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


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 (FOFR2, 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.
markers had all QC inconsistencies resolved and affected
attempted (88%) were approved for analysis. These 353
assessed. With these procedures in place, 353 of 402 markers
the QC genotype and the corresponding sample genotype was
merging genotypic data into PEDIGENE, agreement between
termination of prenatally detected affected pregnancies.18 19
the increased mortality associated with spina bifida, and
NTD pedigrees is hampered by lethality in anencephaly cases,
major causative gene. An alternative approach of using
assessment of candidate genes in NTDs has yet to identify a

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.26 Multiplex reactions using two colour fluorescein
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
collection. DNA for marker genotypes were managed using the
PEDIGENE system22 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 PEDCHECK22 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 hetero-
genity lod scores (Hetlod scores) using HOMOG.27

Power analyses of the multiplex families in the screen were
carried out using SIMLINK28 allowing for a dominant
model and a disease allele frequency of 0.001. When fully
informativeness, 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.29 Genetic marker distance
was based on 10 cM sex average integrated maps from
deCode Genetics30 and the Marshfield Medical Research
Foundation maps. Map order was verified using Map-O-
Mat.31 Marker allele frequencies were estimated from our
dataset using all individuals.32 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 Spairs sharing statistic34 and the exponen-
tial model35 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

www.jmedgenet.com
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 scan 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, lipomyelomenigocele, 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 D7S3056. Deviations from Hardy-Weinberg equilibrium (HWE) (p < 0.06) were tested using the Genetic Data Analysis (GDA) program. 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 cM 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>
<thead>
<tr>
<th>Table 1</th>
<th>Description of samples included in the NTD genomic screen analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td><strong>Broad phenotype criteria</strong></td>
</tr>
<tr>
<td><strong>(n = 44 families)</strong></td>
<td><strong>(n = 17 families)</strong></td>
</tr>
<tr>
<td>Total no. of sampled individuals</td>
<td>292</td>
</tr>
<tr>
<td>Sampled affected individuals</td>
<td>89</td>
</tr>
<tr>
<td>Sampled male participants</td>
<td>138</td>
</tr>
<tr>
<td>Sampled female participants</td>
<td>154</td>
</tr>
<tr>
<td>Affected sibling pairs</td>
<td>21</td>
</tr>
<tr>
<td>Affected half sib pairs</td>
<td>2</td>
</tr>
<tr>
<td>Affected avuncular pairs</td>
<td>12</td>
</tr>
<tr>
<td>Affected first cousin pairs</td>
<td>11</td>
</tr>
<tr>
<td>Affected second cousin pairs</td>
<td>2</td>
</tr>
<tr>
<td>Affected relative pairs &gt; 2nd degree</td>
<td>20</td>
</tr>
</tbody>
</table>

*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 (▼).
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

<table>
<thead>
<tr>
<th>Cytogenetic band</th>
<th>Chromosomal region</th>
<th>Parametric*</th>
<th>Nonparametric†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Broad</td>
<td>Narrow</td>
</tr>
<tr>
<td><strong>Complete dataset</strong></td>
<td></td>
<td>2.45</td>
<td>2.06</td>
</tr>
<tr>
<td>7p22</td>
<td>d7s3056–d7s3051</td>
<td>2.08</td>
<td>1.44</td>
</tr>
<tr>
<td>10q25.3</td>
<td>d10s1731</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>11p15</td>
<td>d11s2362–d11s1999</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>15q26</td>
<td>d15s127–d15s652</td>
<td>–</td>
<td>1.46</td>
</tr>
<tr>
<td>21q22.11</td>
<td>d21s1270</td>
<td>1.31</td>
<td>–</td>
</tr>
<tr>
<td><strong>Without family 8776</strong></td>
<td></td>
<td>2.25</td>
<td>1.78</td>
</tr>
<tr>
<td>10q25.3</td>
<td>d10s1731 (134.96 cM)</td>
<td>2.52</td>
<td>1.78</td>
</tr>
<tr>
<td>11p15.4</td>
<td>d11s2362 (7.27 cM)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>15q26.1</td>
<td>d15s652 (99.92 cM)</td>
<td>–</td>
<td>1.77</td>
</tr>
<tr>
<td>21q22.11</td>
<td>d21s1270 (31.83 cM)</td>
<td>1.38</td>
<td>–</td>
</tr>
</tbody>
</table>

*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.
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 narrow phenotype. The region on 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. 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, 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. 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), breast cancer, and Hirschsprung’s disease (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
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 flum. 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, 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 collection of genes based on positional linkage evidence) 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, 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, GPR1, 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. GDFN family receptor alpha 1 (GPR1) is expressed in embryonic mouse spinal cord. Pax2 belongs to the Pax gene family, long investigated for its role in relation to NTD.

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. Recently, neural cell adhesion molecule 1 (NCAM1), which maps to 11q21.3, has been shown to be associated in NTD singleton families; 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

We are grateful for the participation and continued efforts of all the families in the NTD study. We also thank P Banks, A Boyles, and S Seth for assistance with collecting family data. This work was supported by grants from the National Institutes of Health (HD39948, HD39083, ES11375, NS39818, ES011961, NS26630, HD39195, HD39081) and a Ruth L Kirschstein National Research Service Awards predoctoral fellowship (NS046249).

Authors’ affiliations

E Randolph, D S Enterline, T M George, D G Siegel, E C Melvin, J Allen, I L Floyd, P Hammock, J Mackey, L Mehlthurer, 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 Varela, Virginia Commonwealth University, Richmond, VA, USA

B Iskandar, University of Wisconsin Hospitals, Madison, WI, USA

J Itø, D Mclone, Children’s Memorial Hospital, Chicago, IL, USA

N Lasarsky, P Mack, E Meerpohl, 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 at 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 Health 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


35. McPeek MS, Straths A. Assignment of linkage disequilibrium by the decay of haplotype sharing with application to fine-scale genetic mapping. Am J Hum Genet 1999;65:859–70.


