

University of Dundee

Cohort Profile

Hébert, Harry L.; Shepherd, Bridget; Milburn, Keith; Veluchamy, Abirami; Meng, Weihua; Carr, Fiona

Published in:
International Journal of Epidemiology

DOI:
[10.1093/ije/dyx140](https://doi.org/10.1093/ije/dyx140)

Publication date:
2018

Licence:
CC BY

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

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):

Hébert, H. L., Shepherd, B., Milburn, K., Veluchamy, A., Meng, W., Carr, F., Donnelly, L. A., Tavendale, R., Leese, G., Colhoun, H. M., Dow, E., Morris, A. D., Doney, A. S., Lang, C. C., Pearson, E. R., Smith, B. H., & Palmer, C. N. A. (2018). Cohort Profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS). *International Journal of Epidemiology*, 47(2), 380-381j. <https://doi.org/10.1093/ije/dyx140>

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Cohort Profile

Cohort Profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS)

**Harry L Hébert,¹ Bridget Shepherd,² Keith Milburn,³
 Abirami Veluchamy,^{1,2} Weihua Meng,¹ Fiona Carr,² Louise A Donnelly,²
 Roger Tavendale,² Graham Leese,² Helen M Colhoun,^{1,4} Ellie Dow,²
 Andrew D Morris,⁵ Alexander S Doney,² Chim C Lang,²
 Ewan R Pearson,² Blair H Smith¹ and Colin N A Palmer^{2*}**

¹Division of Population Health Sciences, ²Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, ³Health Informatics Centre Services, Ninewells Hospital & Medical School, University of Dundee, Dundee, UK, ⁴Institute of Genetics & Molecular Medicine and ⁵Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

*Corresponding author. Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, Division of Molecular & Clinical Medicine, Level 5, Mailbox 12, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK. E-mail: c.n.a.palmer@dundee.ac.uk

Editorial decision 26 June 2017; Accepted 7 July 2017

Why was the cohort set up?

The prevalence of diabetes worldwide has been steadily increasing over the past 20 years. In 1997 it was estimated to be 124 million,¹ in 2015 it was estimated to be 415 million among 20–70 year olds, and this is expected to rise to 642 million by 2040.² In the UK, an estimated 4 million people have diabetes either diagnosed or undiagnosed.² This represents a significant burden on health care resources,³ particularly given that type 2 diabetes (T2D) is associated with comorbidities including obesity,⁴ cardiovascular disease,⁵ chronic kidney disease⁶ and neuropathy.⁷ T2D is a complex disorder, caused by a combination of environmental and genetic factors.⁸ Before the first genome-wide association study (GWAS) was conducted for T2D in 2007,⁹ very few genetic loci were known to be involved with T2D. However, linkage and candidate-gene association studies have often failed to replicate findings through lack of power and inadequate knowledge of the underlying biological pathways.^{10,11}

Diabetes Audit and Research in Tayside Scotland (DARTS) started in 1996 as a joint collaboration between

the University of Dundee's Department of Medicine and Medicines Monitoring Unit (MEMO), three Tayside Health Care Trusts (at Ninewells Hospital and Medical School, Perth Royal Infirmary and Stracathro Hospital) and a group of Tayside general practitioners (GPs) with a special interest in diabetes care.¹² Initially supported by the Scottish Home and Health Department, the Wellcome Trust, the Robertson Trust and Tenovus Tayside, the aim of the study was to identify all patients with diabetes within the wider Tayside region, through electronic record linkage, in order to improve health care over and above that which was practical through existing general practice lists alone. In 1998, consenting patients within this electronic database were recruited to the Genetics of DARTS (GoDARTS) study and invited to provide a blood sample for DNA extraction, for research purposes. At the same time, they were invited to provide phenotypic data (clinical and lifestyle factors), through questionnaires and clinical examination. This resource was intended to help identify the relative contribution of specific genetic and environmental factors that are associated with disease onset, progression and response to treatment.^{10,11}

Who is in the cohort?

Patients from the Tayside region of Scotland ($n = 391\,274$ on 1 January 1996¹²) were added to the DARTS register, for clinical purposes, through electronic record linkage on the basis of having diabetes mellitus according to primary and/or secondary care data sources. These included hospital diabetes clinics, mobile diabetes eye units, diabetes prescription databases, the Tayside regional biochemistry database and all diabetes-related hospital discharge records. This electronic record linkage technique has a sensitivity of 97% and is continually being updated, creating a longitudinal dataset of clinical data which is manually validated by a dedicated team of clinicians.¹² Patients with T2D, which comprises around 90% of all diabetes cases, were invited to participate in the GoDARTS study either at diabetes or eye screening clinics or through their GP.

For the pilot phase of the study (GoDARTS1), 2763 patients with T2D were recruited from December 1998 to October 2004. This phase was used to test recruitment processes and the ability to anonymously link patient clinical data from electronic records to the study, and was funded by Tenovus Tayside. As this was the primary aim of the pilot phase, only blood samples were taken at the point of recruitment and no baseline data were recorded.

From October 2004 to May 2009, a second collection (GoDARTS2) was undertaken as part of the Wellcome Trust United Kingdom Type 2 Diabetes Case-Control Collection (WTCCC). A total of 16 146 people were recruited in this phase, including 7989 patients with T2D and 8157 matched healthy controls. We initially invited five matched non-diabetic controls per case from the corresponding GP practice; however, after initial success, this was reduced to two controls per case and on average one of the invited controls accepted. This incidentally included 1292 patients with T2D who had already been recruited in the GoDARTS1 phase. Baseline clinical and lifestyle measurements (Table 1) were recorded for all patients recruited in GoDARTS2. From October 2009 until 2015, an extension to the WTCCC project was granted (GoDARTS3), with 1342 patients with T2D being recruited during this time. Some of these participants had also been recruited to GoDARTS1 ($n = 20$), GoDARTS2 ($n = 513$) or both ($n = 120$), where baseline data did not exist or original DNA quality was poor (Figure 1). This gives a current total GoDARTS cohort of 18 306 participants, 10 149 of whom have T2D ($\sim 44.8\%$ of the DARTS study, representing the diabetic population in Tayside) and 8157 of whom were healthy controls at baseline.

Currently the cohort is in the early stages of GoDARTS4, the fourth phase of the study. In this phase, recruitment is continuing through a number of initiatives including the Scottish Health Research Register (SHARE)/Scottish Diabetes Research Network (SDRN), the Genetics of SHARE

(GoSHARE) and GoDARTS-Scotland. SHARE/SDRN is a register of patients in Scotland who want to participate in medical research and have provided consent for their electronic medical records to be used for research purposes [<http://www.share-sdrn.org>]. GoSHARE is a parallel project which additionally aims to get permission to collect spare blood from people attending for routine clinical tests at hospital or GP clinics, that would otherwise go to waste after the necessary tests had been performed [<http://www.goshare.org.uk/>]. Since the aim is to involve everyone who is resident in the Tayside area, this will inevitably include people with T2D, and they will contribute to the GoDARTS study. GoDARTS Scotland is a sub-study specifically recruiting people who have been diagnosed with T2D in the past 2 years in order to study response to therapies, including metformin.

