

Excess maternal transmission of variants in the *THADA* gene to offspring with type 2 diabetes

Rashmi B. Prasad¹ · Anna Lessmark¹ · Peter Almgren¹ · Györgyi Kovacs² · Ola Hansson¹ · Nikolay Oskolkov¹ · Marta Vitai² · Claes Ladenvall¹ · Peter Kovacs^{3,4} · Joao Fadista¹ · Michael Lachmann⁵ · Yuedan Zhou¹ · Emily Sonestedt¹ · Wenny Poon¹ · Claes B. Wollheim^{1,6} · Marju Orho-Melander¹ · Michael Stumvoll^{3,4} · Tiinamaija Tuomi^{7,8} · Svante Pääbo⁹ · Laszlo Koranyi² · Leif Groop^{1,10}

Received: 26 November 2015 / Accepted: 1 April 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract

Aims/hypothesis Genome-wide association studies (GWAS) have identified more than 65 genetic loci associated with risk of type 2 diabetes. However, the contribution of distorted parental transmission of alleles to risk of type 2 diabetes has been mostly unexplored. Our goal was therefore to search for parent-of-origin effects (POE) among type 2 diabetes loci in families.

Methods Families from the Botnia study ($n=4,211$, 1,083 families) were genotyped for 72 single-nucleotide polymorphisms (SNPs) associated with type 2 diabetes and assessed for POE on type 2 diabetes. The family-based Hungarian Transdanubian Biobank (HTB) ($n=1,463$, >135 families) was used to replicate SNPs showing POE. Association of type 2 diabetes loci within families was also tested.

Results Three loci showed nominal POE, including the previously reported variants in *KCNQ1*, for type 2 diabetes in families from Botnia (rs2237895: $p_{POE}=0.037$), which can be considered positive controls. The strongest POE was seen for rs7578597 SNP in the *THADA* gene, showing excess transmission of the maternal risk allele T to diabetic offspring (Botnia: $p_{POE}=0.01$; HTB $p_{POE}=0.045$). These data are consistent with previous evidence of allelic imbalance for expression in islets, suggesting that the *THADA* gene can be imprinted in a POE-specific fashion. Five CpG sites, including those flanking rs7578597, showed differential methylation between diabetic and non-diabetic donor islets.

Conclusions/interpretation Taken together, the data emphasise the need for genetic studies to consider from which parent an offspring has inherited a susceptibility allele.

Electronic supplementary material The online version of this article (doi:10.1007/s00125-016-3973-9) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

✉ Rashmi B. Prasad
Rashmi.Prasad@med.lu.se

✉ Leif Groop
Leif.Groop@med.lu.se

¹ Lund University Diabetes Centre, Department of Clinical Sciences, Lund University CRC, Skåne University Hospital Malmö, SE-205 02 Malmö, Sweden

² Heart Center Foundation, DRC, Balatonfüred, Hungary

³ Department of Medicine, University of Leipzig, Leipzig, Germany

⁴ IFB Adiposity Diseases, University of Leipzig, Leipzig, Germany

⁵ Santa Fe Institute, Santa Fe, NM, USA

⁶ Department of Cell Physiology and Metabolism, University Medical Center, Geneva, Switzerland

⁷ Folkhälsan Research Centre, Helsinki, Finland

⁸ Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland

⁹ Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

¹⁰ Finnish Institute of Molecular Medicine (FIMM), Helsinki University, Helsinki, Finland

Keywords Families · Genetic association studies · *KCNQ1* · Maternal effects · Methylation · Parental transmission · Parent-of-origin · *THADA* · Type 2 diabetes

Abbreviations

FBAT	Family-based association test
GAD	Glutamic acid decarboxylase autoantibody
GWAS	Genome-wide association studies
HTB	Hungarian Transdanubian Biobank
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance
LD	Linkage disequilibrium
PBL	Peripheral blood lymphocytes
POE	Parent-of-origin effect(s)
SNP	Single-nucleotide polymorphism
TDT	Transmission disequilibrium tests

Introduction

Type 2 diabetes, a lifelong debilitating disorder with rapidly increasing prevalence [1], is characterised by chronic hyperglycaemia due to impaired insulin secretion and sensitivity. This epidemic has been ascribed to a collision between genetic susceptibility and change in the environment [2].

Whereas environmental triggers are well known, it has been more difficult to dissect the genetic causes of type 2 diabetes. Heritability of type 2 diabetes estimated as sibling relative risk (λ_S) is ~ 3 in most unrelated populations but as high as 8 if there are several affected siblings in the family [3, 4]. Lifetime risk of diabetes in an offspring of a parent with type 2 diabetes is $\sim 40\%$ and higher if the mother rather than the father has type 2 diabetes [5]. We have previously observed sex-specific parent-of-origin effects (POE) on intermediate traits like insulin secretion and lipids in offspring of parents with type 2 diabetes [5, 6]. Genome-wide association studies (GWAS) have revealed >65 loci associated with type 2 diabetes and related traits, although they account for less than 15% of the heritability of type 2 diabetes [2, 7–12]. Part of the missing heritability could be explained by POE, wherein a certain allele might confer risk when inherited from one parent but might be neutral or even protective when inherited from the other parent [13]. Such a situation would be missed in regular GWAS where it is assumed that both parental alleles are transmitted equally. The most plausible explanation for POE is methylation and, consequently, imprinting of one of the parental alleles. If this occurs during the fetal period, it could influence development of functional beta cell mass, reducing the ability of an individual to increase

insulin secretion when exposed to an affluent westernised environment. Kong et al reported POE of four imprinted regions, including rs4731702 (*KLF14*), rs2334499 (near *MOB2*), rs231362 and rs2237892 (*KCNQ1*) [14], and this finding has been replicated in subsequent studies [15].

Whereas information on imprinting was the starting point for the above studies, little information is available on whether other single-nucleotide polymorphisms (SNPs) associated with type 2 diabetes or glycaemic traits show POE transmission and could thereby contribute to the excess risk observed with maternally transmitted diabetes. Such studies have been hampered by the paucity of family-based cohorts. To address this question, we assessed POE transmission of 72 SNPs associated with type 2 diabetes from previous GWAS studies (based upon results up to 2012) corresponding to 65 unique loci in two large family studies, the Botnia study from western Finland and southern Sweden and the Hungarian Transdanubian Biobank (HTB).

