



ARTICLE

# Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo

Teppo Varilo<sup>1</sup>, Maris Laan<sup>2,3</sup>, Iiris Hovatta<sup>1</sup>, Victor Wiebe<sup>2</sup>, Joseph D Terwilliger<sup>4</sup> and Leena Peltonen<sup>1</sup>

<sup>1</sup>Department of Human Molecular Genetics, National Public Health Institute, Helsinki, and Department of Medical Genetics, University of Helsinki, Finland; <sup>2</sup>Max-Planck-Institute for Evolutionary Anthropology and Institute of Zoology, University of Munich, Germany; <sup>3</sup>Institute of Molecular and Cell Biology, University of Tartu, Estonia; <sup>4</sup>Department of Psychiatry and Columbia Genome Center, Columbia University, New York, USA

Linkage disequilibrium (LD), non-random association of alleles at closely linked chromosomal loci, has been used as a tool in the identification of disease alleles, and this has led to an improved understanding of pathology in many monogenic Mendelian human diseases. We are currently moving from the mapping and identification of monogenic disease loci to attempts at identifying loci involved in predisposition to multifactorial diseases. In the selection of ascertainment strategies in the studies of these complex diseases, the extent of background LD in different populations is an important consideration. Here, we compare the extent of LD among the alleles of linked loci in a randomly ascertained sample of individuals from the Finnish population and a set of individuals ascertained from the region of Kuusamo, a small sub-population, founded some 13 generations ago, which has experienced very little subsequent immigration. Thirty-three microsatellite loci were genotyped in chromosomal regions on 13q, 19q, 21q, Xq, and Xp. The genetic diversity of these loci was determined separately in the general Finnish sample and in the Kuusamo sample. The X-chromosomal loci are characterised by higher levels of LD in the samples from Kuusamo than in the much larger (and older) general population of Finland, whereas in alleles of autosomal loci very little LD was seen in either of these two samples. *European Journal of Human Genetics* (2000) 8, 604–612.

**Keywords:** linkage disequilibrium; isolated populations; expanded populations; Finnish population; multifactorial diseases; microsatellite markers

## Introduction

The quantification of LD on randomly ascertained chromosomes is an important step in evaluating the potential for gene mapping in different populations. Populations with an unusual history characterised by a high level of genetic drift due to a restricted number of founders which subsequently expanded, and population bottlenecks, reducing the genetic diversity of a population, and/or small population size can provide good conditions for identification of disease alleles.

The existence of high levels of LD over large chromosomal regions is characteristic of populations with reduced haplotype diversity. Such reductions in haplotype diversity are indicative of reduced allelic diversity resulting from the same demographic events, though the reductions in haplotype diversity can be demonstrated with greatly reduced sample sizes, due to the increased informativeness of haplotypes over alleles of single loci as measures of drift. This may lead to greater levels of LD over larger chromosomal regions between anonymous marker loci and disease alleles (true positives), although there may likewise be chromosomal segments shared by chance among random samples (false positives) with higher frequency as well. Population history studies can therefore be of use in both linkage and LD-based mapping projects.<sup>1</sup> Gene mapping in rare monogenic diseases using

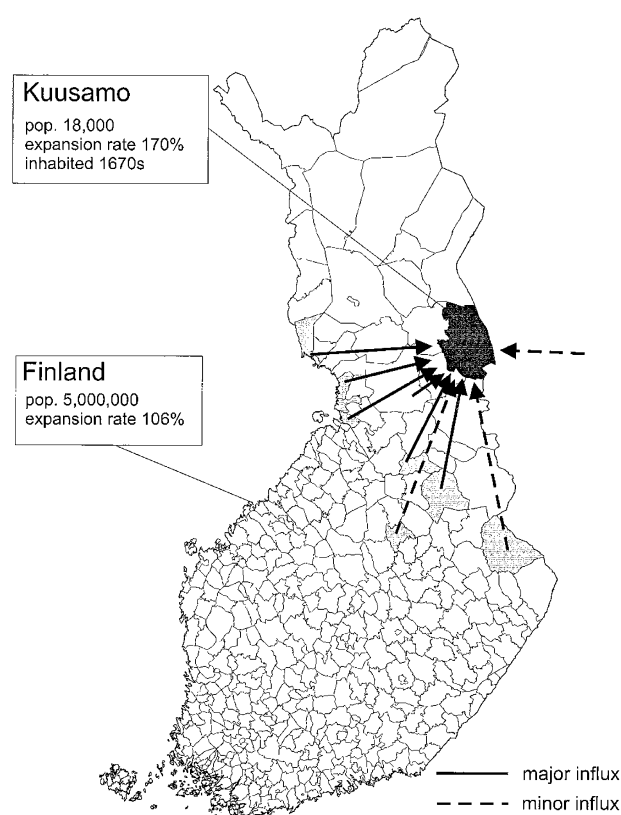
Correspondence: Leena Peltonen, MD, PhD, Professor and Chair, Department of Human Molecular Genetics, UCLA School of Medicine, 695 Charles E Young Drive South, Los Angeles, CA, USA 90095-7088. Tel: 310 794 5631; Fax: 310 794 5446; E-mail: Lpeltonen@mednet.ucla.edu  
Received 17 September 1999; revised 4 February 2000; accepted 24 February 2000

LD has been particularly successful in rapidly expanded isolated populations (with a small effective population size relative to the disease allele frequency) like that of Finland. Rare disease alleles that increased rapidly in copy number during the expansion of the population have exhibited LD with polymorphic loci at chromosomal intervals of up to 13 cM, offering advantages for initial locus identification.<sup>2</sup> For common diseases, however, the situation tends to be less clear-cut due to allelic heterogeneity, the older age of the alleles,<sup>3</sup> and difficulties in defining the disease phenotype.<sup>4</sup> Simulations and some population studies have suggested that less background LD should exist in rapidly growing populations than in populations of constant size.<sup>5–8</sup> However, the situation in smaller and younger sub-populations has not been studied previously. It should be noted that higher levels of LD are coincident with and result from reductions in haplotype diversity, which is a surrogate for allelic diversity. For this latter reason, smaller populations with extensive reductions in haplotype diversity may be expected to have fewer, more easily identifiable disease-predisposing alleles at a reduced number of loci than larger populations with extensive haplotype (and correspondingly allelic) diversity.<sup>4</sup>

