ORIGINAL ARTICLE

Whole genomewide linkage screen for neural tube defects reveals regions of interest on chromosomes 7 and 10

E Rampersaud, A G Bassuk, D S Enterline, T M George, D G Siegel, E C Melvin, J Aben, J Allen, A Aylsworth, T Brei, J Bodurtha, C Buran, L E Floyd, P Hammock, B Iskandar, J Ito, JA Kessler, N Lasarsky, P Mack, J Mackey, D McLone, E Meeropol, L Mehltretter, L E Mitchell, W J Oakes, J S Nye, C Powell, K Sawin, R Stevenson, M Walker, S G West, G Worley, J R Gilbert, M C Speer

....

See end of article for authors' affiliations
....

Correspondence to:
Dr M C Speer, Duke University Medical Center, Box 3445, Durham, NC 27710, USA; marcy@chg.duhs.duke.edu

Neural tube defects (NTDs) are the second most common birth defects (1 in 1000 live births) in the world. Periconceptional maternal folate supplementation reduces NTD risk by 50–70%; however, studies of folate related and other developmental genes in humans have failed to definitively identify a major causal gene for NTD. The aetiology of NTDs remains unknown and both genetic and environmental factors are implicated. We present findings from a microsatellite based screen of 44 multiplex pedigrees ascertained through the NTD Collaborative Group. For the linkage analysis, we defined our phenotype narrowly by considering individuals with a lumbosacral level myelomeningocele as affected, then we expanded the phenotype to include all types of NTDs. Two point parametric analyses were performed using VITESSE and HOMOG. Multipoint parametric and nonparametric analyses were performed using ALLEGRO. Initial results identified chromosomes 7 and 10, both with maximum parametric multipoint lod scores (Mlod) >2.0. Chromosome 7 produced the highest score in the 24 cM interval between D7S3056 and D7S3051 (parametric Mlod 2.45; nonparametric Mlod 1.89). Further investigation demonstrated that results on chromosome 7 were being primarily driven by a single large pedigree (parametric Mlod 2.40). When this family was removed from analysis, chromosome 10 was the most interesting region, with a peak Mlod of 2.25 at D10S1731. Based on mouse human synteny, two candidate genes (*Meox2*, *Twist1*) were identified on chromosome 7. A review of public databases revealed three biologically plausible candidates (*FGFR2*, *GFRA1*, *Pax2*) on chromosome 10. The results from this screen provide valuable positional data for prioritisation of candidate gene assessment in future studies of NTDs.

T he second most common severely disabling birth defects (1 in 1000 live births) in the world are neural tube defects (NTDs), with the highest reported incidence rates being in northern China (3.7 per 1000 live births) and in Ireland (1.0 per 1000 live births). In the USA, incidence increases from the west to the east coast, with the highest rates seen in the Appalachian region.

NTDs result from a failure of neurulation, which occurs around the 28th day after conception, at a time when most women do not know they are pregnant. There are three principal forms: anencephaly, encephalocele, and spina bifida cystica (open spina bifida). Anencephaly is lethal in all cases. Patients with encephalocele may survive but are mentally retarded. Thus, the vast majority of patients seen have spina bifida cystica, which manifests as an open spinal lesion containing spinal tissue, resulting in abnormal innervation beneath the level of lesion, varying degrees of muscle weakness and sensory impairment, and a neurogenic bladder and bowel.

Neural tube defects are caused by a complex interaction between genes and the environment. Several lines of evidence suggest a genetic component to NTDs. Firstly, the estimated recurrence risk in siblings is 2–5%, giving a λ_s value in NTD families of 25–50, representing up to a 50 fold increased risk over that observed in the general population.[1] This risk is increased to 4% for offspring of a person with an NTD. Khoury *et al* have shown that for a recurrence risk to be this high, an environmental teratogen would have to increase the risk at least 100 fold to exhibit the same degree of familial aggregation, indicating that a genetic component is required.[2] Additionally, estimates from small twin studies

indicate a higher concordance rate in monozygotic twins of 7.7% compared to 4.0% for dizygotic twins.[3 4] NTDs are also commonly associated with other known genetic disorders including trisomy 13 and 18, Meckel-Gruber syndrome, and chromosomal rearrangements.[5 6]

In familial cases, NTDs tend to breed true within families; in other words, recurrences in families in which the proband is affected with spina bifida tend to be also spina bifida, and recurrences in families in which the case is anencephalic tend to be also anencephalic.[7–11] However, 30–40% of recurrences involve an NTD phenotype that is different from the proband phenotype. This intrafamily heterogeneity may represent the pleiotropic effect of a common underlying gene, or may suggest that in families with different phenotypic presentations, various forms of NTDs result from different underlying genes.

The most substantial environmental risk factor for NTDs is insufficient periconceptional maternal folate consumption. Adequate folate supplementation reduces NTD recurrence risk by 50–70%,[12–14] yet the recurrence risk is not entirely eliminated,[15] suggesting that additional genetic factors are responsible for the development of NTDs. To date, most genetic studies in humans have focused on evaluating folate related candidate genes, genes in early developmental pathways, and genes from mouse models (reviewed in Juriloff

....

Abbreviations: AD, Alzheimer's disease; GDA, Genetic Data Analysis program; HD, Hirschsprung's disease; Hetlod, heterogeneity lod score; HWE, Hardy-Weinberg equilibrium; Mlod, multipoint lod score; NTD, neural tube defect; QC, quality control; SNP, single nucleotide polymorphism

and Harris[16] and Copp *et al*[17]). Despite extensive efforts, the assessment of candidate genes in NTDs has yet to identify a major causative gene. An alternative approach of using linkage analysis to identify positional candidates in multiplex NTD pedigrees is hampered by lethality in anencephaly cases, the increased mortality associated with spina bifida, and termination of prenatally detected affected pregnancies.[18][19] As a result, a full genomic screen in neural tube defects has not been previously reported.

