

## ORIGINAL ARTICLE

# Five linkage regions each harbor multiple type 2 diabetes genes in the African American subset of the GENNID Study

Sandra J Hasstedt<sup>1</sup>, Heather M Highland<sup>2</sup>, Steven C Elbein<sup>3,\*</sup>, Craig L Hanis<sup>2</sup>, Swapan K Das<sup>3</sup> and the American Diabetes Association GENNID Study Group<sup>4</sup>

We previously localized type 2 diabetes (T2D)-susceptibility genes to five chromosomal regions through a genome-wide linkage scan of T2D and age of diagnosis (AOD) in the African American subset of the GENNID sample. To follow up these findings, we repeated the linkage and association analysis using genotypes on an additional 9203 fine-mapping single nucleotide polymorphisms (SNPs) selected to tag genes under the linkage peaks. In each of the five regions, we confirmed linkage and inferred the presence of  $\geq 2$  susceptibility genes. The evidence of multiple susceptibility genes consisted of: (1) multiple linkage peaks in four of the five regions; and (2) association of T2D and AOD with SNPs within  $\geq 2$  genes in every region. The associated genes included 3 previously reported to associate with T2D or related traits (*GRB10*, *NEDD4L*, *LIPG*) and 24 novel candidate genes, including genes in lipid metabolism (*ACOXL*) and cell–cell and cell–matrix adhesion (*MAGI2*, *CLDN4*, *CTNNA2*).

*Journal of Human Genetics* (2013) 58, 378–383; doi:10.1038/jhg.2013.21; published online 4 April 2013

**Keywords:** association; likelihood; linkage; type 2 diabetes

## INTRODUCTION

With the goal of discovering the genes that contribute to the risk of common diseases, numerous susceptibility loci have been identified using linkage analysis. Despite replication of many of these linkages in a second sample, the responsible genes often remain unknown. Some exceptions include *TBC1D1*, identified as the gene responsible for the obesity linkage on chromosome 4p15–14,<sup>1,2</sup> and *HOXB13*, identified as the gene responsible for the prostate cancer linkage on chromosome 17q21–22.<sup>3</sup> However, for many other common disease linkage signals, the underlying causal genes and variants await discovery.

The expectation that a single gene accounts for a linkage peak may contribute to the difficulty in identifying causal genes. On the contrary, linkage peaks may reflect clusters of causal genes: Martin *et al.*<sup>4</sup> attributed a triglyceride linkage on chromosome 7q36 to variants in five genes, and Christians *et al.*<sup>5,6</sup> found that fine-mapping caused a single quantitative trait locus (QTL) for body size in mice to resolve first into three QTLs, then one of those to split into two. Furthermore, the clustering of causal genes may have stymied the multi-group effort to identify type 2 diabetes (T2D) genes on chromosome1q;<sup>7</sup> two strong associations present in populations of

European ancestry failed replication and confirmation in other ethnic groups. These examples suggest that abandoned linkage findings might yet reveal susceptibility genes if reappraised while considering the possibility of multiple causal genes.

As for other common diseases, T2D gene discovery presents a challenge despite strong evidence of a genetic component.<sup>8</sup> The challenge is even greater in African American (AA) populations where both prevalence and genetic diversity are higher<sup>9</sup> and pathophysiology may differ.<sup>10,11</sup> Genome-wide case-control association studies have identified single nucleotide polymorphisms (SNPs) with small effects on T2D risk,<sup>12</sup> but few of the associated SNPs, identified primarily in European ancestry populations, are replicated in AAs<sup>13</sup> and few SNPs have been identified in AA populations.<sup>14</sup> In general, the widespread rejection of family studies in favor of genome-wide case-control association studies has failed to produce the promised prognostic and diagnostic variants for T2D.<sup>15</sup>

As a resource for the discovery of genes related to T2D and its complications, the American Diabetes Association established the Genetics of NIDDM (GENNID) study. From 1993 to 2003, the GENNID study ascertained families through siblings diagnosed with T2D. We have used this resource to increase understanding of the

<sup>1</sup>Department of Human Genetics, University of Utah, Salt Lake City, UT, USA; <sup>2</sup>Human Genetics Center, School of Public Health, University of Texas Health Science Center, Houston, TX, USA and <sup>3</sup>Section on Endocrinology and Metabolism, Wake Forest University School of Medicine, Winston-Salem, NC, USA  
Correspondence: Dr SJ Hasstedt, Department of Human Genetics, University of Utah, 15 N 2030 E RM 2100, Salt Lake City, UT 84112 5330, USA.  
E-mail: sandy@genetics.utah.edu

<sup>4</sup>See Appendix.

