Original Paper

Meta-Analysis of 23 Type 2 Diabetes Linkage Studies from the International Type 2 Diabetes Linkage Analysis Consortium

Weihua Guan[a]   Anna Pluzhnikov[b]   Nancy J. Cox[b]   Michael Boehnke[a]

for the International Type 2 Diabetes Linkage Analysis Consortium

[a]Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, Mich., [b]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Ill., USA

Abstract

Background: The International Type 2 Diabetes Linkage Analysis Consortium was formed to localize type 2 diabetes predisposing variants based on 23 autosomal linkage scans. **Methods:** We carried out meta-analysis using the genome scan meta-analysis (GSMA) method which divides the genome into bins of ~30 cM, ranks the best linkage results in each bin for each sample, and then sums the ranks across samples. We repeated the meta-analysis using 2 cM bins, and/or replacing bin ranks with measures of linkage evidence: bin maximum LOD score or bin minimum p value for bins with p value <0.05 (truncated p value). We also carried out computer simulations to assess the empirical type I error rates of these meta-analysis methods. **Results:** Our analyses provided modest evidence for type 2 diabetes-predisposing variants on chromosomes 4, 10, and 14 (using LOD scores or truncated p values), or chromosome 10 and 16 (using ranks). Our simulation results suggested that uneven marker density across studies results in substantial variation in empirical type I error rates for all meta-analysis methods, but that 2 cM bins and scores that make more explicit use of linkage evidence, especially the truncated p values, reduce this problem. **Conclusion:** We identified regions modestly linked with type 2 diabetes by summarizing results from 23 autosomal genome scans.

Copyright © 2008 S. Karger AG, Basel

Introduction

Linkage studies are widely used to map genetic variants that predispose to human diseases. However, for complex diseases like type 2 diabetes for which many genetic variants together with environmental and behavioral factors are likely involved [1], mapping disease predisposing variants has proven to be difficult. To date, >20 genome wide linkage studies have been carried out to localize type 2 diabetes predisposing variants [2–28]. In only a few cases have individual studies provided significant linkage results [2, 8, 12, 23].

A likely explanation for these mixed results is that individual type 2 diabetes linkage studies are underpowered to detect predisposing variants with small effects. One approach to address this limitation is to carry out a joint analysis of the primary linkage data from a large number of studies. While a joint analysis of primary data is in principle an attractive approach to increase sample

Michael Boehnke, PhD
Department of Biostatistics, School of Public Health
University of Michigan, 1420 Washington Heights
Ann Arbor, MI 48109-2029 (USA)
Tel. +1 734 936 1001, Fax +1 734 615 8322, E-Mail boehnke@umich.edu

size and statistical power, it is not always possible to obtain data from all relevant studies. Differences in study populations, disease definition, and genetic markers genotyped also might lead to problems with this analytic approach. Therefore, meta-analysis of study results often is chosen as an alternative to joint analysis of the primary data to validate regions suggested by one or more smaller studies and to localize susceptibility variants suggested by weak but consistent evidence across multiple studies.

In this paper, we describe results of meta-analyses of 23 autosomal genome scans from the International Type 2 Diabetes Linkage Analysis Consortium (csg.sph.umich.edu/consortium/). This Consortium was formed to bring together groups around the world who are carrying out linkage studies of type 2 diabetes. Sixteen research groups have participated in the Consortium and have contributed autosomal genome scan data on a total of 23 samples. These studies provided an excellent basis for carrying out a large scale meta-analysis for type 2 diabetes.

We describe results for a genome scan meta-analysis (GSMA) [29] of the Consortium data. In GSMA, chromosomes are divided into bins of approximately equal length (~30 cM), which are then ranked within each genome scan by their maximum linkage signal. The ranks are summed across genome scans at each bin to form the test statistics, with significance evaluated analytically or by permutation.

GSMA provides a simple means for combining data across multiple linkage studies. However, ~30 cM bins provide only limited resolution, and ranking bins results in information loss, since the highest ranking bins may represent very different levels of linkage evidence across studies, and lower ranking bins likely provide limited and nearly equivalent linkage evidence within or across studies. Therefore, we also describe modifications of GSMA that address these limitations by considering more direct measures of linkage evidence and smaller bins, and report the results of these analyses on the Consortium data. Specifically, we repeat the meta-analysis using 2 cM rather than ~30 cM bins, and use alternative scores for a bin's linkage evidence, including maximum bin LOD score and truncated minimum bin linkage p value. The latter is a modified version of Fisher's method [30] suggested by Zaykin et al. [31]. Since Fisher's method is known to have low power in the case in which many tests are not significant [32], Zaykin et al. suggested that only p values less than some pre-chosen significance level be considered. This approach acknowledges that a broad range of p values are nearly equivalently uninformative, notably the approximate half of the genome for which the maximum LOD score is zero. Finally, we carry out computer simulations under null hypothesis of no linkage to compare the type I error rates of these alternative methods.

Materials and Methods

Samples and Sample-Specific Maps

We carried out meta-analysis on the results of autosomal linkage genome scans for 23 samples (table 1). Invitations were sent to all research groups known to be conducting genome-wide linkage studies of Type 2 diabetes, and the 23 samples are from the groups that agreed to participate. Detailed information about these samples is provided in the original references (table 1). These samples include a total of 9,455 affected individuals, and there is no known overlap between samples. For each sample, study investigators provided us with the marker maps (genetic markers and map positions in cM) used for the evaluation of linkage evidence, and files that gave LOD scores and corresponding chromosomal positions. LOD scores were evaluated at equally-spaced points along each chromosome (15 samples), or at each genetic marker position and at a fixed number of equally-spaced points between each marker pair (8 samples). There was substantial variability in genetic markers used and some variability in marker density across the 23 samples (table 1). Different groups used different statistical methods to evaluate the linkage evidence, including ASM LOD scores [33], NPL LOD scores [34], and variance component LOD scores [35].

Map and LOD Score Conversion

All meta-analysis methods we applied assume that linkage results are reported for the same genetic map, or at least that marker maps are well aligned so that a bin corresponds to the same chromosomal region in each study. To minimize deviations from this assumption, we generated a single master genetic map based on the sex-averaged deCODE [36] and Marshfield [37] genetic maps using linear interpolation. For each marker M in a study-specific map, we determined its position x on the master map according to the following procedures. First, if M was included in the deCODE map at position d, $x = d$. Second, if M was not in the deCODE map, but was in the Marshfield map at position m, we identified the nearest flanking markers for M which were in both the deCODE and Marshfield maps. If the flanking marker positions were $d_1 < d_2$ and $m_1 < m_2$ in the deCODE and Marshfield maps, respectively, then by linear interpolation, $x = d_1 + (d_2 - d_1)(m - m_1)/(m_2 - m_1)$. Third, if M was in neither the deCODE nor Marshfield maps, we used linear interpolation based on base pair distance in the August 2001 freeze of the Human Genome Project Working Draft [38]. 3,889 unique markers could be placed using this interpolation scheme; 96 markers could not be placed and were discarded.

We rescaled each study's LOD score evaluation positions in the same manner and calculated linear interpolated LOD score values at those evaluation positions. To obtain the interpolated LOD score L at position x, we found the two nearest flanking positions ($x_1 < x < x_2$) at which LOD scores L_1 and L_2 were reported and calculated the interpolated LOD score as $L = L_1 + (L_2 - L_1)(x - x_1)/(x_2 - x_1)$. At chromosome ends, we used the LOD score at the nearest reported point on that chromosome.

