Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population

Neil Risch[1,9], Deborah de Leon[3], Laurie Ozelius[4,7], Patricia Kramer[6], Laura Almasy[2], Burton Singer[8], Stanley Fahn[3], Xandra Breakefield[4,5] & Susan Bressman[3]

[1]*Department of Epidemiology and Public Health, and* [2]*Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06520, USA*
[3]*Dystonia Clinical Research Center, Neurological Institute, Columbia Presbyterian Medical Center, New York, New York 10032, USA*
[4]*Molecular Neurogenetics Unit, Neuroscience Center (Neurology), Massachusetts General Hospital East, Charlestown, Massachusetts, USA 02129*
[5]*Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA 02115*
[6]*Department of Neurology, Oregon Health Sciences University, Portland, Oregon, USA 97201*
[7]*Neuroscience Program, Harvard Medical School, Boston, Massachusetts, USA 02115*
[8]*Office of Population Research, Princeton University, Princeton, NJ 08544*

[9]*Current address: Department of Genetics, School of Medicine, Stanford University, Stanford, CA 94305*

Correspondence should be addressed to N.R.

We have examined data on six closely linked microsatellite loci on chromosome 9q34 from 59 Ashkenazi Jewish families with idiopathic torsion dystonia (ITD). Our data show that the vast majority (>90%) of early-onset ITD cases in the Ashkenazi population are due to a single founder mutation, which we estimate first appeared approximately 350 years ago. We also show that carriers preferentially originate from the northern part of the historic Jewish Pale of settlement (Lithuania and Byelorussia). The recent origin of this dominant mutation and its current high frequency (between 1/6,000 and 1/2,000) suggest that the Ashkenazi population descends from a limited group of founders, and emphasize the importance of genetic drift in determining disease allele frequencies in this population.

The Ashkenazi Jewish population is subject to a range of genetic diseases unique to this population[1]. Undoubtedly, this is due to its remarkable demographic history, including phenomenal expansions and contractions over the past two millennia, as well as the general practice of endogamy. The expansion of this population in Eastern Europe in the sixteenth to nineteenth centuries probably has few parallels in recent history. Rapid growth of a population from a limited group of founders is precisely the condition under which one would expect to see unique genetic conditions achieve high frequency by chance (genetic drift). Nonetheless, the explanation for the high frequency of genetic conditions, such as Tay Sachs disease, in the Jewish population has been debated for decades with some observers favoring heterozygote advantage[2–7], while others favour drift[8–10].

Recent advances in molecular genetics now make possible a reassessment of this issue. The discovery of highly polymorphic marker loci, specifically short tandem repeat polymorphisms (STRs) densely distributed throughout the genome, affords the opportunity to characterize the history of a particular mutation. Here, we perform such an analysis for one genetic disease found in the Ashkenazi population at high frequency, idiopathic torsion dystonia (ITD). ITD is a movement disorder characterized by twisting movements or postures that may affect any voluntary muscle group. Clinical variation is extreme, ranging from generalized involvement of the limbs and axial muscles, to focal involvement of a single body region. Age at onset also varies dramatically, ranging from early childhood to the eighth decade[11].

ITD has long been noted to occur at a higher frequency in Ashkenazi Jews than either non-Ashkenazi Jews or non-Jews[12–13]. We have noted that early onset (before age 28) appears to define a clinically homogeneous subset in which there is onset in a limb and a tendency for the disorder to generalize[11,14]. It is also this group that appears to show a marked excess in Ashkenazim.

The mode of inheritance of ITD in Ashkenazim has been controversial, with some observers arguing for recessive inheritance[15,16] and others for dominant inheritance[12,17]. In our studies, we clearly demonstrated dominant inheritance, but with low (30%) penetrance; that is, we estimated that only 30% of carriers of this dominant gene are affected[14,18].

The dominant inheritance, but high frequency of this disease in the Ashkenazi population led us to speculate that a single founder mutation might underlie the majority of ITD cases in the current Ashkenazi population[18]. It was possible to verify this prediction through linkage analysis, once we determined that the gene for early onset ITD (*DYT1*) is located on chromosome 9q34 (refs 19,20). When looking for linkage disequilibrium with one of the marker loci, *ASS*, we discovered very strong linkage disequilibrium at a relatively large genetic distance (at least a few centiMorgans)[21]. This finding confirmed the likely single mutational origin of most cases of early onset ITD in the Ashkenazi population, and suggested a recent origin for this mutation.

We have now expanded the number of Ashkenazi families with ITD; in addition, we have added four dinucleotide repeat polymorphisms recently mapped to the 9q34 region[22] to the two (*ASS,ABL*) previously analyzed. These advances enable us to calculate a date of origin of the founder mutation for ITD in the Ashkenazim. Furthermore, we also perform a geographic analysis of

Table 1 Lod Scores between 9q34 Markers and DYT1

θ	Loci *D9S62a/D9S62b/D9S63*	*ASS/ABL*	*D9S64*
0	9.06	4.70	-4.33
.01	8.82	5.86	2.81
.02	8.66	6.12	4.00
.03	8.53	6.27	4.60
.04	8.39	6.34	4.94
.05	8.26	6.35	5.15
.06		6.32	5.26
.07		6.26	5.32
.08		6.18	5.33
.09			5.31
.10			5.26

our oldest carriers to determine a likely location of origin of this mutation, and an epidemiologic analysis to determine the prevalence of the mutation in the current Ashkenazi population. These results suggest that previously reported prevalence rates are severely underestimated. Finally, we discuss the ramifications of our findings for the recent history of the Ashkenazi population. The dominant nature of this disease mutation, a rare phenomenon among Jewish genetic diseases, allows us to determine more directly the effect of genetic drift without the possible complication of heterozygote advantage attendant to recessive diseases.

Linkage analysis between *DYT1* and 9q34 markers

In our previous analysis, we obtained a maximum lod score with *ASS* of 3.49 at a recombination fraction of 0.03. In our new analysis, the maximum lod score with *ASS/ABL* (haplotype) was 6.35 at a recombination fraction of 0.048 (Table 1). A higher lod score of 9.06 at θ=0 was obtained with the proximal three polymorphisms (*D9S62a, D9S62b, D9S63*), indicating no crossovers. The more distal locus, *D9S64*, gave a maximum lod score of 5.33 at θ=0.078. Because the three proximal markers are essentially completely informative, multipoint analysis of the 3 clusters of loci with disease gives a maximum lod of 9.0 with *DYT1* located at the proximal three loci.