Methods

Study population and measurements

The Botnia study was initiated in 1990 in healthcare centres in the Botnia region in western Finland and subsequently extended to other parts of Finland and southern Sweden [5]. For the current study, all individuals from complete trios (DNA from both parents and at least one child) were selected, yielding 4,211 individuals, forming 2,322 trios from 1,083 families after exclusion of patients with type 1 diabetes, glutamic acid decarboxylase autoantibody (GAD) positivity or MODY (Table 1). Of these, 25.7% had a type 2 diabetes diagnosis based on the WHO 1998 criteria. Height, weight and waist and hip circumference were recorded and OGTT performed as reported previously [3]. Blood samples were drawn 10 min before the OGTT and then at 0, 30, 60 and 120 min. Plasma insulin was measured by ELISA (Dako, Ely, UK) [3, 5].

The HTB was initiated in 1992 at the Hungarian Heart Center in Balatonfüred. Samples and data were collected from type 2 diabetes patients and their families. The HTB includes 9,279 individuals from 1,022 families from the west side of the Danube area in Hungary. For the current study, type 2 diabetes, GAD-positive and MODY individuals were excluded and 1,463 individuals were genotyped, of whom 29.3% had a type 2 diabetes diagnosis based on the WHO 1998 criteria (Table 1).

Informed consent was obtained from all the participants. The study protocols were approved by the ethics committees of Helsinki University Central Hospital and Lund University for the Botnia study and by local ethics committees for the HTB.

Table 1 Clinical and metabolic characteristics of the study population

Characteristic	Botnia	HTB
Total no. of individuals	4,212	1,463
<i>n</i> (male sex/female sex)	2,090/2,122	682/781
Age (years) (mean ± SD)	45.53 ± 18.36	49.78 ± 19.89
BMI (kg/m ²) (mean ± SD)	26.03 ± 4.85	26.5 ± 5.45
T2D (<i>n</i>)	1,081	428
T2D + severe IGT	1,183	
T2D + severe IGT + severe IFG	1,095	
T2D + hyperglycaemia	1,723	
Parents		
Total no. of fathers	1,083	320
Total no. of mothers	1,083	320
Offspring		
Total no. of offspring	2,188	364
Total no. of offspring after excluding T1D, MODY	1,982	364
Offspring with T2D	246	153
Offspring with T2D + severe IGT	253	
Offspring with T2D + severe IGT + severe IFG	302	
Offspring with T2D + hyperglycaemia	540	

The study population comprised of families from the Botnia Study (southern Sweden and western Finland) and the HTB

T1D, type 1 diabetes; T2D, type 2 diabetes

Glucose tolerance status

Glucose tolerance and type 2 diabetes was based on the WHO criteria [16]. Hyperglycaemia included type 2 diabetes + impaired fasting glucose (IFG) + impaired glucose tolerance (IGT) (fasting glucose ≥ 6.1 mmol/l or 2 h plasma glucose ≥ 7.8 mmol/l). Normal glucose tolerance was defined as FPG < 6.1 mmol/l and 2 h plasma glucose < 7.8 mmol/l. For intermediate phenotype definitions, see electronic supplementary material (ESM) [Methods](#).

Genotyping

Seventy-two SNPs from 65 loci which have shown consistent association with type 2 diabetes and related traits based upon GWAS data available in 2012 were selected for this study (ESM Table 1). Genotyping was performed on the Sequenom MassARRAY iPLEX Platform. The average genotyping success rate was $>98\%$ with $>98\%$ concordance in $\sim 6.5\%$ replication samples. SNPs rs12779790, rs553668, rs231362 and rs2237892 were genotyped with an allelic discrimination assay-by-design method on ABI 7900 (Applied Biosystems, Warrington, UK). All SNPs were in Hardy–Weinberg equilibrium in unaffected parents ($p > 0.001$) with the exception of SNPs rs17782313 and rs6467136 ($p \geq 0.001$), which were excluded from analyses.

Statistical analyses

Family-based tests for association Association was tested by family-based association tests (FBATs), which accommodate any type of genetic model and family construction [17, 18], and transmission disequilibrium test (TDT) using the PLINK software (<http://pngu.ngh.harvard.edu/~purcell/plink/>) [19, 20]. A conservative power calculation on the present set of families suggests that the power to detect ORs of 1.5 using TDT at levels of type I error of 0.05 under the null hypothesis is 58%, 69% and 63% for risk alleles with frequencies of 0.1, 0.2 and 0.3, respectively [21].

Mendelian errors within families were identified by Pedstats and PLINK. To increase power we used ParentTDT [22]. To obtain empirical *p* values, 10,000 gene-dropping adaptive permutations were performed.

To evaluate the association of SNPs with continuous traits, non-diabetic nuclear families were considered; all measures of glycaemia, insulin secretion and sensitivity were log-transformed. The number of alleles shared identical-by-descent (IBD estimates) was computed using Merlin (Merlin 1.1.2, <http://csg.sph.umich.edu/abecasis/merlin/tour/ibd.html>). Variance components including environmental, polygenic and additive were used to model the phenotypic similarities using maximum likelihood in the orthogonal model [23]. All measurements were adjusted for age, sex and BMI.

Parent-of origin test Parent-of-origin tests were performed to assess distorted transmission of risk alleles from each of the parents separately to the affected offspring using PLINK [20]. The difference between the transmissions of alleles from the father and mother was computed as a Z statistic and resulting p values were subjected to 10,000 gene-drop adaptive permutations to generate empirical p values while controlling for familial relationships.

To assess maternal and paternal effects on continuous traits separately in nuclear families, the orthogonal model was implemented after modelling for environmental, polygenic and additive variances using maximum likelihood approach using qTDT [23]. Analyses for BMI were adjusted for age and sex while other measurements were adjusted for age, sex and BMI. For all analyses, 10,000 Monte Carlo permutations were performed to obtain empirical p values [23]. Haplotype analyses for association with type 2 diabetes and parental asymmetry tests were performed using FAMHAP [24] using three-marker and four-marker haplotypes (for *KCNQ1* variants).

Methylation data Methylation analyses were performed on peripheral blood lymphocytes and islets and expression studies on the latter (see ESM Methods). The *THADA* variant showing a genetic POE was selected for analyses. In trios, β values were used to assess association of methylation with genotype taking parental origin into account. Differences in methylation between the two groups at individual CpG sites were assessed using limma package in R Bioconductor [25] for larger groups and Wilcoxon rank sum test for small test groups. The β values at CpG sites were mapped to the genes of interest and plotted to provide an overview of methylation status for the region of interest. The ratio of methylated CpGs per gene to total number of CpG probes was calculated taking into account genes with fewer CpG probes as follows: (number of differentially methylated CpG probes)/ $\sqrt{\text{total number of CpG probes}}$ to assess for uniqueness in patterns.