Table 1. Characteristics of Type 2 diabetes genome scans

Study	Population	# affected individuals		Marker density[a] cM/marker	Linkage statistic	Ref.	Subgroup[b]
		meta-analysis	simulations				
FUSION 1	European (Finland)	1,155	605	8.4	ASM LOD	11	European/ European American
FUSION 2	European (Finland)	585	389	8.1	ASM LOD	25	
University of Lund Botnia	European (Finland)	233	110	4.6	NPL	3	
University of Lund Sibs	European (Finland)	1,150	507	6.4	NPL	15	
Pasteur Institute Lille	European (France)	433	327	8.8	ASM LOD	13	
United Kingdom	European (UK)	1,223	700	9.0	ASM LOD	17	
Washington University	Israeli Ashkenazi	472	307	7.1	NPL	16	
University of Arkansas EA	European American	127	127	7.4	NPL	8	
GENNID EA1	European American	188	150	7.5	ASM LOD	10, 56	
GENNID EA2	European American	225	121	8.5	ASM LOD	10, 56	
Amish Family Diabetes Study	European American	55	55	9.2	NPL	21	
AADM	West African	402	402	9.1	NPL	26	African/ African American
University of Arkansas AA	African American	341	113	8.7	NPL		
GENNID AA	African American	124	61	8.8	ASM LOD	10, 56	
Wake Forest University	African American	631	407	9.1	ASM LOD	24	
SFBR	Mexican American	74	35	10.2	Variance component LOD	55	Mexican American
University of Chicago 1	Mexican American	424	206	6.9	NPL	2	
GENNID MA1	Mexican American	173	123	7.4	ASM LOD	10, 56	
GENNID MA2	Mexican American	137	100	8.9	ASM LOD	10, 56	
U Texas Health Science Center at San Antonio	Mexican American	440	64	10.1	Variance component LOD	7	
NIDDK Phoenix	Native American	625	529	6.5	NPL	5	
University of Chicago 2	Japanese	188	188	8.4	NPL	22	
GENNID JA	Japanese American	50	19	8.9	ASM LOD	10, 56	

[a] Marker densities were calculated based on the master genetic maps (see Materials and Methods).
[b] Subgroup denotes the group of studies on which the subgroup analyses were performed.

Linkage Meta-Analysis Methods: GSMA

To apply GSMA, we divided the joint autosomal maps into 115 bins of width ~30 cM. Ideally, all bins would be of equal width and any linkage peak would fall within a single bin. In fact, chromosome map lengths are not even multiples of 30 cM, and given 23 studies, it is impossible to assign bins without splitting some linkage peaks. For simplicity, for a chromosome of length L cM, we chose bin width L/b cM, with b chosen to make L/b as near to 30 as possible. Table 2 displays the numbers of bins and corresponding bin widths for each of the 22 autosomes. Bin widths varied from 26.8 cM (chromosome 17) to 36.0 cM (chromosome 22).

For each genome scan, we assigned ranks to bins by maximum linkage signal, so that the bin with the largest maximum linkage signal for that genome scan was assigned rank 115, the second highest rank 114, and so forth. To account partially for differences in study sample sizes, we calculated the score for bin b as $S(b) = \Sigma_i w_i R_i(b)$, where w_i is the square root of number of affected individuals in study i and $R_i(b)$ is the rank of bin b in study i. To estimate the significance for a particular bin score $S(b)$, we used a permutation approach. Specifically, we generated a large number of replicate permutations of the ranks of bins within each study, recalculated the scores $S(b')$ over all bins b' in the genome, and pooled these scores over all permutations and bins to estimate the distribution of $S(b)$. For a bin b with observed score $S(b)$, we estimated its p value as the proportion of observed $S(b')$ values greater than or equal to the observed $S(b)$.

Linkage Meta-Analysis Methods: Alternative Scoring Schemes

To reflect the strength of linkage results more directly, we considered two scoring schemes as alternatives to weighted ranks. First, we used the maximum LOD scores of bins and calculated $S(b) = \Sigma_i \max LOD_i(b)$, where $\max LOD_i(b)$ is the maximum LOD score in bin b in study i. Because the reported linkage statistics (LOD scores) from the 23 studies (table 1) all have the same asymptotic distribution under the null hypothesis (1/2 : 1/2 mixture of χ^2 distribution with 1 degree of freedom and point mass at 0, divided by 2 ln10 [39]), we were able to combine the reported LOD scores directly from individual studies. Had this not been the case, transformation would have been required to put the linkage statistics on a common scale.

Second, we calculated the minimum p values of bins, $p_i(b) = 1/2 \{1 - \chi^2 [2\ln10 \max LOD_i(b)]\}$, where $\chi^2(\cdot)$ is the cumulative chi-square distribution with 1 degree of freedom. We then truncated the minimum p values bins at 0.05

$$p_{it}(b) = \begin{cases} p_i(b), & \text{if} \quad p(b) < 0.05 \\ 1, & \text{if} \quad p(b) \geq 0.05 \end{cases}$$

and calculated $S(b) = \Sigma_i - \log_{10}[p_{it}(b)]$. Truncating p values in this way meant that any $p > 0.05$ was treated as equally uninformative. We evaluated the significance of these two scores by permutation as described above for ranks.

Linkage Meta-Analysis Methods: Bin Width

To obtain higher resolution than that provided by ~30 cM bins, we repeated all three meta-analysis methods using bins of width 2 cM, yielding 1,758 bins for the autosomal genome (table 2).

Table 2. Estimated chromosome lengths in the master map and numbers of bins

Chromosome	Chromosome length, cM	~30 cM bins		2 cM bins
		number of bins	bin width cM	number of bins
1	275	9	30.6	138
2	260	9	28.9	130
3	224	7	32.0	112
4	210	7	30.0	105
5	206	7	29.4	103
6	190	6	31.7	95
7	184	6	30.7	92
8	168	6	28.0	84
9	162	5	32.4	81
10	178	6	29.7	89
11	154	5	30.8	77
12	172	6	28.7	86
13	130	4	32.5	65
14	124	4	31.0	62
15	134	4	33.5	67
16	132	4	33.0	66
17	134	5	26.8	67
18	122	4	30.5	61
19	116	4	29.0	58
20	100	3	33.3	50
21	68	2	34.0	34
22	72	2	36.0	36
Total	3,515	115	30.6	1,758

Subgroup Meta-Analysis

We carried out subgroup analyses on subsets of studies defined by the historical geographic ancestry of their samples. We did so for European/European American, African/African American, and Mexican American (table 1).

Simulation

To compare the meta-analysis methods, we obtained family structures, genotypes, marker maps, and marker allele frequencies for each genome scan from each of the participating research groups. We then used gene dropping to evaluate the validity of the six combinations of three scoring schemes and two bin widths, and to evaluate the impact of differing sets of markers used in the different studies and uneven marker densities within each study.

We simulated pedigree data for the 23 studies under the null hypothesis of no linkage using MERLIN version 0.10.3 [40]. To reduce the computation load, we reduced the complexity of some family sets. Specifically, we restricted samples to nuclear families, and deleted unaffected siblings as needed so that no nuclear family had >8 siblings. The numbers of individuals used in the simulations for each of the 23 studies are listed in table 1. We considered three simulation conditions. In simulation 1, to evaluate the type I error rates in our meta-analysis, we generated genotype data for each study based on the marker map for that study. In

simulation 2, we generated genotype data for all 23 studies based on the marker map for one study (FUSION 1 [11]) to eliminate the difference between individual studies. In simulation 3, we generated genotype data for all 23 studies based on an artificial map in which each autosome is 150 cM long, markers are evenly spaced at 10 cM, and each marker has four equally-frequent alleles. Compared to results in simulation 2, simulation 3 allows us to further assess the impact of (uneven) marker density and informativeness on type I error rates.

For each of the three simulation settings, we generated 1,000 simulated datasets, each comprised of 23 samples, as in the Consortium study. We carried out affected sibling pairs (ASP) linkage analysis with LOD scores evaluated every 2 cM based on the linear model of Kong and Cox [33] using MERLIN, version 0.10.3 [40]. We carried out meta-analysis for each of our statistics computed for ~30 and 2 cM bins, and estimated significance levels (P_i) by permutation using 1,000 permutation iterates per simulated dataset. We estimated the empirical p values for each bin as the proportion of the P_i's less than 0.05 at the corresponding bin for the 1,000 simulation replicates.

Results

Meta-Analysis of Consortium Data

We first carried out meta-analyses using ~30 cM bins. In this analysis, two regions showed significant results at the 0.01 level: chromosome 10 at 119–148 cM (p_{TP} = 0.0020) and chromosome 14 at 31–62 cM (p_{LOD} = 0.0051, p_{TP} = 0.0068). All positions here and later are in the deCODE map. Although the signals did not meet genome-wide significance, they did meet genome-wide suggestive significance, that is, one false positive per genome scan, based on Bonferroni correction. Interestingly, the chromosome 10 region includes the transcription factor 7-like 2 (*TCF7L2*) gene [41], variants in which are known to influence type 2 diabetes susceptibility.