At the point of recruitment, all participants in GoDARTS provide, by invitation, informed consent for their data to be used for research purposes and explicit consent for use in collaboration with industry. This includes allowing their baseline data to be linked anonymously to individual patient medical records including laboratory data, hospital admissions and Scottish Care Information – Diabetes (SCI-Diabetes) data. SCI-Diabetes is a shared electronic patient record which can be accessed by health professionals and researchers to aid the treatment of diabetes patients in Scotland. In this way, longitudinal data can be accessed relating to routine diabetes management, for example glycosylated haemoglobin (HbA1c), fasting insulin and fasting glucose, as well as previous patient diagnoses including diabetic complications. Furthermore, $\sim 95\%$ of patients have consented to being contacted for future studies, aiding research beyond T2D.

How often have they been followed up?

As patients attend a baseline clinic at recruitment, initial measurements are cross-sectional. However, the use of electronic record linkage, which automatically updates patient details and grants access to NHS data as far back as 1987, makes GoDARTS a longitudinal cohort. This is made possible through the use of the community health index (CHI) number, which is a unique numerical identifier issued to each patient on first registration with a GP or admission to a hospital in Scotland. Around 96% of the UK population are estimated to be registered with a GP.¹³ The CHI is a 10-digit number consisting of six digits corresponding to the patient's date of birth (DDMMYY), two digits randomly generated, one digit corresponding to the patient's gender (odd for males, even for females) and one check digit. The CHI number links to live databases which are constantly being updated, such as the Scottish Morbidity Record (SMR) providing data on primary and

Table 1. Summary of baseline data collected and comparison of response rates between cases and controls in GoDARTS

Measure	Response (%)		Notes
	Cases (<i>n</i> = 8698)	Controls (<i>n</i> = 8140)	
Age (years)	99.37	99.53	
Gender	100	100	
Ethnicity	100	100	Caucasian or non-Caucasian
Height (cm)	99.68	100	
Weight (kg)	99.57	99.95	
Waist (cm)	99.57	99.84	
Diastolic blood pressure (mmHg)	1 = 99.90 2 = 99.17	1 = 99.99 2 = 99.83	Two measures taken
Systolic blood pressure (mmHg)	1 = 99.90 2 = 99.17	1 = 99.99 2 = 99.83	Two measures taken
Heart rate (bpm)	1 = 99.75 2 = 98.87	1 = 99.90 2 = 99.50	Two measures taken
Diabetes treatment	99.21	n/a	Diet, tablets or pills
Diabetes medication ^a	4.73	n/a	Dose, date and time last taken
Family history of diabetes ^a	8.24	n/a	
Present smoking status (Amount smoked)	99.71 (85.77) ^b	99.94 (81.80) ^b	
Past smoking status (Amount smoked)	99.61 (95.05) ^b	99.90 (95.22) ^b	
Age started smoking	99.22 ^c	99.52 ^c	
Stopped normal periods (Age periods stopped)	99.53 (90.70) ^b	99.85 (73.85) ^b	
Level of physical activity during work:			
1. Recently	1 = 99.66	1 = 99.89	
2. Past 10 years	2 = 99.69	2 = 99.85	
3. Youth	3 = 99.56	3 = 99.93	
Level of physical activity during travel:			
1. Recently	1 = 99.60	1 = 99.89	
2. Past 10 years	2 = 99.59	2 = 99.84	
3. Youth	3 = 99.47	3 = 99.84	
Level of physical activity during leisure:			
1. Recently	1 = 99.67	1 = 99.93	
2. Past 10 years	2 = 99.57	2 = 99.82	
3. Youth	3 = 99.54	3 = 99.91	
Confirmed type 2 diabetes	99.25	n/a	
Location patient was screened	100	100	Clinic, GP surgery, home or other
HbA1c (%)	93.17	99.82	
Cholesterol (mmol/l)	92.73	99.94	
HDL (mmol/l)	92.72	99.94	
LDL (mmol/l)	84.07	97.02	
Creatinine (μmol/l)	92.99	99.84	
Triglycerides (mmol/l)	92.62	99.84	

n/a, not available.

^aBaseline data only available in participants recruited in GoDARTS3.

^bResponse rate calculated according to the number of positive responses to the main question.

^cResponse rate calculated according to the number of positive responses to present and/or past smoking status.

secondary diagnoses for patients discharged from hospital since 1980, the Tayside echocardiography database providing data on all echocardiograms performed at Ninewells Hospital since 1994, the General Registrar's Office providing mortality data since 1998, the biochemistry database listing all assays performed since 1981 and a database containing all prescriptions dispensed since 1989. This allows identification of an up-to-date record of every individual's health care processes and outcomes and

linkage of corresponding datasets. An anonymization process converts the CHI into a study pro-CHI, to protect the identities and confidentiality of individuals while retaining the ability to link patient data across multiple datasets.

What has been measured?

For GoDARTS1, only blood samples were taken for DNA extraction as this was a pilot phase used to test the ability

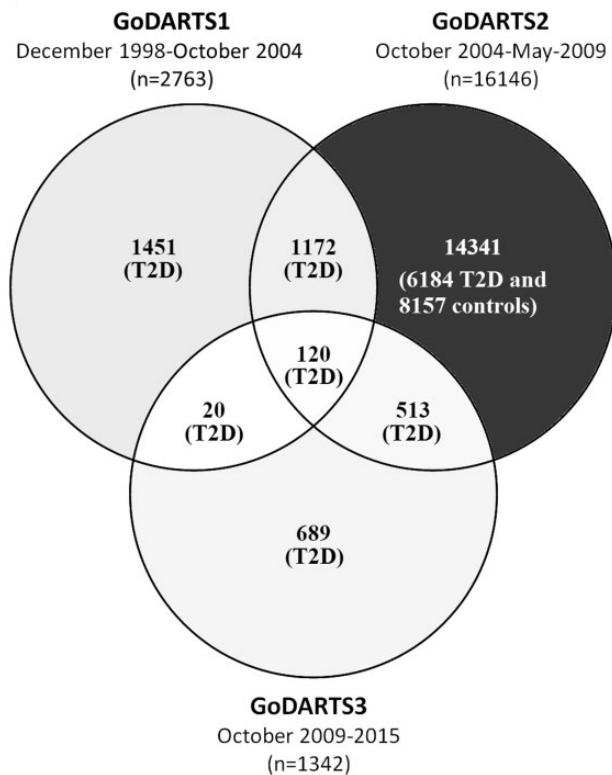


Figure 1. A venn diagram showing the overlap in patient recruitment between GoDARTS1, GoDARTS2 and GoDARTS3.

to link electronic health records anonymously to genetic data. For GoDARTS2 and GoDARTS3, participants completed a lifestyle questionnaire and consented to baseline measurements being recorded at recruitment (Table 1). In addition, during GoDARTS3 urine samples (~80% of recruits) were taken for proteomic and metabolomics analysis and RNA (~30% of recruits) was extracted from blood samples. The lifestyle questionnaire contains items relating to smoking history (present and past status, along with amount and age started where applicable), as well as level of physical activities in three common locations (work/education, travel and home life) over three different time periods in life (recently, past 10 years and youth). In addition, women were asked about their menopausal history. Baseline observations were recorded and included height, weight and waist measurements, as well as heart rate and blood pressure. The patient's recruitment information was recorded including ethnicity, screening location, confirmation of T2D and medication history, as well as family history of diabetes and whether the patient had previously participated in GoDARTS. As baseline data were only recorded for participants recruited in GoDARTS2 and GoDARTS3, there are 1451 participants who were only involved in GoDARTS1 (Figure 1) and do

