulin at registration, %	96·5	33·4	9·9
Metformin at registration, %	10·2	43·8	52·0
History of gestational diabetes, % of women	3·9	11·9	11·3
Non-Scandinavian origin, %	12·5	16·8	23·2

*Only patients older than 18 at registration are included.

Table S2. Patient characteristics in ANDIS using k-means clustering.

	SAID	SIDD	SIRD	MOD	MARD
N	577	1575	1373	1942	3513
Frequency, %	6·4	17·5	15·3	21·6	39·1
Men, %	55·1	64·8	58·8	52·1	62·0
Hba1c at diagnosis, mmol/l	80·03(30·84)	101·85(19·26)	54·07(15·46)	57·70(16·07)	50·08(9·85)
BMI, kg/m^2	27·45(6·44)	28·86(4·77)	33·85(5·24)	35·71(5·43)	27·94(3·44)
Age at diagnosis, years	50·48(17·93)	56·74(11·14)	65·25(9·34)	48·96(9·54)	67·37(8·55)
HOMA2-B	56·71(44·65)	47·64(28·93)	150·47(47·20)	95·03(32·45)	86·59(26·37)
HOMA2-IR	2·16(1·56)	3·18(1·73)	5·54(2·74)	3·35(1·21)	2·55(0·84)
Insulin at registration, %	41·9	29·1	3·7	3·3	1·6
Metformin at registration, %	44·7	77·8	48·8	59·1	44·0
Family history of diabetes, %	59	64	56	70	58
History of gestational diabetes, % of women	10·3	7·5	4·5	21·7	5
Non-Scandinavian origin, %	15·4	26·6	15·1	32·3	16·8

Table S3. Cluster center coordinates in ANDIS.

	Cluster	Hba1c	BMI	Age at onset	HOMA2-B	HOMA2-IR
Women	2/SIDD	1·8702613	-0·2415449	-0·1929637	-0·97446899	0·056469
	3/SIRD	-0·254848	0·5189057	0·3214557	1·35581907	1·1801933
	4/MOD	-0·3003478	0·6683606	-0·9388278	-0·03556857	-0·1405151
	5/MARD	-0·4582762	-0·5854255	0·5980588	-0·14552652	-0·4254893
Men	2/SIDD	1·52185804	-0·4284673	-0·4017103	-0·98397328	-0·1630751
	3/SIRD	-0·39080167	0·5396294	0·4235841	1·29059153	1·1801031
	4/MOD	-0·06915764	1·0305317	-1·0157681	0·15742215	0·1343923
	5/MARD	-0·5367578	-0·4776681	0·5031031	-0·09004338	-0·4233873

Table S4. Classifier performance for cluster assignment based on Euclidian distance to ANDIS cluster centers.

Cluster	Sensitivity	Specificity
ANDIU		
SIDD	1	0·98
SIRD	0·96	0·98
MOD	0·89	0·99
MARD	0·95	0·98
SDR		
SIDD	0·84	0·96
SIRD	0·65	0·94
MOD	0·56	0·86
MARD	0·86	0·92
DIREVA short duration		
SIDD	0·92	0·97
SIRD	0·96	0·95
MOD	0·83	0·97
MARD	0·91	0·98
DIREVA long duration		
SIDD	0·76	0·95
SIRD	0·64	0·92
MOD	0·78	0·92
MARD	0·95	0·91

Table S5. Logistic regression analysis of risk factors for ketoacidosis at diagnosis in ANDIS.

	OR(CI95%)*	P
HOMA2-B	0·66(0·56-0·77)	$1·1x10^{-7}$
HOMA2-IR	0·95(0·81-1·12)	0·5457
Age at onset	0·65(0·59-0·72)	$1·5x10^{-17}$
BMI	0·94(0·84-1·05)	0·2845
HbA1c	2·73(2·47-3·03)	$2·0x10^{-82}$

*OR are for 1 SD change in variable.

Table S6a. Cox regression analysis comparing antidiabetic treatments between clusters in ANDIS.