We next performed the meta-analyses for 2 cM bins to obtain finer resolution (table 3, fig. 1). Consistent with the 30 cM bin analyses, the strongest results for the combined data for the LOD score (p_{LOD}) and truncated p value (p_{TP}) based analyses were obtained on chromosome 14 at ~58 cM (p_{LOD} = 0.0011, p_{TP} = 0.0017), chromosome 10 at ~138 cM (p_{LOD} = 0.0021, p_{TP} = 0.0019), about 4 cM from *TCF7L2*. Other interesting results were obtained on chromosome 4 at ~76 cM (p_{LOD} = 0.0048, p_{TP} = 0.0055) and ~176 cM (p_{LOD} = 0.0129, p_{TP} = 0.0074), but these do not meet the genome-wide suggestive significance based on the Bonferroni correction. All these regions also were identified as being of interest in the weighted rank-based analysis (p_{RANK} < 0.10), with the chromosome 10 region the second strongest rank-based result (p_{RANK} = 0.0055). Table 4 lists the studies that made notable contributions

Table 3. Meta-analysis results of consortium data using 2 cM bins: regions (cM) with p value < 0.05

Chr	Rank$_w$	LOD	TPM
1		58–60 (0.049)	58–60 (0.026)
			162–166 (0.013)
2			0–20 (0.020)
			44–54 (0.024)
		258 (0.038)	258 (0.025)
3		58–64 (0.021)	60–62 (0.044)
	180–186 (0.013)		
4		72–84 (0.0048)	68–86 (0.0055)
	138–140 (0.028)		
	172–182 (0.040)	170–180 (0.013)	170–190 (0.0074)
5		60–66 (0.039)	54–60 (0.040)
6	114–134 (0.018)	146–160 (0.016)	146–158 (0.017)
7	28–44 (0.022)		
8	34–40 (0.019)		
9		24–34 (0.023)	34 (0.049)
10	110–138 (0.0055)	122–156 (0.0021)	122–158 (0.0019)
14		48–68 (0.0011)	46–70 (0.0017)
		98–106 (0.042)	100–106 (0.020)
16	38–88 (0.0020)		
17	34–44 (0.012)	40–52 (0.023)	
20	4–12 (0.026)	60–64 (0.039)	
	34 (0.039)		
21			46–48 (0.044)
22	44–48 (0.037)		

The results are presented separately for different scoring schemes: weighted ranks (Rank$_w$); maximum LOD score (LOD); truncated minimum p value (TPM). The best p values within these regions are given in parentheses.

(LOD >1.0) to each of these meta-analysis linkage signals. Notably, the chromosome 2 signal at 258 cM is 1–2 cM from calpain-10 [42]. Also, the chromosome 3 signal at ~60 cM is ~20 cM from the peroxisome proliferator-activated receptor gamma (*PPARG* [43]) locus.

Combining results over all studies implicitly assumes that genetic mechanisms and allele frequencies are consistent across all groups. In an attempt to decrease genetic heterogeneity, we re-ran the analysis on geographically defined subgroups: European/European American, African/African American, and Mexican American (table 1). The regions identified by at least one method (LOD score, truncated p value, and/or rank) using ~30 cM bins (significant at the 0.01 level) are listed in table 5. Many of these regions identified by subgroup analyses were not identified in the full analyses, consistent with either more power with more homogeneous subsets of the data, or an increased false positive rate owing to more tests. Among
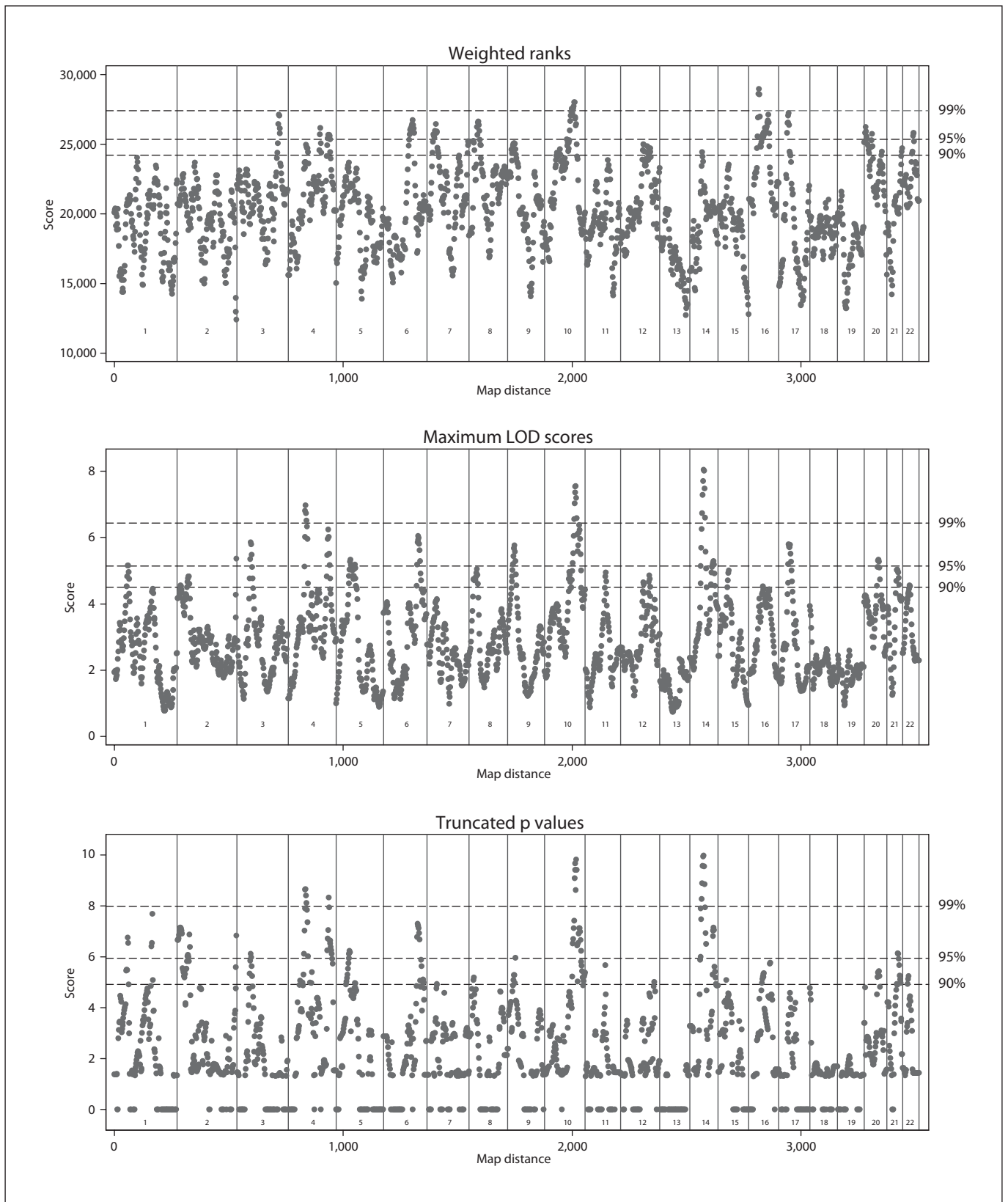
Fig. 1. Meta-analysis results of the IT2DLAC data using 2 cM bins.

Table 4. Individual studies with substantial contributions to linkage signals of meta-analysis

Region	Study	Position[a], cM	LOD, score[b]	Ref.[c]
Chromosome 4 (~76 cM)	AADM	76	1.76	15
	University of Arkansas AA	76	2.77	
	GENNID MA1	104	1.15	
	University of Lund Sibs	92	1.44	
Chromosome 4 (~176 cM)	University of Arkansas AA	158	1.11	7, 16, 19
	University of Lund Botnia	180	1.00	
	Washington University	178	1.26	
Chromosome 10 (~138 cM)	University of Arkansas AA	154	1.04	7, 17, 25
	FUSION 2	136	1.11	
	GENNID JA	136	1.06	
	United Kingdom	108	1.98	
	U Texas Health Science Center at San Antonio	150	2.88	
Chromosome 14 (~58 cM)	AADM	60	1.75	21, 25
	University of Arkansas AA	60	1.52	
	FUSION 2	58	1.98	
	GENNID AA	88	1.08	
	GENNID EA2	38	1.07	
	SFBR	54	1.53	

[a] The positions are in the deCODE map, where a local maximum LOD scores are achieved in the individual study.
[b] The LOD scores are based on linear interpolation of original LOD scores in the individual study (see Materials and Methods).
[c] Publications which showed strong or modest linkage signal (LOD score greater than 1) close to the corresponding regions. They may or may not be consistent with the data applied in our meta-analyses.

these results, the analyses of the African/African American group identified a region on chromosome 6 at 127–158 cM (p_{RANK} = 0.0012, p_{LOD} = 0.0002, p_{TP} = 0.0004) with p values almost reaching the genome-wide significance level using the Bonferroni correction. Two of the four studies included in this subgroup showed linkage signals which may contribute to this finding: GENNID AA samples have an interpolated LOD score of 2.45 at ~148 cM on the same chromosome, and Wake Forest University samples have an interpolated LOD score of 2.25 at ~162 cM. Analyses using 2 cM bins showed the strongest signal at ~152 cM in this region (p_{RANK} = 5.5 × 10^{-5}, p_{LOD} = 4.3 × 10^{-5}, p_{TP} = 5.7 × 10^{-5}).