In the present article, we describe the results of the first genomewide screen of 44 multiplex NTD families collected from a national collaborative ascertainment effort. The data presented represent valuable positional information to assist prioritisation of candidate gene assessment in future studies of neural tube defects.

MATERIALS AND METHODS
Clinical data collection
Probands were identified from a variety of sources including myelodysplasia clinics, annual meetings of the Spina Bifida Association of America (SBAA), and the worldwide web. Ascertainment was carried out by researchers as part of the NTD Collaborative Group (details at end of paper). All individuals identified in this manner were included in the study, and additional information was gathered to confirm their diagnoses. First degree relatives and relatives connecting related affected individuals in extended families were also ascertained. Detailed family histories were obtained and medical records, including operative reports and presurgical *x* ray films, were collected for review of diagnosis by neurosurgeons. Study staff trained in phlebotomy obtained blood samples from affected individuals and related family members at medical centres, clinics, or by visiting participants' homes. In some cases, participants were sent kits by post. The study was overseen by the Duke University Medical Center institutional review board and informed consent was obtained from all participants.

Genotyping methods
DNA was extracted from whole blood using the Puregene system (Gentra Systems, Minneapolis, MN, USA). The DNA samples from study subjects were organised into lists with a standardised order of samples for which the technician was blinded to sex and family composition, with quality control (QC) standards incorporated at specified slots in the list. DNA samples were aliquoted into 96 well plates and genotyping performed using the fast automated angle scan technique method.[20] Multiplex reactions using two colour fluorescence were used to increase the efficiency, lower the cost, and significantly increase the speed of the genomic screen.[21]

To minimise systematic errors due to sample switches, gel loading or running problems, or reading errors in the genotyping procedures, QC samples were added to each gel analysed in the laboratory that by virtue of their positioning ensured they would cover a majority of possible technical errors. In addition to genotyping two control individuals from the Centre d'Étude du Polymorphisme Humain, who were common to all gels, QC samples representing randomly selected duplicated individuals from the dataset were included to allow both within and between gel comparisons; six QC samples were selected per 84 samples analysed. The laboratory technician was blinded to the identity of the QC samples. Data for marker genotypes were managed using the PEDIGENE system[22] in preparation for analysis. Before merging genotypic data into PEDIGENE, agreement between the QC genotype and the corresponding sample genotype was assessed. With these procedures in place, 353 of 402 markers attempted (88%) were approved for analysis. These 353 markers had all QC inconsistencies resolved and affected

genotypes were re-read. The mean genotyping efficiency over all 353 markers was 96.9%.

Error checking
Mendelian pedigree inconsistencies were identified using PEDCHECK[23] and checked by laboratory technicians who were blinded to the pedigree structure. Further verification of interfamilial and intrafamilial genetic relationships was performed using RELPAIR;[24][25] at the beginning of the study using the first 50 genotyped markers, and then later using all 353 genotyped markers.

Linkage analyses
Because of the genetic complexity of NTDs, we used a multifaceted analytic strategy and performed both parametric and nonparametric linkage analysis. Single point parametric linkage analysis was conducted using VITESSE,[26] under dominant and recessive models for affected patients only, and allowing for a disease allele frequency of 0.001. These two point lod scores were used to calculate genetic heterogeneity lod scores (Hetlod scores) using HOMOG.[27]

Power analyses of the multiplex families in the screen were carried out using SIMLINK[28][29] allowing for a dominant model and a disease allele frequency of 0.001. When fully informative, the families were capable of generating an estimated combined mean lod score of 4.745 at $\theta = 0.05$ (SE = 0.05; maximum 10.04) under the broad analysis scheme, and a combined mean lod score of 3.7 at $\theta = 0.05$ (SE = 0.05; maximum 7.51) under the narrow analysis scheme. These results are driven primarily by the existence of a large family in which four affected individuals, mostly related as cousins, are available. When a heterogeneity model that allowed for 50% between pedigree heterogeneity was used, the families were capable of generating an estimated combined mean lod score of 1.34 at $\theta = 0.05$ (SE = 0.04; maximum 8.34) under the broad analysis scheme.

Multipoint parametric and nonparametric linkage analyses were performed using ALLEGRO.[30] Genetic marker distance was based on 10 cM sex average integrated maps from deCode Genetics[31] and the Marshfield Medical Research Foundation maps. Map order was verified using Map-O-Mat.[32] Marker allele frequencies were estimated from our dataset using all individuals.[33] As our sample was comprised of pedigrees of varying size, we assessed identity by descent sharing (lod*) between all pairs of affected individuals within a family using the S_{pairs} sharing statistic[34] and the exponential model[35] as implemented in ALLEGRO. The ALLEGRO program calculates a full likelihood using affected but unsampled individuals (see accompanying article), thus we included ''multiplex by history'' pedigrees (n = 8) in our analyses. These families had one sampled affected person and additional relatives with NTD who were unavailable for sampling. We defined our most interesting regions as those with maximum two point parametric Hetlod scores >2.0, or multipoint parametric Hetlod or nonparametric lod* >1.3. We refer to the maximum multipoint lod score, whether parametric or nonparametric, as the Mlod.