The 'founding' of today's Finnish population is hypothesised to have taken place in the beginning of the first millennium AD with small migrant populations, mostly from the South,<sup>2,9–12</sup> forming settled dwellings in villages. Finland has been inhabited for about 10 000 years, since the last glacial period, but in prehistoric times, only a few thousand natives could have been supported by hunting, fishing, and gathering. Farming was introduced into the southern and south-western coastal areas some 3000–4000 years ago by Eastern Uralic immigrants hypothesised by the dual theory of Finland's habitation, which is supported by analysis of Y-chromosome haplotype variation.<sup>13,14</sup> Founder effects, genetic drift, and subsequent isolation of this population can be seen in the reduced allelic diversity, characterised by the 'Finnish disease heritage', which refers to 35 rare, mostly autosomal recessive, Mendelian disorders over-represented in Finland.<sup>10,15</sup> Causative mutations have so far been identified in 19 of the diseases and 13 more have been mapped to restricted chromosomal regions. The molecular findings in these diseases are consistent with the hypothesis of one major founder mutation responsible for the majority of cases of these diseases in Finland.<sup>2,16</sup>

Fast expansion, increasing the population size in Finland from 250 000 to the present 5 million, began in the 16th century, when groups of farmers, mostly inhabiting a small area of South Savo in south-eastern Finland, moved to the central, western and finally northern parts of the country, clearing the land by fire, and founding regional sub-isolates. Thus, within a century, the inhabited land area of Finland doubled. An excellent example of this isolation by distance is the population of Kuusamo in the wilderness of north-eastern Finland near the present-day border with Russia

(Figure 1). This area is known to have been inhabited at the end of the 17th century by small immigrant groups from the south and west, and 39 of these families survived the great famine of 1695–1697. When parish registers were established in 1718 the population in the 165 houses numbered 615. The population of Kuusamo subsequently grew to its present 18 000 and remained isolated until World War II. Our recent genealogical search has indicated that practically all inhabitants of Kuusamo born before 1939 have ancestors from these 39 families and show numerous genealogical connections.<sup>17</sup> Such sub-populations, established by only a few people, have received attention as promising resources for identifying the genetic components of complex diseases. The small population size, lack of migration, and multiple relationships between individuals in the population could reduce the genetic diversity, and lead to increased levels of



**Figure 1** The main migration movements to Kuusamo (shaded) on the map of Finland. Kuusamo was inhabited at the end of the internal migration movement mainly by families from Ostrobothnia and from South Kainuu, both groups having originated from South Savo in the 1500s. The population growth rates were calculated by  $g \ln p/f$ , where  $g$  = generations from the founding of the population,  $p$  = population size today and  $f$  = population size at foundation (we assumed  $g = 12$ ,  $p = 18\,000$ , and  $f = 80$  for the Kuusamo and  $g = 120$ ,  $p = 5\,000\,000$  and  $f = 10\,000$  for the all-Finland sample, assuming that one generation is equal to 30 years).

linkage disequilibrium (LD). Furthermore, disease alleles which might be found at higher frequency in this isolate due to drift, may allow for association and identity-by-descent (IBD) mapping of alleles involved in more common disease.<sup>1,18</sup>

This study was undertaken to address the question of the extent of LD in randomly ascertained chromosomes from such population isolates. We studied the allelic association of microsatellite loci in samples from the Kuusamo subpopulation, which was founded 300 years ago, and compared it with LD in a randomly ascertained sample of chromosomes from the general Finnish population. The three autosomal regions (Figure 2) were chosen flanking three known disease loci of

the 'Finnish disease heritage' on 13q (variant late infantile neuronal ceroid lipofuscinosis; vLINCL), 19q (congenital nephrosis; CNF) and 21q (autoimmune polyendocrinopathy-candidosis-ectodermal dystrophy; APECED), which were identified through LD extending over wide intervals around the rare disease alleles (consistent with expectations given their recent common ancestry) of 11 cM in the case of vLINCL,<sup>19</sup> 3 cM in APECED<sup>20</sup> and 2.5 cM in CNF,<sup>21</sup> respectively. We also analysed markers spanning two regions of the X chromosome in order to estimate the influence of the smaller effective population size and the smaller number of generations in which recombination could have caused the LD to decay. Our results indicate that the X-chromosomal loci have detectable allelic associations across very short genetic distances in both populations (albeit stronger in the Kuusamo sample), and that for autosomal loci only a very restricted level of background LD could be detected in either sample.

## Materials and methods

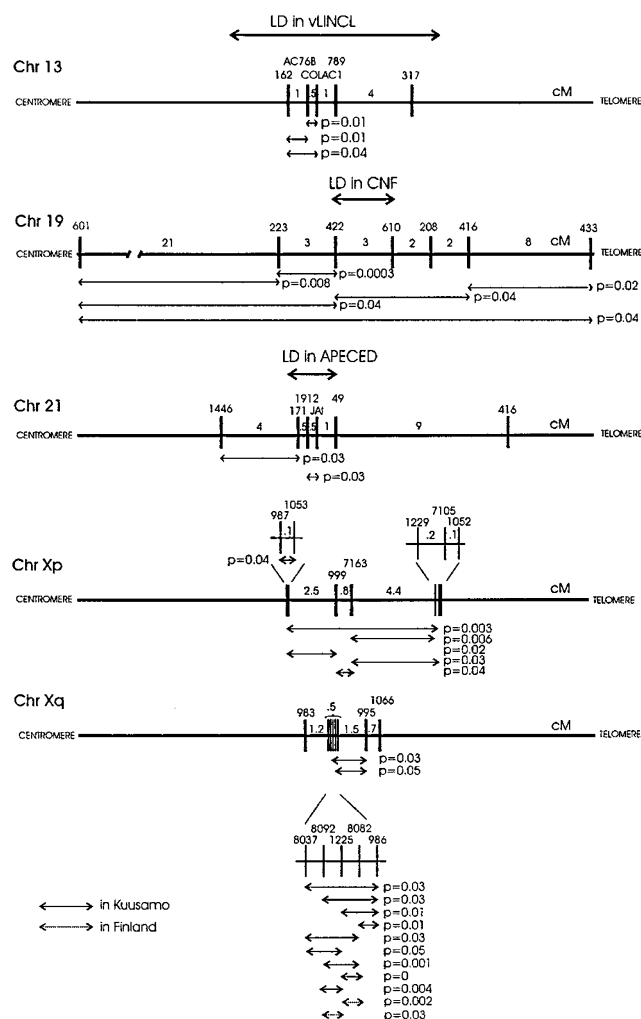
### DNA samples

The random sample from the overall Finnish population consisted of families ( $n = 64$ ), with parents and one child obtained from paternity tests from all over Finland. The Kuusamo sample consisted of families, with parents and one child affected with schizophrenia, collected from across the community for our schizophrenia study in Kuusamo ( $n = 28$ ).<sup>17</sup> X-chromosomal haplotypes were determined for 39 Kuusamo males and for 80 Finnish males originating from these sample sets. Total DNA was extracted from the leukocytes of frozen peripheral venous blood using standard procedures.<sup>22</sup>