Linkage disequilibrium at *DYT1*

Allele frequencies on disease-bearing (D, n=36) and control (N, n=306) Ashkenazi chromosomes for the six microsatellite loci *D9S62a, D9S62b, D9S63, ASS, ABL* and *D9S64* are given in Table 2. Results of χ^2 tests are given in Table 3. As we reported[21], there is a striking excess in the frequency of the '12' allele at *ASS* on *DYT1* chromosomes versus controls. We also reported[21] an excess of the '4' allele at *ABL*, although this is also the most common in controls.

We now confirm this excess, and present results for four new microsatellites. For *D9S62a*, the '2' allele is the most common on *ITD* chromosomes. For *D9S62b*, the '8' allele is found most often on disease chromosomes. *D9S63* reveals a high frequency of the '16' allele on *ITD* chromosomes, while locus *D9S64* displays associations with two different alleles, '2' and '10'. The most likely explanation for association with two alleles at this locus is an early recombination event, perhaps in the first generation after the occurrence of the mutation. An early mutation at this marker locus is also a possibility, but given that the recombination fraction from this locus to *DYT1* is about 4%, recombination seems more likely.

The values of the linkage disequilibrium parameter δ for the six loci are given in Table 3. δ is highest for locus *D9S63*, and decreases slightly for the proximal loci, *D9S62a* and *D9S62b* and distal loci *ASS* and *ABL*, and is lower for *D9S64*. These results suggest that *DYT1* lies very close to locus *D9S63*, but cannot be further localized proximal or distal to this locus.

Full haplotypes for the 36 disease-bearing chromosomes for all six marker loci are presented in Table 4. Of these, 33 carry a '16' allele at *D9S63*. Only one of these does not also carry a '2' or '8' allele at *D9S62 a* and *b*, or a '12' allele at *ASS*. Hence, 32/36=88.9% carry either a '2-8-16' haplotype at *D9S62a,b* and *D9S63*, or a '16-12' haplotype at *D9S63–*

Table 2 Allele frequencies for disease and normal chromosomes

Locus	*D9S62a*		*D9S62b*		*D9S63*		*ASS*		*ABL*		*D9S64*	
Allele	D	N	D	N	D	N	D	N	D	N	D	N
–12									.027	.058		
–10				.010								
–8				.003								
–4				.023								
–2				.007								
0	.027	.039		.076		.196				.105		
2	.861	.029		.046	.056	.021	.139	.428		.029	.278	.067
4	.111	.771	.027	.030		.039		.072	.889	.625		
6		.114	.083	.109		.056	.027	.086	.083	.177	.027	.060
8		.046	.861	.135		.069	.027	.138				.057
10			.027	.313		.023		.059		.004	.333	.057
12				.072		.075	.806	.086			.194	.295
14				.043		.190		.089		.004	.111	.201
16				.049	.917	.118		.039			.027	.174
18				.043	.027	.065		.003				.060
20				.033		.020						.020
22				.003		.023					.027	.007
24				.003		.007						
30				.003								

Table 3 Linkage disequilibrium between 9q34 markers and *DYT1* in the Ashkenazi population

Locus	Allele	χ^2	δ
D9S62a	2	208.3	.857
D9S62b	8	97.8	.839
D9S63	16	122.0	.906
ASS	12	118.4	.787
ABL	4	8.7	.704
D9S64	2	15.2	.556
	10	27.9	

ASS, or both. The frequency of the '2-8-16' haplotype among control chromosomes was 3/306=0.0098; no cases of the '16-12' haplotype occurred.

For the remaining 23 cases, we were able to phase their chromosomes into haplotypes using other family members, although we could not determine which of the two haplotypes contained the disease mutation. Given the strong association with four of the markers (*D9S62a* and *b*, *D9S63* and *ASS*), we were able to determine "probable" haplotype carrier status for most of these individuals. In 22 cases, one of the haplotypes carried either the '2-8-16' combination or the '16-12' combination, or both. The posterior probability that these "associated" haplotypes carry the disease mutation (see Methodology) was in each case at least 99%. For the one case with no associated alleles evident for any of the loci, we arbitrarily selected one of the two chromosomes at random. These 23 cases are also listed in Table 4.

Of the 59 haplotypes listed in Table 4, 55 carry a '16' allele at *D9S63*. Of these 55, only one does not also carry a '2' or '8' allele at *D9S62a* and *b*, or a '12' allele at *ASS*. Twenty seven carry the full haplotype at all six loci (including both '2' and '10' alleles at *D9S64*); a further 15 carry all alleles except those at *D9S64*.

These results suggest that the large majority of early-onset ITD cases in the Ashkenazi population are due to a unique mutation characterized by a haplotype of markers at 9q34 loci. This finding is consistent with our earlier prediction that genetic drift of a single founder mutation was likely to explain the high frequency of this dominant disease in the Jewish population[18]. Four disease chromosomes carried none of the associated alleles at loci *D9S62a*, *D9S62b*, *D9S63* or *ASS*, suggesting that these do not carry the same *DYT1* mutation as the remaining majority. A single disease chromosome carried only the "16" allele at *D9S63*, but no others; it is unclear whether this chromosome bears the same *DYT1* mutation as the others, although unusual clinical features in affected members of this family (neck and eye involvement, later onset) suggest that it may not[23]. In any event, we estimate that at least 91.5% (54/59) of early, limb or neck onset ITD in the Ashkenazi population is due to a single progenitor mutation. We refer to this mutation as *DYT1*^{AJ}.

Dating the mutation

The frequencies of various haplotypes based on the markers given in Table 4 can be used to estimate the original time of appearance of the *DYT1*^{AJ} mutation. A similar type of analysis has recently been performed for the ΔF508 mutation in cystic fibrosis[24,25] and the gene for diastrophic dysplasia in Finland[26].

For this analysis, we use the 59 families from which the haplotypes in Table 4 were derived. The last four chromosomes listed in Table 4 contain no associated alleles, and we assume that these cases either represent a distinct *DYT1* mutation or non-*DYT1* cases. The fifth from last chromosome contains only the '16' allele at *D9S63*, and we eliminate this family from analysis as well. To support this removal, we note that this family is also somewhat atypical clinically[22]. We assume that the remaining 54 families carry the same original mutation. We analyze these remaining 54 families in three ways: (i) by including only a single individual from each family (probands, n=54); (ii) by including all affected individuals from these families (n=96); (iii) by including all affected individuals plus all unaffected individuals who are known to carry the *DYT1*^{AJ} mutation by inheritance (n=180). Analyses (2) and (3) are likely to be more accurate due to the inclusion of a larger number of subjects and the allowance for known recombination events within the families. For example, among the 36 multiplex families, we observed five recombinations on disease chromosomes: two recombinations occurred between *D9S63* and *ASS/ABL*, and three recombination events between *ASS/ABL* and *D9S64*.