Our dataset showed intrafamily phenotypic heterogeneity, meaning that affected individuals within the same family sometimes had different types of NTD. This might suggest that the underlying causative genes for various forms of NTD are different. To account for possible genetic heterogeneity in our sample, we established two phenotype definitions for NTD. The narrow phenotypic definition classified as affected only those individuals presenting with the most common type of NTD—that is, spina bifida with the lesion located at the lumbosacral level (lumbosacral myelomeningocele). The phenotype was then expanded to include all families in

which two or more individuals had an NTD, regardless of phenotypic presentation (broad phenotype definition).

We excluded all cases of spina bifida occulta from our analyses. Additionally, in pedigrees with monozygotic twins, we excluded one twin from each pair in the analysis. Using this criterion, one monozygotic twin in pedigree 8836 was removed. The majority of families in our sample were white (n = 41), but two families were Hispanic and one had mixed with African American and white ethnicity. We performed the genome screen using all families, and then re-analysed the data using only the white families.

RESULTS

Following data cleaning, 44 families and 292 samples (table 1) were included in the genomic screen analysis, consisting of 21 sampled affected full sibling pairs, 12 sampled affected avuncular pairs, and 35 other affected relative pairs. In total, there were 89 sampled affected individuals, the majority of whom had lumbosacral myelomeningocele (n = 50). After exclusion of one monozygotic twin, 49 sampled individuals with lumbosacral myelomeningocele were included in the analysis. The remaining affected individuals had other forms of NTD including anencephaly, cervical myelomeningocele, craniorachischisis, encephalocele, lipomyelomeningocele, rachischisis, and thoracic myelomeningocele. Seventeen of these pedigrees were informative for the narrow phenotypic classification; meaning that at least two affected individuals had lumbosacral myelomeningocele. Three of these pedigrees were "multiplex by history" families.

Results from linkage analysis using all 44 families compared with the 41 white families only were similar, so we combined the results. Fig 1 shows the maximum two point lod scores for all 353 markers arranged in map order under narrow and broad phenotypic definitions. The best two point lod scores were on chromosome 7p22 with a score of 2.31 at D7S513 and at 11p15 with a score of 2.44 at D11S2362. Parametric and nonparametric multipoint linkage results (Mlod >1.3), under both phenotype definitions, are summarised in table 2. For simplification, only multipoint results for the dominant model are presented for the parametric analyses. Five regions of interest were identified from these analyses: chromosomes 7p22, 10q25.3, 11p15, 15q26, and 21q22.11. The best multipoint peak fell in the 24 cM interval spanned by D7S3056 and D7S3051, producing a parametric Mlod score of 2.45 (fig 2A) under the broad phenotype, and demonstrating consistent linkage evidence (Mlod >1.3) across all analyses. Chromosome 10 achieved a parametric Mlod score of 2.08 under the broad phenotype (D10S1731 at 134.96 cM) (fig 2B). The best linkage peak for chromosome 11 (Mlod 1.51) fell between D11S2362 and D11S1999 (fig 2C), using the broad phenotype and a nonparametric analytic approach. Chromosome 15 produced a parametric Mlod score of 1.46 under the narrow classification, between D15S127 and D15S652 (fig 2D). Chromosome 21 had a peak Mlod score of 1.44 at D21S1270 (fig. 2E) under the narrow phenotype.

Our strongest linkage evidence based on two point and multipoint analyses was on chromosome 7. However, as the multipoint peak fell around the most telomeric marker typed (D7S3056), we genotyped additional single nucleotide polymorphisms (SNPs) in the region to verify our findings. We selected "assay on demand" SNPs (Applied Biosystems) having a minimum allele frequency >0.30. We analysed 13 SNPs (table 3), four of which are located distal to the most telomeric marker D7S3056. Deviations from Hardy-Weinberg equilibrium (HWE) (p<0.06) were tested using the Genetic Data Analysis (GDA) program.[36] Two markers (HCV8511072, RS852423) were not in HWE, hence pairwise linkage disequilibrium (LD) was calculated for all markers using the composite LD measure as implemented in GDA, which does not assume HWE.

Because current implementations of multipoint linkage analysis are unable to account for intermarker LD, the least informative marker in a pair of markers in LD was excluded from the multipoint analysis. Thus, HCV8511072 and RS852423, which were in LD with RS1636166 (p = 0.002) and RS1470539 (p = 0.06), respectively, were eliminated from further analysis. With the addition of the SNPs to the original microsatellite marker data, we were able to confirm that our linkage peak mapped to marker D7S641 located 17.41 cM from the telomere (Mlod 2.49 under the broad phenotype; Mlod 2.09 under the narrow phenotype).

An examination of family specific parametric scores identified a single pedigree (family 8776) primarily contributing to the linkage results on chromosome 7. Under the broad phenotype, family 8776 produced an Mlod score of 2.40 at marker D7S513 located 22.6 cMs from pter. Under the narrow phenotype, which in this family led to the exclusion of an individual affected with a fatty filum (100), family 8776

Table 1 Description of samples included in the NTD genomic screen analysis

Participants	Broad phenotype criteria* (n = 44 families)	Narrow phenotype criteria† (n = 17 families)
Total no. of sampled individuals	292	123
Sampled affected individuals	89	34
Sampled male participants	138	59
Sampled female participants	154	64
Affected sibling pairs	21	9
Affected half sibling pairs	2	0
Affected avuncular pairs	12	3
Affected first cousin pairs	11	4
Affected second cousin pairs	2	2
Affected relative pairs >2nd degree	20	6