### Microsatellite typing

The Finnish and Kuusamo families were genotyped on 13q, 19q and 21q with five, seven and six microsatellites, respectively, and the males were genotyped on Xq13 and Xp22 with eight and seven dinucleotide markers, respectively (Tables A and B on our web site). The loci and primers were either chosen from the Génethon Resource Centre<sup>23</sup> or were obtained from refined genetic maps of chromosomal regions flanking the disease alleles,<sup>20,21,24</sup> T. Klockars and P. Björnses, personal communication (1997). For the autosomal regions, the initial dinucleotide marker was chosen as near to the disease locus as possible, and the other loci studied were selected over gradually increasing distances (Figure 2). The two X-chromosomal regions differ in the density of the markers studied; on Xq13, 8 microsatellites map to a region of < 4 cM,<sup>25</sup> whereas on Xp22, 7 markers are located with 15 cM (Figure 2).

The X-chromosomal markers were typed as described.<sup>7</sup> For the autosomal markers, PCR was performed in microtiter wells in a volume of 15  $\mu$ l containing 10 ng template DNA, 5 pmol primers, 0.2 mM of each of the nucleotides, 20 mM



**Figure 2** Genetic maps and allelic associations between microsatellite loci on chromosomes 13, 19, 21, and X. Above the line the interval is indicated which reveals LD in disease chromosomes.<sup>19,21,20</sup> Below the line the interval of significant pairwise LD ( $P < 0.05$ ) seen in the general study populations of Finland and Kuusamo is given. The very long pairwise LDs seen below with weak evidence of association are expected to be due to the rate of false positives.

Tris-HCl (pH 8.8), 1.5 mM MgCl<sub>2</sub>, 15 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1% Tween, 0.01% gelatin, and 0.1 U Taq polymerase (Promega; Falkenberg, Sweden). An initial denaturation at 94°C for 3 min was followed by 30–35 (depending on the primers) cycles of denaturation at 94°C for 30 s, annealing at 46–60°C (depending on the primers) for 30 s and extension at 72°C for 30 s, and a final elongation stage at 72°C for 5 min. Tetra-nucleotide amplicons were electrophoresed on an automated laser fluorescence DNA sequencer (Pharmacia LKB A.L.F.; Uppsala, Sweden) and the molecular sizes resolved with allele links software (Pharmacia Biotech). The dinucleotide loci, amplified with primers radiolabelled at the 5' end with 32P g-ATP, were separated by electrophoresis with 1400 V for 2–4 h in 5% denaturing polyacrylamide sequencing gels. After drying, the gels were subjected to autoradiography on Kodak X-OMAT film for 1–7 days and the alleles were determined from the autoradiographs. Haplotypes were constructed manually, assuming the haplotypes in each family to be those which can be explained with a minimum number of obligate recombinations.

### Data analysis

Allele frequencies for each microsatellite locus were estimated by gene counting<sup>26</sup> and the locus diversity was estimated as

$$1 - \sum_i p_i^2$$

where  $p_i$  is the estimated frequency of the  $i$ th allele at the locus.<sup>27</sup> To describe the extent of non-random allelic association between pairs of loci, the tail probability ( $P$  value) of Fisher's exact test was determined using GenePop 1.0.<sup>28</sup> Differences in allele frequency distributions at the microsatellite loci were evaluated by exact tests for population differentiation,<sup>29</sup> as determined by GenePop 1.0.<sup>28</sup> We also applied a more powerful likelihood ratio test for LD between the loci, see Terwilliger and Ott,<sup>30</sup> and Göring and Terwilliger,<sup>31</sup> as described in more detail in the randomisation test section.

### Randomisation test

In order to compare the strength of LD between Kuusamo and the global Finnish sample, it was necessary to equilibrate the sample sizes used in order to avoid concluding that the larger population had more LD simply because there is more power to detect LD in a larger sample. To this end, a bootstrap analysis was performed,<sup>32,33</sup> in which collections of 50 haplotypes from each population were sampled (with replacement) per bootstrap replicate. For each pair of marker loci, in each sample of 50 haplotypes, a likelihood ratio test was performed to compare the haplotype frequencies allowing for LD to the haplotype frequencies obtained, assuming absence of LD<sup>31,32</sup> to be

$$\Lambda = 21n \frac{L(\hat{p}_{ij})}{L(p_{ij} = \hat{p}_i \hat{p}_j)}.$$

10 000 bootstrap samples were generated in order to estimate the distribution of  $L$  for a sample of 50 haplotypes only (39 for the X-chromosomal data because of the smaller available sample sizes). Next, another bootstrap was performed, in which a set of 50 haplotypes (39 for the X chromosome) from each population was simulated by independently sampling (with replacement) one allele at each marker locus to generate 50 haplotypes (39 in X) with no LD. The statistic  $\Lambda$  was computed for each bootstrap replicate to estimate the distribution of  $\Lambda$  under both the null ( $\Lambda_0$ ) and alternative ( $\Lambda_1$ ) hypotheses. In effect,  $f_{\Lambda_0}$  is a bootstrap estimate of the distribution of the statistic  $\Lambda$  under the null hypothesis of no LD. If under alternative hypothesis we had a single statistic,  $\Lambda_1$ , then  $P(\Lambda_0 > \Lambda_1)$  would be a bootstrap estimate of the  $P$  value. However, because we performed 10 000 subsamples from the larger total sample of haplotypes, we do not have a single value for the statistic  $\Lambda_1$ , but rather a distribution of values for this statistic. Thus we computed the expected  $P$  value per subsample of 50 replicates by integrating over this entire distribution, by  $E[P(\Lambda_0 > \Lambda_1)] = \int_0^\infty P(\Lambda_0 > \Lambda_1) df_{\Lambda_1}$ , which we estimated by comparing all the simulated replicates of  $\Lambda_0$  to  $\Lambda_1$  by the double sum