For this analysis, we considered the loci *D9S62a*, *D9S62b* and *D9S63* as defining the progenitor haplotype carrying the *DYT1*^{AJ} mutation. These loci, and in particular *D9S63* must be extremely close to the *DYT1* locus as few haplotypes do not carry associated alleles at these loci. In fact, none of the first 54 haplotypes listed in Table 5 is missing the "16" allele at *D9S63*. We use the locus *ASS* as our first locus next to (distal to) the *DYT1* locus. For the recombination fraction between *DYT1* and *ASS* (θ_1) we use the estimated recombination fraction and support interval for the proximal cluster of marker loci *D9S62a/D9S62b/D9S63* to the cluster *ASS/ABL*, namely θ_1=0.018 (0.007,0.033). These values are likely to be far more accurate

Table 4 Disease chromosomes at six microsatellite loci

Locus						N		
D9s62a	*D9S62b*	*D9S63*	*ASS*	*ABL*	*D9s64*	Def.	Prob.	Total
2	8	16	12	4	10	9	5	14
2	8	16	12	4	2	7	5	12
2	8	16	12	4	12	6	0	6
2	8	16	12	4	14	3	2	5
2	8	16	12	4	16	1	2	3
2	8	16	12	4	4	0	1	1
2	8	16	12	6	10	1	0	1
2	8	16	2	4	14	1	1	2
2	8	16	2	4	16	0	1	1
2	8	16	2	4	22	1	0	1
2	8	16	2	4	10	0	1	1
2	8	16	2	-12	2	1	0	1
2	8	16	-2	4	16	0	1	1
2	4	16	12	4	2	1	1	2
4	8	16	12	4	2	1	1	2
4	12	16	12	4	2	0	1	1
0	6	16	2	4	10	1	0	1
4	6	18	2	4	10	1	0	1
4	6	2	6	6	12	1	0	1
4	10	2	8	6	6	1	0	1
0	6	12	4	6	12	0	1	1

Def., phased, determined by inheritance; Prob., phased, not determined by inheritance (see Methodology).

Table 5 Estimates of the age of the DYT1^{AJ} mutation (in generations)

Loci	Group	Proportion recombinant	Generations (SI)[a]
D9S63–ASS	Probands	7/54 (.130)	8.4 (4.6, 21.8)
	All affecteds	13/96 (.135)	8.8 (4.8, 22.8)
	Affecteds & carriers	24/180 (.133)	8.7 (4.7, 22.4)
ASS–D9S64	Probands	15/47 (.319)	19.5 (10.6, 37.5)
	All affecteds	27/83 (.325)	19.9 (10.8, 38.4)
	Affecteds & Carriers	45/156(.288)	17.1 (8.3, 33.0)
D9S63–ASS–D9S64	Probands		13.1 (8.4, 23.5)
	All affecteds		13.4 (8.7, 24.0)
	Affecteds & carriers		12.3 (7.9, 22.1)

[a]SI, Support interval based on 1-lod unit support intervals for the recombination fractions.

due to the greater information on marker to marker linkage than disease to marker linkage. The disease locus is presumed to be very tightly linked to the three proximal loci.

We also considered the *D9S64* locus, distal to *ASS*, with a recombination fraction (support interval) to *ASS* of 0.023 (0.012,0.042). We combine the '2' and '10' alleles and consider them as a single allele for this analysis.

Our attempts to date the *DYT1*^{AJ} mutation are shown in Table 5. Consistent with the observed strong linkage disequilibrium over a sizeable genetic distance, the mutation is estimated to be of recent origin. *ASS* gives an estimate of about nine generations, with a support interval of five to 22 generations. The more distant locus, *D9S64*, gives larger estimates—about 17 generations using all affecteds and carriers. Even though there is a difference in the estimates based on these two loci, their support intervals, based on uncertainty of the underlying recombination fractions, show a large region of overlap (8.3 to 22.4 generations). The estimate based on both loci simultaneously is about 12 to 13 generations, with a support interval of about 8 to 22 generations. Assuming an average generation length of 25 years, the time since origin would be about 300 years, with a support interval of 200 to 550 years. Assuming an average year of birth of 1950 for all affecteds and carriers in these pedigrees, we can then place the approximate year of origin for the *DYT1*^{AJ} mutation as 1650, with a range of 1400 to 1750. In any event, these data provide compelling evidence for a recent origin for this mutation, which should also have implications for its distribution within the Ashkenazi population.

Geographic origin of carriers

A number of authors have attempted to trace the geographic origin of Ashkenazi patients with ITD. For example, Zeman and Dyken[12] noted that "the highest concentration has been found in Southern Lithuania and the adjacent areas of Byelorussia and Poland. The Ashkenazi Jews from this region account for 56% of those recorded cases on whom pertinent information was available." Eldridge[16] considered ITD to be recessive in the Ashkenazi population, and therefore traced ancestry on both sides of the family. Among 116 grandparents of cases, he found only a slight excess from the northern part of the Pale (such as Lithuania and surrounding regions). It is interesting that several of the first Ashkenazi cases described at the turn of the century came from Lithuania[27].

Because we have identified a unique haplotype associated with the vast majority of early-onset ITD cases, we can identify the side of the family, and in some cases the oldest living ancestor, who carries this mutation-bearing haplotype, even though such individuals are often unaffected. Hence, we can provide a far more specific geographic analysis for mutation carriers.