\*Deceased.

Received 17 October 2012; revised 26 February 2013; accepted 1 March 2013; published online 4 April 2013

genetics of T2D by applying linkage and family-based association analysis to the AA subset of the GENNID sample. Genome-wide linkage analysis using genotypes on 5914 SNPs identified chromosomal regions that potentially harbor risk genes for T2D and age of T2D diagnosis (AOD). The strongest signal for T2D occurred on chromosome 2 at 95–121 megabases (Mb), with weaker support for AOD in the same region and for T2D at 68–95 Mb. Both T2D and AOD also showed linkage on chromosome 13 at 19–30 Mb, but the linkage was limited to T2D on chromosome 7 at 50–79 Mb and to AOD on chromosome 18 at 31–65 Mb.<sup>16</sup> Other analyses inferred pleiotropy with triglyceride in the chromosome 2p region<sup>17</sup> and pleiotropy with obesity in the chromosome 13 region.<sup>18</sup>

Herein we present linkage and association analyses on the AA subset of the GENNID sample using genotypes on 9203 fine-mapping SNPs added to the genome-scan SNPs used previously. Our first goal was to more precisely localize the T2D-susceptibility genes in each of the five chromosomal regions identified in the genome scan (chromosome 2 at 68–95 and 95–121 Mb, chromosome 7 at 50–79 Mb, chromosome 13 at 19–30 Mb and chromosome 18 at 31–65 Mb). Our second goal was to test for the association of the gene-based fine-mapping SNPs with T2D and AOD in order to identify genes related to T2D in AAs.

## SUBJECTS AND METHODS

The GENNID study ascertained families through sibling pairs each with a T2D diagnosis.<sup>19</sup> During Phase 1, extended family members were also studied; one site ascertained AAs. During Phase 2, data collection beyond the sibling pair was limited to parents, or, if parents were unavailable, unaffected siblings; five sites ascertained AAs. During Phase 3, only affected sibling pairs and trios were studied; five additional sites ascertained AAs. In total, 1496 AA members of 580 pedigrees were studied at 10 sites. Data cleaning and re-evaluation of T2D diagnoses by current criteria reduced the analysis sample to 1344 members of 524 pedigrees;<sup>16</sup> for this study, we selected an informative subset of 1077 members of 415 pedigrees. This study was approved by the Institutional Review Board at each participating institution.

T2D diagnoses, originally following National Diabetes Data Group criteria, were re-evaluated using current criteria before all analyses; 84 affected and 13 unaffected sample members were re-assigned as unknown. AOD was reported on a standardized questionnaire. Body mass index (BMI) was computed from height and weight obtained from physical examination.

We selected a subset of the sample for fine mapping by excluding unaffected individuals aged <53 years, the age by which onset had occurred in 75% of cases, and then any individual, either affected or unaffected, who consequently had no family members in the sample. This sample subset maintained significant lod scores in each of the linkage regions and produced power >80% to detect association. Power to produce a *P*-value of 0.0001 was estimated from simulation of 1000 replicates of a SNP with minor allele frequency 0.4 and heterozygous effect a 10% increase in penetrance at age 50 years for T2D or 3-year increase in AOD.

For fine-map genotyping, we used SNAGGER<sup>20</sup> (<http://snagger.sourceforge.net/>) to select tagSNPs within genes in the linkage regions on chromosomes 2, 7, 13 and 18. Using HapMap phases 1 and 2 (<http://hapmap.ncbi.nlm.nih.gov>), variants in the region were tagged at linkage disequilibrium (LD)  $r^2 \geq 0.7$ . Each SNP selected had a minor allele frequency >0.1 and Illumina design score >0.4; SNP pairs had minimum spacing of 60 base pairs (bp). Center for Inherited Disease Research (CIDR) performed the genotyping using the Illumina iSelect chip (Illumina, Inc, San Diego, CA, USA). The genetic map locations of the SNPs in centiMorgans (cM) were obtained from the Rutgers Combined Linkage-Physical Map of the Human Genome (<http://compgen.rutgers.edu/maps>); locations of SNPs not mapped directly were estimated using physical positions from NCBI dbSNP Build 123 in a smoothing calculation (Conway Institute Bioinformatics Service; <http://integrin.ucd.ie>). Genotyping errors were identified using Pedcheck<sup>21</sup> and MERLIN;<sup>22</sup> we zeroed 4978 genotypes of 710 SNPs, 74 identified by Pedcheck