Disease association using case–control data The *THADA* variant showing POE was also assessed for association with type 2 diabetes, metabolic, anthropometric and dietary traits in case–control studies from MCC, PPP and the Leipzig cohorts. See ESM Methods.

Results

Family-based replication of variants previously associated with type 2 diabetes in case–control studies

We first assessed whether the effects of the reported loci could be replicated in our families. Of the 72 type 2 diabetes risk

variants tested (ESM Table 2) in 4,189 individuals from 1,083 families from the Botnia study, 18 variants showed association with type 2 diabetes and 16 with hyperglycaemia (type 2 diabetes + IFG + IGT) in at least one of the tests (TDT or FBAT) (Table 2 and ESM Table 2 a, b). With the exception of rs9470794 (intron of *ZFAND3*), the direction of risk was concordant with previously reported GWAS studies for all phenotypes tested. The rs163184 (*KCNQ1*) variant showed a strong association with type 2 diabetes ($p_{FBAT}=0.0028$, $p_{TDT}=0.002$) and hyperglycaemia ($p_{FBAT}=0.021$, $p_{TDT}=0.014$) in all the tests performed. rs7903146 (*TCF7L2*) and rs10885122 (near *ADRA2A*) showed an association with type 2 diabetes ($p_{FBAT}=0.007$ and 0.021 and $p_{TDT}=0.01$ and 0.03 , respectively) in all the tests, with the former also being associated with hyperglycaemia ($p_{COM}=0.012$). rs10010131 (*WFS1*) and rs6017317 (*HNF4A*) were associated with type 2 diabetes and hyperglycaemia in FBAT and TDT ($p_{FBAT}=0.016$ and 0.024 and $p_{TDT}=0.02$ and 0.02 , respectively) (Table 2).

The rs7957197 in *OASL* was associated with risk of type 2 diabetes ($p_{FBAT}=0.028$) (Table 2) whereas rs864745 variant in *JAZF1* and rs1801282 in the *PPARG* gene was associated with hyperglycaemia ($p_{COM}=0.019$) (Table 2) in at least one of the tests.

POE on risk of type 2 diabetes and hyperglycaemia

We first studied locus-by-locus POE transmissions of all selected alleles to offspring (first, with type 2 diabetes and, second, with hyperglycaemia) among 4,189 individuals from 1,083 families from the Botnia study.

Of the 72 SNPs tested, variants in *THADA*, *KCNQ1* and *CRY2* showed nominal POE on transmission to diabetic offspring. Additionally, other variants in *KCNQ1* showed POE only on transmission to offspring with type 2 diabetes and can therefore be considered as a positive control, whereas those in *PRC1* and near *DGKB/TMEM5*, and *CDC123/CAMK1D* genes showed POE on transmission to hyperglycaemic offspring only. In the following sections, we have analysed these variants in more detail (Table 3).

KCNQ1 Consistent with previously reported data [14, 15], the G allele in rs163184 showed an increased transmission from mothers to diabetic offspring ($p_{MAT}=0.008$) and to hyperglycaemic offspring ($p_{MAT}=0.004$, $p_{POE}=0.016$) (Table 3 and Fig. 1). The rs2237895 SNP showed a stronger maternal than paternal transmission of risk allele C to diabetic offspring ($p_{MAT}=0.04$, $p_{POE}=0.037$) and hyperglycaemic offspring ($p_{MAT}=0.026$, $p_{POE}=0.018$) (Table 3 and Fig. 1). Another *KCNQ1* intronic variant rs2237892 showed excess maternal transmission of risk allele C to hyperglycaemic offspring ($p_{MAT}=0.041$) (Table 3 and Fig. 1). The four SNPs were in weak-to-moderate linkage disequilibrium (LD)

Table 2 Replication of association with type 2 diabetes and hyperglycaemia

CHR	SNP	Gene	Location	R/O	FBAT		TDT		ParenTDT				
					Var(S)	Z	T:U	OR	<i>P</i> _{TDT}	A:U_PAR	<i>P</i> _{PAR}	<i>P</i> _{COM}	
Type 2 diabetes													
11	rs163184	<i>KCNQ1</i>	Intron	G/T	51.750	2.989	0.0028**	124:81	1.53	0.002**	204:184	0.38	0.008**
10	rs7903146	<i>TCF7L2</i>	Intronic/promoter	T/C	49.500	2.701	0.0069**	106:69	1.41	0.01**	191:119	0.00013***	0.0003***
4	rs10010131	<i>WFS1</i>	Intron	G/A	64.306	2.390	0.016*	124:88	1.4	0.02*	202:190	0.59	0.851
10	rs10885122	<i>ADRA2A</i>	Intergenic	G/T	23.000	2.294	0.021*	60:39	1.53	0.03*	112:87	0.08	0.13
20	rs6017317	<i>HNF4A</i>	Intergenic	G/T	33.500	2.246	0.024*	89:63	1.41	0.02*	155:160	0.79	0.857
12	rs7957197	<i>OASL/TCF1/HNF1A</i>	Intron of OASL	T/A	56.750	2.190	0.028*	102:76	1.34	0.520	143:149	0.74	0.065
Hyperglycaemia													
4	rs10010131	<i>WFS1</i>	Intron	G/A	136.361	2.698	0.006**	247:193	1.28	0.019*	161:182	0.32	0.693
20	rs6017317	<i>HNF4A</i>	Intergenic	G/T	92.250	2.447	0.014*	190_144	1.31	0.017*	115:134	0.26	0.774
11	rs163184	<i>KCNQ1</i>	Intron	G/T	120.361	2.294	0.021*	254:199	1.28	0.014*	191:149	0.05	0.0005***
7	rs864745	<i>JAZF1</i>	Intron	T/C	116.694	1.682	0.092	261:224	1.17	0.08	192_167	0.25	0.035*
10	rs7903146	<i>TCF7L2</i>	Intronic/promoter	T/C	93.750	1.704	0.084	194:158	1.22	0.06	152:111	0.019*	0.012*
3	rs1801282	<i>PPARG</i>	Coding-missense	C/G	72.194	1.432	0.14	143:118	1.21	0.15	111:74	0.008**	0.019*

p* < 0.05, *p* < 0.01, ****p* < 0.001

A:U, no of risk alleles in the affected; *P*_{FBAT}: *p* value for the *Z* statistic; T:U, no. of risk alleles transmitted to untransmitted; *P*_{COM}: empirical *p* value for the parenTDT test; *P*_{PAR}: *p* value for the parental asymmetry test; *P*_{TDT}: pointwise permuted *p* value for the TDT test; R/O, risk allele/other allele; Var(S), empirical variance; Z, S normalised using E(S) and Var(S)