*Includes families with at least two affected individuals with any type of NTD; all individuals with an NTD coded as affected. †Families having at least two affected individuals with lumbosacral myelomeningocele; only individuals with lumbosacral myelomeningocele coded as affected. "Multiplex by history" pedigrees are also included.
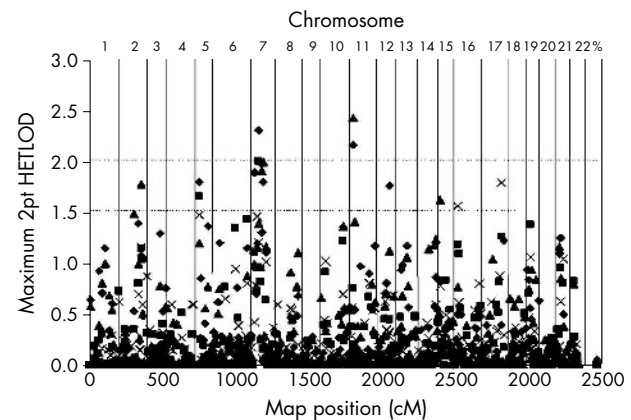
Figure 1. Two point maximum parametric lod scores, allowing for genetic heterogeneity (Hetlod). Results for all 353 markers typed are shown for four analytic schemes: broad phenotype (n = 44) assuming a dominant genetic model (♦), narrow phenotype (n = 17) assuming a dominant genetic model (■), broad phenotype (n = 44) assuming a recessive genetic model (▲), and narrow phenotype (n = 17) assuming a recessive genetic model (×).
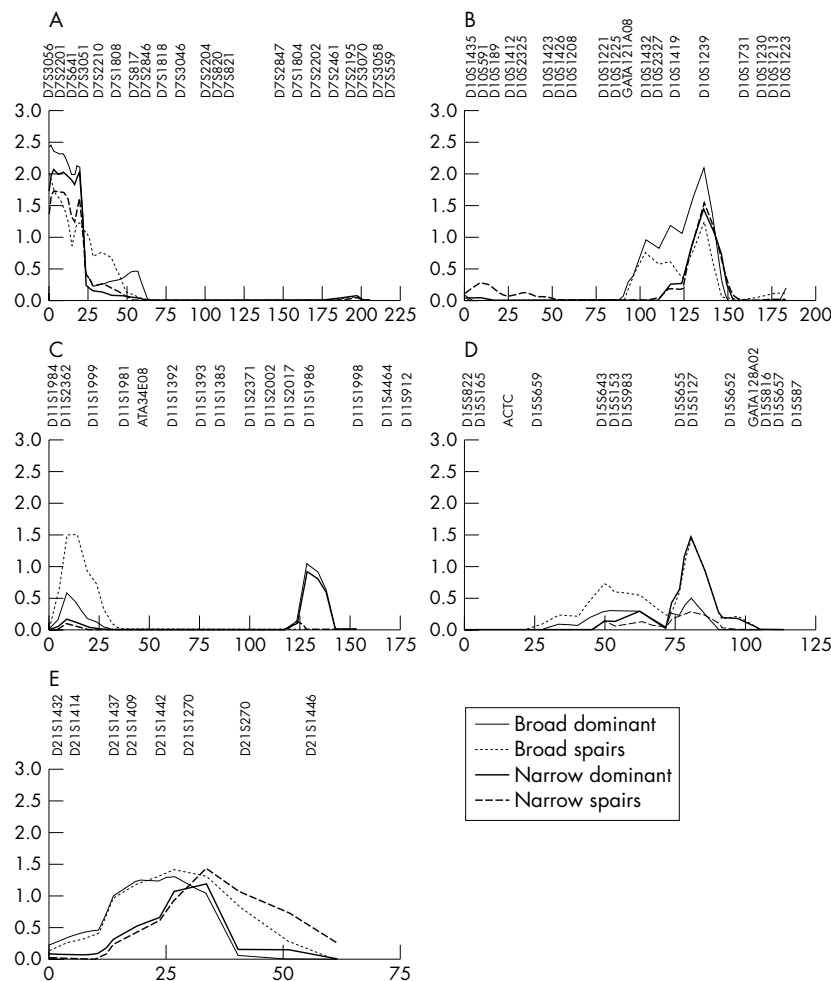
Figure 2 Multipoint lod score curves for Mlod >1.3 from chromosomes: (A) 7, (B) 10, (C) 11, (D) 15, and (E) 21 under the broad (44 families) and narrow (17 families) phenotypic classification. Parametric Hetlod scores and nonparametric sPairs lod* are shown. Marker names and positions are given at the top of each graph.

produced an Mlod of 2.10 at marker D7S513. Fig3 shows segregation of a shared disease haplotype (solid black bars) among affected individuals, spanning a 20 cM interval flanked by markers D7S3056–D7S2557. A crossover between D7S641 and D7S2201 occurs in individual 0126, an unaffected member of the pedigree, and potentially narrows the

shared region to the 9 cM interval between D7S3056 and D7S641.

To further confirm that family 8776 was indeed driving the parametric linkage results, we removed the pedigree from the dataset and reran the analysis using the remaining 43 multiplex families. With the removal of family 8776, the

Table 2 ALLEGRO multipoint lod scores >1.3 for parametric (dominant model) and nonparametric linkage analyses using a broad and a narrow phenotype criteria for all families and after removal of family 8776

Cytogenetic band	Chromosomal region	Parametric*		Nonparametric†	
		Broad	Narrow	Broad	Narrow
Complete dataset					
7p22	d7s3056-d7s3051	2.45	2.06	1.89	1.72
10q25.3	d10s1731	2.08	1.44	–	1.54
11p15	d11s2362-d11s1999	–	–	1.51	–
15q26	d15s127-d15s652	–	1.46	–	1.41
21q22.11	d21s1270	1.31	–	1.43	1.44
Without family 8776					
10q25.3	d10s1731 (134.96 cM)	2.25	1.78	(1.2)	1.73
11p15.4	d11s2362 (7.27 cM)	–	–	1.32	–
15q26.1	d15s652 (99.92 cM)	–	1.77	–	–
21q22.11	d21s1270 (31.83 cM)	1.38	–	1.32	–

*Parametric Hetlod scores; †pairs scoring statistic under an exponential model. cM, DECODE cm. n = 44 for broad phenotype and n = 17 for narrow phenotype in the complete dataset; n = 43 and n = 16 for broad and narrow, respectively, without family 8776.