and 4904 by Merlin. Genotypes for 24 SNPs, available on <95% of the sample following data cleaning, were retained after confirmation of minimal effects on the results. Multi-point identity-by-descent probabilities were computed at each cM using MERLIN,<sup>22</sup> treating as haplotypes SNP sets with pairwise LD  $r^2 > 0.7$ . We identified 279, 175, 133 and 259 SNP sets on chromosomes 2, 7, 13 and 18, respectively.

Likelihood analysis, as implemented in jPAP,<sup>23,24</sup> was used for univariate linkage analysis of unadjusted T2D (uT2D), T2D adjusted for BMI (bT2D) and AOD and for bivariate association analysis of uT2D and AOD. AOD was adjusted for gender and modeled as a normal density with mean, s.d. and gender effect as parameters. T2D risk was modeled to account for AOD in affected pedigree members, while allowing for censored observations<sup>24</sup> with age, gender and BMI (for bT2D only) effects and penetrance as parameters. For each trait, additional parameters included heritability, a QTL effect in linkage analysis and an additive SNP effect in association analysis.

We applied variance components linkage analysis using the univariate model for each trait in conjunction with the identity-by-descent probabilities. At each cM across the five regions, all parameters were estimated for AOD while only heritability and QTL effect were estimated for uT2D and bT2D with all the other parameters for those traits fixed at estimates obtained upon maximizing the likelihood while correcting the likelihood for the ascertainment of each pedigree through an affected sib pair. No further ascertainment correction was made in the linkage analyses. The lod score at each cM was computed as the common logarithm of the ratio of the maximized likelihoods with the QTL effect estimated to the maximized likelihood, with the QTL effect set to zero.

We assessed each fine-mapping SNP for association with uT2D and AOD by coding the SNP genotype as an additive covariate and testing its effect on both traits in a bivariate model. *P*-values were determined using a two-degree-of-freedom  $\chi^2$  statistic computed as twice the natural logarithm of the ratio of the maximized likelihood with covariate effects estimated for both T2D and AOD to the maximized likelihood, with both effects set equal to zero. We controlled for multiple testing separately within each of the five linkage regions by specifying a false discovery rate<sup>25,26</sup> of 0.05 accounting for 1291, 1780, 1819, 1336 and 2977 fine-mapping SNPs for chromosomes 2p, 2q, 7, 13 and 18, respectively.

We tested for an independent effect of each secondary associated SNP, conditional on the SNP within the same gene that attained the highest significance, by comparing the maximized likelihood estimating T2D and AOD covariate effects for both SNPs simultaneously to the maximized likelihood estimating T2D and AOD covariate effects for the most significant SNP alone. *P*-values were again determined using a two-degree-of-freedom  $\chi^2$  statistic. *P* < 0.05 after Bonferroni correction for the number of secondary SNPs in the gene supported the independence of the secondary SNP.

## RESULTS

We genotyped 9203 SNPs within five regions encompassing a total of 133 cM and 128 Mb (Table 1) and merged them with 244 (36 duplicates) genome-scan SNPs. Spacing between SNPs averaged 0.0145 cM (13 910 bp) and ranged from 0.0 to 5.3 cM and from 26 bp to 6 Mb.

To comply with the sample size limitations of the CIDR genotyping platform, we selected a subset of the original linkage sample to be

**Table 1** Fine-mapping regions by chromosome, genetic location, physical position and number of SNPs

Chromosome	Location (cM)	Position (Mb)	Number of SNPs
2p	91.29–111.78	68.41–89.05	1291
2q	111.79–136.02	95.54–121.75	1780
7	72.60–92.94	49.82–79.08	1819
13	0.00–25.25	19.50–31.54	1336
18	62.74–105.59	32.67–66.72	2977

Abbreviation: SNP, single nucleotide polymorphism.