We were able to identify at least country or countries of origin in Europe for the oldest carrier in 44 families (Table 6). Many individuals traced their ancestry to the historic Pale of settlement, roughly defined as the western part of Russia, Lithuania and Poland, where Jews were confined to live by the Russian czars by laws imposed in 1791 and 1835 (see Fig. 1). For 19 cases, Russia was the single country of origin, in seven cases Poland was; in five cases, the origin was either Russia or Poland. For an additional two cases, the origin was either Russia or a Central or Western European country, for three cases either Poland or a Central or Western European country, and for 6 cases the origin involved only Central or Western European

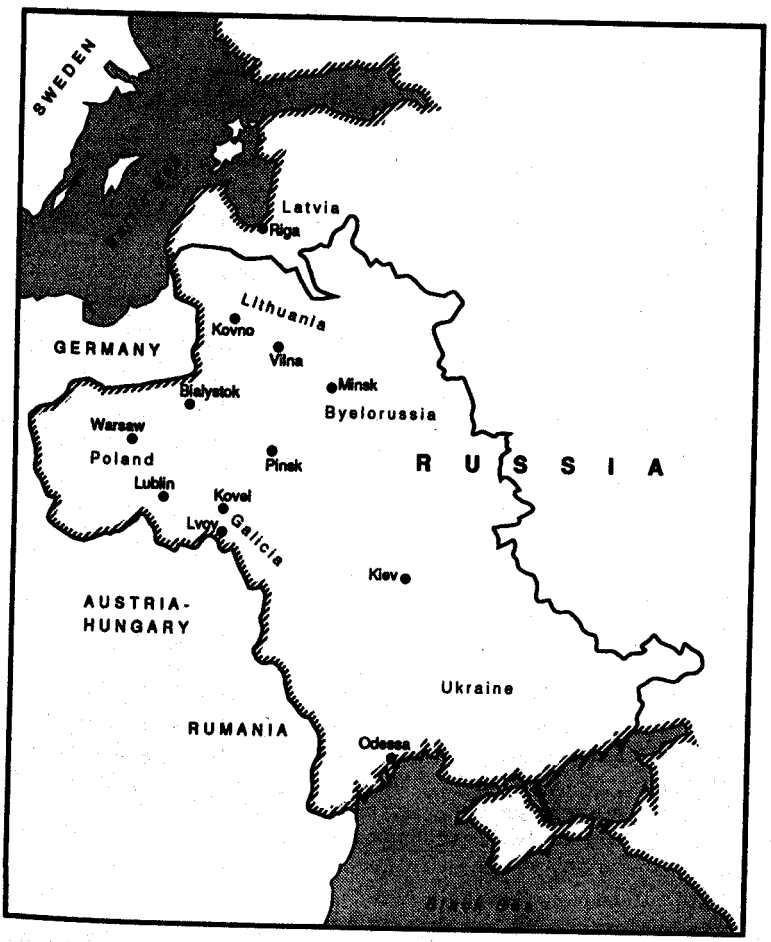


Fig. 1 The Jewish Pale Settlement in Russia. Adapted from M. Gilbert, *Atlas of Jewish History*, Dorset Press, 3rd edition, 1985.

Table 6 Geographic origins of carriers

Country	Region	City or town	N
Russia	Lithuania	Vilna	1
		Riga	1
		Kovna	1
		Bialystok	1
		?	2
	Byelorussia	Minsk	2
		Minsk/Pinsk	1
		Kopost	1
	Ukraine	Kiev	1
		Odessa	1
	?		7
Poland	Galicia	Kovel	1
		Lvov	1
	?		7
Russia/Poland	?		5
Russia/Other[a]	?		2
Poland/Other	?		3
Other			6

[a]Other includes Romania, Hungary, Czechoslovakia, Austria-Hungary, Germany, France

countries. For those carriers from Russia or Poland for which a specific region could be identified, 6 were from Lithuania, 4 from Byelorussia, 2 from the Ukraine, and 2 from Galicia. These results are certainly consistent with a preponderance of carriers deriving from the northern part of the Pale, namely Lithuania and Byelorussia, as noted previously[12]. Hence, the likely first appearance of the *DYT1*^{AJ} mutation was in Lithuania or Byelorussia, from which it spread to the rest of the Pale through migration over the centuries. In further support of this thesis is the observation that the few oldest carriers we could identify, those born in the early to mid 1800s, all came from Lithuania or Byelorussia.

Estimating the gene frequency

Estimates of the prevalence of ITD in the Ashkenazi population have varied from 1/36,000 (ref. 12), 1/17,000 (ref. 16), and 1/23,000 (ref. 17). Furthermore, Zilber *et al.* [17] estimated the prevalence to be 1/15,000 for Ashkenazi Jews of Eastern European origin (and 1/23,000 overall). These values are likely to be severe underestimates. The ITD cases included were only the most severely affected, representing a small percentage of the total patient population. Furthermore, many ITD cases are undiagnosed.

We attempted to derive a more realistic estimate of the prevalence of ITD in the Ashkenazi population — more specifically, an estimate of the frequency of the unique mutation described above — using the patient database of the Dystonia Clinical Research Center (DCRC) at Columbia Presbyterian Medical Center. We used the same definition of disease as described above - namely idiopathic (non-secondary) dystonia with age of onset at or before age 28, and symptom onset in a limb, trunk or neck. We restricted our analysis to the New York metropolitan area encompassed by the four telephone area codes 212, 718, 914 and 516; this includes all of New York City, Nassau and Suffolk County, Westchester, Rockland, Orange and Putnam counties.

Based on the patient database, there were 73 affected individuals meeting the above criteria living in the area in 1990. From the American Jewish Yearbook[28], the total number of Ashkenazi Jews (assuming 95% of the Jews are Ashkenazi) living in this area in 1990 is estimated at 1,466,800. Hence, the crude point prevalence rate is 4.98 per 100,000, or 1 in 20,093. This prevalence is clearly an underestimate, because it only includes patients examined at the DCRC and the denominator is not age-adjusted, and includes children too young to have had onset of the disease.

An accurate estimate of true prevalence is difficult to obtain. However, in a prior family study[14], we noted that half of the affected individuals examined had not had a prior diagnosis of ITD. This same observation has also been made by others[29]. Thus, it is safe to double the prevalence estimate to 1/10,047. The median age at onset of ITD in our sample is around 12 years of age. At least 15% of the population is under that age[30], so again it is safe to divide the estimated prevalence by 0.90, giving 1/9,042. To obtain the gene frequency from this prevalence estimate, we divide by the penetrance (0.30) to get a carrier frequency of 1/2,713. Of the 59 carriers in Table 4, only 54 appear to carry the *DYT1*^{AJ} mutation, or 91.5%. Hence, the carrier frequency for *DYT1*^{AJ} is 1/2,965, corresponding to a gene frequency of 1/5,929~1/6,000.