Table 3 Physical map location of SNPs used to verify chromosome 7 linkage signal. Microsatellite marker locations (**bold**) are also shown for reference

Marker name	Position (bp)*
RS1636166	2 685 704
RS1357310	3 196 845
HCV8511072†	3 713 727
RS1470539	4 173 026
D7S3056	**4 237 377**
RS4720437	4 778 704
RS852423†	5 278 704
D7S2201	**5 374 993**
RS2302334	5 778 704
HCV25632983	6 225 487
HCV3219225	6 832 651
RS1544465	7 329 057
RS12847	7 836 531
D7S641	**8 458 102**
RS886505	8 496 394
RS721689	8 967 586
D7S513	**11 395 687**

*Ensembl marker location; †markers removed from analyses following tests of pairwise intermaker LD.

most significant region was on chromosome 10 with a parametric Mlod score of 2.25 at 134.96 cM at D10S1731 under the broad phenotype definition (table 2). While still of interest, the lod scores for chromosomes 11 and 21 were somewhat reduced. The region on chromosome 11 produced a nonparametric Mlod score of 1.32 at 7.27 cM, under the narrow phenotype. The region on chromosome 21 produced a parametric Mlod of 1.38 and a nonparametric Mlod of 1.32 at 31.83 cM, under the broad phenotype. Chromosome 15 had a slightly higher Mlod score of 1.77 at D15S652 under the narrow phenotype. Furthermore, our results showed that after this family was excluded, no evidence in favour of linkage remained on chromosome 7 (Mlod<1.0).

DISCUSSION

Genomic screens using traditional linkage analysis approaches are dependent on the availability of multiplex pedigrees; most genomic screens are large, with more than 100 families included.[37][38] However, screens of as few as 31–80 pedigrees have identified regions of interest in complex disease. Importantly, the initial genomic screen in late onset Alzheimer disease included only 31 multiplex pedigrees,[38] yet identified a region of interest on chromosome 19 that was subsequently found to harbour ApoE, a major susceptibility gene for Alzheimer disease. In systemic lupus erythematosus, 14 pedigrees with 34 affected individuals identified a lod score of 6.2 in a region on chromosome 15.[39] Thus, small sample sizes in genomic screens can identify regions of interest when the genetic effect is strong. In NTDs, samples from affected individuals in multiplex families can be difficult to obtain because of the high mortality associated with the condition and terminations of affected pregnancies. Thus, no other genome wide screens in NTD have been reported.

Additionally, in some common complex diseases, identification of rare families demonstrating mendelian or near-mendelian inheritance patterns has led to the identification of important loci through traditional linkage analysis approaches. Classic examples of such diseases are Alzheimer's disease (AD),[40] breast cancer,[41] and Hirschsprung's disease[42] (HD). This approach can potentially identify mendelian forms of the disease that present phenotypically the same as non-mendelian forms, such as in AD and breast cancer. In those disorders, the identified loci were specific to a mendelian subset of the disease and did not extend to the sporadic non-mendelian forms. Alternatively, susceptibility genes mapped in multiplex families may also be found to increase risk for the common non-mendelian forms of a disease, as is the case with HD.

In this report of a microsatellite based genome screen of 44 multiplex NTD families, the most significant linkage result was on chromosome 7, which produced consistently high lod
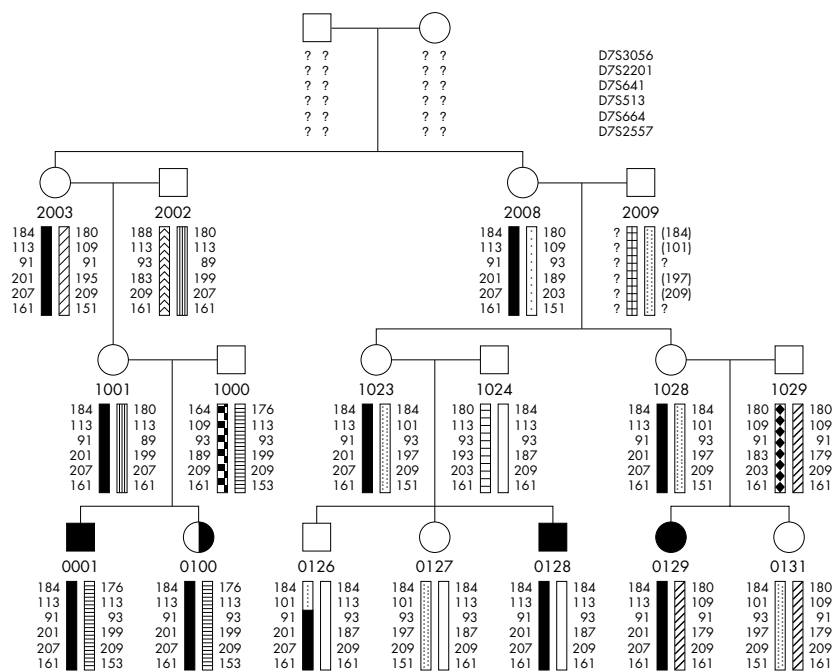
Figure 3 Pedigree diagram of family 8776 demonstrating segregation of the 184-113 haplotype in individuals affected with lumbosacral myelomeningocele. Individual 0100 has a fatty filum (an NTD variant) and also carries the associated haplotype.