**Table 2** Characteristics of the sample by gender and T2D status

T2D status	Age/AOD (years)					BMI (kgm <sup>-2</sup> )				
	N	Mean	s.d.	Minimum	Maximum	N	Mean	s.d.	Minimum	Maximum
<i>Gender</i>										
<i>Male</i>										
Affected	339	45.45	11.56	18	92	339	30.18	5.83	16.78	51.42
Unaffected	40	67.15	10.11	53	90	40	27.29	5.07	17.81	43.82
<i>Female</i>										
Affected	634	45.35	12.12	15	84	634	33.96	7.28	16.78	54.60
Unaffected	64	66.39	10.32	53	95	64	30.82	7.44	17.29	49.90

Abbreviations: AOD, age of diagnosis; BMI, body mass index; T2D, type 2 diabetes.

informative for both linkage and association. The genotyped subset included 1077 members of 415 pedigrees that ranged from 2 to 11 members (1–8 were genotyped). In this subset, AOD and BMI differed little from the complete sample, but the unaffected sample members, selected for a minimum age of 53 years, were older (Table 2).

We previously identified five linkage peaks within four chromosomal regions from autosomal scans of T2D and AOD; the chromosome 2 region contained two peaks separated by 30 Mb and the centromere.<sup>16,18</sup> Using updated genetic map positions for the genome-scan SNPs, we repeated the linkage analyses on the smaller sample: the genome-scan lod scores for all the five peaks (Table 3) generally agreed with our published lod scores for bT2D and AOD<sup>16</sup> and for uT2D.<sup>18</sup>

Upon adding the fine-mapping SNPs, linkage evidence remained in all five regions but strengthened only on chromosome 2 (Table 3). On chromosome 18, only AOD supported linkage, although uT2D provided weak support (lod = 2.62) upon elimination of sample members with BMI > 45 kg m<sup>-2</sup>. The most remarkable effect of fine-mapping was the splitting of the lod score curves into multiple peaks in all of the regions except on chromosome 18.

Additional evidence that multiple susceptibility genes contribute to the linkage signals derived from the identification of associated SNPs that reside in multiple genes within each of the five regions, including on chromosome 18 (Table 4). The number of associated SNPs ranged from 4 on chromosome 13 to 17 on chromosome 2q; the number of associated genes ranged from 2 on chromosome 13 to 9 on chromosome 7. Although 20 of the 27 associated genes were identified through a single SNP, five genes were identified through 2 or 3 SNPs and *ARHGAP25* and *DPP10* were identified by 8 and 10, respectively. However, six associated SNP pairs exceeded LD  $r^2 \geq 0.7$ : *ARHGAP25* (rs6714065 and rs7605681), *CTNNA2* (rs968820 and rs1368915), *DPP10* (rs843417 and rs1823267, rs10204212 and rs13432035, rs4848376 and rs11694256), and *HIP1* (rs1179625 and rs1179622). Nevertheless, conditional association analysis supported the independence of  $\geq 2$  SNPs in *ARHGAP25*, *POLR1B*, *DPP10* and *MTUS2* (Supplementary Table S1).

## DISCUSSION

The fine-mapping of five chromosomal regions allowed us not only to confirm our genome-scan linkages but also to infer the presence of multiple T2D-susceptibility genes in each of the regions. Two types of evidence supported the presence of multiple susceptibility genes underlying the linkage peaks. First, except chromosome 18, each lod score peak split into at least two peaks when fine-mapping SNPs

**Table 3** Lod scores by trait in each of the five chromosomal regions

Chromosome	SNPs <sup>a</sup>	Location (cM)	Position (Mb)	uT2D	bT2D	AOD
2p	GS	108	85	5.04	4.42	<2
	FM	100	75	6.33	3.61	2.20
		105	80	6.93	5.63	2.65
2q	GS	130	115	5.51	5.66	<2
	FM	114	100	7.56	5.72	2.78
		129	114	4.99	4.58	<2
		133	119	3.12	4.20	<2
		136	121	4.45	4.57	<2
7	GS	80	63	5.50	3.74	<2
	FM	81	67	4.26	2.99	<2
		89	76	2.83	2.70	<2
		93	79	2.71	2.46	<2
13	GS	17	27	6.86	3.25	2.36
	FM	1	20	4.62	3.66	<2
		8	23	4.75	3.50	<2
		14	26	7.51	3.89	2.47
		21	29	6.60	3.12	2.49
18	GS	71	42	<2	<2	3.22
	FM	69	41	<2	<2	3.23

Abbreviations: AOD, age of diagnosis; bT2D, type 2 diabetes (T2D) adjusted for body mass index; FM, fine map; GS, genome scan; SNP, single nucleotide polymorphism; uT2D, unadjusted T2D.