Results for the several meta-analysis methods are correlated. To assess this correlation, we calculated the Pearson correlations on the logarithm of the p values between the different pairs of methods for the analysis of all 23 genome scans. Not surprisingly, the correlation is strongest for the methods based on LOD scores and truncated p values (r = 0.88 and 0.89 for 2 and ~30 cM bins, respectively), and less strong for those based on LOD scores and weighted ranks (0.55 and 0.63), or p values and weighted ranks (0.36 and 0.54).

Simulation Results: Type I Error

To assess the validity of the different meta-analysis methods, we carried out a computer simulation. Results are summarized for each meta-analysis score and bin width (table 6). The bin-wise type I errors are calculated as the proportions of simulation replicates with p value < 0.05, and are then averaged over all bins on the genome. Given study-specific maps with different marker densities (simulation 1) and ~30 cM bins, the average type I error rates are modestly inflated beyond the nominal 0.05 level, but the point-wise type I errors can be either inflated or deflated, resulting in fairly large variations among bins. Inflation and variation are greatest for ranks and least for truncated p values, and greater for ~30 cM bins than for 2 cM bins. Given a constant map across all studies (simulation 2), inflation of the average type I error rates and the variations are reduced, particularly for 2 cM bins. The truncated p value based method still outperforms the other two in both average type I error rates and variation. The inflation is eliminated for 2 cM bins and the variations are greatly reduced for all the methods, comparable to those obtained in simulation 3. Given the artificial setting of identical length chromosomes and

Table 5. Regions of interest in subgroup analyses using ~30 cM bins

Subgroup	Chromosome	Region, cM	p_{RANK}	p_{LOD}	p_{TP}
European/European American	6	95–127	0.0039	0.032	0.041
African/African American	4	60–90	0.153	0.0017	0.0044
	6	127–158	0.0012	0.0002	0.0004
	6	158–190	0.0017	0.0028	0.0035
	10	119–148	0.0069	0.055	0.0088
	14	62–93	0.0014	0.0013	0.0025
	16	33–66	0.0038	0.036	0.019
Mexican American	3	64–96	0.0002	0.0017	0.0016
	15	34–67	0.0033	0.040	0.036

The regions include bins identified by at least one of the three methods at 0.01 significance level.

The results are presented separately for different scoring schemes: weighted ranks (p_{RANK}); maximum LOD score (p_{LOD}); truncated minimum p value (p_{TP}).

Table 6. Averaged Size (± standard deviations) over all bins on the genome for meta-analysis methods under different simulation settings: simulation based on actual studies' marker maps (setting 1), based on a single study (FUSION 1) marker map (setting 2), and based on an artificial, evenly spaced marker map (setting 3)

Significance	Method		Setting		
			1	2	3
Bin-wise[a]	$Rank_w$	30 cM	0.058 ± 0.038	0.055 ± 0.027	0.051 ± 0.007
	LOD		0.055 ± 0.030	0.054 ± 0.026	0.051 ± 0.008
	TPM		0.053 ± 0.026	0.052 ± 0.021	0.050 ± 0.008
	$Rank_w$	2 cM	0.052 ± 0.024	0.051 ± 0.007	0.050 ± 0.007
	LOD		0.051 ± 0.017	0.051 ± 0.008	0.050 ± 0.007
	TPM		0.051 ± 0.014	0.050 ± 0.007	0.050 ± 0.007
Genome-wide suggestive[b]	$Rank_w$	30 cM	1.1 ± 1.1 ($\times 10^{-2}$)	1.0 ± 0.74 ($\times 10^{-2}$)	0.95 ± 0.32 ($\times 10^{-2}$)
	LOD		1.0 ± 0.84 ($\times 10^{-2}$)	0.99 ± 0.72 ($\times 10^{-2}$)	0.93 ± 0.32 ($\times 10^{-2}$)
	TPM		0.99 ± 0.69 ($\times 10^{-2}$)	0.93 ± 0.58 ($\times 10^{-2}$)	0.93 ± 0.28 ($\times 10^{-2}$)
	$Rank_w$	2 cM	0.71 ± 1.1 ($\times 10^{-3}$)	0.56 ± 0.78 ($\times 10^{-3}$)	0.71 ± 0.83 ($\times 10^{-3}$)
	LOD		0.51 ± 0.77 ($\times 10^{-3}$)	0.57 ± 0.77 ($\times 10^{-3}$)	0.60 ± 0.74 ($\times 10^{-3}$)
	TPM		0.51 ± 0.79 ($\times 10^{-3}$)	0.57 ± 0.81 ($\times 10^{-3}$)	0.61 ± 0.77 ($\times 10^{-3}$)

The results are presented separately for different scoring schemes: weighted ranks ($Rank_w$); maximum LOD score (LOD); truncated minimum p value (TPM).

[a] Bin-wise significance level is .05.

[b] Genome-wide suggestive significance levels are 1/(number of bins) using the Bonferroni correction, which are 0.87×10^{-2} (setting 1 and 2) and 0.91×10^{-2} (setting 3) for 30-cM bins, and 0.57×10^{-3} (setting 1 and 2) and 0.60×10^{-3} (setting 3) for 2-cM bins.

identical marker maps across studies (simulation 3), type I error rates are consistent with nominal levels for all methods. Similar results were obtained evaluated for genome-wide suggestive type I error rates.

While simulation 1 tries to reproduce the scenario of our meta-analyses of the consortium data, simulation 2 unifies the sets of markers used in different studies. Because the simulations were carried out under the null hypothesis of no linkage, the positions of markers should not affect the linkage or meta-analysis results. Difference in marker sets in fact reflects the variation of marker distributions across studies, and then impacts the results through effect of marker densities. Elimination of this difference gives us a better opportunity to evaluate the

Table 7. Sizes at four interesting regions for meta-analysis methods (using 2 cM bins) under simulation setting 1 (simulation based on actual studies' marker maps)

	Chromosome			
	4 ∼76 cM	4 ∼176 cM	10 ∼138 cM	14 ∼58 cM
Rank$_w$	0.086	0.070	0.037	0.063
LOD	0.080	0.060	0.040	0.078
TPM	0.067	0.049	0.039	0.071

The results are presented separately for using different scoring schemes: weighted ranks (Rank$_w$); maximum LOD score (LOD); truncated minimum p value (TPM).

Fig. 2. Sizes of bins in simulation 2 using 30 cM bins and truncated p values as the test scores.

effect of uneven marker densities. As an example, the results of simulation 2 using 30 cM bins (and truncated p values as the test statistics) are shown in figure 2. The graph shows that several bins on chromosome 6, 11, 12, and 20 have inflated type I error rates >0.10, and the two bins on chromosome 22 have empirical type I error rates <0.02. This observation corresponds to the unusually high or low marker densities on the five chromosomes. The average marker distances on chromosomes 6, 11, 12, 20, and 22 are 5.3, 5.9, 5.9, 3.9, and 22.6 cM, respectively, compared to 8.5 to 11.7 cM on other chromosomes, suggesting that bins with higher marker densities tend to produce larger test statistics and more false-positive results. In contrast, for the 2 cM bins, the p values in these regions vary between 0.030 and 0.073, with most values close to the nominal value of 0.05. In simulation 3, when the markers are evenly spaced on the genome, inflation of the average type I errors disappears and variation is substantially reduced. This suggests that uneven marker spacing may play a role in departures of type I errors from nominal levels. Furthermore, the somewhat better performance of 2 cM compared to ∼30 cM bins may reflect a dilution of the effect of uneven marker density by the increased number of bins.