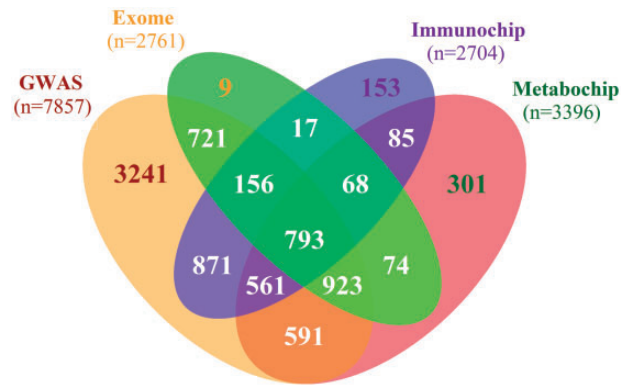


Figure 2. A venn diagram showing the overlap of T2D cases genotyped between different platforms. Overall, 8564 cases out of a possible 10 149 have been genotyped on at least one platform, with 7857 having genome-wide data.

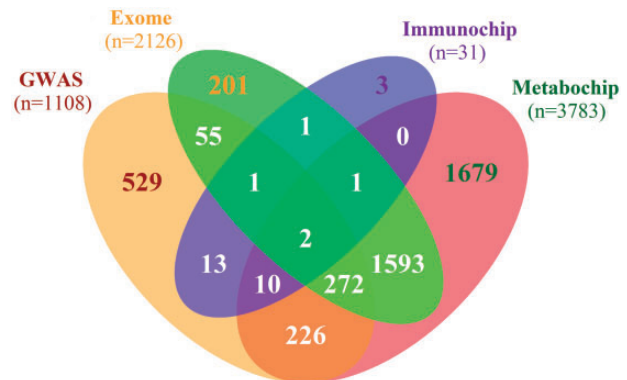


Figure 3. A venn diagram showing the overlap of controls genotyped between different platforms. Overall, 4586 controls out of a possible 8157 have been genotyped on at least one platform, with 1108 having genome-wide data.

not have these data. Furthermore, there are 17 healthy control participants from GoDARTS2 who are missing baseline data, meaning that a total of 16 838 patients have these available, including 8698 cases and 8140 controls.

As well as phenotypic data, genetic data are available for 8564 T2D cases (Figure 2) and 4586 controls (Figure 3) after quality control. Samples have been genotyped across five platforms. GWAS data have been obtained for 7857 T2D cases and 1108 controls, using the Affymetrix Genome-Wide Human SNP Array 6.0 and the Illumina HumanOmni Express. The Affymetrix GWAS chip contains 932 979 single nucleotide polymorphisms (SNPs), and the Illumina GWAS chip contains 731 296. This has allowed for imputation of additional and missing genotypes by SHAPEIT¹⁴ and IMPUTE2¹⁵ using the 1000 Genomes reference panel.¹⁶ In addition, 707 T2D cases and 3478 controls have been genotyped using custom genotyping arrays from Illumina. These include the ImmunoChip, Cardio-MetaboChip (MetaboChip) and Human Exome array. The ImmunoChip contains

196 524 genetic markers from loci that have previously been associated with at least one of 13 autoimmune diseases, including T1D,¹⁷ and the MetaboChip contains 196 725 SNPs from loci that have prior evidence of association with T2D, coronary artery disease/myocardial infarction and 21 related traits.¹⁸ The specific criteria by which markers on the Cardio-MetaboChip and the ImmunoChip have been chosen makes these platforms a cost-effective means of replicating and fine-mapping known loci and discovering novel loci by virtue of overlapping biological mechanisms between the related traits. The Human Exome Array contains 247 870 genetic markers from across the exome, allowing for studies to focus on identifying protein-altering variants.¹⁹

What has it found?

Baseline and follow-up epidemiology

Baseline clinical and demographic statistics are summarized in Table 2. Overall, 53.33% of the cohort are male, which is similar to the proportion represented in DARTS (52.83%), with the proportion being higher in cases (56.38%) compared with controls (50.08%). The majority of the cohort are Caucasian (99.70%) and the median age at recruitment was higher in cases (67 years) compared with controls (60 years). This observation is also apparent when the cases and controls are further dichotomized into males (66 vs 62 years) and females (68 vs 58 years). The cohort contains data on a number of continuous traits known to be associated with T2D. For example, median body mass index (BMI) (30.6 vs 26.6 kg/m²), resting heart rate (1st = 73 vs 68 bpm), creatinine (89 vs 87 µmol/l) and triglyceride (1.880 vs 1.315 mmol/l) levels were all higher in cases compared with controls. Furthermore, there was a higher proportion of past smokers among those with T2D (63.14% vs 53.56%).

As of 2014, mortality data have shown that the number of deaths at 9 years after recruitment was 2587 out of 10 149 among the cases and the Kaplan–Meier survival probability is 70.0%, whereas among the controls the number of deaths was 851 out of 8157 and the Kaplan–Meier survival probability is 88.2% (Figure 4). Control group mortality data do not go beyond this, as recruitment of controls did not begin until GoDARTS2 (approximately 7 years after the start of GoDARTS1); however, the number of deaths after 16 years among the cases was 2941 (out of 10 149) with a Kaplan–Meier survival probability of 53.5%.

According to SCI-Diabetes data, the number of people initially recruited as controls at baseline, but who went on to develop diabetes, is 650 (out of 8157) and the Kaplan–Meier cumulative incidence probability is 8.3% (Figure 5). Also captured were self-reported physical activity data,

and these can be seen to successfully stratify the effect of the fat mass and obesity-associated protein gene (*FTO*) risk allele, rs9939609, where the genetic association with BMI is largely attenuated in active individuals, as has been observed in large meta-analyses (Figure 6).

Research output

Over 100 studies have been published using GoDARTS either as the primary study cohort or as part of a larger meta-analysis or replication. The following is a summary of important studies, in all of which GoDARTS has been involved. A full and up-to-date list of studies can be found at [www.researcherid.com/rid/K-9448-2016] (Researcher ID: K-9448-2016).