A lod drop > 1 for  $\geq 1$  traits separates locations.

<sup>a</sup>Lod score computed using GS or FM SNPs.

were added to the linkage analysis. Second, in every region, including chromosome 18,  $\geq 2$  potential susceptibility genes were identified through the association with T2D and AOD of fine-mapping SNPs residing in those genes.

Multiple susceptibility genes may often underlie linkage peaks for common diseases. In fact, Martin *et al.*<sup>4</sup> attributed many failures to identify causal variants to the incorrect assumption that a single or a limited number of variants are responsible for a linkage signal. As evidence, they identified variants in five genes that fully accounted for a linkage peak for plasma triglyceride level. If small effects are typical of common disease-susceptibility genes, their detection through linkage analysis may be limited to locations where the genes cluster. The corollary is that isolated genes may elude detection by linkage analysis. However, fine-mapping using next-generation sequencing may provide sufficient power necessary to detect even isolated genes.

Although chromosome 18 failed to show T2D linkage and chromosome 7 failed to show AOD linkage, we nevertheless expect

**Table 4** Genes by chromosomal region that associate with uT2D and AOD in bivariate analysis, as identified through FDR <0.05 for SNPs within the genes

Chromosome	Position (Mb)	Gene	SNPs <sup>a</sup>	Nominal P <sup>b</sup>
2p	69	<i>CNRIP1</i>	rs6746740	0.0000915
	69	<i>ARHGAP25</i>	rs6714065, rs7605681, rs7579103, rs13418802, rs2311413, rs6546439, rs10190115, rs4854520	0.000000075
2q	75	<i>TACR1</i>	rs4853119	0.000217
	80	<i>CTNNA2</i>	rs968820, rs13394481, rs1368915	0.0000504
	100	<i>AFF3</i>	rs13031524	0.000158
	103	<i>TMEM182</i>	rs917965	0.000170
	111	<i>ACOXL</i>	rs7583450	0.000444
	113	<i>TTL</i>	rs7569407	0.000131
	113	<i>POLR1B</i>	rs13395779, rs1545133	0.0000105
	114	<i>IL36B</i>	rs6744874	0.0000442
	116	<i>DPP10</i>	rs10194806, rs843417, rs10204212, rs1823267, rs10165355, rs958457, rs13432035, rs4848376, rs11694256, rs7569681	0.0000598
7	51	<i>GRB10</i>	rs7794604	0.0000164
	69	<i>AUTS2</i>	rs17141989	0.0000896
	71	<i>WBSCR17</i>	rs599574	0.0000572
	71	<i>CALN1</i>	rs10234063	0.000246
	73	<i>CLDN4</i>	rs4640941	0.0000621
	73	<i>WBSCR28</i>	rs11770052	0.0000248
	75	<i>HIP1</i>	rs1179625, rs2705788, rs1179622	0.0000374
	77	<i>PION</i>	rs11771461	0.0000208
	78	<i>MAGI2</i>	rs6465973	0.0000958
13	28	<i>USP12</i>	rs7331742, rs17085261	0.0000576
	30	<i>MTUS2</i>	rs9508344, rs4616138	0.0000544
18	35	<i>CELF4</i>	rs949345	0.0000664
	47	<i>LIPG</i>	rs11875600	0.0000168
	56	<i>NEDD4L</i>	rs12327065	0.0000224
	57	<i>CCBE1</i>	rs7245331	0.0000467
	66	<i>CCDC102B</i>	rs3905374	0.0000779

Abbreviations: AOD, age of diagnosis; FDR, false discovery rate; SNP, single nucleotide polymorphism; uT2D, unadjusted type 2 diabetes.

<sup>a</sup>SNPs listed in order of increasing nominal *P*.<sup>b</sup>Smallest *P* for any SNP in the gene.

that the susceptibility genes residing in all the regions both increase T2D risk and decrease AOD. Variation between regions in the information available undoubtedly affects the relative strength of the corresponding lod scores for T2D or AOD. As evidence, the elimination of T2D cases with BMI > 45 kg m<sup>-2</sup> revealed a weak T2D linkage on chromosome 18 and fine-mapping revealed AOD linkages on chromosomes 2p and 2q. The lod scores are generally lower for AOD than for T2D as expected as AOD reflects onset age imprecisely and the familial correlation of AOD may partially result from temporal clustering of T2D diagnoses among relatives. Nevertheless, AOD proved sufficiently accurate to produce linkage evidence in four of the chromosomal regions.