In simulation 1, the four regions (chromosome 4 at ∼76 cM and at ∼176 cM, 10 at ∼138 cM, and 14 at ∼58 cM) identified in the meta-analyses have acceptable empirical type I errors between 0.037 and 0.086 for all the three methods when using 2 cM bins (table 7). These results suggest that linkage evidence may be modestly overstated on chromosomes 4 (∼76 cM) and 14 (∼76 and ∼176 cM), and modestly understated on chromosome 10 (∼138 cM).
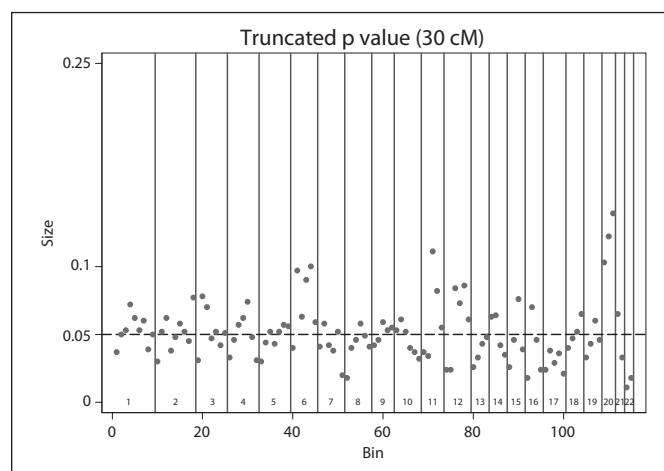
Discussion

This meta-analysis of all 23 type 2 diabetes autosomal genome scans provides modest evidence for linkage on several chromosomes, notably 14 (∼58 cM in the deCODE map), 10 (∼138 cM), 4 (∼76 cM and ∼176 cM), and 16 (∼44 cM). When we restricted our attention to geographically/racially defined subgroups in an effort to decrease genetic heterogeneity, we found evidence for several additional chromosomal regions, notably chromosomes 1 (∼164 cM in Europeans/European Americans, and ∼144 cM in Africans/African Americans), 2 (∼246–258 cM in Mexican Americans), and 6 (∼116 cM in Europeans/European Americans, and ∼152 cM in Africans/African Americans). Many of these regions are only identified in one subgroup and so may reflect true etiologic differences by group or a combination of false positives and false negatives.

Two regions which have received considerable attention in the type 2 diabetes genetic linkage literature owing to repeated reports of genetic linkage are 1q [beginning with 5, 8] and 20q [beginning with 4, 44, 45]. While neither of these regions was among the most interesting in this meta-analysis, both yielded some evidence for linkage: for chromosome 1 at ∼166 cM, $p_{TP} = 0.013$, and for chromosome 20 at ∼62 cM, $p_{TP} = 0.071$ and $p_{LOD} = 0.039$. In subgroup analyses on European/European American samples both regions are significant at the 0.01 level, but are not interesting in African/African American or Mexican American samples (all p > 0.10).

Several genome-wide association studies [46–51] have recently been published on susceptibility loci for type 2 diabetes. Except the *TCF7L2* gene which is identified by all the six studies and is one of the strongest signals in our meta-analysis, the other main signals found in these association studies are not close to the signals in meta-analysis. But there are intriguing overlaps between some modestly associated single nucleotide polymorphisms (SNPs) and our meta-analysis peaks. For example, our signal on chromosome 4 (~76 cM) is ~1 Mb from SNP rs282705 reported by Sladek et al. [46] (p = 9.0 × 10^{-6}), and 5 Mb from SNP SNP_A-4299379 reported by Saxena et al. [48] (p = 7.2 × 10^{-5}). Our signal on chromosome 14 (~58 cM) is ~8 Mb from SNPs rs1256517 and rs1256526 in Sladek et al. [46] (p = 4.7 × 10^{-6}). Our signal on chromosome 16 (~44 cM) is ~1 Mb from a modestly associated SNP rs724466 in Steinthorsdottir et al. [50] (p = 2.7 × 10^{-5}), and ~8 Mb from one of a modestly associated SNP rs10521095 in Scott et al. [47] (p = 3.8 × 10^{-5}).

Demenais et al. [52] conducted a meta-analysis of four type 2 diabetes linkage scans of European samples using the GSMA method with unweighted ranks. These four studies: University of Lund Botnia, University of Lund Sibs, United Kingdom, and Pasteur Institute Lille, are included in our analyses, although with slight differences in the samples. Demenais et al. found six regions on the genome to be nominally significant at 0.05 level. Their region on chromosome 16 (29–58 cM) directly overlaps one of our top four signals (chromosome 16 ~44 cM). Four more of their regions on chromosomes 1 (145–174 cM), 2 (58–87 cM), 6 (116–145 cM), and 17 (29–58 cM) are also at least nominally significant at the 0.05 level in our analyses (table 3), and the chromosome 1, 16, and 17 regions also are significant at the 0.01 level in our subgroup analyses of 11 European/European American samples (table 5).

While these results are of interest, it should be noted that the evidence for linkage at individual locations and overall is at best modest. For the ~30 cM bins, the smallest p values for the truncated p value, LOD score, and weighted rank methods were 0.0020, 0.0051, and 0.0138, respectively. None of these p values met the threshold for genome-wide significance using the Bonferroni correction. Indeed, among 115 tests based on the ~30 cM bins, 2 were significant at the 0.01 level and 3 at the 0.05 level for the truncated p value method; 1 and 6 for the analyses based on LOD scores, and 0 and 9 for those based on weighted ranks. 1.15 and 5.75 such results would be expected at the 0.01 and 0.05 levels in the absence of linkage information. Similarly, the minimum p value observed

for 2 cM bins for the truncated p value based meta-analysis was 0.0017, while those for the LOD score and weighted rank based analyses were 0.0011 and 0.0020. Among 1,758 tests based on the 2 cM bins, 18 were significant at the 0.01 level and 94 at the 0.05 level for the truncated p value method; these numbers were 18 and 83 for the analyses based on LOD scores and 8 and 90 for those based on weighted ranks. 17.58 and 87.90 such results would be expected at the 0.01 and 0.05 levels even in the absence of linkage information. Again, no excess of small p values was observed.

The GSMA method proposed by Wise et al. [29] provides an approach to combine the results of multiple linkage studies to identify susceptibility loci for the disease of interest. It suggests the use of ranks to unify various measures of linkage results and the use of bins (~30 cM) to obtain independence of test statistics. Our simulations have shown that GSMA results in appropriate type I error rates under the conditions of homogeneous marker sets and constant marker density (simulation 3), but results in modestly inflated type I error rates on average and some variability in these rates by location under more realistic settings (simulations 1 and 2). The alternative scoring schemes of maximum LOD scores and truncated p values provided similar but slightly better type I error control. The truncated p value method not only gives the best average type I error rate compared to the nominal value, but also the smallest standard deviations among the methods using the three scores (table 6).

Ranks allow the advantage of easy conversion of different types of linkage scores (such as LOD scores, mean IBD sharing, or p values) into a uniform scale, but do not directly reflect the strength of the corresponding linkage signals. Even under the null hypothesis of no linkage anywhere in the genome, high ranks still must be assigned and often will be assigned to regions with little evidence for linkage. Since most linkage studies report statistics that are of (approximately) known asymptotic distribution, conversion to a common scale such as LOD scores or p values should usually be possible, as was the case for the 23 samples in the Consortium data.

Analysis based on LOD scores preserves the magnitude of the linkage results from each study. Analysis based on truncated p values emphasizes the gap between 'significant' (smaller than a cut-off value) and 'insignificant' (greater than a cut-off value) results by replacing the latter by the extreme value of 1. Both methods make use of more information than the ranks. In addition, since the bins falling into the same linkage peak are obviously correlated, the truncated p value based analysis narrows

the width of the peak by truncating 'insignificant' results into the baseline and hence reduces the extent of correlation among the test statistics. This may explain the better control of the type I error rates using this method. Forabosco et al. [53] showed that only using strongly significant results, such as LOD score >1, 2, or 3, will result in loss of power in GSMA. In our analyses, we use the more modest threshold value of 0.05 for the p values, which corresponds to a LOD score of 0.59. We also conducted the meta-analyses using a cut-off p value of 0.10 (results not shown), and the results are similar to those presented here with no additional interesting regions identified. Our analyses results from using the LOD scores (and so no p value truncation) and truncated p values are also highly correlated, with Pearson correlation of 0.88 between the logarithm of the p values.

We also carried out meta-analyses and simulations using ranks without weighting by sample size and minimum p values without truncation as measures of linkage signals (results not presented). Unweighted ranks are expected to perform less well than the other methods since they ignore differences in the amount of information contained in individual studies. (Non-truncated) minimum p values produced results highly correlated with those using the maximum LOD scores, with the Pearson correlations 0.96 and 0.92 in ~30 cM and 2 cM bin methods respectively, and identified three out of the four best regions; the region ~76 cM on chromosome 4 was not identified at 0.05 significance level by this method.