scores across different analysis schemes (parametric, broad phenotype Mlod 2.45; parametric, narrow phenotype Mlod 2.06; nonparametric, broad phenotype Mlod 1.89; nonparametric, narrow phenotype Mlod 1.72). One family (8776) was identified as primarily driving the linkage results on chromosome 7 and appears to segregate a common region on 7p22 among all affected individuals, including one individual with an NTD variant, fatty filum. Using only affected individuals, the region of interest spans 20 cM and is flanked by markers D7S3056 and D7S2557. A crossover in an unaffected individual (126) in this family allowed us to potentially narrow the candidate region and restrict it to the 9 cM region between D7S3056 and D7S641. However, crossovers in unaffected individuals should be treated with caution, as it is never clear whether the individual harbours some phenotypic variant, undetected because it is asymptomatic.

A review of public databases, using our internally developed DAS-Ensembl integration and locally developed scripts,[43] found 72 genes from NCBI and 63 ESTs from Ensembl in the region surrounding our highest linkage peak on chromosome 7 in family 8776, representing an unbiased collection of genes based on positional linkage evidence. While none of these is an obvious candidate for NTD, evaluation of syntenic regions between the mouse and human genomes (NCBI/OMIM Davis mouse to human homology maps) reveals an additional two genes (*Meox2, Twist1*) in regions on mouse chromosome 12 that are homologous to human 7p21.1–22.2. *Meox2* is a homeobox gene that is expressed in the somites during neurulation,[44] while mouse embryos homozygous for *Twist1* have been shown to develop neural tube defects.

When family 8776 was eliminated from the linkage analysis, regions of interest on chromosomes 10, 11, 15, and 21 were identified. The best linkage evidence was at 10q25.3, which produced a parametric multipoint lod score of 2.25. Three candidate genes (*FGFR2, GFRA1,* and *Pax2*) map close to 10q25.3. Fibroblast growth factor receptor 2 (*FGFR2*) is expressed in the spinal cord of the chick embryo in the stages between gastrulation and limb bud formation.[45] GDNF family receptor alpha 1 (*GFRA1*) is expressed in embryonic mouse spinal cord.[46] Pax2 belongs to the *Pax* gene family, long investigated for its role in relation to NTD.[47–49]

Several plausible NTD candidates map to the short arm on chromosome 21 (*CBS, RFC1*, and *NCAM2*). Cystathionine beta-synthase (*CBS*) and reduced folate carrier protein 1 (*RFC1*) are both important players in the folate metabolism pathway.[50] Recently, neural cell adhesion molecule 1 (*NCAM1*), which maps to 11q21.3, has been shown to be associated in NTD singleton families,[51] making the *NCAM2* gene on chromosome 21 potentially interesting. No obvious candidate genes in the regions on chromosomes 11 and 15 are apparent.

Our approach to evaluating these data will be to maximise the amount of information we extract from the genomic screen by carefully characterising regions of interest, continuing to add multiplex pedigrees as they become available, expanding the phenotypic classifications to increase the sample size, and integrating data from other important lines of investigation involving biologically plausible candidate genes, such as those from mouse models of NTDs and genes involved in folate metabolism. These data represent an important and useful tool for narrowing the search for candidate genes for NTDs.

ACKNOWLEDGEMENTS

. . . .

Authors' affiliations

E Rampersaud, D S Enterline, T M George, D G Siegel, E C Melvin, J Allen, L E Floyd, P Hammock, J Mackey, L Mehltretter, S G West, G Worley, J R Gilbert, M C Speer, Duke University Medical Center, Durham, NC, USA
A G Bassuk, J A Kessler, Northwestern University's Feinberg School of Medicine, Departments of Pediatrics and Neurology and Children's Memorial Hospital, Chicago, IL, USA
J Aben, C Powell, Children's Rehabilitation Service, Birmingham, AL, USA
A Aylsworth, University of North Carolina, Chapel Hill, NC, USA
T Brei, C Buran, K Sawin, Indiana University School of Medicine, Indianapolis, IN, USA
J Bodurtha, Virginia Commonwealth University, Richmond, VA, USA
B Iskandar, University of Wisconsin Hospitals, Madison, WI, USA
J Ito, D McLone, Children's Memorial Hospital, Chicago, IL, USA
N Lasarsky, P Mack, E Meeropol, Shriner's Hospital, Springfield, IL, USA
L E Mitchell, Institute of Biosciences and Technology, Texas A&M University Health Science Center, Houston, TX, USA
W J Oakes, University of Alabama, Birmingham, USA
J S Nye, Johnson & Johnson, Princeton, USA
R Stevenson, Greenwood Genetic Center Greenwood, USA
M Walker, University of Utah, Salt Lake City, USA

Competing interests: none declared

Broad phenotype includes all types of NTDs; narrow phenotype is restricted to individuals with lumbosacral myelomeningocele. Chromosomes 7 and 11 have Mlod >2.0.