	Events	Censored	% events	HR(CI95%)	P
Sustained insulin					
1/SAID	260	194	57·3	26·87(21·17-34·11)	$4{\cdot}3x10^{-161}$
2/SIDD	381	908	29·6	10·97(8·73-13·77)	$2{\cdot}5x10^{-94}$
3/SIRD	73	1121	6·1	2·03(1·49-2·76)	$7{\cdot}0x10^{-6}$
4/MOD	97	1539	5·9	1·90(1·43-2·53)	$1{\cdot}0x10^{-5}$
5/MARD	92	2806	3·2	1	-
Metformin					
1/SAID	267	177	60·1	0·83(0·73-0·95)	0·0053
2/SIDD	1152	81	93·4	2·56(2·38-2·76)	$2{\cdot}7x10^{-138}$
3/SIRD	859	295	74·4	1·17(1·08-1·27)	$8{\cdot}5x10^{-5}$
4/MOD	1402	185	88·3	1·67(1·56-1-79)	$1{\cdot}8x10^{-48}$
5/MARD	2007	826	70·8	1	-
Oral treatment other than metformin					
1/SAID	36	419	7·9	1·00(0·70-1·42)	0·9871
2/SIDD	339	950	26·3	3·99(3·36-4·74)	$3{\cdot}0x10^{-56}$
3/SIRD	162	1034	13·5	2·10(1·71-2·58)	$1{\cdot}1x10^{-12}$
4/MOD	264	1372	16·1	2·36 (1·97-2·83)	$2{\cdot}1x10^{-20}$
5/MARD	212	2687	7·3	1	-
Reaching treatment goal					
1/SAID	347	109	76·1	0·51(0·46-0·57)	$5{\cdot}5x10^{-32}$
2/SIDD	792	499	61·3	0·33(0·30-0·36)	$1{\cdot}3x10^{-162}$
3/SIRD	1051	146	86·9	0·90(0·84-0·97)	0·0055
4/MOD	1355	284	82·7	0·71(0·67-0·76)	$1{\cdot}3x10^{-23}$
5/MARD	2622	283	88·8	1	-

The table information from the Swedish Drug Prescription Registry of patient's pick-up of medication from the pharmacy. Sustained insulin is defined as more than 6 months on insulin treatment. Reaching treatment goal is defined as the first time point after diagnosis the HbA1c is below 52 mmol/mol.

Table S6b. P-values for pairwise comparisons of cox-regressions in table 6a.

	SAID vs SIDD	SAID vs SIRD	SAID vs MOD	SIDD vs SIRD	SIDD vs MOD	SIRD vs MOD
Sustained insulin	<0·0001*	<0·0001*	<0·0001*	<0·0001*	<0·0001*	<0·0001*
Metformin	<0·0001*	<0·0001*	<0·0001*	0·0002*	<0·0001*	<0·0001*
Oral treatment other than metformin	<0·0001*	0·0001*	<0·0001*	0·9748	<0·0001*	<0·0001*
Reaching treatment goal	<0·0001*	<0·0001*	<0·0001*	<0·0001*	<0·0001*	<0·0001*

*Significant after adjustment for multiple comparisons (10 tests).

Table S7a. Cox regression analysis comparing risk of kidney complications in ANDIS.

		Events	Censored	Events (%)	HR(CI95%)[1]	p[1]	HR(CI95%)[2]	p[2]
CKD 3A	1/SAID	37	534	6·5	0·91(0·65-1·27)	0·5691	0·86(0·61-1·20)	0·3646
	2/SIDD	123	1411	8·0	1·34(1·09-1·64)	0·0058	1·23(1·00-1·51)	0·0553
	3/SIRD	298	1037	22·3	2·41(2·08-2·79)	$1·4x10^{-31}$	1·56(1·34-1·82)	$6·4x10^{-9}$
	4/MOD	55	1828	2·9	1·17(0·86-1·60)	0·3102	1·06(0·77-1·44)	0·7307
	5/MARD	454	2950	13·3	1	-	1	-
CKD 3B	1/SAID	11	557	1·9	1·01(0·55-1·87)	0·9749	0·92(0·49-1·70)	0·7805
	2/SIDD	42	1492	2·7	1·73(1·21-2·48)	0·0029	1·21(0·83-1·76)	0·3196
	3/SIRD	118	1217	8·8	3·34(2·59-4·30)	$8·3x10^{-21}$	1·86(1·44-2·41)	$3·0x10^{-6}$
	4/MOD	19	1821	1·0	1·73(1·01-2·96)	0·0452	1·49(0·86-2·57)	0·1512
	5/MARD	128	3270	3·8	1	-	1	-
Macro-albuminuria	1/SAID	13	379	3·3	2·47(1·30-4·69)	0·0057	2·52(1·37-4·79)	0·0047
	2/SIDD	41	970	4·1	2·53(1·63-3·94)	3·7x10-5	2·71(1·74-4·20)	$9·0x10^{-6}$
	3/SIRD	46	770	5·6	2·89(1·92-4·35)	$3·4x10^{-7}$	2·45(1·62-3·71)	$2·2x10^{-5}$
	4/MOD	29	1077	2·6	2·31(1·35-3·95)	0·0023	2·19(1·28-3·75)	0·0041
	5/MARD	47	1964	2·3	1	-	1	-
ESRD	1/SAID	3	482	0·6	1·16(0·35-3·84)	0·8057	1·07(0·33-3·55)	0·9076
	2/SIDD	16	1438	1·1	2·58(1·36-4·91)	0·0038	2·68(1·42-5·07)	0·0025
	3/SIRD	25	1306	1·9	3·12(1·82-5·36)	$3·8x10^{-5}$	2·00(1·15-3·49)	0·0147
	4/MOD	8	1313	0·6	2·37(0·96-5·83)	0·0614	2·04(0·82-5·05)	0·1231
	5/MARD	28	3203	0·9	1	-	1	-

[1] Cox regressions adjusted for sex and age at onset.