$$\frac{1}{N} \sum_i \sum_j I[\Lambda_0(i) > \Lambda_1(j)],$$

where  $\Lambda_0(i)$  is the value of the statistic  $\Lambda_0$  in the  $i$ -th subsampled bootstrap replicate, and  $N$  is the total number of comparisons ( $10\,000 \times 10\,000 = 100\,000\,000$ ). In our randomisation subsample test, this estimated probability would correspond to the chance of finding as much LD as was observed in a random subsample of 50 haplotypes from the dataset if there were no LD between the loci, and is analogous to a normalised ' $P$  value' to facilitate comparing the LD observed in the isolate and in the general Finnish samples. All the bootstrap replicates were made independently for the two samples, and the resulting statistical results are summarised in Table 1.

### Results

Only small differences in the number of alleles were found between the Kuusamo subisolate and Finland as a whole (Table A on our website). Nine out of 18 autosomal markers and eight out of 15 X-chromosomal markers revealed a smaller number of alleles in the Kuusamo sample than in the all-Finland sample. The allele frequency distributions were significantly different for 11/18 autosomal and 14/15 X-chromosomal microsatellites at the  $P < 0.05$  level, and 4/18 autosomal and 8/15 X-chromosomal markers at the  $P < 0.001$  level.

Both the Finnish and Kuusamo study samples revealed an almost complete absence of evidence for LD in our samples for all three autosomal regions studied (Figure 2; Table 1). For

Table 1 Allelic association between loci

<i>Locus 1</i>	<i>Locus 2</i>	<i>Kuusamo Fisher exact test</i>	<i>Finns Fisher exact test</i>	<i>Kuusamo LRT test whole sample</i>	<i>Finland LRT test whole sample</i>	<i>Kuusamo LRT test reduced subsample</i>	<i>Finland LRT test reduced subsample</i>	<i>Intermarker distance (cM)</i>
D13S317	D13S162	<u>0.06315</u>	0.59614	<b>0.0293</b>	0.6138	<u>0.0525</u>	0.2427	7.5
D13S317	AC76B	0.87146	<u>0.07162</u>	0.4139	0.2334	0.2322	0.4545	5.5
D13S317	COLAC1	0.49996	0.66511	0.2265	0.3390	0.1728	0.3968	5
D13S317	D13S789	0.49427	0.78805	0.2330	0.7013	0.2013	0.2853	4
D13S162	D13S789	0.28617	0.46174	0.1271	0.7798	0.1195	0.3794	3.5
D13S162	COLAC1	0.37433	0.17351	0.1062	0.3128	<b>0.0416</b>	0.4841	2.5
D13S162	AC76B	<u>0.07216</u>	0.67021	<b>0.0222</b>	0.8153	<b>0.0098</b>	0.5847	2
AC76B	D13S789	0.15263	0.69323	<b>0.0475</b>	0.5735	<u>0.0609</u>	0.4899	1.5
COLAC1	D13S789	0.74549	<b>0.00504</b>	0.3253	<b>0.0026</b>	0.2571	0.1885	1
AC76B	COLAC1	<b>0.00723</b>	0.42192	<b>0.0036</b>	0.4608	<b>0.0144</b>	0.4314	0.5
D19S601	D19S433	0.11067	0.19844	<u>0.0559</u>	0.3747	<b>0.0437</b>	0.3698	39
D19S601	D19S416	0.99019	0.64239	0.6084	0.4634	0.1636	0.4178	31
D19S601	D19S208	0.90546	0.61204	0.6416	0.8188	0.1731	0.5373	29
D19S601	D19S610	0.89103	0.16380	0.6114	0.4018	0.2305	0.2674	27
D19S601	D19S422	0.53500	<u>0.08095</u>	0.1789	<u>0.0546</u>	<b>0.0378</b>	0.1501	24
D19S601	D19S223	<b>0.03440</b>	0.56832	<b>0.0113</b>	0.4120	<b>0.0081</b>	0.3379	21
D19S223	D19S433	0.77861	0.35463	0.5259	<u>0.0883</u>	0.1668	0.1277	18