Table 3 Parental specific transmissions to diabetic offspring and hyperglycaemic offspring in families from the Botnia study

CHR	SNP	Gene	Location	R/O	T:U_PAT	p_{PAT}	T:U_MAT	p_{MAT}	Z_{POE}	p_{POE}
Type 2 diabetes										
2	rs7578597	THADA	Coding–missense	T:C	09:16	0.16	17:09	0.08	2.06	0.010**
11	rs11605924	CRY2	Intron	A:C	69:47	0.04*	60:62	0.85	-1.592	0.011*
11	rs2237895	KCNQ1	Intron	C:A	56.5:55.5	0.92	68.5:46.5	0.040*	-1.379	0.037*
11	rs163184	KCNQ1	Intron	G:T	67:50	0.12	62:36	0.008**	0.89	0.37
Hyperglycaemia										
2	rs7578597	THADA	Coding–missense	T:C	21.5:32.5	0.13	31.5:26.5	0.51	1.53	0.07
11	rs11605924	CRY2	Intron	A:C	128:99	0.054	135:129	0.71	-1.163	0.06
11	rs2237895	KCNQ1	Intron	C:A	112:113	0.95	135:101	0.026*	-1.596	0.018*
11	rs163184	KCNQ1	Intron	G:T	129.5: 115.5	0.37	124.5: 83.5	0.004**	1.494	0.016*
11	rs2237892	KCNQ1	Intron	C:T	24:21	0.65	20:09	0.041*	1.329	0.10
7	rs2191349	DGKB/TMEM195	Intergenic	T:G	108:105	0.84	100:130	0.047*	1.521	0.033*
15	rs8042680	PRCI	Intron	A:C	118.5:84.5	0.017*	117.5:120.5	0.84	1.887	0.013*
10	rs12779790	CDC123,CAMK1D	Intergenic	G:A	68.5:46.5	0.040*	64.5:71.5	0.54	1.915	0.009**

Paternal, maternal transmission tests and parent-of-origin tests are shown

* $p < 0.05$, ** $p < 0.01$

p_{MAT} : asymptotic p value for maternal transmission test; p_{PAT} : asymptotic p value for paternal transmission test; p_{POE} : permuted pointwise p value for the parent-of-origin test; R/O = risk allele/other allele; T:U PAT, no. of risk alleles transmitted:untransmitted from the mother; T:U PAT, no. of risk alleles transmitted:untransmitted from the father; Z_{POE} : Z score for difference in paternal vs maternal ORs

suggesting that they may represent independent signals (ESM Fig. 1). Haplotype associations with three-marker (rs163184, rs2237895, rs2237892) and four-marker

haplotypes showed no association, although parental asymmetry tests were significant for haplotypes based on four markers (Table 4).

Fig. 1 POE of *KCNQ1* on type 2 diabetes and hyperglycaemia.

(a, b) Transmission of maternal and paternal alleles of rs2237895 to diabetic offspring (a) and hyperglycaemic offspring (b). Risk allele C shows a statistically significant maternal effect and POE in both. (c, d) Transmission of maternal and paternal alleles of rs163184 to diabetic offspring (c) and hyperglycaemic offspring (d). Risk allele G shows statistically significant maternal effect. Black bars, risk alleles; white bars, non-risk alleles.

* $p < 0.05$ and ** $p < 0.01$

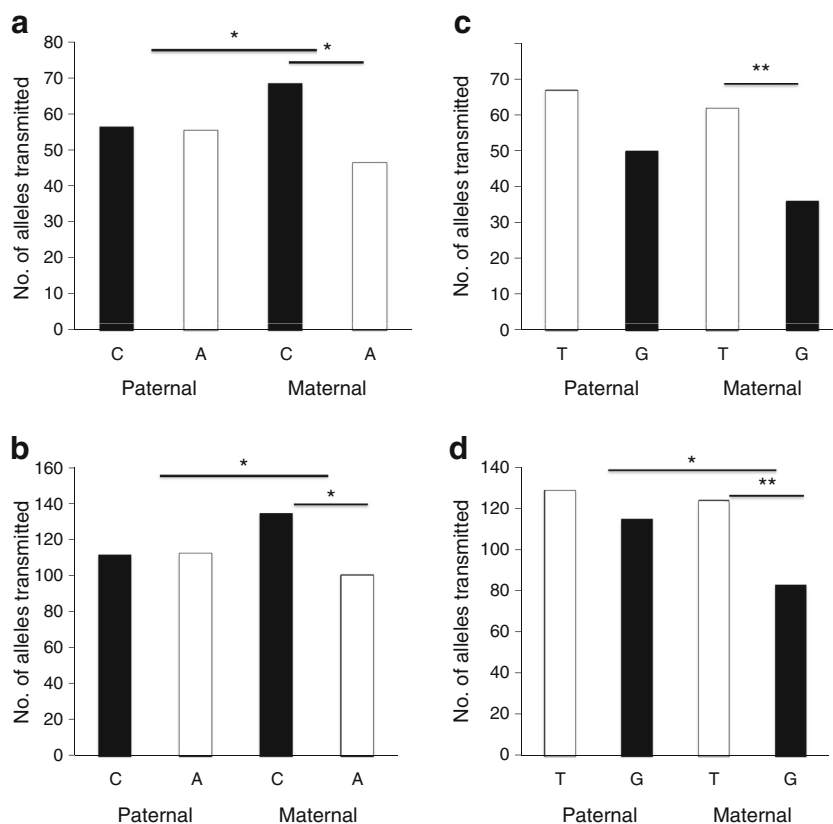


Table 4 Haplotype analyses based on rs2237892, rs163184 and rs2237895 and on rs231362, rs2237892.rs163184 and rs2237895