GoDARTS began in the pre-GWAS era, with its first study being published in 2002.²⁰ At this time candidate-gene studies were conducted, and these were particularly successful in replicating associations at the peroxisome proliferator-activated receptor (PPAR) transcription factor family, which had previously been associated with a range of phenotypes including T2D. In particular, two variants (rs1801282 and rs3856806) in *PPARG* were shown to have opposing effects on body weight, with the non-synonymous rs1801282 (Pro12Ala) associated with lower BMI and the synonymous rs3856806 (C1431T) associated with higher BMI.²⁰ This provided an explanation for previous inconsistencies found between this gene and BMI. Similar findings were reported with respect to T2D susceptibility²¹ and myocardial infarction,²² with haplotype analysis confirming the protective effect of rs1801282 with these phenotypes in contrast to the risk effect of rs3856806. In addition, rs1801282 was shown to have an opposing effect to another polymorphism, rs10865710 (C681G), with respect to height and weight in pre-pubertal children.²³ Further studies identified similar attenuations in the related genes *PPARA* with myocardial infarction,²⁴ and replicated an association at *PPARD* with reduced adult height.²⁵

In 2007, the GoDARTS study became part of the UK T2D genetics consortium collection (UKT2DGCC), which formed the main replication cohort for the WTCCC. This involved large-scale studies investigating quantitative traits including height and obesity, in addition to T2D. One of the first published studies using the UKT2DGCC found an association in the *FTO* gene with obesity. This effect was observed from age 7 years and older.²⁶ Another related study identified associations at the *MC4R* gene with BMI and childhood obesity among 7–11-year-olds.²⁷ GoDARTS was involved in the seminal publications describing the identification of many genes for type 2 diabetes, including the first descriptions of *CDKAL1* and

Table 2. Comparison between cases and controls in baseline measurements

Measure	Cases <i>n</i> = 8698	Controls <i>n</i> = 8140	Overall
Gender (% male)	56.38	50.08	53.33
Age (years)	67	60	64
Males	66	62	65
Females	68	58	63
Ethnicity (% Caucasian)	99.68	99.72	99.70
BMI (median) (kg/m ²)	30.6	26.6	28.4
Males	30.0	27.0	28.4
Females	31.5	26.2	28.4
Height (cm)	168	168.5	168
Males	174	175	175
Females	159	162	160
Weight (kg)	86.05	76.20	81.3
Males	90.4	83.2	87
Females	79.8	68.5	73
Waist (cm)	104	93	99
Males	106	98	102
Females	101	86	93
Current smokers (%)	16.37	16.35	16.36
Males	16.42	16.49	16.45
Females	16.31	16.20	16.25
Current amount smoked (packs per week)	5	5	5
Males	6	5	5
Females	5	5	5
Past smokers (%)	63.14	53.56	58.50
Males	70.24	59.80	65.49
Females	53.97	47.30	50.52
Past amount smoked (packs per week)	7	7	7
Males	7	7	7
Females	7	6	7
Age started smoking (years)	16	16	16
Males	16	16	16
Females	17	17	17
Resting pulse 1 (bpm)	73	68	70
Males	72	66	69
Females	75	70	72
Resting pulse 2 (bpm)	73	68	71
Males	72	66	69
Females	75	70	72
Systolic blood pressure 1 (mm Hg)	141	135	139
Males	141	138	140
Females	141	131	136
Diastolic blood pressure 1 (mm Hg)	77	80	78
Males	78	81	79
Females	76	78	77
Systolic blood pressure 2 (mm Hg)	140	134	137

(continued)

Table 2. Continued

Measure	Cases <i>n</i> = 8698	Controls <i>n</i> = 8140	Overall
Males	140	137	139
Females	140	130	135
Diastolic blood pressure 2 (mm Hg)	75	78	77
Males	76	80	78
Females	74	77	76
HbA1c (%)	7.1	5.5	6.0
Males	7.1	5.5	6.0
Females	7.2	5.5	5.9
Cholesterol (mmol/l)	4.30	5.24	4.72
Males	4.19	5.12	4.57
Females	4.44	5.36	4.90
HDL (mmol/l)	1.28	1.57	1.41
Males	1.21	1.41	1.30
Females	1.39	1.76	1.57
LDL (mmol/l)	2.01	2.93	2.43
Males	1.97	2.90	2.37
Females	2.05	2.97	2.51
Creatinine (μmol/l)	89	87	88
Males	95	95	95
Females	82	79	80
Triglycerides (mmol/l)	1.880	1.315	1.59
Males	1.87	1.47	1.69
Females	1.88	1.19	1.49

Median values given for all continuous data.

CDKN2A/CDKN2B variants and risk for diabetes^{28–32} and also the original report of the association of the *HMGGA2* gene with height.^{33,34} Smaller-scale candidate-gene studies have also been conducted to good effect in T2D, as demonstrated by the identification of the *WFS1* gene.³⁵ T2D-related traits have also been studied with 10 novel loci associated with fasting glucose levels,^{29,36} IGF1 influencing fasting insulin²⁹ and GIPR associated with 2-h glucose levels after oral glucose challenge.³²

In addition to GWAS of anthropometric traits and T2D susceptibility, GoDARTS has pioneered the use of electronic medical records for the study of drug efficacy. This was initially used to study response to sulphonylureas³⁷ and statin lipid-lowering agents.³⁸ As part of the WTCCC2 consortium, GoDARTS served as a discovery cohort for the GWAS of response to statins and metformin.³⁹ The metformin analysis found a novel association of glycaemic response to metformin at a locus including the ataxia telangiectasia mutated (*ATM*) gene, which provided new clues to the mechanism of action of this mysterious drug.^{40,41} More recently, the GoDARTS cohort was the main discovery cohort for a large Metformin Genetics Consortium study (MetGen) that showed a variant in *SLC2A2* to be consistently associated with altered glycaemic response to metformin.⁴² Other investigated

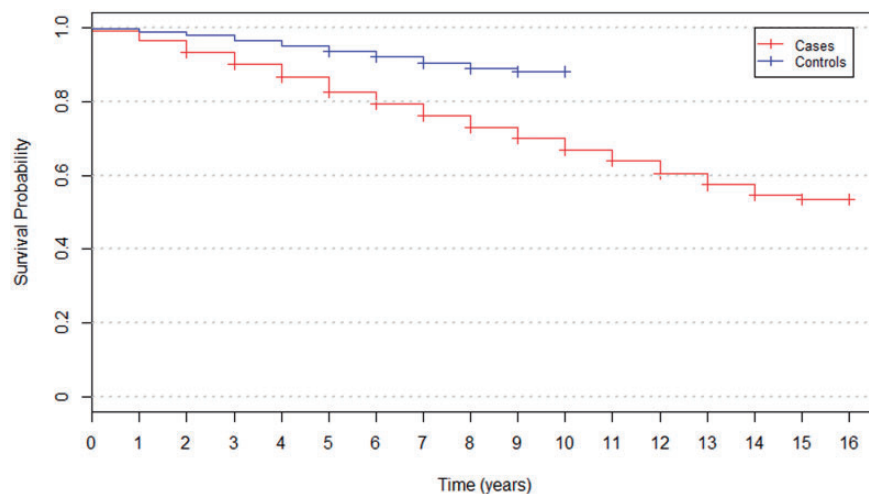


Figure 4. A Kaplan-Meier plot comparing survival rate since baseline recruitment in cases and controls.