[2] Cox regression adjusted for sex, age at onset and first eGFR.

CKD3A was defined as eGFR <60 mL/min/1·73m^{2}and CKD3B as eGFR <45 mL/min/1·73m^{2}for more than 90 days. End-stage renal disease (ESRD) was defined as at least one eGFR below 15 mL/min/1·73m^2. Macroalbuminuria was defined as at least two out of three consecutive visits with albumin excretion rate (AER) ≥200 µg/min or AER ≥300 mg/24 h or albumin-creatinine ratio (ACR) ≥25/35 mg/mmol for men/women. Median duration at first eGFR 53(IQR 1-286) days.

Table S7b. P-values for pairwise comparisons of cox-regressions in table 7a adjusted for sex and age at onset of diabetes.

	SAID vs SIDD	SAID vs SIRD	SAID vs MOD	SIDD vs SIRD	SIDD vs MOD	SIRD vs MOD
CKD 3A	0·1402	<0·0001*	0·4460	<0·0001*	0·5171	<0·0001*
CKD 3B	0·2125	0·0003*	0·1947	0·0002*	0·7823	0·0205
Macroalbuminuria	0·7615	0·4761	0·7744	0·5351	0·9938	0·5935
ESRD	0·2077	0·1455	0·2867	0·7674	0·9141	0·7479

*Significant after adjustment for multiple comparisons (10 tests).

Table S7c. P-values for pairwise comparisons of cox-regressions in table 7a adjusted for sex, age at onset of diabetes and first eGFR.

	SAID vs SIDD	SAID vs SIRD	SAID vs MOD	SIDD vs SIRD	SIDD vs MOD	SIRD vs MOD
CKD 3A	0·1584	0·0016*	0·5548	0·0123	0·4314	0·0096
CKD 3B	0·5777	0·0356	0·2474	0·0142	0·3755	0·4469
Macroalbuminuria	0·6627	0·8675	0·9308	0·6989	0·6443	0·9199
ESRD	0·1511	0·3658	0·3294	0·2968	0·6293	0·7815

*Significant after adjustment for multiple comparisons (10 tests).

Table S8a. Cox regression analysis comparing risk of cardiovascular disease between clusters in ANDIS.

		Events	Censored	Events (%)	HR(CI95%)[1]	p[1]	HR(CI95%)[2]	p[2]
CE	1/SAID	20	531	3·6	0·69(0·48-1·00)	0·0476	0·86(0·60-1·23)	0·4031
	2/SIDD	73	1417	4·9	0·95(0·76-1·17)	0·6134	0·98(0·79-1·22)	0·8730
	3/SIRD	80	1067	7·0	1·60(1·30-1·97)	$1·2x10^{-5}$	1·20(0·96-1·49)	0·1047
	4/MOD	60	1749	3·3	0·72(0·57-0·91)	0·0049	1·08(0·84-1·40)	0·5418
	5/MARD	196	2797	6·5	1	-	1	-
Stroke	1/SAID	8	538	1·5	0·61(0·34-1·08)	0·099	0·74(0·42-1·32)	0·3063
	2/SIDD	42	1421	2·9	1·27(0·93-1·73)	0·1295	1·35 (0·99-1·83)	0·0593
	3/SIRD	37	1052	3·4	1·68(1·22-2·30)	0·0015	1·17(0·85-1·63)	0·3344
	4/MOD	14	1769	0·8	0·36(0·23-0·57)	$9·0x10^{-6}$	0·64(0·40-1·03)	0·0661
	5/MARD	137	2748	4·7	1	-	1	-

[1] Cox regression adjusted for sex.

[2] Cox regressions adjusted for sex and age at onset.

Coronary events (CE) were defined by ICD-10 codes I21, I20, I251, I253, I254, I255, I256, I257, I258, I259. Stroke was defined by ICD-10 codes I60, I61, I63 and I64.

Table S8b. P-values for post hoc pairwise comparisons of cox-regressions in table 8a adjusted for sex.