Haplotype	T	NT	RATIO	Paternal	Maternal	Ratio
rs2237892, rs163184 and rs2237895						
C T A	54.99	82.975	0.663	40.495	29.495	1.373
C G C	90.99	68.975	1.319	34.99	44	0.795
C G A	34.01	23.025	1.477	12.505	21.505	0.581
T T A	9	12	0.75	8	2	4
C T C	1.01	3.025	0.334	1.01	0	13505.12
rs231362, rs2237892.rs163184 and rs2237895						
G C T A	33.96	58.116	0.584	17.348	15.612	1.111
G C G C	48.507	40.373	1.201	20.701	27.807	0.744
A C G C	68.475	60.599	1.13	28.866	28.609	1.009
A C T A	35.023	31.856	1.099	23.563	16.459	1.432
A C G A	12.953	12.54	1.033	6.993	5.96	1.173
G C G A	18.064	9.488	1.904	3.511	14.553	0.241
A T T A	6.548	6	1.091	7.548	0	–
G C T C	1.016	1.023	0.994	1.016	0	11960.78
G T T A	1.452	4	0.363	0.452	1	0.452
A C T C	0.001	2.004	0.001	0.001	0	46.014

For haplotype analyses based on rs2237892, rs163184 and rs2237895: test for association $p=0.056$; parental asymmetry test $p=0.11$ (besthap: TTA). For haplotype analyses based on rs231362, rs2237892.rs163184 and rs2237895: test for association $p=0.067$; parental asymmetry test, $p=0.033$ (besthap: ATTA)

NT, haplotypes not transmitted; T, haplotypes transmitted

THADA In the Botnia families, the risk T allele of the *THADA* variant rs7578597 showed a nominal POE on type 2 diabetes risk with excess transmission from mothers to diabetic offspring ($p_{POE}=0.01$) (Table 3).

To assess whether this effect was restricted to the diabetic phenotype, we examined parental specific transmissions: (1) to all 1,945 offspring regardless of phenotype and (2) to trios with non-diabetic offspring using permutation to obtain empirical p values. The POE was neither seen when all 1,945 offspring were considered nor when the analysis was restricted to non-diabetic offspring trios, suggesting that POE was related to the hyperglycaemic phenotype (ESM Table 3).

The same pattern was seen in the Hungarian families who showed higher transmission of T allele from the mother while the other allele, C, showed a higher transmission from father to type 2 diabetic offspring with a statistically significant replication for the POE offspring ($p_{POE}=0.045$) (Table 5 and Fig. 2). A combined analysis of the two studies showed a significant POE for the SNP ($p_{POE}=0.0006$) (Table 5 and Fig. 2).

Given the bidirectional effect of each of the parental alleles, we also tested whether the SNP or the region would show differences in DNA methylation pattern in peripheral blood lymphocytes (PBL) from a subset of trios and in human pancreatic islets from unrelated cadaver donors. No clear difference in methylation pattern was seen in blood between carriers

of the maternal and paternal T allele but the low frequency (6.3%) of the minor (non-risk) C allele precluded statistical tests in the numbers studied. In contrast, 12 out of 46 CpG sites spanning the *THADA* gene tested showed differential methylation between diabetic and non-diabetic donor islets. Five CpG sites remained after correction for multiple testing, including the two CpG sites flanking rs7578597 (cg03647861 [$p=0.035$] and cg25938803 [$p=0.033$]) (Fig. 3). To assess how unusual this pattern of methylation was, we calculated the ratio of differentially methylated CpG probes to the square root of total number of CpG probes, accounting also for the poorly covered genes, considering the bias in the probe distribution on the chip. From this, we found that *THADA* had a p value of 0.075, with a ratio of 1.8, and scored right in between *KCNQ1* (ratio=4.5, $p=0.0021$) and *KLF14* (ratio=0.4, $p=0.58$) (ESM Figs 2 and 3). We also found differences in methylation at the same CpG sites between diabetic carriers of CT and TT genotypes, with CT genotype carriers showing lower methylation than TT genotype carriers (ESM Fig. 4).

Next, we tested whether the SNP would influence the expression of *THADA* or a nearby gene *PLEKHH2* in *cis* in four human tissues (skeletal muscle, adipose tissue, blood and islets of Langerhans) as the nine-base-pair insertion introduces a binding site for the transcription factor C/EBP. No significant effect of the variant was observed on expression of these two genes (37–47 individuals analysed per tissue; heterozygote $n=6$, homozygote derived $n=36$, $p>0.05$).

Table 5 Parental specific transmissions of risk allele T in the missense coding SNP rs7578597 in *THADA* to diabetic offspring

Study	R/O	T:U_PAT	p_{PAT}	T:U_MAT	p_{MAT}	Z_{POE}	p_{POE}
Botnia	T:C	09:16	0.16	17:09	0.08	2.06	0.010**
HTB ^a	T:C	02:08	0.057	05:02	0.25	2	0.045*
Combined ^a	T:C	11:24	0.027	22:11	0.055	2.841	0.0006***

Paternal, maternal transmission tests and parent-of-origin tests are shown

^a Based on a single test

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

p_{MAT} , asymptotic p value for maternal transmission test; p_{PAT} , asymptotic p value for paternal transmission test; p_{POE} , permuted pointwise p value for the parent-of-origin test; R/O, risk allele/other allele; T:U MAT, no. of risk alleles transmitted:untransmitted from the mother; T:U PAT, no. of risk alleles transmitted:untransmitted from the father; Z_{POE} , Z score for difference in paternal vs maternal ORs

We did not observe any significant POE of *THADA* variants on glucose, insulin secretion and action, lipids, BMI, waist-to hip circumference ratio, blood pressure or liver enzymes in the Botnia families (ESM Table 4). To expand the number of potential phenotypes, we also explored whether there would be differences in 27 different phenotypes between unrelated C and T allele carriers, but none of them reached significance (ESM Table 5), with the exception of a robust association with type 2 diabetes in a case-control cohort ($n = 38,069$ individuals) (Fig. 4).