D19S422	D19S433	0.43490	<u>0.07243</u>	0.2134	<b>0.0031</b>	0.1123	<u>0.0892</u>	15
D19S610	D19S433	0.71949	0.50744	0.3078	0.1369	0.2331	0.3179	12
D19S208	D19S433	0.89257	0.27764	0.4948	0.4895	0.1906	0.3932	10
D19S223	D19S416	0.90479	0.61120	0.5425	0.1521	0.1365	0.1673	10
D19S416	D19S433	0.20681	<b>0.04941</b>	<b>0.0180</b>	<b>0.0055</b>	<b>0.0231</b>	0.1224	8
D19S223	D19S208	0.40709	0.74165	0.1669	0.8731	<u>0.0819</u>	0.4538	8
D19S422	D19S416	0.40082	0.12463	0.1026	<b>0.0037</b>	<b>0.0437</b>	<u>0.0943</u>	7
D19S223	D19S610	0.18774	0.16232	<u>0.0544</u>	0.1618	<u>0.0808</u>	0.2095	6
D19S422	D19S208	0.36394	0.88904	<u>0.1966</u>	0.8918	<u>0.0904</u>	0.5122	5
D19S610	D19S416	0.90662	0.28098	0.2926	<u>0.0860</u>	0.1910	0.1623	4
D19S422	D19S610	0.99024	<b>0.01522</b>	0.6001	<b>0.0016</b>	0.2007	0.1109	3
D19S223	D19S422	<b>0.00030</b>	0.20704	<b>0.0001</b>	0.1145	<b>0.0003</b>	<u>0.0575</u>	3
D19S208	D19S416	0.65921	0.34148	0.2361	0.5583	<u>0.0680</u>	0.4374	2
D19S610	D19S208	0.40577	0.61976	0.1710	0.6568	0.1201	0.5097	2
D21S416	D21S1446	0.35278	0.15082	0.4300	0.1123	0.2129	0.2708	15
D21S416	D21S171	0.47205	0.78315	0.2994	0.6048	0.1297	0.3489	11
D21S416	D21S1912	0.94871	<u>0.09045</u>	0.7298	<b>0.0273</b>	0.4159	<u>0.0861</u>	10.5
D21S416	JAI	0.44158	0.63430	0.1687	0.4037	0.1687	0.2430	10
D21S416	D21S49	0.55351	0.51755	0.2218	0.2500	0.1071	0.2213	9
D21S49	D21S1446	0.59559	0.11363	0.5030	0.1373	0.2405	0.2522	6
JAI	D21S1446	0.27191	0.66414	0.2683	0.3143	0.2748	0.2846	5
D21S191	D21S1446	0.16820	0.17924	<u>0.0624</u>	<b>0.0450</b>	<u>0.0942</u>	0.1885	4.5
D21S171	D21S1446	<b>0.00869</b>	0.53128	<b>0.0022</b>	0.4669	<b>0.0266</b>	0.4514	4
D21S49	D21S171	0.68500	0.79993	0.4680	0.2931	0.1312	0.2442	2
D21S49	D21S1912	0.92714	0.18192	0.6236	<b>0.0119</b>	0.3290	<u>0.0728</u>	1.5
JAI	D21S171	0.40758	0.21255	0.1475	<b>0.0342</b>	0.1133	0.1450	1
D21S49	JAI	0.23182	0.28765	0.2215	<b>0.0458</b>	0.1821	0.1452	1
D21S191	D21S171	0.32524	0.65044	0.1418	<u>0.0791</u>	<u>0.0631</u>	0.1266	0.5
JAI	D21S1912	0.74544	<b>0.00199</b>	0.5764	<b>0.0002</b>	0.3216	<b>0.0254</b>	0.5
DXS1052	DXS987	0.81636	0.61387	0.5131	0.2553	0.3344	0.3062	8.1
DXS1052	DXS1053	0.63329	0.13666	0.3154	<b>0.0318</b>	0.2476	0.2322	8
DXS7105	DXS987	<b>0.00974</b>	0.25416	<b>0.0114</b>	<b>0.0328</b>	<b>0.00323</b>	<u>0.0983</u>	8
DXS7105	DXS1053	0.33876	0.10328	0.1382	<b>0.0007</b>	<u>0.0890</u>	<u>0.0888</u>	7.9
DXS1229	DXS987	<u>0.06070</u>	0.41379	0.1027	0.3034	0.1028	0.4580	7.8
DXS1229	DXS1053	<b>0.02656</b>	0.15421	0.1400	<u>0.0857</u>	0.1610	0.3616	7.7
DXS1052	DXS999	0.98045	0.10202	0.7788	<b>0.0471</b>	0.5190	0.1683	5.5
DXS7105	DXS999	0.44935	0.88330	0.2032	0.1098	0.1457	0.1312	5.4
DXS1229	DXS999	0.49280	<b>0.00704</b>	0.2899	<b>0.0068</b>	0.2864	0.1748	5.2
DXS1052	DXS7163	0.89261	0.67907	0.6633	0.4578	0.4237	0.4555	4.7
DXS7105	DXS7163	<b>0.01369</b>	0.95066	<b>0.0244</b>	0.6194	<b>0.0334</b>	0.3709	4.6
DXS1229	DXS7163	<b>0.00073</b>	<b>0.01556</b>	<b>0.0007</b>	<b>0.0338</b>	<b>0.0061</b>	0.2775	4.4
DXS7163	DXS987	0.15700	0.99017	<u>0.0594</u>	0.8872	<u>0.0722</u>	0.4166	3.4