The NTD Collaborative Group in the USA includes the following centres: Duke University Medical Center (Durham, NC), Children's Rehabilitation Service (Birmingham, Alabama), University of Alabama (Birmingham, AL), University of North Carolina (Chapel Hill, NC), Carolinas Medical Center (Charlotte, NC), Northwestern University's Feinberg School of Medicine, Departments of Pediatrics and Neurology, and Children's Memorial Hospital (Chicago, IL),Institute of Biosciences and Technology, Texas A&M University System Heath Center (Houston TX, Indiana University School of Medicine (Indianapolis, IN), University of Wisconsin Hospitals (Madison, WI), Virginia Commonwealth University (Richmond, VA), University of Utah (Salt Lake City, UT), and Shriner's Hospital (Springfield, MA)

REFERENCES

1 **Risch N**. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 1990;**46**:229–41.
2 **Khoury MJ**, Wagener DK. Epidemiological evaluation of the use of genetics to improve the predictive value of disease risk factors. *Am J Hum Genet* 1995;**56**:835–44.
3 **Elwood JM**, Little J, Elwood JH. *Epidemiology and control of neural tube defects*. Oxford: Oxford University Press, 1992.
4 **Li H**, Zhao Y, Li S, Xu Y, Huang B, Cui M, Zheng G. (Epidemiology of twin and twin with birth defects in China). *Zhonghua Yi Xue Za Zhi* 2002;**82**:164–7.
5 **Hume RF**, Drugan A, Reichler A, Lampinen J, Martin LS, Johnson MP, Evans MI. Aneuploidy among prenatally detected neural tube defects. *Am J Med Genet* 1996;**61**:171–3.
6 **Kennedy D**, Chitayat D, Winsor EJT, Silver M, Toi A. Prenatally diagnosed neural tube defects: Ultrasound, chromosome, and autopsy or postnatal findings in 212 cases. *Am J Med Genet* 1998;**77**:317–21.
7 **Toriello HV**, Higgins JV. Possible causal heterogeneity in spina bifida cystica. *Am J Med Genet* 1985;**21**:13–20.
8 **Hall JG**, Keena BA. Adjusting recurrence risks for neural tube defects based on B.C. data. *Am J Hum Genet* 1986;**39**:A64.
9 **Frecker MF**, Fraser FC, Heneghan WD. Are 'upper' and 'lower' neural tube defects aetiologically different? *J Med Genet* 1988;**25**:503–4.
10 **Drainer E**, May HM, Tolmie JL. Do familial neural tube defects breed true? *J Med Genet* 1991;**28**:605–8.
11 **Garabedian BH**, Fraser FC. Upper and lower neural tube defects: an alternate hypothesis. *J Med Genet* 1993;**30**:849–1.
12 **Milunsky A**, Jick H, Jick SS, Bruell CL, MacLaughlin DS, Rothman KJ, Willett W. Multivitamin/folic acid supplementation in early pregnancy reduces the prevalence of neural tube defects. *JAMA* 1991;**262**:2847–52.
13 **MRC Vitamin Study Research Group**. Prevention of neural tube defects: Results of the Medical Research Council Vitamin Study. *Lancet* 1991;**338**:131–7.

14 **Centers for Disease Control and Prevention**. Recommendations for the use of folic acid to reduce the number of cases of spina bifida and other neural tube defects. *Morbid Mortal Wkly Rep* 1992;**41**:1–7.

15 **Chatkupt S**, Skurnick JH, Jaggi M, Mitruka K, Koenigsberger MR, Johnson WG. Study of genetics, epidemiology, and vitamin usage in familial spina bifida in the United States in the 1990s. *Neurology* 1994;**44**:65–69.

16 **Juriloff DM**, Harris MJ. Mouse models for neural tube closure defects. *Hum Mol Genet* 2000;**9**:993–1000.

17 **Copp AJ**, Greene ND, Murdoch JN. The genetic basis of mammalian neurulation. *Nat Rev Genet* 2003;**4**:784–93.

18 **Harris M**. Why are the genes that cause risk of human neural tube defects so hard to find? *Teratology* 2001;**63**:165–6.

19 **Melvin EC**, George TM, Worley G, Franklin A, Mackey J, Viles K, Shah N, Drake CR, Enterline DS, McLone D, Nye J, Oakes WJ, McLaughlin C, Walker ML, Peterson P, Brei T, Buran C, Aben J, Ohm B, Bermans I, Qumsiyeh M, Vance J, Pericak-Vance MA, Speer MC. Genetic studies in neural tube defects. NTD Collaborative Group. *Pediatr Neurosurg* 2000;**32**:1–9.

20 **Vance JM**, Ben Othmane K. Methods of Genotyping. In: Haines JL, Pericak-Vance MA, eds. *Approaches to gene mapping in complex human diseases*. New York: Wiley-Liss, 1998:213–28.

21 **Vance JM**, Slotterbeck B, Yamaoka L, Haynes C, Roses AD, Pericak-Vance MA. A fluorescent genotyping system with emphasis on multiple users, flexibility, cost and high output. Genome Mapping and Sequencing, CSH Laboratory, 1996;**239**.

22 **Haynes C**, Speer MC, Peedin M, Roses AD, Haines JL, Vance JM, Pericak-Vance MA. PEDIGENE: A comprehensive data management system to facilitate efficient and rapid disease gene mapping. *Am J Hum Genet* 1995;**57**:A193.

23 **O'Connell JR**, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998;**63**:259–66.

24 **Epstein MP**, Duren WL, Boehnke M. Improved inference of relationship for pairs of individuals. *Am J Hum Genet* 2000;**67**:1219–31.

25 **Boehnke M**, Cox NJ. Accurate inference of relationships in sib-pair linkage studies. *Am J Hum Genet* 1997;**61**:423–9.

26 **O'Connell JR**. Rapid multipoint linkage analysis via inheritance vectors in the Elston-Stewart algorithm. *Hum Hered* 2001;**51**:226–40.

27 **Ott J**. Linkage analysis and family classification under heterogeneity. *Ann Hum Genet* 1983;**47**:311–20.

28 **Boehnke M**. Estimating the power of a proposed linkage study: a practical computer simulation approach. *Am J Hum Genet* 1986;**39**:513–27.