Few of the associated genes have previously been reported to associate with T2D or related traits. The exceptions include *GRB10* with T2D,<sup>27</sup> *NEDD4L* with diabetic nephropathy<sup>28</sup> and *LIPG* with lipid metabolism.<sup>29</sup> Novel candidates reported herein include a cytokine (*IL36B*) and genes involved in lipid metabolism (*ACOXL*) and cell–cell and cell–matrix adhesions (*MAGI2*, *CLDN4*, *CTNNA2*). Interestingly, the candidates also included genes involved in Williams–Beuren syndrome (*WBSCR28*, *WBSCR17*, *CLDN4*),<sup>30</sup> whose sufferers have a high prevalence of diabetes and pre-diabetes.<sup>31</sup> Another unusual candidate is *DPP10*, which is reportedly associated with asthma.<sup>32</sup> *DPP10* is related to *DPP4*, whose inhibitors are oral anti-

hyperglycemics used in T2D therapy,<sup>33</sup> and *DPP6*, which contributes to a triglyceride linkage;<sup>4</sup> however, unlike *DPP4*, *DDP10* lacks serine protease/dipeptidyl peptidase activity.<sup>34</sup>

Support for the functionality of these genes derived from evidence that their expression levels are affected by our associated SNPs or proxies with high LD. We tested 14 SNPs in the 5 genes for which expression levels from transformed lymphocytes were measured on a subset of 160 sample members; nominal significance was obtained for one: *P* = 0.0148 for an effect of rs7579103 on *ARHGAP2* expression levels. In other populations and various tissues,<sup>35</sup> expression levels of *ARHGAP2*, *AFF3*, *POLR1B*, *HIP1*, *LIPG* and *NEDD4L* showed evidence of an effect of an associated SNP or its proxy (*P* < 1 × 10<sup>-5</sup>).

For further insight into the nature of the associations, we haplotyped sets of 15–30 fine-mapping SNPs, each encompassing ≥ 2 of the associated SNPs reported in Table 4. Haplotypes that attained *P* < 0.00001 are shown in Supplementary Table S2. For *ARHGAP25*, two haplotypes associated with T2D: the first extended across four of the associated SNPs and contained the risk alleles for all four; the second extended across six of the associated SNPs and contained the risk alleles for all six but differed from the first haplotype for alleles at other SNPs. We propose that each haplotype harbors a distinct causal variant, despite sharing associations with the same risk alleles, demonstrating the complexity of disease



associations. Similar patterns occurred for haplotypes in other genes (Supplementary Table S2).

Admixed populations offer both advantages and challenges for genetic studies. The advantages include the opportunity to exploit information on local ancestry to localize disease-susceptibility genes. The challenges include the presence of genetic heterogeneity within the sample. As with any such analysis, confirmation of these findings awaits replication in another sample.

In summary, linkage and association analysis using genotypes on 9203 fine-mapping SNPs added to five chromosomal regions confirmed each linkage and identified potential susceptibility genes in each of the regions. In addition, every region appeared to harbor at least two T2D-susceptibility genes.

## ACKNOWLEDGEMENTS

This project was funded by NIH Grant DK071100. This study was supported by the American Diabetes Association. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C.

## WEB RESOURCES

GENNID [http://professional.diabetes.org/Diabetes\\_Research.aspx?typ=18&cid=64380](http://professional.diabetes.org/Diabetes_Research.aspx?typ=18&cid=64380)

SNAGGER <http://snagger.sourceforge.net/>

CIDR <http://www.cidr.jhmi.edu/>

Rutgers Combined Linkage-Physical Map of the Human Genome <http://compgen.rutgers.edu/RutgersMap/Default.aspx>

Conway Institute Bioinformatics Service, University of Dublin, Ireland <http://integrin.ucd.ie/cgi-bin/rs2cm.cgi>