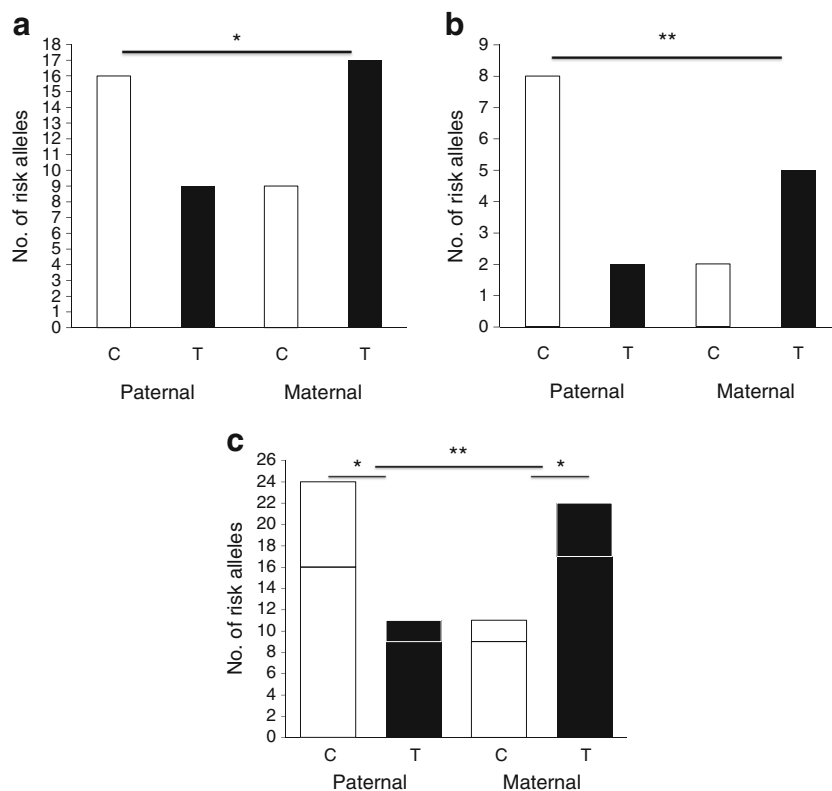
Variants in *CRY2* and *PRC1*, and near *DGKB/TMEM5* and *CDC123/CAMK1D* Variants in *CRY2* showed POE on type 2 diabetes and hyperglycaemia whereas those in *PRC1*

and near *DGKB/TMEM5*, and *CDC123/CAMK1D* showed POE on hyperglycaemia, with the risk allele being transmitted from the father (Table 3).

Discussion

Heritability of a disease is determined by segregation of disease-promoting alleles in families. Given the paucity of family materials, most genetic studies have been carried out in outbred populations of unrelated individuals where heritability estimates cannot be calculated in a classical way. The present study is one of the first family-based studies assessing specific parental transmissions of alleles in relation to type 2

Fig. 2 POE of rs7578597 (*THADA*) on type 2 diabetes. Risk allele T shows an excess transmission from the mother while the C allele shows an opposite trend with excess transmission from the father to diabetic offspring in Botnia (a), Hungary (b) and Botnia (lower bars) and Hungary (upper bars) combined (c). POE is significant in both studies and in the combined analyses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



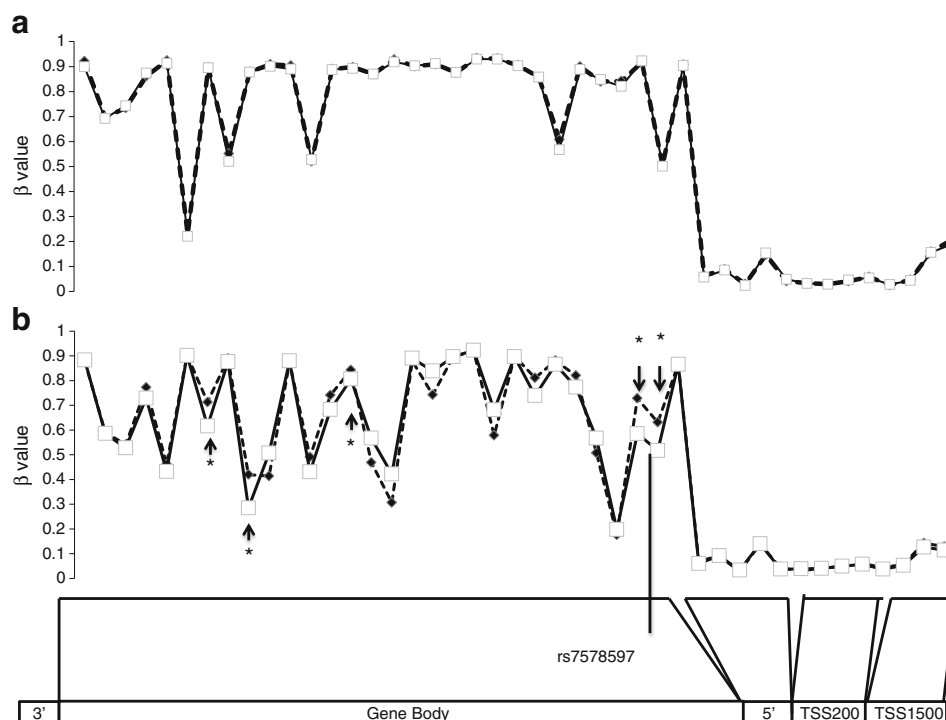


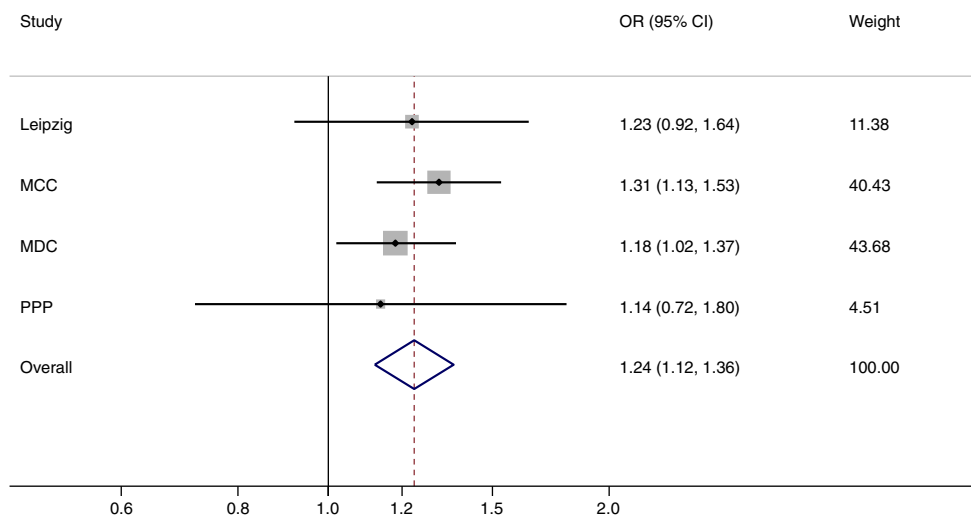
Fig. 3 Methylation status for CpG sites spanning the *THADA* gene in PBLs from trios (Botnia) (a) and human islets (b). Forty-three CpG sites spanning the *THADA* gene were tested for differences in methylation status between cases (with type 2 diabetes) and controls, mapped to gene locations. β values showing degree of methylation are plotted on the y-axis. White squares with black lines, cases; black diamonds with dotted

lines, controls. None of the CpG sites showed any difference in methylation status between cases and controls in the PBLs. Five CpG sites in the gene body region of donor islets showed a significant difference in methylation status ($\Delta\beta$), as indicated by the arrows. Interestingly, two CpG sites flanking the SNP rs7578597 were among them. * p_{adj} (p value after adjusting for multiple testing) <0.05

diabetes risk. Several previous observations have pointed to the possibility of distorted parental transmission of metabolic traits. We [5] and others [26, 27] have demonstrated higher risk of type 2 diabetes in offspring of diabetic mothers than in those of diabetic fathers. Furthermore, we reported that a parental history of type 2 diabetes influenced the insulin response to an oral glucose load, with male offspring of diabetic mothers showing the lowest insulin values [5].