29 **Ploughman LM**, Boehnke M. Estimating the power of a proposed linkage study for a complex genetic trait. *Am J Hum Genet* 1989;**44**:543–51.

30 **Gudbjartsson DF**, Jonasson K, Frigge ML, Kong A. Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 2000;**25**:12–13.

31 **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**:241–7.

32 **Matise TC**, Sachidanandam R, Clark AG, Kruglyak L, Wijsman E, Kakol J, Buyske S, Chui B, Cohen P, de Toma C, Ehm M, Glanowski S, He C, Heil J, Markianos K, McMullen I, Pericak-Vance MA, Silbergleit A, Stein L, Wagner M, Wilson AF, Winick JD, Winn-Deen ES, Yamashiro CT, Cann HM, Lai E, Holden AL. A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set. *Am J Hum Genet* 2003;**73**:271–84.

33 **Broman KW**. Estimation of allele frequencies with data on sibships. *Genet Epidemiol* 2001;**20**:307–15.

34 **Whittemore AS**, Halpern J. A class of tests for linkage using affected pedigree members. *Biometrics* 1994;**50**:118–27.

35 **McPeek MS**, Strahs A. Assignment of linkage disequilibrium by the decay of haplotype sharing with application to fine-scale genetic mapping. *Am J Hum Genet* 1999;**65**:858–75.

36 **Weir BS**. *Genetic data analysis II: methods for discrete population genetic data*. Sunderland: Sinaur Associates, 1996.

37 **Hauser ER**, Crossman DC, Granger C, Haines JL, Jones CJH, Mooser V, Middleton L, Roses AD, Hauser MA, Pericak-Vance MA, Vance JM, Kraus WE. A genome-wide scan in 438 families with early-onset coronary artery disease. *Am J Hum Gen* 2002;**71**:459.

38 **Pericak-Vance MA**, Bebout JL, Gaskell PC, Yamaoka LH, Hung WY, Alberts MJ, Walker AP, Bartlett RJ, Haynes CS, Welsh KA, Earl NL, Heyman A, Clark CM, Roses AD. Linkage studies in familial Alzheimer's disease: Evidence for chromosome 19 linkage. *Am J Hum Genet* 1991;**48**:1034–50.

39 **Namjou B**, Nath SK, Kilpatrick J, Kelly JA, Reid J, James JA, Harley JB. Stratification of pedigrees multiplex for systemic lupus erythematosus and for self-reported rheumatoid arthritis detects a systemic lupus erythematosus susceptibility gene (SLER1) at 5p15.3. *Arthritis Rheum* 2002;**46**:2937–45.

40 **Strittmatter WJ**, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, Roses AD. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci USA* 1993;**90**:1977–81.

41 **Newman B**, Austin MA, Lee M, King MC. Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. *Proc Natl Acad Sci USA* 1988;**85**:3044–8.

42 **Angrist M**, Kauffman E, Slaugenhaupt SA, Matise TC, Puffenberger EG, Washington SS, Lipson A, Cass DT, Reyna T, Weeks DE. A gene for Hirschsprung disease (megacolon) in the pericentromeric region of human chromosome 10. *Nat Genet* 1993;**4**:351–6.

43 **Xu H**, Vance JM, Stenger JE. Integrating genomic laboratory data with whole genome annotation using Ensembl-DAS. *Am J Hum Genet* 2002;**71**:393–1301A.

44 **Mankoo BS**, Skuntz S, Harrigan I, Grigorieva E, Candia A, Wright CV, Arnheiter H, Pachnis V. The concerted action of Meox homeobox genes is required upstream of genetic pathways essential for the formation, patterning and differentiation of somites. *Development* 2003;**130**:4655–64.

45 **Walshe J**, Mason I. Expression of FGFR1, FGFR2 and FGFR3 during early neural development in the chick embryo. *Mechanisms of Development* 2000;**90**:103–10.

46 **Garces A**, Haase G, Airaksinen MS, Livet J, Filippi P, deLapeyriere O. GFRalpha 1 is required for development of distinct subpopulations of motoneuron. *J Neurosci* 2000;**20**:4992–5000.

47 **Chatkupt S**, Hol FA, Shugart YY, Geurds MPA, Stenroos ES, Koenigsberger MR, Hamel BCJ, Johnson WG, Mariman ECM. Absence of linkage between familial neural tube defects and PAX3 gene. *J Med Genet* 1995;**32**:200–4.

48 **Hol FA**, Geurds MPA, Chatkupt S, Shugart YY, Balling R, SchranderStumpel CTR, Johnson WG, Hamel BCJ, Mariman ECM. PAX genes and human neural tube defects: An amino acid substitution in PAX1 in a patient with spina bifida. *J Med Genet* 1996;**33**:655–60.

49 **Volcik KA**, Blanton SH, Kruzel MC, Townsend IT, Tyerman GH, Mier RJ, Northrup H. Testing for genetic associations with the PAX gene family in a spina bifida population. *Am J Med Genet* 2002;**110**:195–202.

50 **Ramsbottom D**, Scott JM, Molloy A, Weir DG, Kirke PN, Mills JL, Gallagher PM, Whitehead AS. Are common mutations of cystathionine beta-synthase involved in the aetiology of neural tube defects? *Clin Genet* 1997;**51**:39–42.

51 **Deak KL**, Boyles AL, Etchevers HC, Melvin EC, Siegel DG, Graham FL, Slifer S, Enterline DS, George TM, Vekemans M, McClay D, Bassuk AG, Kessler JA, Linney E, Gilbert JR, Speer MC, NTD Collaborative Group. SNPs in the neural cell adhesion molecule 1 (NCAM1) may be associated with human neural tube defects. *Hum Genet* 2005;**117**:113–42.