Besides considering alternative scoring schemes, we also considered narrower bin widths. Although ~30 cM bins can contain some linkage peaks and potentially reduce the correlation between adjacent bins, it is still difficult to develop a good binning approach to keep all linkage peaks from being split into two or more bins. Further, wide bins decrease the resolution of the linkage results, and may attenuate the distinction between weaker and stronger linkage signals. We therefore decided to use 2 cM bins as an approximation of continuous evaluation of the linkage statistic. An advantage of the narrow bins is that most of the bins contain zero or one markers in the linkage studies we collected. The effect of marker densities is likely minimized, consistent with the better control of type I errors using narrower bins in simulations 1 and 2.

Since LOD scores were not calculated at every 2 cM in every study, we interpolated the LOD scores linearly (see 'Materials and Methods'). To evaluate the impact of interpolation, we calculated the LOD scores at every 10 cM instead, using the same simulated pedigrees and geno-

types as those in our simulations, and interpolated the scores at every 2 cM. The interpolated scores are very close to the actual values and should not have changed the power of GSMA noticeably.

Because the test statistics (ranks, LOD scores, or truncated p values) are permuted within study to assess the significances, splitting a linkage peak into multiple bins, which we shall call correlated bins, causes dependence of bins and may bias the estimated p values from permutation. Using 2 cM bins obviously does increase the number of correlated bins. However, and more important, it does not change the proportion of correlated bins compared to that using ~30 cM bins. For example, we simulated 100 chromosomes of 150 cM under the null hypothesis, and counted the number of bins belonging to the same linkage peak; here a peak is defined as the contiguous bins with LOD >0. The proportion of correlated bins was 0.429 for 2 cM bins, and 0.508 for 30 cM bins. For the truncated p values with cut-off at 0.05, that is, LOD >0.59, the proportions dropped to 0.043 and 0.058 for 2 cM and 30 cM bins, respectively.

A more important issue with 2 cM bins is the problem of multiple testing. With more bins, a substantially more stringent threshold for p values will be required to declare genome-wide significance of any bin. On the other hand, the 2 cM bins do provide higher resolution for the interesting regions, which can be more easily compared with other linkage studies and genome-wide association studies. The 2 cM bin-based methods also control the type I error rates better than the ~30 cM-based methods, especially when there is substantial variability in marker density. Therefore, we suggest an initial screening using ~30 cM bins followed by using 2 cM bins to localize linkage signals and control for false positives.

To obtain more accurate evaluation of significance, we may need to know the multi-dimensional distributions of the test statistics under the same settings of our meta-analyses of the Consortium data. Ideally, if we could use the complete pedigree information in simulation 1 without the pedigree reduction described in 'Materials and Methods', it would provide an estimate for the desired distribution, given that the number of simulation replicates is sufficiently large. A similar approach was suggested by Wise [54] to combine candidate region studies with genome-wide studies using the GSMA method. If we are willing to ignore the differences between the Consortium data and the reduced data for simulation, we can estimate the p value of each bin as the proportion of simulation replicates in which the score of the bin in simulation was greater than that in our meta-

analyses. The same regions were then identified on chromosome 10 and 14 using either ~30 cM bins or 2 cM bins, and on chromosome 4 using 2 cM bins at the 0.01 significance level, with p values similar to those obtained by permutations.

When data are available, a joint linkage analysis may also be performed by pooling pedigree information across the same 23 individual studies. Its results should be similar to ours from the meta-analyses described here, but some differences can also be expected. The meta-analysis considers only positive linkage signals, while in the joint analysis, deficiencies of sharing in some studies may cancel excess sharing in other studies. In the meta-analysis, individual studies can apply various methods of analysis, including parametric/non-parametric and variance component linkage analyses, and might consider auxiliary phenotype data. Although this flexibility is appreciated when original genotype data cannot be accessed, it may add heterogeneity among the individual studies' results to be summarized by the meta-analysis. On the other hand, while the joint analyses can make a better use of information, they are sometimes limited by computational powers needed for the large sample size of the pooled sample.

In summary, our meta-analyses of the International Type 2 Diabetes Linkage Analysis Consortium data suggested evidence for linkage on regions of chromosomes 4, 10, 14, and 16 for type 2 diabetes, with no signal reaching genome-wide significance, but those on chromosomes 10 and 14 reaching a level expected to occur by chance once per genome. Subgroup analyses are consistent with the possibility that genetic heterogeneity of the collected samples may be a cause for the modest linkage signals. Computer simulations showed that variations of the marker density within and between studies could result in modestly inflated or deflated type I errors in current analyses. Smaller bins (2 cM) and alternative test statistics more directly based on the linkage evidence, such as LOD scores or truncated p values, may help to draw the type I errors towards the nominal value, but more sophisticated approaches need to be considered to correct this problem fully.

International Type 2 Diabetes Linkage Analysis Consortium

AADM:
Charles N. Rotimi, the National Human Genome Center, Department of Community and Family medicine, Howard University, Washington, DC, USA

Adebowale Adeyemo, the National Human Genome Center, Department of Community and Family medicine, Howard University, Washington, DC, USA

Guanjie Chen, the National Human Genome Center, Department of Community and Family medicine, Howard University, Washington, DC, USA

Yuanxiu Chen, the National Human Genome Center, Department of Community and Family medicine, Howard University, Washington, DC, USA

Amish Family Diabetes Study:
Toni I. Pollin, University of Maryland School of Medicine, Baltimore, MD, USA

Wen-Chi Hsueh, School of Medicine, University of California, San Francisco, CA, USA

Braxton D. Mitchell, University of Maryland School of Medicine, Baltimore, MD, USA

Alan R. Shuldiner, University of Maryland School of Medicine, Baltimore, MD, USA

FUSION:
Michael Boehnke, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

Weihua Guan, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

Francis S. Collins, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

Jaakko Tuomilheto, National Public Health Institute, Helsinki, Finland

GlaxoSmithKline (for GENNID):
Margaret G. Ehm, GlaxoSmithKline, Pharmacogenetics, Research Triangle Park, NC, USA

Pamela L. St. Jean, GlaxoSmithKline, Pharmacogenetics, Research Triangle Park, NC, USA

Michael J. Wagner, Institute for Pharmacogenomics and Individualized Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Daniel K. Burns, GlaxoSmithKline, Pharmacogenetics, Research Triangle Park, NC, USA

NIDDK Phoenix:
Robert L. Hanson, Phoenix Epidemiology and Clinical Research Branch, National Institute OF Diabetes and Digestive and Kidney Diseases, Phoenix, AZ, USA

Clifton Bogardus, Phoenix Epidemiology and Clinical Research Branch, National Institute OF Diabetes and Digestive and Kidney Diseases, Phoenix, AZ, USA

William C. Knowler, Phoenix Epidemiology and Clinical Research Branch, National Institute OF Diabetes and Digestive and Kidney Diseases, Phoenix, AZ, USA

Pasteur Institute Lille:
Christian Dina, UMR8090, Genomics and Molecular Physiology of Metabolic diseases, Centre National de la Recherche Scientifique, Lille, France
Cécile Lecoeur, UMR8090, Genomics and Molecular Physiology of Metabolic diseases, Centre National de la Recherche Scientifique, Lille, France
Philippe Froguel, UMR8090, Genomics and Molecular Physiology of Metabolic diseases, Centre National de la Recherche Scientifique, Lille, France

Southwest Foundation for Biomedical Research (SFBR):
Jean W. MacCluer, Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX, USA
John Blangero, Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX, USA
Braxton D. Mitchell, Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX, USA
Shelley A. Cole, Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX, USA

United Kingdom:
Mark I. McCarthy, Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford UK
Andrew T. Hattersley, Institute of Biomedical and Clinical Science, Peninsula Medical School, Exeter, UK
Tim M. Frayling, Institute of Biomedical and Clinical Science, Peninsula Medical School, Exeter, UK
Steven Wiltshire, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

University of Arkansas:
Steven C. Elbein, Department of Medicine, University of Arkansas for Medical Sciences, College of Medicine and Central Arkansas Veterans Healthcare System, Little Rock, AR, USA
Sandy Hasstedt, Department of Genetics, University of Utah School of Medicine, Little Rock, AR, USA

University of Chicago:
Nancy J. Cox, Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA
Anna Pluzhnikov, Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA

University of Lund and Helsinki:
Cecilia Lindgren, Clinical Research Centre, Karolinska University Hospital, Huddinge, Sweden; Diabetes Research Laboratories, Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, University of Oxford, Headington, Oxford, UK
Peter Almgren, Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmö, Sweden
Tiinamaija Tuomi, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland
Leif Groop, Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmö, Sweden

U Texas Health Science Center at San Antonio:
Michael P. Stern, Division of Clinical Epidemiology, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA
Ravindranath Duggirala, Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX, USA
Rector Arya, Division of Clinical Epidemiology, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

Wake Forest University:
Donald Bowden, Wake Forest University School of Medicine, Winston-Salem NC, USA
Barry Freedman, Wake Forest University School of Medicine, Winston-Salem NC, USA

Washington University:
M. Alan Permutt, Washington University School of Medicine, St. Louis, MO, USA
Brian Suarez, Washington University School of Medicine, St. Louis, MO, USA
Jon Wasson, Washington University School of Medicine, St. Louis, MO, USA
Benjamin Glaser, the Hadassah Hebrew University Medical Center, Israel

References

1 Bennett PH, Bogardus C, Tuomilehto J, Zimmett P: Epidemiology and natural history of NIDDM; in Alberti KGMM, DeFronzo RA, Keen H, Zimmett P (eds): Nonobese and Obese. New York, Wiley, 1992, pp 147–176.