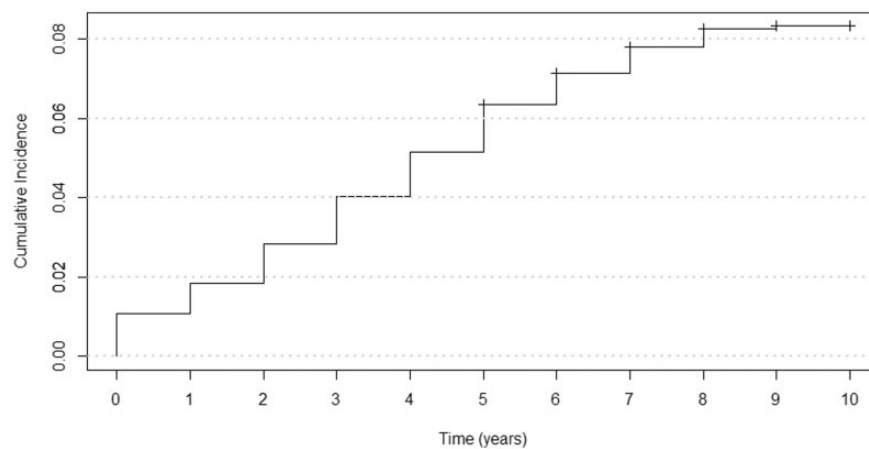


Figure 5. A Kaplan-Meier cumulative incidence plot of diabetes in the GoDARTS baseline controls group.

phenotypes in GoDARTS include response to thiazolidinediones⁴³ and sulphonylureas,^{37,44} and adverse reactions to statins, metformin and angiotensin-converting enzyme (ACE) inhibitors.^{40,45}

Linkage of the GoDARTS study to the Tayside echocardiography database has allowed the identification of genetic variants associated with left ventricular hypertrophy,⁴⁶ as well as the association of both high (>10%) and low (<6%) HbA1c levels with risk of incident heart failure.⁴⁷

From around 2012, the focus of studies has switched from genome-wide analysis to more targeted approaches. One of these methods involves genotyping with the MetaboChip. This has been influential in the discovery and fine-mapping of loci in diabetes- and cardiovascular disease-related phenotypes. In particular, 17 novel T2D loci^{48,49} have been discovered and a further 39 genetic regions have undergone variant localization and genomic annotation to identify the causative mutations.⁵⁰ Similar success has been seen in coronary artery disease, with 25

loci being identified.^{51,52} Glycaemic and anthropometric traits have also been studied successfully, with loci being identified with fasting insulin, fasting glucose, height and obesity.^{53–57}

Another targeted approach that has been increasingly used is exome sequencing and follow-up studies using the Illumina Exome chip. This has been used to capture rare disease-associated variants lying within the protein coding region of the genome, which are hypothesized to make up much of the missing heritability in common diseases. Due to their protein-altering nature, disease-associated polymorphisms discovered in these regions are likely to be causative, and the biological pathways can be more easily elucidated than other study methods. Exome chip genotyping in GoDARTS contributed to the demonstration that, loss-of-function variants in *APOC3*⁵⁸ and *ANGPTL4*⁵⁹ are simultaneously associated with lower triglyceride levels and reduced risk of coronary artery disease (CAD). Similarly, inactivating mutations in *NPC1L1* have a

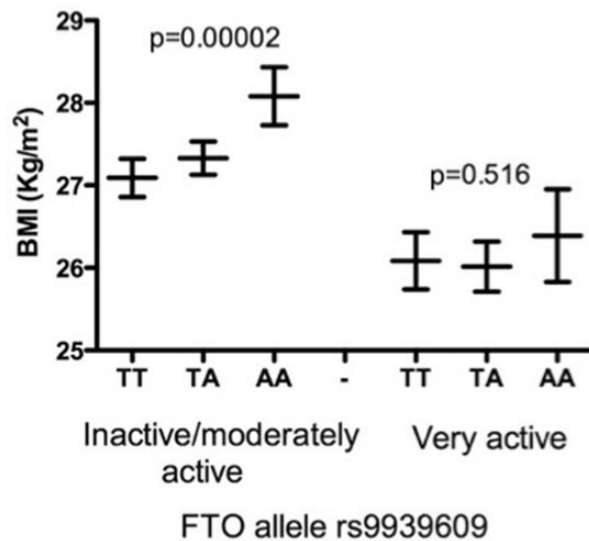


Figure 6. A plot showing that increased physical activity successfully stratifies the association of the obesity risk allele rs9939609 in *FTO* with BMI.

protective effect on CAD risk and reduced low-density lipoprotein (LDL) cholesterol level.⁶⁰ Whereas other phenotypes related to glycaemic traits,^{61,62} lipids⁶³ and adiponectin⁶⁴ have successfully identified loci, a large exome sequencing/exome chip study has revealed that coding variation does not play a large role in T2D susceptibility.⁶⁵

In more recent years, GoDARTS has been used in Mendelian randomisation studies to establish the causal relationship between a variable and disease. This has been used to demonstrate a role for sex hormone-binding globulin⁶⁶ and to rule out a causal effect between circulating triglycerides and adiponectin in T2D.^{67,68} In addition, adiposity and adiponectin have been implicated in moderating CVD risk.^{69,70} Other studies have been conducted, and a more comprehensive list of disease phenotypes can be found in Table 3.

What are the strengths and weaknesses?

The main strengths of GoDARTS are: its large size [including 10 149 participants with T2D (~51% of people with T2D in Tayside) and 8157 controls]; the availability of rich genetic and phenotypic data; the ability to link patient genetic and baseline data to routine electronic medical records; and the existing consent for use of these for research and for future contact for possible research participation, especially the potential for recruitment by genotype studies. Consent for future contact has allowed further studies to take place. One of these is DOLORisk, an EU Horizon 2020-funded project that will be re-phenotyping

Table 3. Summary of phenotypes studied using the GoDARTS cohort

Phenotype	Study reference
ACEi-induced cough	(71)
Adiponectin levels	(64)
ARM D and diabetic retinal disease in T2D	(72)
Diabetic atherogenic lipid profile and myocardial infarction	(73)
Diabetic chronic kidney disease	(74)
Diabetic glomerular filtration rate	(75)
Diabetic left ventricular hypertrophy	(46)
Diabetic myocardial infarction	(76, 77)
Diabetic nephropathy	(78)
Diabetic neuropathic pain	(79, 80)
Diabetic retinopathy	(81)
Diabetic smoking-related cardiovascular morbidity	(82)
Diabetic statin intolerance	(83)
Glycaemic response to metformin	(84–86)
Glycaemic response to sulphonylureas	(37, 87, 88)
Glycaemic traits	(29, 61, 89)
Intolerance to metformin in T2D	(90)
LDLc response to statin therapy and CAD during statin treatment	(91)
Lipid response to statin therapy	(40, 45, 92)
Prostate cancer	(93)
Serum triglyceride level, insulin resistance and T2D in severe obesity	(94)
Serum urate concentration, excretion and gout	(95)
Stroke in T2D	(96)
T2D	(31, 35, 50, 98)

ACEi, angiotensin-converting enzyme inhibitor; ARM D, age-related macular degeneration; CAD, coronary artery disease; LDLc, low-density lipoprotein cholesterol; T2D, type 2 diabetes.

participants for neuropathic pain and related traits, to identify possible risk factors [<http://dolorisk.eu/>]. As a result, the GoDARTS cohort is rich in longitudinal phenotypic data, such as biochemistry, prescribing, morbidity and demography. In addition the linkage of the study to participants' individual electronic medical records, which are constantly updated, means the cohort is not limited by loss-to-follow up bias that can beset other longitudinal studies. This linkage is made possible through the use of the CHI number which allows patient data to remain anonymous. The availability of a large range of clinical and demographic data allows a large range of diabetes-related phenotypes to be investigated, on both genome-wide and more targeted scales, and also provides a means to control for common confounders such as BMI, smoking, age and blood pressure. The large number of samples recruited at baseline (8697 T2D cases and 8141 controls) provides the

statistical power with which to identify genetic variants conferring susceptibility to disease, both as a stand-alone cohort and as part of meta-analyses. Furthermore, RNA has been collected from patient blood samples, which will enable gene expression studies to be conducted in the future.