	SAID vs SIDD	SAID vs SIRD	SAID vs MOD	SIDD vs SIRD	SIDD vs MOD	SIRD vs MOD
CE	0·2208	0·0009*	0·8973	0·0012*	0·1142	<0·0001*
Stroke	0·2661	0·0457	0·0271	0·1801	<0·0001*	<0·0001*

*Significant after adjustment for multiple comparisons (10 tests).

Table S8c. P-values for post hoc pairwise comparisons of cox-regressions in table 8a adjusted for sex and age.

	SAID vs SIDD	SAID vs SIRD	SAID vs MOD	SIDD vs SIRD	SIDD vs MOD	SIRD vs MOD
CE	0·5919	0·1850	0·3780	0·2269	0·5981	0·5813
Stroke	0·4180	0·6918	0·2652	0·5399	0·0217	0·0739

*Significant after adjustment for multiple comparisons (10 tests).

Table S9a. Cox regression analysis comparing risk of kidney complications between clusters in SDR.

		Events	Censored	Events (%)	HR(CI95%)[1]	p[1]	HR(CI95%)[2]	p[2]
CKD 3A	1/SAID	24	140	14·6	1·08(0·70-1·66)	0·7414	1·01(0·65-1·56)	0·9710
	2/SIDD	85	235	26·6	1·50(1·14-1·96)	0·0035	1·13(0·85-1·49)	0·3961
	3/SIRD	118	141	45·6	1·87(1·48-2·38)	$2·4x10^{-7}$	1·39(1·09-1·76)	0·0082
	4/MOD	43	272	13·7	1·41(0·97-2·06)	0·0726	1·47(1·00-2·16)	0·0482
	5/MARD	161	382	29·7	1	-	1	-
CKD 3B	1/SAID	14	149	8·6	1·25(0·70-2·22)	0·4527	1·03(0·58-1·84)	0·9208
	2/SIDD	51	269	15·9	1·75(1·21-2·52)	0·0028	1·24(0·85-1·81)	0·2621
	3/SIRD	69	190	26·6	2·19(1·58-3·04)	$3·0x10^{-6}$	1·55(1·11-2·16)	0·0098
	4/MOD	18	297	5·7	1·15(0·66-2·03)	0·6199	1·12(0·63-1·99)	0·6910
	5/MARD	76	466	14·0	1	-	1	-
Macro-albuminuria	1/SAID	4	147	2·6	0·35(0·12-1·03)	0·0564	0·33(0·11-0·98)	0·0467
	2/SIDD	29	273	9·6	1·27(0·74-2·19)	0·3891	1·25(0·73-2·15)	0·4192
	3/SIRD	28	202	12·2	2·18(1·31-3·63)	0·0026	1·91(1·14-3·19)	0·0135
	4/MOD	27	276	8·9	1·31(0·72-2·38)	0·3737	1·34(0·74-2·45)	0·3373
	5/MARD	32	473	6·3	1	-	1	-
ESRD	1/SAID	5	158	3·1	1·89(0·68-5·27)	0·2249	1·73(0·61-4·88)	0·3014
	2/SIDD	17	289	5·6	2·26(1·11-4·58)	0·0238	1·87(0·91-3·81)	0·0868
	3/SIRD	32	222	12·6	4·89(2·68-8·93)	$2·4x10^{-7}$	3·61(1·96-6·64)	$3·9x10^{-5}$
	4/MOD	8	305	2·6	1·74(0·68-4·42)	0·2448	1·73(0·68-4·44)	0·2522
	5/MARD	16	524	3·0	1	-	1	-

[1]Cox regressions adjusted for sex and age at onset.

[2]Cox regression adjusted for sex· age at onset and first eGFR.

CKD3A was defined as eGFR < 60 mL/min/1·73m^2 and (CKD 3B) as eGFR < 45 mL/min/1·73m^2 for more than 90 days. End-stage renal disease (ESRD) was defined as at least one eGFR below 15 mL/min/1·73m^2. Macroalbuminuria was defined as at least two out of three consecutive visits with albumin excretion rate (AER) ≥200 µg/min or AER ≥300 mg/24 h or albumin-creatinine ratio (ACR) ≥25/35 mg/mmol for men/women. Median duration at first eGFR=100(IQR 54-173) days.

Table S9b. P-values for pairwise comparisons of cox-regressions in table 9a adjusted for sex and age at onset.

	SAID vs SIDD	SAID vs SIRD	SAID vs MOD	SIDD vs SIRD	SIDD vs MOD	SIRD vs MOD
CKD 3A	0·1533	0·0139	0·3454	0·1226	0·6750	0·1229
CKD 3B	0·2634	0·0557	0·7114	0·2304	0·1026	0·0171
Macroalbuminuria	0·0153	0·0010	0·0126	0·0564	0·9023	0·0981
ESRD	0·7250	0·0524	0·8947	0·0133	0·5635	0·0187

Table S9c. P-values for pairwise comparisons of cox-regressions in table 9a adjusted for sex, age at onset and first eGFR.