This estimate requires the assumption that all diagnosed cases of ITD in the New York Metropolitan area have been seen at the DCRC. In fact, it is likely that not even a majority of these cases have been seen there. If we assume that one-third of all ITD cases are seen at the DCRC, the gene frequency estimate becomes 1/2,000. Thus, the true gene frequency is likely to lie in the range of 1/6,000 to 1/2,000.

Implications for Ashkenazi history

There is little doubt that the Ashkenazi population has gone through several severe bottlenecks as well as growth periods in its history. The remarkable array of genetic diseases that appear at high frequency in this population[1] is a legacy to this checkered demographic past. The likely explanation for the high frequency of otherwise rare autosomal recessive diseases has been debated for decades; while some have argued for a selective advantage for heterozygous carriers, others have argued for the predominant effect of genetic drift.

We believe that our results for ITD in this population offers a distinct opportunity to address this issue, as ITD is an autosomal dominant disease. The usual consideration of heterozygote advantage, applied for recessive disorders, does not apply here. Dominant mutations are directly subject to selection. For ITD, there is also selection against those who carry the *DYT1*^{AJ} mutation; severely affected individuals tend not to reproduce. However, the expression of this gene is highly variable, with only a small proportion of carriers having a severe form. We had previously estimated the overall penetrance at 30%[14,18]; this includes both mild and severe cases. Perhaps only a third of those with clinical symptoms would be categorized as severely affected, or 10% of gene carriers. Hence, while selection is operating against carriers, it is relatively mild due to the low probability of a severe and debilitating manifestation. This fact, as well as its recent origin, probably explains the relatively high frequency of a potentially severe disease in this population.

An intriguing finding is the apparently high gene frequency in this population, in conjunction with its recent origin. If we take account of selection operating against this mutation, its frequency in earlier generations would be even higher. At face value, this observation would lead to the conclusion that the current day Ashkenazi population likely derives from a relatively small group of founders from the recent past—perhaps 300 years ago, but no more than probably 500 years ago. Is this conclusion consistent with the known history of this population?

Detailed accounts of the number of Ashkenazi Jews living in Central and Eastern Europe in the Middle Ages and after do not exist, although several authors have speculated as to these numbers. Perhaps the best census analysis of the Ashkenazi Jews over the past 800 years is given by Weinryb[31]. The earliest reliable estimate for the number of Jews in Eastern Europe in this period is from the census of 1765, which calculated 430,000 Jews in Poland, and another 130,000 in the Ukraine[31]. It is also known that by 1900, there were 5,000,000 Jews living in the Pale, a population expansion of approximately 10-fold in only 135 years. Assuming an average generation of 25 years would give a growth rate of about 1.5-fold per generation. Assuming a similar growth rate for the 100 years prior to 1765 would give a population size of about 100,000 in 1665. This figure is not inconsistent with other facts about this period; for example, Weinryb[31] notes that "the century after 1667 seems to have been a time of considerable Jewish population growth."

Estimates of population size from the beginning of the 16th century to the mid 17th century are highly speculative, although it is generally assumed that this period was also one of considerable growth. Weinryb[31] suggests that there were 10–12,000 Jews living in Poland (and Lithuania) in 1500, which expanded to some 80–100,000 by the end of the century, and about 150–170,000 half a century later. The seminal event in terms of the Jewish population at this point in time was the Cossack massacres in 1648. Although all parts of Jewish settlement in Eastern Europe were affected, from Lithuania to the Ukraine, the Jewish population of the Ukraine was most devastated, with the total destruction of many Jewish communities. The overall effect of the Cossack uprising on the Jewish population, in terms of number killed, is controversial. Some report that the majority of the population was destroyed; for example, Fraikor[8] suggests only about 20,000 remained. Weinryb[31] contends that such claims are exaggerated, although perhaps as much as 25% of the existing population was killed. This would place the remaining Jewish population in the Pale region in the mid 1600s at approximately 100,000, consistent with the number observed at the 1765 census and the apparent growth rate during this period.

Our analysis of linkage disequilibrium with the *DYT1*[AJ] mutation suggests its first appearance to be around 1650, with a lower support limit of around 1400. Estimates of the size of the Jewish population in this period range from about 10,000 to 100,000. Thus, when the mutation first appeared, the carrier frequency would be only 1/10,000 to 1/100,000. Then how is it possible to explain such a high carrier frequency in today's Ashkenazi population?

We believe the answer lies in consideration of the specific dynamics of population growth among the Ashkenazim during this period of time, namely that not all individuals reproduced equally. Weinryb[31] comments: "more children survived to adulthood in affluent families than in less affluent ones. A number of genealogies of business leaders, prominent rabbis, community leaders, and the like—generally belonging to the more affluent classes—show that such people often had four, six, sometimes even eight or nine children who reached adulthood. On the other hand, there are some indications that poorer families tended to be small ones. It should also be added that overcrowding, which favors epidemics, was more prevalent among the poorer classes ... In short, the number of children surviving among Polish Jews seems to have varied considerably from one social level to another."

The tremendous population growth taking place in the Ashkenazi population through the 16th to 19th century did not occur uniformly, but was probably restricted to the wealthier individuals who had large numbers of children who survived. Thus, historically, the most important determinant of fertility in the Ashkenazi population was social class. We speculate that disease genes found at high frequency in the current Ashkenazi population were fortuitously carried by those of higher social class.

Thus, the current day Ashkenazi population primarily descends from a much smaller group of individuals (the wealthier) in the 17th to 18th centuries than the total population of the time. The number of individuals from whom current Ashkenazi Jews descend (i.e. effective population size) could be as low as a few thousand or less as recently as 500 years ago.

The enormous expansion in the Ashkenazi population in the 16th to 19th centuries has few analogies in the present day, with the possible exception of the French Canadians, another group known for a range of genetic diseases at high frequency. We speculate that founder effects resulting from the dynamics of population growth in the Jewish population, as described above, may well explain most, if not all the genetic diseases observed at high frequency in the Ashkenazi population today.

This conclusion can be tested further by examining linkage disequilibrium for the other genetic diseases which occur at a high frequency in the Ashkenazi population. Our contention would be that most, if not all, could be traced back to a similar time point in history. This would also explain the dramatic difference between the Ashkenazi and non-Ashkenazi (Sephardic) Jewish populations with regard to the specific genetic diseases found at high frequency. Preliminary data for three other Ashkenazi (recessive) diseases—familial dysautonomia[32], Bloom syndrome[33] and familial hyperinsulinism (M.A. Permutt, personal communication)—all show strong linkage disequilibrium over a large genetic distance, supporting a recent origin for these diseases as well.