Table 1 continues on next page

Table 1 Continued

Locus 1	Locus 2	Kuusamo Fisher exact test	Finns Fisher exact test	Kuusamo LRT test whole sample	Finland LRT test whole sample	Kuusamo LRT test reduced subsample	Finland LRT test reduced subsample	Intermarker distance (cM)
DXS7163	DXS1053	0.54823	0.17491	0.2602	<b>0.0125</b>	0.1413	0.2298	3.3
DXS999	DXS987	<u>0.05866</u>	0.56337	<b>0.0017</b>	<b>0.0305</b>	<b>0.0154</b>	<u>0.0929</u>	2.6
DXS999	DXS1053	0.13284	0.59290	<b>0.0490</b>	<b>0.0193</b>	<u>0.0669</u>	0.1568	2.5
DXS7163	DXS999	<b>0.01355</b>	<b>0.03075</b>	<b>0.0047</b>	<b>0.0011</b>	<b>0.0378</b>	0.1276	0.8
DXS1052	DXS1229	0.79337	0.71488	0.6795	0.8038	0.4266	0.5602	0.3
DXS7105	DXS1229	<u>0.07829</u>	0.12016	<u>0.0934</u>	<u>0.0540</u>	0.1160	0.3312	0.2
DXS1053	DXS987	0.20916	0.31160	<u>0.0787</u>	<b>0.0169</b>	<b>0.0428</b>	0.1657	0.1
DXS1052	DXS7105	0.73787	0.94328	0.3529	0.6543	0.3709	0.4324	0.1
DXS1066	DXS983	0.29384	0.33848	0.1388	0.7491	0.1728	0.4584	3.9
DXS995	DXS983	0.53920	0.50813	0.2569	0.5761	0.1819	0.3940	3.2
DXS1066	DXS8037	0.86263	0.66055	0.4700	0.9052	0.3959	0.4302	2.7
DXS1066	DXS8092	0.58927	0.34406	0.3155	0.2562	0.2162	0.2056	2.5
DXS1066	DXS8082	0.53716	0.38662	0.2453	0.4400	0.2358	0.4740	2.5
DXS1066	DXS1225	<b>0.02430</b>	0.12499	<b>0.0048</b>	<b>0.0492</b>	<u>0.0642</u>	0.2149	2.5
DXS1066	DXS986	0.93546	0.23339	0.7479	0.2249	0.3997	0.3722	2.2
DXS995	DXS8037	0.74544	0.87376	0.4111	0.4593	0.2151	0.4010	2
DXS995	DXS8092	<b>0.02919</b>	0.11506	<b>0.0154</b>	<b>0.0001</b>	<b>0.0469</b>	<u>0.0680</u>	2
DXS995	DXS8082	0.19488	0.12749	<u>0.0674</u>	0.2540	<u>0.0743</u>	0.2618	2
DXS995	DXS1225	<b>0.01854</b>	0.15388	<b>0.0097</b>	<b>0.0054</b>	<b>0.0261</b>	<u>0.0980</u>	2
DXS986	DXS983	0.22534	0.82905	<u>0.0684</u>	0.8308	<u>0.0903</u>	0.4896	1.7
DXS995	DXS986	0.25640	0.72858	<u>0.0827</u>	0.2457	0.1016	0.3285	1.5
DXS1225	DXS983	<u>0.09532</u>	0.63015	<b>0.0249</b>	0.3435	<u>0.0581</u>	0.2651	1.5
DXS8082	DXS983	0.23839	0.56539	<u>0.0593</u>	0.6508	<u>0.0849</u>	0.4989	1.5
DXS8092	DXS983	0.16374	0.31404	<u>0.0646</u>	<u>0.0519</u>	<u>0.0993</u>	0.1545	1.5
DXS8037	DXS983	0.22476	0.68306	0.1819	0.7961	0.1175	0.4406	1.2
DXS1066	DXS995	0.35926	0.42012	0.1872	0.3271	0.1757	0.4335	0.7
DXS1225	DXS986	<b>0.00279</b>	0.39330	<b>0.0020</b>	<b>0.0368</b>	<b>0.0140</b>	0.1165	<0.5
DXS8037	DXS986	<b>0.01609</b>	0.61956	<b>0.0080</b>	0.4157	<b>0.0336</b>	0.3039	<0.5
DXS8037	DXS8092	0.15204	0.18043	<b>0.0006</b>	0.1499	0.1023	<u>0.0823</u>	<0.5
DXS1225	DXS8092	<b>0.00182</b>	0.28304	<b>0.0003</b>	<b>0.0000</b>	<b>0.0042</b>	<b>0.0257</b>	<0.5
DXS8037	DXS8082	<b>0.00436</b>	0.23770	<b>0.0340</b>	<b>0.0040</b>	<b>0.0265</b>	0.3181	<0.5
DXS1225	DXS8082	<b>0.00000</b>	<b>0.00000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0025</b>	<0.5
DXS8082	DXS986	<b>0.00183</b>	<u>0.09156</u>	<b>0.0009</b>	<u>0.0538</u>	<b>0.0098</b>	0.2774	<0.5
DXS8092	DXS986	<b>0.03738</b>	0.33130	<b>0.0142</b>	<b>0.0028</b>	<b>0.0328</b>	<u>0.0783</u>	<0.5
DXS8082	DXS8092	<b>0.00030</b>	<u>0.07444</u>	<b>0.0000</b>	<b>0.0005</b>	<b>0.0009</b>	<u>0.0785</u>	<0.5
DXS8037	DXS1225	<u>0.07797</u>	0.83624	<b>0.0182</b>	<u>0.0720</u>	<b>0.0475</b>	0.2255	<0.5

The first two columns give the results of the Fisher exact test, which determines the probability of a dataset that is as unlikely as or more unlikely than the dataset observed under the assumption of no linkage disequilibrium (GenePop 1.0 software<sup>26</sup>). The next two columns are the *P* values obtained from a likelihood ratio test comparing the alternative hypothesis of linkage disequilibrium with the null hypothesis of the absence of linkage disequilibrium. While both tests attempt to disprove the null hypothesis of the absence of linkage disequilibrium, the relevant statistics are different. In the Fisher exact test, the inference is based on how unlikely the data are when the null hypothesis is true, whereas in the likelihood ratio test, the inference is based on how many more times more likely a dataset is under the alternative hypothesis than under the null hypothesis. In general this test is more powerful than the Fisher exact test when there are multi-allelic marker loci; and when a marker locus has two alleles their properties are the same. The *P* values presented are similarly 'exact', giving the probability of a dataset resulting in a statistic as large as or larger than that found in the observed data. The fifth and sixth columns are also based on the likelihood ratio tests, adjusted for sample size, as described in the text. There were more Finnish than Kuusamo families in the study, so there is more power to detect LD in the all-Finland sample, even if there were no differences in the existing levels of LD. These two columns are the result of a bootstrap subsampling analysis where samples of 50 autosomes or 39 X chromosomes (fewer X because of the reduced sample size available) from each population were sampled to determine the expected *P* values in each population when the sample sizes were fixed to be equal (see text for details). The bootstrap procedure was applied separately to the Finnish and the Kuusamo data. Significant allelic associations ( $P < 0.05$ ) are in bold, and associations suggestive of LD ( $P < 0.1$ ) are underlined.

the all-Finland sample, only a single pair of loci gave weak evidence of LD on 21q ( $< 0.5$  cM;  $P = 0.025$ ), which does not survive a Bonferroni correction for multiple testing. For the Kuusamo study sample, a single pair of loci gave 'highly significant' evidence of LD on 19q (intermarker interval 3 cM;  $P$  approximately 0.0003), and weak evidence of association was detected for three locus pairs on 13q, five pairs on

19q and one pair on 21q. No distance-correlated pattern of background LD was observed in either the Finnish or the Kuusamo samples.

In contrast to autosomes, the Xq13 and Xp22 regions studied revealed different patterns of LD for the Finns and the Kuusamo population (Figure 2; Table 1). For the Finns, the detected level of background LD on the X chromosome

was similar to that on the autosomes, with a single significantly associated locus pair (out of a total of 28) on Xq13 (intermarker interval  $<0.5$  cM;  $P$  approximately 0.0025), which does not survive Bonferroni correction, and not a single pair out of 21 at Xp22. In the Kuusamo population, the Xp22 region revealed slightly increased levels of LD with six associated locus pairs over distances 0.1–8 cM ( $P = 0.003$ – $0.04$ ) in the reduced subsample. At Xq13, a region characterised by an extremely high physical/genetic distance ratio, a large number of allelic associations were detected in the Kuusamo samples. Of the 28 polymorphic marker pairs tested, 11 showed LD at the 0.05 significance level, four of which have  $P < 0.01$ , of which two had  $P < 0.001$ . Most of these associations were found between tightly-linked pairs of loci ( $< 1$  cM).