2 Hanis CL, Boerwinkle E, Chakraborty R, Ellsworth DL, Concannon P, Stirling B, Morrison VA, Wapelhorst B, Spielman RS, Gogolin-Ewens KJ, Shepard JM, Williams SR, Risch N, Hinds D, Iwasaki N, Ogata M, Omori Y, Petzold C, Rietzch H, Schroder HE, Schulze J, Cox NJ, Menzel S, Boriraj VV, Chen X, Lim LR, Lindner T, Mereu LE, Wang YQ, Xiang K, Yamagata K, Yang Y, Bell GI: A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. Nat Genet 1996;13:161–166.

3 Mahtani MM, Widen E, Lehto M, Thomas J, McCarthy M, Brayer J, Bryant B, Chan G, Daly M, Forsblom C, Kanninen T, Kirby A, Kruglyak L, Munnelly K, Parkkonen M, Reeve-Daly MP, Weaver A, Brettin T, Duyk G, Lander ES, Groop LC: Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. Nat Genet 1996;14:90–94.

4 Zouali H, Hani E, Philippi A, Vionnet N, Beckmann J, Demenais F, Froguel P: A susceptibility locus for early-onset non-insulin dependent (type 2) diabetes mellitus maps to chromosome 20q, proximal to the phosphoenolpyruvate carboxykinase gene. Hum Mol Genet 1997;6:1401–1408.

5 Hanson RL, Ehm MG, Pettitt DJ, Prochazka M, Thompson DB, Timberlake D, Foroud T, Kobes S, Baier L, Burns DK, Almasy L, Blangero J, Garvey WT, Bennett PH, Knowler WC: An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. Am J Hum Genet 1998;63:1130–1138.

6 Pratley RE, Thompson DB, Prochazka M, Baier L, Mott D, Ravussin E, Sakul H, Ehm MG, Burns DK, Foroud T, Garvey WT, Hanson RL, Knowler WC, Bennett PH, Bogardus C: An autosomal genomic scan for loci linked to prediabetic phenotypes in Pima Indians. J Clin Invest 1998;101:1757–1764.

7 Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP: Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. Am J Hum Genet 1999;64:1127–1140.

8 Elbein SC, Hoffman MD, Teng K, Leppert MF, Hasstedt SJ: A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians. Diabetes 1999;48:1175–1182.

9 Hegele RA, Sun F, Harris SB, Anderson C, Hanley AJ, Zinman B: Genome-wide scanning for type 2 diabetes susceptibility in Canadian Oji-Cree, using 190 microsatellite markers. J Hum Genet 1999;44:10–14.

10 Ehm MG, Karnoub MC, Sakul H, Gottschalk K, Holt DC, Weber JL, Vaske D, Briley D, Briley L, Kopf J, McMillen P, Nguyen Q, Reisman M, Lai EH, Joslyn G, Shepherd NS, Bell C, Wagner MJ, Burns DK: Genomewide search for type 2 diabetes susceptibility genes in four Americans populations. Am J Hum Genet 2000;66:1871–1881.

11 Ghosh S, Watanabe RM, Valle TT, Hauser ER, Magnuson VL, Langefeld CD, Ally DS, Mohlke KL, Silander K, Kohtamäki K, Chines P, Balow J, Birznieks G, Chang J, Eldridge W, Erdos MR, Karanjawala ZE, Knapp JI, Kudelko K, Martin C, Morales-Mena A, Musick A, Musick T, Pfahl C, Porter R, Rayman JB, Rha D, Segal L, Shapiro S, Sharaf R, Shurtleff B, So A, Tannenbaum J, Te C, Tover J, Unni A, Welch C, Whiten R, Witt A, Blaschak-Harvan J, Douglas JA, Duren WL, Epstein MP, Fingerlin TE, Kaleta HS, Lange EM, Li C, McEachin RC, Stringham HM, Trager E, White PP, Eriksson J, Toivanen L, Vidgren G, Nylund SJ, Tuomilehto-Wolf E, Ross EH, Demirchyan E, Hagopian WA, Buchanan TA, Tuomilehto J, Bergman RN, Collins FS, and Boehnke M: The Finland-United States Investigation of Non-Insulin-Dependent Diabetes Mellitus Genetics (FUSION) Study. I. An autosomal genome scan for genes that predispose to Type 2 diabetes. Am J Hum Genet 2000;67:1174–1185.

12 Klupa T, Malecki MT, Pezzolesi M, Ji L, Curtis S, Langefeld CD, Rich SS, Warren JH, Krolewski AS: Further evidence for a susceptibility locus for type 2 diabetes on chromosome 20q131-q132. Diabetes 2000;49:2212–2216.

13 Vionnet N, Hani El H, Dupont S, Gallina S, Francke S, Dotte S, De Matos F, Durand E, Lepretre F, Lecoeur C, Gallina P, Zekiri L, Dina C, Froguel P: Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chro-

mosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21-q24. Am J Hum Genet 2000;67:1470–1480.

14 Luo TH, Zhao Y, Li G, Yuan WT, Zhao JJ, Chen JL, Huang W, Luo M: A genome-wide search for type II diabetes susceptibility genes in Chinese Hans. Diabetologia 2001;44:501–506.

15 Parker A, Meyer J, Lewitzky S, Rennich JS, Chan G, Thomas JD, Orho-Melander M, Lehtovirta M, Forsblom C, Hyrkko A, Carlsson M, Lindgren C, Groop LC: A gene conferring susceptibility to type 2 diabetes in conjunction with obesity is located on chromosome 18p11. Diabetes 2001;50:675–680.

16 Permutt MA, Wasson JC, Suarez BK, Lin J, Thomas J, Meyer J, Lewitzky S, Rennich JS, Parker A, DuPrat L, Maruti S, Chayen S, Glaser B: A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. Diabetes 2001;50:681–685.

17 Wiltshire S, Hattersley AT, Hitman GA, Walker M, Levy JC, Sampson M, O'Rahilly S, Frayling TM, Bell JI, Lathrop GM, Bennett A, Dhillon R, Fletcher C, Groves CJ, Jones E, Prestwich P, Simecek N, Subba-Rao PV, Wishart M, Foxon R, Howell S, Smedley D, Cardon LR, Menzel S, McCarthy MI: A genome-wide scan for loci predisposing to type 2 diabetes in a UK population (the Diabetes UK Warren 2 Repository): Analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. Am J Hum Genet 2001;69:553–569.

18 Busfield F, Duffy DL, Kesting JB, Walker SM, Lovelock PK, Good D, Tate H, Watego D, Marczak M, Hayman N, Shaw JT: A genomewide search for type 2 diabetes-susceptibility genes in indigenous Australians. Am J Hum Genet 2002;70:349–357.

19 Lindgren C, Mahtani M, Widén E, McCarthy M, Kirby A, Daly M, Lehto M, Thomas J, Glucksman S, Brayer J, Kanninen T, Almgren P, Tuomi T, Groop LC& Lander ES: A genome-wide search for susceptibility loci linked to type 2 diabetes in Finnish families (the Botnia Study). J Hum Mol Genetics 2002;70:509–516.

20 Mori Y, Otabe S, Dina C, Yasuda K, Populaire C, Lecoeur C, Vatin V, Durand E, Hara K, Okada T, Tobe K, Boutin P, Kadowaki T, Froguel P: Genome wide search for type 2 diabetes in Japanese affected sib-pairs confirms susceptibility genes on 3q, 15q and 20q and identifies new candidate loci on 7p and 11p. Diabetes 2002;51:1247–1255.