Methodology

Study subjects. Subjects were ascertained from a computerized database of patients seen by members of the Movement Disorders Group at Columbia Presbyterian Medical Center and from the Dystonia Medical Research Foundation (by advertisement). Since ITD appears to be a heterogeneous disorder, we restricted our study to families in which at least one (or the only) affected had onset of symptoms prior to age 28 and dystonia first involved a limb, the trunk or neck. Our previous family study[14] as well as the age of onset distribution for ITD in the Ashkenazim[11] suggest these are reasonable criteria for subgrouping. Further genetic analysis also supports this separation[25].

All affected individuals have Ashkenazi Jewish ancestry. The vast majority are 100% Ashkenazi, while a few cases have a Sephardic or non-Jewish parent or grandparent(s). In no case was the disease inherited through a non-Ashkenazi ancestor. All consenting family

members were examined by neurologists experienced in diagnosing dystonia. On-site examination included evaluation of all body regions at rest and action; examinations were also videotaped for subsequent blinded review by at least two non-examining neurologists to confirm the diagnosis. These methods have been described in detail elsewhere[14,20]. Only individuals classified as having definite dystonia were included as affected for these analyses. Venous blood samples were obtained from all participating family members for subsequent DNA analysis (described below).

A total of 59 families were ascertained; 39 are multiplex, containing at least two affected relatives, while the remaining 20 are simplex, containing only a single affected. Some of these multiplex families have been described previously[20,21]. For genetic marker analyses, we supplemented our family material with the Venezuelan Reference Pedigree (VRP)[34] to determine distances between markers.

DNA analysis. DNA was extracted from lymphoblast lines[35] or whole blood according to methods described elsewhere[36]. Analysis of $(GT)_n$ polymorphisms was carried out on genomic DNA by using oligonucleotide primer pairs described in ref 22.

Genetic marker map. Six microsatellite loci from chromosome 9q34 were studied. For *D9S62*, two microsatellite polymorphisms exist, which we call *D9S62a* (*32/37* in ref. 22) and *D9S62b* (*1290/1301* in ref. 37). The order of the six microsatellite loci is as follows: centromere–*D9S62a*–*D9S62b*–*D9S63*–*ASS*–*ABL*–*D9S64*–qter. The three loci *D9S62a, D9S62b* and *D9S63* are very tightly linked, with no observed recombinant between them in at least several hundred meioses; hence, the genetic distance between these loci is quite small. Similarly, the loci *ASS* and *ABL* are very tightly linked, with no observed recombinant in several hundred meioses.

Linkage analysis was performed between the proximal cluster of three loci (*D9S62a, D9S62b, D9S63*) and the cluster of two loci *ASS/ABL*. For both clusters haplotypes were formed and assumed to constitute a single locus. Linkage analysis was also performed between *ASS/ABL* and *D9S64*. The computer program LIPED was used for this analysis[28]. The maximum likelihood estimate for the recombination fraction, θ_{max}, between the proximal cluster and *ASS/ABL* is 0.018 (maximum lod=99.72), with a 1-lod unit support interval of 0.007 to 0.033; θ_{max} between *ASS/ABL* and *D9S64* is 0.023 (maximum lod=88.43), with a 1-lod unit support interval of 0.012 to 0.042.

Linkage analyses. We updated our previous linkage analysis[20] between *DYT1* and chromosome 9q34 markers in the Ashkenazi population by supplementing our original 12 families with 23 additional informative multiplex families and inclusion of the newly discovered microsatellite markers proximal and distal to *ASS*. The assumptions underlying the present analysis are the same as those employed in our previous analysis[20], namely autosomal dominant inheritance of a rare allele with penetrance 30% by age 35. A linear age of onset correction for penetrance was employed, ranging from a low of 0% risk before age 5 to a final risk of 30% by age 35. All linkage calculations were performed using the computer program LIPED, Ver. 3 (ref. 38). For this analysis, we again combined loci that are extremely close into single haplotypes for analysis; specifically, the loci *D9S62a, D9S62b* and *D9S63*, were merged into a single haplotype, as were *ASS* and *ABL*.

Allelic association and linkage disequilibrium. We previously reported allelic association between alleles at the *ASS* and *ABL* loci and *DYT1* in the Ashkenazi population[21]. Here we update these findings with an enlarged sample and the new microsatellite markers. From the 39 multiplex families, we were able to identify a disease bearing chromosome in 36 (an example of a non-informative family is one that contains only two affected sibs who share chromosome 9q34 regions from both parents identical by descent). In the 36 families, there were 5 cases of recombination between *DYT1* and some of the markers. For these cases, we used the predominant (or oldest) haplotype occurring in the family. From our total of 59 families, we also constructed a collection of control chromosomes from individuals marrying into the families who had no direct relationship to the affecteds of the family; we also required that these individuals be of 100% Ashkenazi Jewish origin. We supplemented this sample with other unrelated Ashkenazi individuals from a genetic study of dysautonomia[32]. In this fashion we were able to obtain a total of 306 Ashkenazi control chromosomes for this analysis. Statistical comparison of allele frequencies between disease and normal chromosomes was based on a chi-square test for a 2×2 table, with alleles classified into two groups, one for the associated allele, and all others combined into a single group.

To assess the degree of linkage disequilibrium, the most appropriate measure to employ is $\delta=(p_D - p_N)/(1-p_N)$, where p_D is the frequency of the associated allele on disease chromosomes and p_N the frequency of the same allele on normal chromosomes. The parameter δ was originally defined by Bengtsson and Thomson[39] and is also described further elsewhere[40]. δ represents an estimate of the proportion of disease chromosomes bearing the original associated allele. Here we assume that a disease chromosome not bearing an original associated allele (either due to a different mutation or through recombination) may still carry that allele (by "state") with probability p_N. Since the proportion of chromosomes carrying the same progenitor mutation is the same across all loci, differences in δ across loci should primarily represent effects of recombination (although mutation at the marker loci may also be an issue). Hence, δ can be used to determine the most likely location of the disease locus among a collection of linked loci (Risch and Devlin, manuscript submitted).