	SAID vs SIDD	SAID vs SIRD	SAID vs MOD	SIDD vs SIRD	SIDD vs MOD	SIRD vs MOD
CKD 3A	0·6273	0·1620	0·1760	0·1582	0·2364	0·8687
CKD 3B	0·5345	0·1657	0·9394	0·2432	0·5871	0·1940
Macroalbuminuria	0·0132	0·0017*	0·0084	0·1363	0·7830	0·2659
ESRD	0·8829	0·1378	0·9749	0·0331	0·8982	0·1042

*Significant after adjustment for multiple comparisons (10 tests).

Table S10a. Cox regression analysis comparing risk of diabetic retinopathy between clusters in SDR.

	Cluster	Events	Censored	Events (%)	HR(CI95%)[1]	p[1]	HR(CI95%)[2]	p[2]
DR	1/SAID	85	57	59·9	0·90(0·75-1·09)	0·2813	1·06(0·87-1·29)	0·5647
	2/SIDD	169	96	63·8	1·32(1·14-1·52)	0·0002	1·33(1·15-1·54)	0·0001
	3/SIRD	62	117	34·6	0·77(0·62-0·95)	0·0141	0·67(0·54-0·84)	0·0004
	4/MOD	134	139	49·1	1·01(0·86-1·18)	0·9065	1·14(0·96-1·34)	0·1330
	5/MARD	165	252	39·6	1	-	1	-

[1]Cox regressions adjusted for sex.

[2]Cox regression adjusted for sex and age at onset.

Table S10b. P-values for post hoc pairwise comparisons of cox-regressions in table 10a.

	SAID vs SIDD	SAID vs SIRD	SAID vs MOD	SIDD vs SIRD	SIDD vs MOD	SIRD vs MOD
DR[1]	0·0047*	0·3356	0·4208	0·0003*	0·0228	0·0754
DR[2]	0·0946	0·0111	0·6203	<0·0001*	0·1778	0·0014*

[1]Cox regressions adjusted for sex.

[2]Cox regression adjusted for sex and age at onset.

*Significant after adjustment for multiple comparisons (10 tests).

Table S11a. Cox regression analysis comparing risk of cardiovascular disease between clusters in SDR.

	Cluster	Events	Censored	Events (%)	HR(CI95%)[1]	p[1]	HR(CI95%)[2]	p[2]
CE	1/SAID	13	147	8·1	0·51(0·33-0·80)	0·0034	0·75(0·48-1·17)	0·1989
	2/SIDD	54	262	17·1	1·09(0·84-1·41)	0·5278	1·11(0·86-1·43)	0·4322
	3/SIRD	55	181	23·3	1·76(1·36-2·27)	$1{\cdot}5x10^{-5}$	1·10(0·84-1·44)	0·4903
	4/MOD	36	278	11·5	0·73(0·54-0·98)	0·0358	1·31(0·95-1·82)	0·0999
	5/MARD	102	423	19·4	1	-	1	-
Stroke	1/SAID	11	143	7·1	0·91(0·55-1·50)	0·6970	1·38(0·83-2·28	0·2155
	2/SIDD	20	295	6·3	0·79(0·53-1·18)	0·2555	0·80(0·53-1·19)	0·2630
	3/SIRD	26	205	11·3	1·72(1·20-2·48)	0·0033	1·04(0·71-1·51)	0·8586
	4/MOD	12	280	4·1	0·48(0·29-0·78)	0·0028	0·91(0·54-1·53)	0·7263
	5/MARD	58	460	11·2	1	-	1	-

[1] Cox regression adjusted for sex.

[2] Cox regressions adjusted for sex and age at onset.

Coronary events (CE) were defined by ICD-10 codes I21, I20, I251, I253, I254, I255, I256, I257, I258, I259. Stroke was defined by ICD-10 codes I60, I61, I63.

Table S11b. P-values for post hoc pairwise comparisons of cox-regressions in table 11a adjusted for sex.

	SAID vs SIDD	SAID vs SIRD	SAID vs MOD	SIDD vs SIRD	SIDD vs MOD	SIRD vs MOD
CE	0·0149	0·0001*	0·2750	0·0121	0·0640	<0·0001*
Stroke	0·7262	0·0737	0·1246	0·0093	0·1646	0·0002

*Significant after adjustment for multiple comparisons (10 tests).

Table S11c. P-values for post hoc pairwise comparisons of cox-regressions in table 11a adjusted for sex and age at onset.