## Discussion

We identified slightly more LD in Kuusamo than in Finland as a whole, at least on the X-chromosome, if not the autosomes. According to theoretical expectations, not only the founder population size and the time since expansion, but also the growth rate, can influence the amount of LD present in a growing population.<sup>5</sup> This could indicate that the level of haplotype diversity on the chromosomes of the founder population may have been relatively high, in which case the Kuusamo founder population size of approximately 39 families at the end of the 17th century is still too large to find significant levels of LD. In the Kuusamo population, we observed only single, randomly associated pairs of loci on the autosomes, and even on the X chromosome, with a smaller number of recombinations and smaller effective population size, the LD patterns were preserved systematically only over very short ( $< 1$  cM) genetic distances. The reduction in the number of alleles in the Kuusamo subsample was observed only for the X-chromosomal markers, the difference in the autosomal markers being small. Also the population growth rate over the past 300 years, which reached 170% per generation for Kuusamo and 135% for the whole of Finland, is expected to break down background LD in the founder population. Furthermore, for Kuusamo, this 170% represents the growth rate over the entire expansion period of 300 years after its foundation, whereas, for the Finns as a whole, the population growth has been spread over 3000 years with a mean increase of 106% per generation.

The prognosis of the success of both linkage and LD mapping depends on the demographic history of the population.<sup>34</sup> Expanded and relatively isolated populations, like the Finns, have proved suitable for successful mapping of rare monogenic diseases.<sup>20,35–37</sup> According to the results presented here, only a very small level of LD is likely to exist between alleles of microsatellite loci on randomly ascertained chromosomes, rendering the interference from the background LD negligible, whereas there may be long conserved haplotypes flanking rare disease alleles, as the alleles are clonally

identical from some single common ancestral chromosome due to linkage in the population-as-pedigree.<sup>4,34</sup> When a sample is ascertained randomly, and not on the basis of the sharing of a rare (eg disease-predisposing) allele, associations with the surrounding markers are not expected, but when the ascertainment is conditional on the presence of a rare allele, and all the copies of that allele are clonal, there may be considerable LD. Information on LD from neutral alleles in anonymous marker loci and from disease alleles has been useful in identifying the locus that has an allele causing the disease, and in fine mapping the precise genomic location of the disease allele.<sup>19,35,38</sup> However, it has been claimed that in complex traits, with multiple (perhaps common) disease-predisposing DNA variants at the population level, representing multiple rare mutations or a few common, and therefore old, mutations, LD may not be such a useful tool.<sup>3</sup> As shown here for autosomal markers in Kuusamo and Finland, both representing expanded populations, the interval revealing LD in alleles in the general population is negligible. This is in accord with a previous study<sup>7</sup> in which seven dinucleotide loci spanning approximately 4 cM on Xq13 revealed LD with only 1–3 of the 21 pairs of loci in three expanded populations: those of Estonia, Finland, and Sweden. However, as a representative of a small, constant population (which went through repeated bottlenecks but never subsequently expanded), in the Saami, LD was detected in 25/36 pairs of loci, over genetic distances as large as 5.6–5.8 cM (three pairs). Thus, in small, constant populations, greatly reduced allelic and haplotype diversity tends to be found over wide areas, as LD can be generated rapidly through the process of genetic drift.<sup>7,8,39</sup>

In Finnish sub-populations arising from the old region of early settlement, we have earlier observed that in alleles associated with complex diseases, such as essential hypertension and multiple sclerosis (MS), LD is detectable only over a greatly restricted interval. In such cases, LD was observed to span some 30–40 kb: examples are the angiotensin II receptor type-1 gene in hypertension and the HLA region of the golli-MBP gene in MS.<sup>40–42</sup> Our data should not discourage scientists from pursuing studies of samples from such genetic isolates. Because of the reductions in genetic, cultural, and environmental heterogeneity in these populations, phenotypic observations are better able to predict the underlying disease-predisposing genotypes, and since the affected individuals in these populations are most certainly related if one looks far enough back into the genealogy, their predisposing alleles may often be IBD. This is borne out by the observations in our initial studies of schizophrenia in the Kuusamo population, where genealogical knowledge of the population structure resembling a very large pedigree offered many advantages. The linkage analysis provided the most significant evidence to date of a region in which a putative disease locus might reside, whereas LD analysis of the same markers in a random cohort of case and control individuals produced little or no such evidence. LD studies of random

markers in potential study populations can give an idea of the allelic diversity as well as of the background LD, and so offer some indication of the power of linkage analysis, not only for LD analysis, but also for mapping the loci involved in disease aetiology.

#### Electronic-Database Information

Authors' web site for additional article data, <http://oxy-gen.ktl.fi/molbio/wwwpub/>

#### Acknowledgements

We are grateful to the Finnish families for providing the data for this study. We are also indebted to Matti Lukka, Liisa Moilanen and Merja Nissi for blood samples, Marjo Kestilä for a primer, and Soili Johansson, Anne Jokiahio, Juha Paloneva, Teemu Perheentupa and Arja Terola for their excellent technical assistance. We thank Dr Svante Pääbo for the continuous discussions on the topic of linkage disequilibrium. The financial support of the Finnish Cultural Foundation, the Ulla Hjelt fund of the Foundation for Pediatric Research, the Alexander-von-Humboldt Stiftung (to ML) and the Hitchings-Elion Fellowship from the Burroughs Wellcome Fund (to JT) is acknowledged.