The TDT is robust to population stratification but suffers from low power if the frequency of the risk allele is low as the analyses are restricted to heterozygous parents. Combining the parental phenotypes in the form of the parental discordance tests somewhat enhances the power. We also applied the FBAT, which allows analysis of complex family structures and inclusion of all offspring. Only about one-quarter of variants previously associated with type 2 diabetes/

Fig. 4 Forest plot of the meta-analysis of the association with type 2 diabetes in the four studies: Leipzig, Malmö Case Control (MCC), Malmö Diet and Cancer (MDC) and Botnia Prevalence, Prediction and Prevention of Diabetes (PPP). A significant association was observed for the rs7578597, with an OR of 1.24 (95% CI 1.12, 1.36, $p=1.96 \times 10^{-5}$)



hyperglycaemia in unrelated individuals showed association with type 2 diabetes in at least one of the tests in the present study. While this is most likely due to limited power, it cannot be ruled out that common variants detected in GWAS explain little of clustering of type 2 diabetes seen in some families.

Among the reported GWAS associations, the strongest signal has been the *TCF7L2* SNP rs7903146 [28]. This was also the strongest signal for association with type 2 diabetes in our parent-of-origin (POE) analysis (Table 2). Rare mutations in the *HNF1A*, *HNF4A* and *WFS1* genes account for a substantial proportion of cases with familial MODY, while common variants are associated with type 2 diabetes [9, 29, 30]. Interestingly, variants in these genes were among those showing association with type 2 diabetes in our studied families.

For POE, we found no global signal when all type 2 diabetes risk loci were considered collectively. However, in a locus-by-locus analysis, we found nominal evidence for POE at the *KCNQ1* and *THADA* loci. Here, we used *KCNQ1* as positive control as variants in *KCNQ1* have shown consistent parental specific associations with type 2 diabetes and hyperglycaemia in previous studies [9, 31]. Consistent with the findings of previous studies [14, 15] the risk allele showed a maternal transmission to affected offspring. We also observed significant differences in methylation status between diabetic and non-diabetic pancreatic islets, but we were not able to analyse which parental allele was methylated and possibly imprinted. *KCNQ1* is known to be an imprinted gene, but this seems to be restricted to a relatively brief window of time during the fetal period. In a previous study, we observed that *KCNQ1* and *KCNQ1OT1* were mono-allelically expressed in fetal tissues but bi-allelically expressed in adult islets [32]. It is thus possible that fetal programming of genes involved in beta cell function could result in reduced functional beta cell mass, which could predispose to type 2 diabetes by aggravating later in life the capacity to compensate for increased needs imposed by insulin resistance and obesity.

One of the key findings of the present study was an indication for POE with maternal transmission of the non-synonymous risk allele T of SNP rs7578597 in the *THADA* gene. Notably, the maternal effect was restricted to hyperglycaemic offspring and was not seen in non-diabetic offspring, indicative of a disease-specific effect. Imprinting often occurs through DNA methylation, so to explore patterns of methylation at the *THADA* locus we assessed the degree of DNA methylation in both PBLs and human islets. In both tissues tested, the gene showed distinctive patterns of methylation. Notably, five CpG sites out of 43 sites analysed (11.6%), including those flanking the rs7578597 SNP, showed a difference in the methylation status between diabetic and non-diabetic donors islets (more than expected by chance alone); no difference was seen in blood. A link between methylation and imprinting of the *THADA* gene is further supported by data showing allelic imbalance for *THADA* expression

in adult human pancreatic islets [33]. However, expression of the *THADA* gene is not restricted to beta cells but is also seen in alpha and exocrine cells [34].

The parent-of-origin tests are based on the TDT, which is robust but suffers from reduced power when being analysed for transmission of rare alleles from heterozygous parents. While this parent-specific association was nominally significant in the families from Botnia, replication in families from Hungary, supported by methylation and allelic imbalance in expression in the islets, lends further support to the parent-of-origin associations.

The function of *THADA* (thyroid adenoma associated), which contains an ARM repeat for protein–protein interactions, is not known but it has been shown to be the target of translocations in thyroid adenomas [35] suggesting a potential role in growth. Disruption of orthologues of *THADA* led to sucrose-dependent toxicity in drosophila [36]. The *THADA* gene has also been ascribed an interesting role during evolution; *THADA* is located in a chromosomal region that was suggested to be subject to positive selection from Neanderthals to modern humans, as the Neanderthal genome was depleted of derived alleles [37].

If *THADA* was a ‘thrifty’ gene, positively selected during evolution, which phenotype would it influence? To address this question we related *THADA* variants to different phenotypes, including glycaemic, both in the trios and in unrelated individuals. Among possible ‘thrifty traits’, determinants of energy metabolism and muscle function have often been reported [38, 39]. However, of the traits tested (ESM Tables 4, 5), no clear POE was observed. Interestingly, a recent study showed *THADA* to be one of the strongest signals for cold adaptation [40], but such phenotypes were not available in the trios.

While there are very few reported imprinted genes, the true number of human imprinted genes is not known. Recent studies in mice and other species have shown that the actual number of imprinted genes is much higher than originally thought and that factors like tissue specificity and temporal effects of imprinting status need to be taken into account [41]. While *THADA* and *KCNQ1* are examples of genes with such potential functions, future research should find an answer to the question: what benefits might such POE have for the offspring?

Acknowledgements We thank the families for their participation in the Botnia study, Malmö Diabetes Registry, Malmö Diet and Cancer study, The Leipzig cohort and the HTB and the study groups for clinically studying the patients. Human islets were provided by the Nordic network for clinical islets transplantation by the courtesy of O. Korsgren, Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. Sincere thanks are given to M. Sterner, G. Gremesperger, H. Mattisson, J. Kravic, U. Blom-Nilsson and J. Postma from Lund University Diabetes Centre, Malmö for all the help in the laboratory and for database and administrative support.