Calculation of disease chromosome carrier probability for singleton cases. For the phased singleton cases, there are two haplotypes, which we label A and B. Let M_A represent the marker information on chromosome A, and similarly for M_B. Let D represent the disease mutation-bearing chromosome, so that $P(M_A|A=D)$ represents the probability of the marker information on chromosome A given that A carries the disease mutation, and similarly for B. Then the posterior probability that A carries the disease mutation is simply R/(R+1), where

$$R = \frac{P(M_A|A=D)}{P(M_A|A\neq D)} \times \frac{P(M_B|B=D)}{P(M_B|B\neq D)}$$

For the haplotype "2-8-16" and/or "16-12", the frequency on disease chromosomes was 32/36 versus 3/306 on normal chromosomes. Hence, if A carries one of these haplotypes, the first ratio is (32/36) ÷ (3/306)=90.7. In every case, the second ratio is greater than 1, because all other haplotypes were more common on control chromosomes than disease chromosomes. Hence, R≥90.7, so the posterior probability that A carries the disease mutation is greater than 90.7/91.7=99%.

Dating the mutation. A simple formula relates the probability (Q) that a disease-mutation-bearing chromosome does not carry a progenitor marker allele as a function of the number of generations (g) since the appearance of the mutation, the frequency for that allele on normal chromosomes (p_N), and the recombination fraction (θ) between the disease and marker locus: $Q=(1-(1-\theta)^g)(1-p_N)$. Hence, if $\hat{Q}$ represents the observed frequency of disease chromosomes not carrying the progenitor marker allele, an estimate of the number of generations since the appearance of the mutation is

$$\hat{g} = \frac{\log[1-\hat{Q}/(1-p_N)]}{\log(1-\theta)} \times \frac{\log(\delta)}{\log(1-\theta)} \qquad (1)$$

where δ is the linkage disequilibrium measure defined previously. This is essentially the same as formula 7 of ref. 24, and a similar formula was given in ref. 25.

This formula applies to the locus closest to the disease locus. For a second marker locus next to the first marker locus, a somewhat more complicated analysis is required. For that we consider only disease chromosomes that carry the progenitor allele at the first marker locus. Among such chromosomes, we determine the proportion that do not carry the progenitor allele at the second marker locus. This proportion can be substituted for Q in formula (1) (where θ is now the recombination fraction between marker loci 1 and 2) and a similar δ calculated to obtain another estimate of g from the second marker locus. This estimate will not be entirely accurate, however, because some chromosomes bearing a progenitor allele at locus 1 will actually be recombinant for this locus, but by chance have recombined with an allele of the same type. If the associated allele at marker 1 has a low frequency on normal chromosomes (p_N), then this possibility has little influence on the estimate of g.

It is also possible to use both marker loci simultaneously to estimate g. Suppose that allele "1" is the progenitor allele for locus 1 and "1" is the progenitor allele for locus 2. Considering loci 1 and 2 simultaneously, there are 3 haplotype classes of interest. Let a represent the haplotype "1-1" at loci 1 and 2; let b represent the class of haplotypes "1-x" for loci 1 and 2, where "x" is any allele other than a "1"; let c represent the class of haplotypes "x-y", where again "x" is any allele other than "1" at locus 1 and "y" is any allele at locus 2. Class c represents cases of recombination between DYT1 and locus 1; class b represents cases of recombination between locus 1 and locus 2 plus cases of recombination between DYT1 and locus 1, where the recombined chromosome still carries allele "1" at locus 1 by chance; and class a represents all other cases. Let Θ_1 be the recombination fraction from the disease locus to marker locus 1, and Θ_2 the recombination fraction between marker loci 1 and 2. Let $p_{1.}$ be the population frequency of allele "1" at marker locus 1, $p_{.1}$ the population frequency of allele "1" at marker locus 2, and p_{11} the population frequency of haplotype "1-1". Then, employing the same assumptions as for the derivation of formula (1), we obtain the probabilities for the three haplotype classes c, b, and a as

$$P(c) = (1-(1-\Theta_1)^g)(1-p_{1.}) \qquad (2)$$
$$P(b) = (1-(1-\Theta_1)^g)(p_1-p_{11}) + (1-\Theta_1)^g(1-(1-\Theta_2)^g)(1-p_{.1})$$
$$P(a) = (1-(1-\Theta_1)^g)p_{11} + (1-\Theta_1)^g[(1-\Theta_2)^g + (1-(1-\Theta_2)^g)(p_{.1})]$$

If n_c, n_b and n_a represent the observed number of haplotypes of class c, b and a respectively, then an estimate of g can be obtained by maximizing the log likelihood

$$\ln L = n_c \ln P(c) + n_b \ln P(b) + n_a \ln P(a)$$

using the formulas (2).

Estimates of g derived using the formulas above are subject to two sources of statistical fluctuation - (1) sampling variance on the proportions of recombinant-type disease chromosome haplotypes (e.g. on the numbers n_a, n_b and n_c above), and (2) uncertainty in the estimates of the recombination fractions. The first is difficult to calculate, as the observed chromosomes are not independent events and hence a binomial variance is inappropriate. Here we focus on the second source of variance, uncertainty in the θ values. We do so by using the lod score tables to define a 1-lod unit support interval[41] for each of the recombination fractions; for the final analysis using both marker loci, we use a joint-lod unit support surface for the two recombination fractions. Using the upper and lower bound of the support intervals (or by obtaining extreme values on the support surface) we obtain a support interval on our estimate of g.

Acknowledgements

We thank Kathleen DeSanctis for technical assistance, Kenneth Lange and Bronya Keats for helpful advice, and the numerous families who contributed to this study. This work was supported by grants from the Dystonia Medical Research Foundation (to SF, SB, XB and PK) and NIH grants NS26656 (to SF and SBB), NS28384 (to XOB), and HG00348 (to NR).

Received 2 November 1994; accepted 4 January 1995.