	SAID vs SIDD	SAID vs SIRD	SAID vs MOD	SIDD vs SIRD	SIDD vs MOD	SIRD vs MOD
CE	0·1991	0·2156	0·0837	0·9637	0·4534	0·4584
Stroke	0·1466	0·4390	0·3311	0·3864	0·7220	0·7377

*Significant after adjustment for multiple comparisons (10 tests).

Table S12. Strongest genetic associations with specific ANDIS clusters (reaching p<0.01).

SNP	Candidate gene		SAID		SIDD		SIRD		MOD		MARD		Difference cluster 2-5
			N=313		N=676		N=603		N=727		N=1646		
		MAF	OR	P	OR	P	OR	P	OR	P	OR	P	P
rs2854275	*HLA_DQB1*	0·13	2·05(1·69-2·56)	$5·7x10^{-10}$	0·82(0·66-1·00)	0·0777	0·97(0·79-1·19)	0·9028	1·11(0·92-1·33)	0·2657	0·87(0·76-1·01)	0·1446	0·0988
rs6467136	*GCC1-PAX4*	0·46	1·42(1·19-1·69)	$7·8x10^{-5}$	1·17(1·04-1·32)	0·0216	1·06(0·94-1·21)	0·3460	1·17(1·04-1·32)	0·0132	1·27(1·16-1·39)	$4·2x10^{-7}$	0·1342
rs7578326	*IRS1*	0·36	0·78(0·65-0·93)	0·0059	0·99(0·87-1·11)	0·5914	0·94(0·83-1·11)	0·3612	1·06(0·94-1·20)	0·4228	0·94(0·86-1·03)	0·1373	0·3140
rs8090011	*LAMA1*	0·37	1·25(1·06-1·48)	0·0069	1·12(0·99-1·23)	0·0526	1·08(0·95-1·23)	0·1473	1·11(0·98-1·25)	0·0645	1·05(0·96-1·15)	0·1686	0·7714
rs10010131	*WFS1*	0·43	1·28(1·07-1·53)	0·0069	1·18(1·04-1·34)	0·0174	1·07(0·94-1·22)	0·3613	1·25(1·11-1·42)	0·0006	1·14(1·04-1·25)	0·0046	0·3020
rs7903146	*TCF7L2*	0·26	1·17(0·97-1·40)	0·0766	1·51(1·33-1·71)	$2·8x10^{-10}$	1·00(0·87-1·15)	0·8626	1·38(1·21-1·56)	$5·7x10^{-7}$	1·41(1·28-1·55)	$1·1x10^{-12}$	$9·6x10^{-6}$
rs10401969	*TM6SF2*	0·09	0·75(0·58-0·97)	0·0376	0·69(0·58-0·83)	0·0002	0·62(0·52-0·75)	$3·1x10^{-6}$	0·89(0·73-1·07)	0·2603	0·77(0·67-0·89)	0·0005	0·0233
rs4402960	*IGF2BP2*	0·29	1·04(0·87-1·24)	0·5013	1·23(1·08-1·40)	0·0002	1·01(0·88-1·16)	0·5279	1·04(0·92-1·18)	0·3089	1·22(1·11-1·33)	$2·1x10^{-6}$	0·0117
rs10811661	*CDKN2B*	0·15	0·87(0·70-1·08)	0·2421	1·33(1·11-1·59)	0·0014	0·98(0·83-1·17)	0·8494	0·99(0·84-1·16)	0·9221	1·18(1·04-1·33)	0·0054	0·0149
rs243088	*BCL11A*	0·47	1·23(1·04-1·45)	0·0126	1·20(1·07-1·35)	0·0025	1·22(1·08-1·35)	0·0008	1·25(1·11-1·40)	0·0001	1·14(1·04-1·24)	0·0024	0·4122
rs1111875	*HHEX/IDE*	0·41	1·16(0·98-1·38)	0·1044	1·21(1·07-1·37)	0·0045	1·05(0·92-1·19)	0·5104	0·94(0·84-1·06)	0·3139	1·11(1·02-1·22)	0·0228	0·0106
rs7607980	*COBL1*	0·13	1·12(0·87-1·45)	0·3197	1·12(0·93-1·34)	0·2059	1·32(1·07-1·61)	0·0044	0·99(0·83-1·17)	0·9746	1·21(1·06-1·39)	0·0027	0·0757
rs7647305	*SFRS10*	0·19	0·85 (0·68-1·07)	0·1319	0·95(0·81-1·11)	0·3808	0·80(0·68-0·95)	0·0063	0·93(0·79-1·08)	0·2637	0·92(0·82-1·03)	0·0982	0·4302
rs5219	*KCNJ11*	0·38	1·05(0·88-1·25)	0·6114	1·18(1·04-1·34)	0·0121	1·03(0·90-1·18)	0·6737	1·28(1·13-1·44)	0·0001	1·10(1·01-1·21)	0·0324	0·0453
rs864745	*JAZF1*	0·49	1·06(0·90-1·25)	0·6120	0·91(0·80-1·02)	0·0674	0·93(0·82-1·05)	0·1419	0·81(0·72-0·91)	0·0002	0·94(0·87-1·03)	0·1355	0·1095
rs7202877	*BCAR1*	0·12	0·89(0·70-1·14)	0·2758	1·29(1·06-1·57)	0·0286	1·08(0·89-1·31)	0·5906	1·35(1·11-1·64)	0·0037	1·11(0·97-1·27)	0·1675	0·1644
rs11708067	*ADCY5*	0·24	0·92(0·75-1·13)	0·4376	0·86(0·74-1·00)	0·0517	0·86(0·73-1·00)	0·0444	0·86(0·75-0·99)	0·0326	0·79(0·71-0·88)	$1·5x10^{-5}$	0·5809
rs516946	*ANK1*	0·22	0·98(0·81-1·20)	0·8582	1·18(1·02-1·37)	0·0312	1·13(0·97-1·32)	0·0930	1·03(0·90-1·18)	0·6120	1·21(1·08-1·34)	0·0004	0·2477
rs243021	*BCL11A*	0·47	1·05(0·89-1·24)	0·4778	1·04(0·93-1·18)	0·2957	1·05(0·92-1·19)	0·4500	1·04(0·92-1·16)	0·4706	1·14(1·05-1·24)	0·0026	0·3261
rs11063069	*CCND2*	0·20	0·83(0·66-1·04)	0·1109	1·17(1·01-1·36)	0·0217	1·11(0·94-1·30)	0·1628	1·11(0·96-1·29)	0·1522	1·15(1·03-1·28)	0·0084	0·9144
rs340874	*PROX1*	0·46	1·05(0·89-1·24)	0·5605	1·02(0·91-1·15)	0·6375	0·97(0·86-1·10)	0·7497	1·04(0·92-1·17)	0·4695	1·12(1·03-1·22)	0·0084	0·1441