A weakness of the GoDARTS cohort is the missing baseline data of GoDARTS1 patients who were not subsequently recruited again in GoDARTS1 or 2. This is due to GoDARTS1 being a pilot phase of the study, and consequently baseline data were not collected at this stage. However, linkage is still possible for these samples. Patients were also not necessarily recruited at the point of diagnosis, which creates heterogeneity in the effects of disease duration on the serum samples obtained. Another weakness concerns the lifestyle questionnaire that was administered to patients. As this was self-completed by the patients, physical activity, smoking and female menopausal history are subject to recall bias.

Where can I find out more?

Further information about this cohort can be found at [<http://diabetesgenetics.dundee.ac.uk/>]. Access to the dataset is available to researchers worldwide and access requests, together with the general management of the resource, are handled by the Access Group. More details on the application and collaboration process can be found at [<http://diabetesgenetics.dundee.ac.uk/Community.aspx>].

Profile in a nutshell

- GoDARTS was set up in 1998 in order to study the genetics underpinning type 2 diabetes (T2D) susceptibility, diabetes complications and patient response to therapy.
- The study is a branch of the pre-existing Diabetes Audit and Research in Tayside Scotland study, which was set up to identify all patients within the Tayside area with diabetes, through electronic record linkage, to provide better care over and above existing registries.
- As of 2014, the study has 18 306 participants aged 16–98, of whom 10 149 have T2D and 8157 are controls. Baseline data are available for 16 838 (8698 cases and 8140 controls), and 8564 T2D cases and 4586 controls have genetic data.
- Baseline data collection includes a self-completed lifestyle questionnaire containing items on physical activity, smoking history, and menopausal history for women. In addition, clinical observations were

recorded and blood and urine samples were taken. Baseline data are linked to existing NHS records, providing morbidity, mortality and prescribing data by electronic record linkage to enable long-term follow-up.

- Consent has been provided by ~95% of participants to be re-contacted regarding possible participation in future studies. Information on collaboration and data access can be found at [<http://diabetesgenetics.dundee.ac.uk/>]. A list of GoDARTS publications can be found at [www.researcherid.com/rid/K-9448-2016].

Funding

The Wellcome Trust United Kingdom Type 2 Diabetes Case Control Collection (supporting GoDARTS) was funded by the Wellcome Trust (072960/Z/03/Z, 084726/Z/08/Z, 084727/Z/08/Z, 085475/Z/08/Z, 085475/B/08/Z) and as part of the EU IMI-SUMMIT programme. C.P. has received grant funding from the Wellcome Trust to develop the GoDARTS cohort. E.P. holds a WT New Investigator award (102820/Z/13/Z). B.H.S., C.N.A.P. and W.M. are members of the DOLORisk consortium which is funded by the European Commission Horizon 2020 (ID:633491), and are partly supported by this grant. H.L.H. and A.V. are supported by DOLORisk.

Acknowledgements

We are grateful to all the participants in this study, the general practitioners, the Scottish School of Primary Care for their help in recruiting the participants and to the whole team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The study complies with the Declaration of Helsinki. We acknowledge the support of the Health Informatics Centre, University of Dundee, for managing and supplying the anonymized data and NHS Tayside, the original data owner.

Conflict of interest: None declared.

References

1. Amos AF, McCarty DJ, Zimmet P. The rising global burden of diabetes and its complications: estimates and projections to the year 2010. *Diabet Med* 1997;14:S1–5.
2. International Diabetes Federation. *IDF Diabetes Atlas*. 7th edn. Brussels: International Diabetes Federation, 2015.
3. Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabet Med* 2012;29:855–62.
4. Abdullah A, Peeters A, de Courten M, Stoelwinder J. The magnitude of association between overweight and obesity and the risk of diabetes: a meta-analysis of prospective cohort studies. *Diabetes Res Clin Pract* 2010;89:309–19.
5. Sarwar N, Gao P, Seshasai SR *et al.*; Emerging Risk Factors Collaboration. Diabetes mellitus, fasting glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet* 2010;375:2215–22.
6. Maric-Bilkan C. Obesity and diabetic kidney disease. *Med Clin North Am* 2013;97:59–74.