MERLIN <http://www.sph.umich.edu/csg/abecasis/Merlin>

PedCheck <http://watson.hgen.pitt.edu/register/docs/pedcheck.html>

jPAP <http://hasstedt.genetics.utah.edu/>

HapMap <http://hapmap.ncbi.nlm.nih.gov/>

- 1 Stone, S., Abkevich, V., Hunt, S. C., Gutin, A., Russell, D. L., Neff, C. D. *et al.* A major predisposition locus for severe obesity, at 4p15-p14. *Am. J. Hum. Genet.* **70**, 1459–1468 (2002).
- 2 Stone, S., Abkevich, V., Russell, D. L., Riley, R., Timms, K., Tran, T. *et al.* TBC1D1 is a candidate for a severe obesity gene and evidence for a gene/gene interaction in obesity predisposition. *Hum. Mol. Genet.* **15**, 2709–2720 (2006).
- 3 Ewing, C. M., Ray, A. M., Lange, E. M., Zuhlke, K. A., Robbins, C. M., Tembe, W. D. *et al.* Germline mutations in HOXB13 and prostate-cancer risk. *N. Engl. J. Med.* **366**, 141–149 (2012).
- 4 Martin, L. J., Kissebah, A. H. & Olivier, M. Accounting for a quantitative trait locus for plasma triglyceride levels: utilization of variants in multiple genes. *PLoS ONE* **7**, e34614 (2012).
- 5 Christians, J. K. & Keightley, P. D. Fine mapping of a murine growth locus to a 1.4-cM region and resolution of linked QTL. *Mamm. Genome* **15**, 482–491 (2004).
- 6 Christians, J. K., Hoeflich, A. & Keightley, P. D. PAPP2, an enzyme that cleaves an insulin-like growth-factor-binding protein, is a candidate gene for a quantitative trait locus affecting body size in mice. *Genetics* **173**, 1547–1553 (2006).
- 7 Prokopenko, I., Zeggini, E., Hanson, R. L., Mitchell, B. D., Rayner, N. W., Akan, P. *et al.* Linkage disequilibrium mapping of the replicated type 2 diabetes linkage signal on chromosome 1q. *Diabetes* **58**, 1704–1709 (2009).
- 8 Das, S. K. & Elbein, S. C. The genetic basis of type 2 diabetes. *Cellscience* **2**, 100–131 (2006).
- 9 Elbein, S. C. Evaluation of polymorphisms known to contribute to risk for diabetes in African and African-American populations. *Curr. Opin. Clin. Nutr. Metab. Care* **10**, 415–419 (2007).
- 10 Smith, L. M., Yao-Borengasser, A., Starks, T., Tripputi, M., Kern, P. A. & Rasouli, N. Insulin resistance in African-American and Caucasian women: differences in lipotoxicity, adipokines, and gene expression in adipose tissue and muscle. *J. Clin. Endocrinol. Metab.* **95**, 4441–4448 (2010).
- 11 Elbein, S. C., Kern, P. A., Rasouli, N., Yao-Borengasser, A., Sharma, N. K. & Das, S. K. Global gene expression profiles of subcutaneous adipose and muscle from glucose-tolerant, insulin-sensitive, and insulin-resistant individuals matched for BMI. *Diabetes* **60**, 1019–1029 (2011).
- 12 Wheeler, E. & Barroso, I. Genome-wide association studies and type 2 diabetes. *Brief Funct. Genomics* **10**, 52–60 (2011).
- 13 Langberg, K. A., Ma, L., Sharma, N. K., Hanis, C. L., Elbein, S. C., Hasstedt, S. J. *et al.* Single nucleotide polymorphisms in JAZF1 and BCL11A gene are nominally associated with type 2 diabetes in African-American families from the GENNID study. *J. Hum. Genet.* **57**, 57–61 (2012).
- 14 Palmer, N. D., McDonough, C. W., Hicks, P. J., Roh, B. H., Wing, M. R., An, S. S. *et al.* A genome-wide association search for type 2 diabetes genes in African Americans. *PLoS One* **2**, e29202 (2012).
- 15 McCarthy, M. I. & Zeggini, E. Genome-wide association studies in type 2 diabetes. *Curr. Diab. Rep.* **9**, 164–171 (2009).
- 16 Elbein, S. C., Das, S. K., Hallman, D. M., Hanis, C. L. & Hasstedt, S. J. Genome-wide linkage and admixture mapping of type 2 diabetes in African American families from the American Diabetes Association GENNID (Genetics of NIDDM) Study Cohort. *Diabetes* **58**, 268–274 (2009).
- 17 Hasstedt, S. J., Hanis, C. L. & Elbein, S. C. Univariate and bivariate linkage analysis identifies pleiotropic loci underlying lipid levels and type 2 diabetes risk. *Ann. Hum. Genet.* **74**, 308–315 (2010).