Funding The study was supported by grants from the Swedish Research Council (Research grant Dnr 521-2010-3490 to LG; a Linnaeus Centre of Excellence Grant for the Lund University Diabetes Centre, Dnr 349-2006-237), a Strategic Research Area grant (EXODIAB, Dnr. 2009-1039), a European Research Council Advanced Researcher Grant to LG (GENETARGET_T2D, GA 269045) and funding from the Sigrid Jusélius Foundation, the Folkhälsan Research Foundation and the Academy of Finland. The work carried out in Leipzig, Germany was supported by a grant from the Deutsche Forschungsgemeinschaft (SFB 1052 ‘Obesity mechanisms’, B03).

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

Contribution statement RBP, AL, PA, GK, PK, YZ, ES, JF, NO, MV, CBW, OH, MOM, TT, SP, LK and LG conceived and designed the experiments. RBP, AL, JF, PA, WP, ML, NO, MS, CL, MS and OH performed the experiments and analysed the data. RBP, AL, PA, PK, CBW and LG wrote the paper. All authors contributed to the critical revision for important intellectual content, read and approved the final manuscript. RBP and LG are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

References

- WHO (2015) World Health Statistics report. Available from www.who.int/gho/publications/world_health_statistics/en/, accessed 14 July 2015
- Lyssenko V, Jonsson A, Almgren P et al (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 359:2220–2232
- Almgren P, Lehtovirta M, Isomaa B et al (2011) Heritability and familiarity of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* 54:2811–2819
- Hemminki K, Li X, Sundquist K, Sundquist J (2010) Familial risks for type 2 diabetes in Sweden. *Diabetes Care* 33:293–297
- Groop L, Forsblom C, Lehtovirta M et al (1996) Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects. *Diabetes* 45:1585–1593
- Prokopenko I, Poon W, Magi R et al (2014) A central role for GRB10 in regulation of islet function in man. *PLoS Genet* 10:e1004235
- Saxena R, Hivert MF, Langenberg C et al (2010) Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 42:142–148
- Saxena R, Voight BF, Lyssenko V et al (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1336
- Voight BF, Scott LJ, Steinthorsdottir V et al (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42:579–589
- Zeggini E, Scott LJ, Saxena R et al (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40:638–645
- Manolio TA, Collins FS, Cox NJ et al (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753
- Prasad RB, Groop L (2015) Genetics of type 2 diabetes-pitfalls and possibilities. *Genes (Basel)* 6:87–123
- Eichler EE, Flint J, Gibson G et al (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 11:446–450
- Kong A, Steinthorsdottir V, Masson G et al (2009) Parental origin of sequence variants associated with complex diseases. *Nature* 462:868–874
- Hanson RL, Guo T, Muller YL et al (2013) Strong parent-of-origin effects in the association of KCNQ1 variants with type 2 diabetes in American Indians. *Diabetes* 62:2984–2991
- Alberti KGMM, Zimmet PZ, Consultation W (1998) Definition, diagnosis and classification of diabetes mellitus and its complications part 1. Diagnosis and classification of diabetes mellitus - provisional report of a WHO consultation. *Diabetic Med* 15:539–553
- Horvath S, Wei E, Xu X, Palmer LJ, Baur M (2001) Family-based association test method: age of onset traits and covariates. *Genet Epidemiol* 21(suppl 1):S403–S408
- Horvath S, Xu X, Laird NM (2001) The family based association test method: strategies for studying general genotype–phenotype associations. *Eur J Hum Genet* 9:301–306
- Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–516
- Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19:149–150
- Purcell S, Sham P, Daly MJ (2005) Parental phenotypes in family-based association analysis. *Am J Hum Genet* 76:249–259
- Abecasis GR, Cookson WO, Cardon LR (2000) Pedigree tests of transmission disequilibrium. *Eur J Hum Genet* 8:545–551
- Becker T, Knapp M (2004) Maximum-likelihood estimation of haplotype frequencies in nuclear families. *Genet Epidemiol* 27:21–32
- Smyth GK (2005) Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (eds) *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer, New York, pp 397–420
- Pettitt DJ, Lawrence JM, Beyer J et al (2008) Association between maternal diabetes in utero and age at offspring’s diagnosis of type 2 diabetes. *Diabetes Care* 31:2126–2130
- Pettitt DJ, McKenna S, McLaughlin C, Patterson CC, Hadden DR, McCance DR (2010) Maternal glucose at 28 weeks of gestation is not associated with obesity in 2-year-old offspring: the Belfast Hyperglycemia and Adverse Pregnancy Outcome (HAPO) family study. *Diabetes Care* 33:1219–1223
- Grant SF, Thorleifsson G, Reynisdottir I et al (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 38:320–323
- Kooner JS, Saleheen D, Sim X et al (2011) Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* 43:984–989
- Sandhu MS, Weedon MN, Fawcett KA et al (2007) Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* 39:951–953
- Yasuda K, Miyake K, Horikawa Y et al (2008) Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 40:1092–1097
- Travers ME, Mackay DJ, Dekker Nitert M et al (2013) Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets. *Diabetes* 62:987–992

33. Fadista J, Vikman P, Laakso EO et al (2014) Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc Natl Acad Sci U S A* 111:13924–13929
34. Bramswig NC, Everett LJ, Schug J et al (2013) Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming. *J Clin Invest* 123:1275–1284
35. Kloth L, Belge G, Burchardt K et al (2011) Decrease in thyroid adenoma associated (THADA) expression is a marker of dedifferentiation of thyroid tissue. *BMC Clin Pathol* 11:13
36. Pendse J, Ramachandran PV, Na J et al (2013) A Drosophila functional evaluation of candidates from human genome-wide association studies of type 2 diabetes and related metabolic traits identifies tissue-specific roles for dHHEX. *BMC Genomics* 14:136
37. Green RE, Krause J, Briggs AW et al (2010) A draft sequence of the Neandertal genome. *Science* 328:710–722
38. Almind K, Kahn CR (2004) Genetic determinants of energy expenditure and insulin resistance in diet-induced obesity in mice. *Diabetes* 53:3274–3285
39. Stannard SR, Johnson NA (2004) Insulin resistance and elevated triglyceride in muscle: more important for survival than 'thrifty' genes? *J Physiol* 554:595–607
40. Cardona A, Pagani L, Antao T et al (2014) Genome-wide analysis of cold adaptation in indigenous Siberian populations. *PLoS One* 9: e98076
41. Gregg C, Zhang J, Butler JE, Haig D, Dulac C (2010) Sex-specific parent-of-origin allelic expression in the mouse brain. *Science* 329: 682–685