Maximum likelihood estimation using geographically matched non-diabetic controls (N=2754).

Table S13. SNPs included in genetic risk scores.

SNP	Effect allele	Weight	Chr	location	Locus
Insulin resistance					
rs459193	G	0·012	5	*ANKRD55*	intergenic
rs13389219	C	0·012	13	*GRB14*	intergenic
rs2943641	C	0·009	2	*IRS1*	intergenic
rs12970134	A	0·008	18	*MC4R*	intergenic
rs1801282	C	0·016	3	*PPARG*	coding - missense
Insulin secretion					
rs10811661	G	0·009	9	*CDKN2B*	intergenic
rs10830963	G	0·039	11	*MTNR1B*	intron
rs1111875	T	0·004	10	*HHEX/IDE*	intergenic
rs11708067	A	0·023	3	*ADCY5*	intron
rs13266634	G	0·016	8	*SLC30A8*	coding - missense
rs1552224	A	0·017	11	*CENTD2*	intergenic
rs163184	G	0·009	11	*KCNQ1*	intron
rs17168486	T	0·013	7	*DGKB*	intergenic
rs2191349	T	0·017	7	*DGKB/TMEM195*	intergenic
rs4402960	T	0·012	3	*IGF2BP2*	intron
rs4607517	A	0·025	7	*GCK*	intergenic
rs5219	T	0·001	11	*KCNJ11*	coding - missense
rs560887	T	0·042	2	*G6PC2/ABCB11*	intron
rs7034200	A	0·042	9	*GLIS3*	intron
rs7756992	G	0·010	6	*CDKAL1*	intron
rs7903146	T	0·020	10	*TCF7L2*	intron / promoter
T2D					
rs10010131	G	0·16	4	*WFS1*	intron
rs10401969	C	0·122	19	*SUGP1/CILP2*	intron
rs10423928	A	0·09	19	GIPR	intron
rs10811661	T	0·02	9	*CDKN2B*	intergenic
rs10842994	C	0·095	12	*KLHDC5*	intergenic
rs10885122	G	0·04	10	*ADRA2A*	intergenic
rs10923931	T	0·122	1	*NOTCH2*	intron
rs11063069	G	0·095	12	*CCND2*	intergenic
rs11071657	A	0·03	3	*FAM148B*	intergenic
rs1111875	C	0·131	10	*HHEX/IDE*	intergenic
rs1153188	A	0·08	12	*DCD*	intergenic
rs11605924	A	0·04	11	*CRY2*	intron
rs11634397	G	0·058	15	*ZFAND6*	intergenic
rs11708067	A	0·12	3	*ADCY5*	intron
rs12571751	A	0·077	10	*ZMIZ1*	intron
rs12779790	G	0·104	10	*CDC123/CAMK1D*	intergenic
rs13266634	C	0·19	8	*SLC30A8*	coding - missense
rs1531343	C	0·1	12	*HMGA2*	intron of pseudogene
rs1552224	A	0·131	11	*CENTD2*	intergenic