21 Hsueh W-C, St. Jean PL, Mitchell BD, Pollin TI, Knowler WC, Ehm MG, Bell CJ, Sakul H, Wagner MJ, Burns DK, Shuldiner AR: Genome-wide and fine-mapping linkage studies of type 2 diabetes and glucose traits in the Old Order Amish: Evidence for a new diabetes locus on chromosome 14q11 and confirmation of a locus on chromosome 1q21-q24. Diabetes 2003;52:550–557.

22 Iwasaki N, Cox NJ, Wang YQ, Schwarz PE, Bell GI, Honda M, Imura M, Ogata M, Saito M, Kamatani N, Iwamoto Y: Mapping genes influencing type 2 diabetes risk and BMI in Japanese subjects. Diabetes 2003:52:209–213.

23 Ng MC, So WY, Cox NJ, Lam VK, Cockram CS, Critchley JA, Bell GI, Chan JC: Genome-wide scan for type 2 diabetes loci in Hong Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25. Diabetes 2004;53:1609–1613.

24 Sale MM, Freedman BI, Langefeld CD, Williams AH, Hicks PJ, Colicigno CJ, Beck SR, Brown WM, Rich SS, Bowden DW: A genome-wide scan for type 2 diabetes in African-American families reveals evidence for a locus on chromosome 6q. Diabetes 2004;53:830–837.

25 Silander K, Scott LJ, Valle TT, Mohlke KL, Stringham HM, Wiles KR, Duren WL, Doheny KF, Pugh EW, Chines P, Narisu N, White PP, Fingerlin TE, Jackson AU, Li C, Ghosh S, Magnuson VL, Colby K, Erdos MR, Hill JE, Hollstein P, Humphreys KM, Kasad R, Lambert J, Lazaridis KN, Lin G, Morales-Mena A, Patzkowski K, Pfahl C, Porter R, Rha D, Segal L, Suh YD, Tovar J, Unni A, Welch C, Douglas JA, Epstein MP, Hauser ER, Hagopian W, Buchanan TA, Watanabe RM, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A large set of Finnish affected sibling pair families with type 2 diabetes suggests susceptibility loci on chromosomes 6, 11, and 14. Diabetes 2004;53:821–829.

26 Rotimi CN, Chen G, Adeyemo AA, Furbert-Harris P, Guass D, Zhou J, Berg K, Adegoke O, Amoah A, Owusu S, Acheampong J, Agyenim-Boateng K, Eghan BA Jr, Oli J, Okafor G, Ofoegbu E, Osotimehin B, Abbiyesuku F, Johnson T, Rufus T, Fasanmade O, Kittles R, Daniel H, Chen Y, Dunston G, Collins FS: A genome-wide search for type 2 diabetes susceptibility genes in West Africans: the Africa America Diabetes Mellitus (AADM) Study. Diabetes 2004;53:838–841.

27 Xiang K, Wang Y, Zheng T, Jia W, Li J, Chen L, Shen K, Wu S, Lin X, Zhang G, Wang C, Wang S, Lu H, Fang Q, Shi Y, Zhang R, Xu J, Weng Q: Genome-wide search for type 2 diabetes/impaired glucose homeostasis susceptibility genes in the Chinese: Significant linkage to chromosome 6q21-q23 and chromosome 1q21-q24. Diabetes 2004;53(1):228–234.

28 Zhao JY, Xiong MM, Huang W, Wang H, Zuo J, Wu GD, Chen Z, Qiang BQ, Zhang ML, Chen JL, Ding W, Yuan WT, Xu HY, Jin L, Li YX, Sun Q, Liu QY, Boerwinkle E, Fang FD: An autosomal genomic scan for loci linked to type 2 diabetes in northern Han Chinese. J Mol Med 2005:209:209–215.

29 Wise LH, Lanchbury JS, Lewis CM: Meta-analysis of genome searches. Ann Hum Genet 1999;63:263–272.

30 Fisher RA: Statistical Methods for Research Workers, ed 4. Edinburgh, Oliver and Boyd, 1932.

31 Zaykin DV, Zhivotovsky LA, Westfall PH, Weir BS: Truncated product method for combining p values. Genet Epidemiol 2002; 22:170–185.

32 Huber P: Robust statistical procedures. SIAM Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, 1977.

33 Kong A, Cox NJ: Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 1997;61:1179–1188.

34 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 1996;58:1347–1363.

35 Amos CI: Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 1994;54:535–543.

36 Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K: A high resolution recombination map of the human genome. Nat Genet 2002;31:241–247.

37 Broman KW, Murray JC, Sheffield VC, White RL, Weber JL: Comprehensive human genetic maps: Individual and sex-specific variation in recombination. Am J Hum Genet 1998;63:861–869.

38 Lander ES, et al: Initial sequencing and analysis of the human genome. Nature 2001:409: 860–921.

39 Self SG, Liang K-Y: Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 1987;82:605–610.

40 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002;30:97–101.

41 Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, Styrkarsdottir U, Magnusson KP, Walters GB, Palsdottir E, Jonsdottir T, Gudmundsdottir T, Gylfason A, Saemundsdottir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdottir U, Gulcher JR, Kong A, Stefansson K: Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 2006;38:320–323.

42 Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI: Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet 2003;26:163–175.

43 Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES: The common PPARg Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet 2000;26: 76–80.

44 Bowden DW, Sale M, Howard TD, Qadri A, Spray BJ, Rothschild CB, Akots G, Rich SS, Freedman BI: Linkage of genetic markers on human chromosomes 20 and 12 to NIDDM in Caucasian sib pairs with a history of diabetic nephropathy. Diabetes 1997;46:882–886.

45 Ji L, Malecki M, Warram JH, Yang Y, Rich SS, Krolewski AS: New susceptibility locus for NIDDM is localized to human chromosome 20q. Diabetes 1997;46:876–881.

46 Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881–885.

47 Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants. Science 2007;316:1341–1345.

48 Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S: Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. Science 2007; 316:1331–1336.

49 Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ,

Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT: Replication of Genome-Wide Association Signals in U.K. Samples Reveals Risk Loci for Type 2 Diabetes. Science 2007;316: 1336–1341.

50 Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, Baker A, Snorradottir S, Bjarnason H, Ng MC, Hansen T, Bagger Y, Wilensky RL, Reilly MP, Adeyemo A, Chen Y, Zhou J, Gudnason V, Chen G, Huang H, Lashley K, Doumatey A, So WY, Ma RC, Andersen G, Borch-Johnsen K, Jorgensen T, van Vliet-Ostaptchouk JV, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Rotimi C, Gurney M, Chan JC, Pedersen O, Sigurdsson G, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K: A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet 2007;39:770–775.

51 Salonen JT, Uimari P, Aalto JM, Pirskanen M, Kaikkonen J, Todorova B, Hypponen J, Korhonen VP, Asikainen J, Devine C, Tuomainen TP, Luedemann J, Nauck M, Kerner W, Stephens RH, New JP, Ollier WE, Gibson JM, Payton A, Horan MA, Pendleton N, Mahoney W, Meyre D, Delplanque J, Froguel P, Luzzatto O, Yakir B, Darvasi A: Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. Am J Hum Genet 2007;81:338–345.

52 Demenais F, Kanninen T, Lindgren CM, Wiltshire S, Gaget S, Dandrieux C, Almgren P, Sjögren M, Hattersley A, Dina C, Tuomi T, McCarthy MI, Froguel P, Groop LC: A meta-analysis of four European genome screens (GIFT Consortium) shows evidence for a novel region on chromosome 17p11.2-q22 linked to type 2 diabetes. Hum Mol Genet 2003;12:1865–1873.

53 Forabosco P, Ng MY, Bouzigon E, Fisher SA, Levinson DF, Lewis CM: Data acquisition for meta-analysis of genome-wide linkage studies using the genome search meta-analysis method. Hum Hered 2007;64:74–81.

54 Wise LH: Inclusion of candidate region studies in meta-analysis using the genome screen meta-analysis method: application to asthma data. Genet Epidemiol 2001;21:S160–S165.

55 Comuzzie AG, Hixson JE, Almasy L, Mitchell BD, Mahaney MC, Dyer TD, Stern MP, MacCluer JW, Blangero J: A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. Nat Genet 1997;15:273–275.

56 Raffel L, Robbins D, Norris J, Boerwinkle E, DeFronzo R, Elbein S, Fujimoto W, Hanis CL, Kahn SE, Permutt MA, Chiu KC, Cruz J, Ehrmann DA, Robertson RP, Rotter JI, Buse J: The GENNID Study: a resource for mapping the genes that cause NIDDM. Diabetes Care 1996;19:864–872.