- 18 Hasstedt, S. J., Hanis, C. L., Das, S. K. & Elbein, S. C. Pleiotropy of type 2 diabetes with obesity. *J. Hum. Genet.* **56**, 491–495 (2011).
- 19 Raffel, L. J., Robbins, D. C., Norris, J. M., Boerwinkle, E., DeFronzo, R. A., Elbein, S. C. *et al.* The GENNID Study. A resource for mapping the genes that cause NIDDM. *Diabetes Care* **19**, 864–872 (1996).
- 20 Edlund, C. K., Lee, W., Hi, L. D., Van Den Berg, D. J. & Conti, D. V. Snagger: A user-friendly program for incorporating additional information for tagSNP selection. *BMC Bioinformatics* **9**, 174 (2008).
- 21 O'Connell, J. R. & Weeks, D. E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **63**, 259–266 (1998).
- 22 Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97–101 (2002).
- 23 Hasstedt, P. A. P. Document-driven software for genetic analysis. *Genet. Epidemiol.* **29**, 255 (2005).
- 24 Hasstedt, S. J. & Thomas, A. Detecting pleiotropy and epistasis using variance components linkage analysis in jPAP. *Hum. Hered.* **72**, 258–263 (2011).
- 25 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
- 26 Benjamini, Y., Heller, R. & Yekutieli, D. Selective inference in complex research. *Philos. Transact. R. Soc. A Math. Phys. Eng. Sci.* **367**, 4255–4271 (2009).
- 27 Di Paola, R., Wojcik, J., Succurro, E., Marucci, A., Chandalia, M., Padovano, L. *et al.* GRB10 gene and type 2 diabetes in whites. *J. Intern. Med.* **267**, 132–133 (2010).
- 28 McDonough, C. W., Bostrom, M. A., Lu, L., Hicks, P. J., Langefeld, C. D., Divers, J. *et al.* Genetic analysis of diabetic nephropathy on chromosome 18 in African Americans: linkage analysis and dense SNP mapping. *Hum. Genet.* **126**, 805–817 (2009).
- 29 Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R. *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* **40**, 161–169 (2008).
- 30 Schubert, C. The genomic basis of the Williams-Beuren syndrome. *Cell Mol. Life Sci.* **66**, 1178–1197 (2009).
- 31 Pober, B. R., Wang, E., Caprio, S., Petersen, K. F., Brandt, C., Stanley, T. *et al.* High prevalence of diabetes and pre-diabetes in adults with Williams syndrome. *Am. J. Med. Genet. C Simin. Med. Genet.* **15**, 291–298 (2010).
- 32 Torgerson, D. G., Capurso, D., Mathias, R. A., Graves, P. E., Hernandez, R. D., Beaty, T. H. *et al.* Resequencing candidate genes implicates rare variants in asthma susceptibility. *Am. J. Hum. Genet.* **90**, 273–281 (2012).
- 33 Duez, H., Cariou, B. & Staels, B. DPP-4 inhibitors in the treatment of type 2 diabetes. *Biochem. Pharmacol.* **83**, 823–832 (2012).
- 34 Qi, S. Y., Riviere, P. J., Trojnar, J., Junien, J. L. & Akinsanya, K. O. Cloning and characterization of dipeptidyl peptidase 10, a new member of an emerging subgroup of serine proteases. *Biochem. J.* **373** (Pt 1), 179–189 (2003).
- 35 Grundberg, E., Small, K. S., Hedman, A. K., Nica, A. C., Buil, A., Keidson, S. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)

## APPENDIX

### American diabetes association GENNID study group

Genetic material collected by, and families characterized by, the American Diabetes Association GENNID Study Group, which includes Eric Boerwinkle, PhD, University of Texas Health Science Center; John Buse, MD, PhD, University of North Carolina; Ralph DeFronzo, MD, University of Texas Health Science Center; David Ehrmann, MD, University of Chicago; Steven C. Elbein, MD,

University of Utah/University of Arkansas; Wilfred Fujimoto, MD and Steven E. Kahn, MB, ChB, University of Washington; Craig L. Hanis, PhD, University of Texas Health Science Center; Richard A. Mulivor, PhD and Jeanne C. Beck, PhD, Coriell Cell Repositories; Jill Norris, PhD, University of Colorado School of Medicine; M. Alan Permutt, MD and Philip Behn, MD, Washington University School of Medicine; Leslie Raffel, MD, Cedars-Sinai Medical Center; and David C. Robbins, MD, Medlantic Research Institute.