rs163184	G	0·28	11	*KCNQ1*	intron
rs17168486	T	0·104	7	*DGKB*	intergenic
rs174550	T	0·04	11	*FADS1*	intron
rs17782313	C	0·06	18	*MC4R*	intergenic
rs1801282	C	0·11	3	*PPARG*	coding - missense
rs2296172	G	0·095	1	*MACF1*	coding - missense
rs231362	G	0·08	11	*KCNQ1*	intron
rs243088	T	0·068	2	*BCL11A*	intergenic
rs2796441	G	0·068	9	*TLE1*	intergenic
rs2943641	C	0·095	2	*IRS1*	intergenic
rs340874	C	0·07	1	*PROX1*	intergenic
rs35767	G	0·04	12	*IGF1*	nearGene-5
rs3923113	A	0·07	2	*GRB14*	intergenic
rs4402960	T	0·122	3	*IGF2BP2*	intron
rs4457053	G	0·08	5	*ZBED3*	intron of ZBED3-AS1
rs459193	G	0·077	5	*ANKRD55*	intergenic
rs4607103	C	0·09	3	*ADAMTS9-AS2*	intron
rs4607517	A	0·068	7	*GCK*	intergenic
rs516946	C	0·09	8	*ANK1*	intron
rs5219	C	0·068	11	*KCNJ11*	coding - missense
rs553668	A	0·42	10	*ADRA2A*	UTR-3
rs560887	T	0·03	2	*G6PC2/ABCB11*	intron
rs6017317	G	0·086	20	*FITM2/R3HDML/HNF4A*	intergenic
rs6467136	G	0·11	7	*GCC1-PAX4*	intergenic
rs6815464	C	0·122	4	*MAEA*	intron
rs7034200	A	0·095	9	*GLIS3*	intron
rs7138803	A	0·11	12	*BCDIN3D/FAIM2*	intergenic
rs7178572	G	0·068	15	*HMG20A*	intergenic
rs7202877	T	0·113	16	*BCAR1*	intergenic
rs7578597	T	0·15	2	*THADA*	coding - missense
rs7593730	C	0·105	2	*RBMS1/ITGB6*	intron
rs7607980	T	0·14	2	*COBLL1*	coding - missense
rs7754840	C	0·27	6	*CDKAL1*	intron
rs7756992	G	0·2	6	*CDKAL1*	intron
rs7903146	T	0·37	10	*TCF7L2*	intron / promoter
rs7944584	G	0·01	11	MADD	intron
rs7957197	T	0·08	12	*OASL/TCF1/HNF1A*	intron of OASL
rs7961581	C	0·09	12	*TSPAN/LGR5*	intergenic
rs8042680	A	0·07	15	*PRC1*	intron
rs8090011	G	0·122	18	*LAMA1*	intron
rs8108269	G	0·03	19	*GIPR*	intergenic
rs831571	C	0·086	3	*PSMD6*	intergenic
rs864745	T	0·1	7	*JAZF1*	intron
rs896854	T	0·06	8	*TP53INP1*	intron
rs9470794	C	0·113	6	*ZFAND3*	intron

Table S14. Genetic risk score analysis.

	T2D		Insulin secretion		Insulin resistance	
	OR	P	OR	P	OR	P
SAID	1·22(1·09-1·37)	0·0008	0·97(0·86-1·09)	0·5914	1·04(0·92-1·17)	0·5248
SIDD	1·45(1·33-1·58)	$7{\cdot}3x10^{-17}$	1·11(01·02-1·20)	0·0143	1·08(0·99-1·18)	0·0770
SIRD	1·07(0·98-1·16)	0·1602	0·98(0·89-1·08)	0·6509	1·06(0·97-1·16)	0·1886
MOD	1·28(1·18-1·39)	$4{\cdot}6x10^{-9}$	1·16(1·08-1·27)	0·0002	1·05(0·96-1·14)	0·2825
MARD	1·44(1·35-1·54)	$3{\cdot}0x10^{-29}$	1·16(1·10-1·23)	$1{\cdot}0x10^{-6}$	1·06(0·99-1·12)	0·0773

OR are for 1 SD change in genetic risk score.