#### References

- Lander ES, Schork NJ: Genetic dissection of complex traits. *Science* 1994; **265**: 2037–2048.
- Peltonen L, Pekkariinen P, Aaltonen J: Messages from an isolate: Lessons from the Finnish gene pool. *Biol Chem* 1995; **376**: 697–704.
- Terwilliger JD, Weiss KM: Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr Opin Biotechnol* 1998; **9**: 578–594.
- Terwilliger JD, Göring HHH: Gene mapping in the 20th and 21st centuries: Statistical methods, data analysis, and experimental design. *Hum Biol* 2000; **72**: 63–132.
- Stalkin M: Linkage disequilibrium in growing and stable populations. *Genetics* 1994; **137**: 331–336.
- Crouau-Roy B, Service S, Slatkin M, Freimer N: A fine-scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence analysis. *Hum Mol Genet* 1996; **5**: 1131–1137.
- Laan M, Pääbo S: Demographic history and linkage disequilibrium in human populations. *Nat Genet* 1997; **17**: 435–438.
- Terwilliger JD, Zöllner S, Laan M, Pääbo S: Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'Drift Mapping' in small populations with no demographic expansion. *Hum Hered* 1998; **48**: 138–154.
- Nevanlinna HR: The Finnish population structure. A genetic and genealogical study. *Hereditas* 1972; **71**: 195–236.
- De la Chapelle A: Disease gene mapping in isolated human populations: the example of Finland. *J Med Genet* 1993; **30**: 857–865.
- Lahermo P, Sajantila A, Sistonen P *et al*: The genetic relationship between the Finns and the Finnish Saami (Lapps): Analysis of nuclear DNA and mtDNA. *Am J Hum Genet* 1996; **58**: 1309–1322.
- De la Chapelle A, Wright FA: Linkage disequilibrium mapping in isolated populations: The example of Finland revisited. *Proc Natl Acad Sci USA* 1998; **95**: 12416–12423.
- Kittles RA, Perola M, Peltonen L *et al*: Dual origins of Finns revealed by Y chromosome haplotype variation. *Am J Hum Genet* 1998; **62**: 1171–1179.
- Lahermo P, Savontaus ML, Sistonen P *et al*: Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. *Eur J Hum Genet* 1999; **7**: 447–458.
- Norio R, Nevanlinna HR, Perheentupa J: Hereditary diseases in Finland. *Ann Clin Res* 1973; **5**: 109–141.
- Peltonen L, Jalanko A, Varilo T: Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* 1999; **10**: 1913–1923.
- Hovatta I, Varilo T, Suvisaari J *et al*: A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation suggesting multiple susceptibility loci. *Am J Hum Genet* 1999; **65**: 1114–1124.
- Sheffield VC, Stone EM, Carmi R: Use of isolated inbred human populations for identification of disease genes. *Trends Genet* 1998; **14**: 391–396.
- Varilo T, Savukoski M, Norio R, Santavuori P, Peltonen L, Järvelä I: Tracing ancestors for the CLN5 mutation: Genealogical and linkage disequilibrium analysis in the Finnish population. *Am J Hum Genet* 1996; **58**: 506–512.
- Aaltonen J, Björns P, Sandkuijl L, Perheentupa J, Peltonen L: An autosomal locus causing autoimmune disease: Autoimmune polyglandular disease type I assigned to chromosome 21. *Nat Genet* 1994; **8**: 83–87.
- Männikkö M, Kestilä M, Holmberg C *et al*: Fine mapping and haplotype analysis of the locus for congenital nephrotic syndrome on chromosome 19q13.1. *Am J Hum Genet* 1995; **57**: 1377–1383.
- Vandenplas S, Wiid I, Grobler-Rabie A *et al*: Blot hybridization analysis of genomic DNA. *J Med Genet* 1984; **21**: 164–172.
- Gyapay G, Morrisette J, Vignal A *et al*: The 1993–94 Génethon human genetic linkage map. *Nat Genet* 1994; **7**: 246–339.
- Klockars T, Savukoski M, Isosomppi J *et al*: Efficient construction of a physical map by fiber-fish of the CLN5 region: Refined assignment and long-range contig covering region on 13q22. *Genomics* 1996; **35**: 71–78.
- Mumm S, Jermak C, Waeltz P *et al*: 22-Mb integrated physical and genetic map based on YAC/STS content spanning the interval DXS1125–DXS95 in human Xq12–q21.31. *Gene* 1998; **208**: 147–156.
- Smith CAB: Counting methods in genetical statistics. *J Roy Stat Soc Ser B* 1957; **15**: 153–192.
- Weir BS: *Genetic Data Analysis II*. Sinauer: Sunderland, MA, 1996.
- Raymond M, Rousset F: GenePop., version 1.0 Institut des Sciences de l'Évolution, Université de Montpellier, France, 1994.
- Raymond M, Rousset F: An exact test for population differentiation. *Evolution* 1995; **49**: 1280–1283.
- Terwilliger JD, Ott J: *Handbook of Human Genetic Linkage*. Johns Hopkins University Press: Baltimore, 1994.
- Göring HHH, Terwilliger JD: Linkage analysis in the presence of errors IV: Joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. *Am J Hum Genet* 2000; **66**: 1310–1327.
- Efron B: *The Jackknife, the Bootstrap, and other Resampling Plans*. Society for Industrial and Applied Mathematics: Philadelphia, PA, 1982.
- Terwilliger JD, Ott J: A multisample bootstrap approach to the estimation of maximized-over-models lod score distributions. *Cytogenet Cell Genet* 1992; **59**: 142–144.
- Terwilliger JD: On the resolution and feasibility of genome scanning approaches to unraveling the genetic components of multifactorial phenotypes. In: Rao DC, Province M, (eds), *Genetic Dissection of Complex Traits: Challenges for the New Millennium*. Academic Press: in press.
- Hästbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E: Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* 1992; **2**: 204–211.
- Nikali K, Suomalainen A, Terwilliger J, Koskinen T, Weissenbach J, Peltonen L: Random search for shared chromosomal regions in four affected individuals: the assignment of a new hereditary ataxia locus. *Am J Hum Genet* 1995; **56**: 1088–1095.



- 37 Savukoski M, Klockars T, Holmberg V, Santavuori P, Lander ES, Palotie L: *CLN5*, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis (vLINCL). *Nat Genet* 1998; **19**: 286–288.
- 38 Mäkelä-Bengs P, Järvinen N, Vuopala K *et al*: Assignment of the disease locus for the lethal congenital contacture syndrome to a restricted region of chromosome 9q34 by genome scan using five affected individuals. *Am J Hum Genet* 1998; **63**: 506–516.
- 39 Laan M, Pääbo S: Mapping genes by drift-generated linkage disequilibrium. *Am J Hum Genet* 1998; **63**: 654–656.
- 40 Kainulainen K, Perola M, Terwilliger J *et al*: Evidence for involvement of the angiotensin II receptor type 1 locus in essential hypertension. *Hypertension* 1999; **33**: 844–849.
- 41 Tienari PJ, Wikström J, Koskimies S, Partanen J, Palo J, Peltonen L: Reappraisal of HLA in multiple sclerosis: close linkage in multiplex families. *Eur J Hum Genet* 1993; **1**: 257–268.
- 42 Tienari PJ, Kuokkanen S, Pastinen T *et al*: Golli-MBP gene in multiple sclerosis susceptibility. *J Neuroimmunol* 1998; **81**: 158–167.