7. Bartoli F, Carrà G, Crocamo C *et al.* Association between depression and neuropathy in people with type 2 diabetes: a meta-analysis. *Int J Geriatr Psychiatry* 2016;**31**:829–36.
8. Medici F, Hawa M, Ianari A, Pyke DA, Leslie RD. Concordance rate for type II diabetes mellitus in monozygotic twins: actuarial analysis. *Diabetologia* 1999;**42**:146–150.
9. Sladek R, Rocheleau G, Rung J *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;**445**:881–85.
10. Dorajoo R, Liu J, Boehm BO. Genetics of Type 2 diabetes and clinical utility. *Genes (Basel)* 2015;**6**:372–84.
11. Bonnefond A, Froguel P. Rare and common genetic events in type 2 diabetes: what should biologists know?. *Cell Metab* 2015;**21**:357–68.
12. Morris AD, Boyle DI, MacAlpine R *et al.* The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. DARTS/MEMO Collaboration. *BMJ* 1997;**315**:524–28.
13. Smith BH, Campbell A, Linksted P *et al.* Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol* 2013;**42**:689–700.
14. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2011;**9**:179–81.
15. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;**5**:e1000529.
16. Auton A, Brooks LD, Durbin RM *et al.*; 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;**526**:68–74.
17. Cortes A, Brown M. Promise and pitfalls of the Immunochip. *Arthritis Res Ther* 2011;**13**:101.
18. Voight BF, Kang HM, Ding J *et al.* The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet* 2012;**8**:e1002793.
19. Guo Y, He J, Zhao S *et al.* Illumina human exome genotyping array clustering and quality control. *Nat Protoc* 2014;**9**:2643–62.
20. Doney A, Fischer B, Frew D *et al.* Haplotype analysis of the PPARGgamma Pro12Ala and C1431T variants reveals opposing associations with body weight. *BMC Genet* 2002;**3**:21.
21. Doney AS, Fischer B, Cecil JE *et al.* Association of the Pro12Ala and C1431T variants of PPARG and their haplotypes with susceptibility to Type 2 diabetes. *Diabetologia* 2004;**47**:555–58.
22. Doney AS, Fischer B, Leese G, Morris AD, Palmer CN. Cardiovascular risk in type 2 diabetes is associated with variation at the PPARG locus: a Go-DARTS study. *Arterioscler Thromb Vasc Biol* 2004;**24**:2403–07.
23. Cecil JE, Fischer B, Doney AS *et al.* The Pro12Ala and C-681G variants of the PPARG locus are associated with opposing growth phenotypes in young schoolchildren. *Diabetologia* 2005;**48**:1496–502.
24. Doney AS, Fischer B, Lee SP, Morris AD, Leese G, Palmer CN. Association of common variation in the PPARA gene with incident myocardial infarction in individuals with type 2 diabetes: a Go-DARTS study. *Nucl Recept* 2005;**3**:4.
25. Burch LR, Zhou K, Donnelly LA *et al.* A single nucleotide polymorphism on exon-4 of the gene encoding PPAR δ is associated with reduced height in adults and children. *J Clin Endocrinol Metab* 2009;**94**:2587–93.
26. Frayling TM, Timpson NJ, Weedon MN *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007;**316**:889–94.
27. Loos RJ, Lindgren CM, Li S *et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008;**40**:768–75.
28. Zeggini E, Scott LJ, Saxena R *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008;**40**:638–45.
29. Dupuis J, Langenberg C, Prokopenko I *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010;**42**:105–16.
30. Zeggini E, Weedon MN, Lindgren CM *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007;**316**:1336–41.
31. Voight BF, Scott LJ, Steinthorsdottir V *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010;**42**:579–89.
32. Saxena R, Hivert MF, Langenberg C *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 2010;**42**:142–48.
33. Weedon MN, Lettre G, Freathy RM *et al.* A common variant of HMG2 is associated with adult and childhood height in the general population. *Nat Genet* 2007;**39**:1245–50.
34. Weedon MN, Lango H, Lindgren CM *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 2008;**40**:575–83.
35. Sandhu MS, Weedon MN, Fawcett KA *et al.* Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* 2007;**39**:951–53.
36. Prokopenko I, Langenberg C, Florez JC *et al.* Variants in MTNR1B influence fasting glucose levels. *Nat Genet* 2009;**41**:77–81.
37. Pearson ER, Donnelly LA, Kimber C *et al.* Variation in TCF7L2 influences therapeutic response to sulfonylureas: a GoDARTS study. *Diabetes* 2007;**56**:2178–82.
38. Donnelly LA, Palmer CN, Whitley AL *et al.* Apolipoprotein E genotypes are associated with lipid-lowering responses to statin treatment in diabetes: a Go-DARTS study. *Pharmacogenet Genomics* 2008;**18**:279–87.
39. Zhou K, Bellenguez C, Spencer CC, *et al.*; GoDARTS and UKPDS Diabetes Pharmacogenetics Study Group, Wellcome Trust Case Control Consortium 2. Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. *Nat Genet* 2011;**43**:117–20.
40. Postmus I, Trompet S, Deshmukh HA *et al.* Pharmacogenetic meta-analysis of genome-wide association studies of LDL cholesterol response to statins. *Nat Commun* 2014;**5**:5068.
41. van Leeuwen N, Nijpels G, Becker ML *et al.* A gene variant near ATM is significantly associated with metformin treatment

- response in type 2 diabetes: a replication and meta-analysis of five cohorts. *Diabetologia* 2012;**55**:1971–77.
42. Zhou K, Yee SW, Seiser EL *et al.* Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin. *Nat Genet* 2016;**48**:1055–59.
 43. Dawed AY, Donnelly L, Tavendale R *et al.* CYP2C8 and SLCO1B1 variants and therapeutic response to thiazolidinediones in patients with type 2 diabetes. *Diabetes Care* 2016;**39**:1902.
 44. Zhou K, Donnelly L, Burch L *et al.* Loss-of-function CYP2C9 variants improve therapeutic response to sulfonylureas in type 2 diabetes: a Go-DARTS study. *Clin Pharmacol Ther* 2010;**87**: 52–56.
 45. Donnelly LA, Doney AS, Dannfald J *et al.* A paucimorphic variant in the HMG-CoA reductase gene is associated with lipid-lowering response to statin treatment in diabetes: a GoDARTS study. *Pharmacogenet Genomics* 2008;**18**:1021–26.
 46. Parry HM, Donnelly LA, Van Zuydam N *et al.* Genetic variants predicting left ventricular hypertrophy in a diabetic population: a Go-DARTS study including meta-analysis. *Cardiovasc Diabetol* 2013;**12**:109.
 47. Parry HM, Deshmukh H, Levin D *et al.* Both high and low HbA1c predict incident heart failure in type 2 diabetes mellitus. *Circ Heart Fail* 2015;**8**:236–42.
 48. Morris AP, Voight BF, Teslovich TM *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012;**44**:981–90.
 49. DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, South Asian Type 2 Diabetes (SAT2D) Consortium *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* 2014;**46**: 234–44.
 50. Gaulton KJ, Ferreira T, Lee Y *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet* 2015;**47**:1415–25.
 51. Nikpay M, Goel A, Won HH *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015;**47**:1121–30.
 52. Deloukas P, Kanoni S, Willenborg C *et al.*; CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* 2013;**45**:25–33.
 53. Scott RA, Lagou V, Welch RP *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 2012;**44**:991–1005.
 54. Berndt SI, Gustafsson S, Magi R *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* 2013;**45**: 501–12.
 55. Wood AR, Esko T, Yang J *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014;**46**:1173–86.
 56. Shungin D, Winkler TW, Croteau-Chonka DC *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 2015;**518**:187–96.
 57. Locke AE, Kahali B, Berndt SI *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;**518**: 197–206.
 58. Crosby J, Peloso GM, Auer PL *et al.*; TG and HDL Working Group of the Exome Sequencing Project, National Heart Lung and Blood Institute. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med* 2014;**371**: 22–31.
 59. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortium Investigators. Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. *N Engl J Med* 2016;**374**:1134–44.
 60. Stitzel NO, Won HH, Morrison AC *et al.*; Myocardial Infarction Genetics Consortium Investigators. Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med* 2014;**371**:2072–82.
 61. Mahajan A, Sim X, Ng HJ *et al.* Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. *PLoS Genet* 2015;**11**:e1004876.
 62. Albrechtsen A, Grarup N, Li Y *et al.* Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* 2013;**56**:298–310.
 63. Kanoni S, Masca NGD, Stirrups KE *et al.* Analysis with the exome array identifies multiple new independent variants in lipid loci. *Hum Mol Genet* 2016;**25**:4094–106.
 64. Warren LL, Li L, Nelson MR *et al.* Deep resequencing unveils genetic architecture of ADIPOQ and identifies a novel low-frequency variant strongly associated with adiponectin variation. *Diabetes* 2012;**61**:1297–301.
 65. Fuchsberger C, Flannick J, Teslovich TM *et al.* The genetic architecture of type 2 diabetes. *Nature* 2016;**536**:41–47.
 66. Perry JR, Weedon MN, Langenberg C *et al.* Genetic evidence that raised sex hormone binding globulin (SHBG) levels reduce the risk of type 2 diabetes. *Hum Mol Genet* 2010;**19**: 535–44.
 67. Yaghootkar H, Lamina C, Scott RA *et al.* Mendelian randomization studies do not support a causal role for reduced circulating adiponectin levels in insulin resistance and type 2 diabetes. *Diabetes* 2013;**62**:3589–98.
 68. De Silva NM, Freathy RM, Palmer TM *et al.* Mendelian randomization studies do not support a role for raised circulating triglyceride levels influencing type 2 diabetes, glucose levels, or insulin resistance. *Diabetes* 2011;**60**:1008–18.
 69. Fall T, Hagg S, Magi R *et al.* The role of adiposity in cardiometabolic traits: a Mendelian randomization analysis. *PLoS Med* 2013;**10**:e1001474.
 70. Persson J, Strawbridge RJ, McLeod O *et al.* Sex-specific effects of adiponectin on carotid intima-media thickness and incident cardiovascular disease. *J Am Heart Assoc* 2015;**4**: e001853.
 71. Mosley J, Shaffer C, Van Driest S *et al.* A genome-wide association study identifies variants in KCNIP4 associated with ACE inhibitor-induced cough. *Pharmacogenomics J* 2015;doi: 10.1038/tj.2015.51.
 72. Doney A, Leese G, Olson J, Morris A, Palmer CNA. The Y402H variant of complement factor H is associated with age-related