1. Goodman, R.M. *Genetic Disorders among the Jewish People.* (Johns Hopkins University Press, Baltimore, 1978).
2. Motulsky, A.G. Possible selective effects of urbanization on Ashkenazi Jews. In *Genetic Diseases among Ashkenazi Jews*, eds Goodman, R.A. & Motulsky, A.G. 301–314 (Raven Press, New York, 1979).
3. Chakravarti, A. & Charaborty, R. Elevated frequency of Tay-Sachs disease among Ashkenazi Jews unlikely by genetic drift alone. *Am. J. hum. Genet.* **30**, 256–261 (1978).
4. Yokoyama, S. Role of genetic drift in the high frequency of Tay-Sachs disease among Ashkenazi Jews. *Ann. hum. Genet.* **43**, 133–136 (1979).
5. Rotter, J.I. & Diamond, J.M. What maintains the frequencies of human genetic diseases? *Nature* **329**, 289–290 (1987).
6. Jorde, L. Genetic diseases in the Ashkenazi population: Evolutionary considerations. In *Genetic Diversity Among Jews: Diseases and Markers at the DNA Level*. eds Bonne-Tamir, B. & Adam, A., 305–318 (Oxford University Press, New York, 1992).
7. Diamond, J.M. Jewish lysosomes. *Nature* **368**, 291–292 (1994)
8. Fraikor, A.L. Tay-Sachs disease: genetic drift among the Ashkenazim Jews. *Soc Bio* **24**, 117–134 (1977).
9. Wagener, D., Cavalli-Sforza, L.L. & Barakat, R. Ethnic variation of genetic disease: roles of drift for recessive lethal genes." *Am. J. hum. Genet.* 30, 262–270 (1978).
10. Chase, G.A. & McKusick, V.A. Controversy in human genetics: founder effect in Tay-Sachs disease. *Am. J. hum. Genet.* **24**, 339–340 (1972).
11. Fahn, S. Generalized dystonia: concept and treatment. *Clin. Neuropharmacol.* **9**, 537–548 (1986.).
12. Zeman, W. & Dyken, P. Dystonia musculorum deformans: clinical, genetic and pathoanatomical studies. *Psychiatr. Neurol. Neurochir.* **70**, 77–121 (1967).
13. Korczyn, A.D. *et al.* Torsion dystonia in Israel. *Ann. Neurol.* **8**, 387–391 (1980).
14. Bressman, S.B. *et al.* Idiopathic dystonia among Ashkenazi Jews: evidence for autosomal dominant inheritance. *Ann. Neurol.* **26**, 612–620 (1989).
15. Eldridge, R. The torsion dystonias: literature review: genetic and clinical studies. *Neurology* **20**, 1–78 (1970).
16. Eldridge, R. & Gottlieb, R. The primary hereditary dystonias: genetic classification of 768 families and revised estimate of gene frequency, autosomal recessive form, and selected bibliography. *Adv. Neurol.* **14**, 457–474 (1976).
17. Zilber, N. *et al.* Inheritance of idiopathic torsion dystonia among Jews. *J. Med. Genet.* **21B**, 13–20 (1984).
18. Risch, N. *et al.* Segregation analysis of idiopathic torsion dystonia in Ashkenazi Jews suggests autosomal dominant inheritance. *Am. J. Hum. Genet.* **46**, 533–538 (1990).
19. Ozelius, L. *et al.* Human gene for torsion dystonia located on chromosome 9q32-34. *Neuron* **2**, 1427–1434 (1989).
20. Kramer, P.L. *et al.* Dystonia gene in Ashkenazi Jewish population is located on chromosome 9q32-34. *Ann. Neurol.* **27**, 114–120 (1990).
21. Ozelius, L. *et al.* Strong allelic association between the torsion dystonia gene (DYT1) and loci on chromosome 9q34 in Ashkenazi Jews. *Am. J. Hum. Genet.* **50**, 619–628 (1992).
22. Kwiatkowski, D. *et al.* Construction of a GT polymorphism map of human 9q. *Genomics* **12**, 229–240 (1992).
23. Bressman, S.B. *et al.* Idiopathic torsion dystonia in Ashkenazi Jews: clinical characterization of a founder mutation. *Ann. Neurol.* **36**, 771–777 (1994).
24. Serre, J.L. *et al.* Studies of RFLP closely linked to the cystic fibrosis locus throughout Europe lead to new considerations in population genetics. *Hum. Genet.* **84**, 449–454 (1990).
25. Cox, T.K. *et al.* Mapping of the cystic fibrosis gene using putative ancestral recombinants. *Am. J. Hum. Genet.* **45**, A136 (1989).
26. Hastbacka, J. *et al.* Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genet.* **2**, 204–211 (1992).
27. Schwalbe, W. Eine eigentumliche tonische Krampffoum mit hysterischen Symptonen. Medicin und chirurgie [Dissertation]. (Universitats-Buchdrukerei von Gustav Schade, Berline, 1908).
28. American Jewish Yearbook (The Jewish Publication Society of America, Philadelphia, 1992).
29. Fletcher, N.A., Harding, A.E. & Marsden, C.D. A genetic study of idiopathic torsion dystonia in the United Kingdom. *Brain* **113**, 379–395 (1990).
30. U.S. Bureau of the Census. *Statistical Abstract of the United States: 1993* (113th edition). (Washington, D.C., 1993).
31. Weinryb, B.D. *The Jews of Poland. A social and Economic History of the Jewish Community of Poland from 1100 to 1800.* (The Jewish Publication Society of America, Philadelphia, 1972).
32. Blumenfeld, A. *et al.* Localization of the gene for familial dysautonomia on chromosome 9 and definition of DNA markers for genetic diagnosis. *Nature Genet.* **4**, 160–164 (1993).
33. Ellis, N.A. *et al.* Linkage disequilibrium between the *FES, D15S127,* and *BLM* Loci in Ashkenazi Jews with Bloom Syndrome. *Am. J. hum. Genet.* **55**, 453–460 (1994).
34. Haines, J.L. *et al.* A genetic linkage map of chromosome 17. *Genomics* **8**, 1–6 (1990).
35. Anderson, M. & Gusella, J. Use of cyclosporin A establishing Epstein-Barr virus-transformed human lymphoblastoid cell lines. *In Vitro* **29**, 856–858 (1984).
36. Breakefield, X. *et al.* Linkage analysis in a family with dominantly-inherited torsion dystonia: exclusion of the pro-opiomelanocortin and glutamic acid decarboxylase genes and other chromosomal regions using DNA polymorphisms. *J. Neurogenetics.* **3**, 159–175 (1986).
37. Henske, E. *et al.* A high-resolution linkage map of human 9q34.1. *Genomics* **17**, 587–591 (1993).
38. Ott, J. A computer program for linkage analysis of general human pedigrees. *Am. J. Hum. Genet.* **28**, 528–529 (1976).
39. Bengtsson, B.O. & Thomson, G. Measuring the strength of associations between HLA antigens and diseases. *Tissue Antigens* **18**, 356–363 (1981).
40. Thomson, G. A review of theoretical aspects of HLA and disease associations. *Theor. Pop. Biol.* **20**, 168–208 (1981).
41. Ott, J. *Analysis of Human Genetic Linkage.* (The Johns Hopkins University Press, Baltimore. 1985).