- macular degeneration but not with diabetic retinal disease in the Go-DARTS study. *Diabet Med* 2009;26:460–65.
73. Doney A, Dannfald J, Kimber CH *et al*. The FTO gene is associated with an atherogenic lipid profile and myocardial infarction in patients with type 2 diabetes: a Genetics of Diabetes Audit and Research Study in Tayside Scotland (Go-DARTS) study. *Circ Cardiovasc Genet* 2009;2:255–59.
 74. Looker HC, Colombo M, Hess S *et al*. Biomarkers of rapid chronic kidney disease progression in type 2 diabetes. *Kidney Int* 2015;88:888–96.
 75. Deshmukh HA, Palmer CNA, Morris AD, Colhoun HM. Investigation of known estimated glomerular filtration rate loci in patients with type 2 diabetes. *Diabet Med* 2013;30:1230–35.
 76. Doney A, Fischer B, Lee S, Morris A, Leese G, Palmer CNA. Association of common variation in the PPARA gene with incident myocardial infarction in individuals with type 2 diabetes: a Go-DARTS study. *Nucl Recept* 2005;3:4.
 77. Doney A, Fischer B, Leese G, Morris A, Palmer CNA. Cardiovascular risk in type 2 diabetes is associated with variation at the PPARG locus: a Go-DARTS study. *Arterioscler Thromb Vasc Biol* 2004;24:2403–07.
 78. Alkayyali S, Lajer M, Deshmukh H *et al*. Common variant in the HMG2A gene increases susceptibility to nephropathy in patients with type 2 diabetes. *Diabetologia* 2013;56:323–29.
 79. Meng W, Deshmukh HA, van Zuydam NR *et al*. A genome-wide association study suggests an association of Chr8p21.3 (GFRA2) with diabetic neuropathic pain. *Eur J Pain* 2015;19:392–99.
 80. Meng W, Deshmukh HA, Donnelly LA; Wellcome Trust Case Control Consortium 2 (WTCCC2), Surrogate markers for Micro-and Macro-vascular hard endpoints for Innovative diabetes Tools (SUMMIT) study group. A genome-wide association study provides evidence of sex-specific involvement of chr1p35.1 (ZSCAN20-TLR12P) and chr8p23.1 (HMGB1P46) with diabetic neuropathic pain. *EBioMedicine* 2015;2:1386–93.
 81. Liu Y, Wang M, Morris A, Doney A, Leese G, Pearson E *et al*. Glycemic exposure and blood pressure influencing progression and remission of diabetic retinopathy: a longitudinal cohort study in GoDARTS. *Diabetes Care* 2013;36:3979–84.
 82. Doney A, Lee S, Leese G, Morris A, Palmer CNA. Increased cardiovascular morbidity and mortality in type 2 diabetes is associated with the glutathione S transferase theta-null genotype: a Go-DARTS study. *Circulation* 2005;111:2927–34.
 83. Donnelly LA, Doney A, Tavendale R *et al*. Common nonsynonymous substitutions in SLCO1B1 predispose to statin intolerance in routinely treated individuals with type 2 diabetes: a Go-DARTS study. *Clin Pharmacol Ther* 2011;89:210–16.
 84. Zhou K, Donnelly L, Yang J *et al*. Heritability of variation in glycaemic response to metformin: a genome-wide complex trait analysis. *Lancet Diabetes Endocrinol* 2014;2:481–87.
 85. Zhou K, Bellenguez C, Spencer CCA *et al*. GoDARTS and UKPDS Diabetes Pharmacogenetics Study Group, Wellcome Trust Case Control Consortium 2. Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. *Nat Genet* 2011;43:117–20.
 86. Zhou K, Donnelly LA, Kimber CH *et al*. Reduced-function SLC22A1 polymorphisms encoding organic cation transporter 1 and glycemic response to metformin: a GoDARTS study. *Diabetes* 2009;58:1434–39.
 87. Grant R, Wexler D. Loss-of-function CYP2C9 variants: finding the correct clinical role for Type 2 diabetes pharmacogenetic testing. *Expert Rev Cardiovasc Ther* 2010;8:339–43.
 88. Zhou K, Donnelly L, Burch L *et al*. Loss-of-function CYP2C9 variants improve therapeutic response to sulfonylureas in type 2 diabetes: a Go-DARTS study. *Clin Pharmacol Ther* 2010;87:52–56.
 89. Scott RA, Lagou V, Welch RP *et al*. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 2012;44:991–1005.
 90. Dujic T, Zhou K, Donnelly LA, Tavendale R, Palmer CNA, Pearson E. Association of organic cation transporter 1 with intolerance to metformin in type 2 diabetes: a GoDARTS study. *Diabetes* 2015;64:1786–93.
 91. Donnelly LA, van Zuydam N, Zhou K *et al*. Robust association of the LPA locus with LDLc lowering response to statin treatment in a meta-analysis of 30,467 individuals from both randomised control trials and observational studies and association with coronary artery disease outcome during statin treatment. *Pharmacogenet Genomics* 2013;23:518–25.
 92. Donnelly LA, Palmer CNA, Whitley AL *et al*. Apolipoprotein E genotypes are associated with lipid-lowering responses to statin treatment in diabetes: a Go-DARTS study. *Pharmacogenet Genomics* 2008;18:279–87.
 93. Gudmundsson J, Sulem P, Steinthorsdottir V *et al*. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 2007;39:977–83.
 94. Palmer CNA, Maglio C, Pirazzi C *et al*. Paradoxical lower serum triglyceride levels and higher type 2 diabetes mellitus susceptibility in obese individuals with the PNPLA3 148M variant. *PLoS One* 2012;7:e39362.
 95. Vitart V, Rudan I, Hayward C *et al*. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat Genet* 2008;40:437–42.
 96. Palmer CNA, Kimber CH, Doney A *et al*. Combined effect of inflammatory gene polymorphisms and the risk of ischemic stroke in a prospective cohort of subjects with type 2 diabetes: a Go-DARTS study. *Diabetes* 2010;59:2945–48.
 97. Kimber CH, Doney A, Pearson E *et al*. TCF7L2 in the Go-DARTS study: evidence for a gene dose effect on both diabetes susceptibility and control of glucose levels. *Diabetologia* 2007;50:1186–91.
 98. Perry JRB, Voight BF, Yengo L *et al*. Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared with obese cases. *PLoS Genet* 2012;8:e1002741.