EXPERIMENTAL and MOLECULAR MEDICINE, Vol. 38, No. 6, 662-667, December 2006

Heterozygosities of 735 microsatellite markers and background linkage disequilibrium in the Korean population

Kyung-A Lee[1] and Jong-Won Kim[2,3]

[1]Department of Laboratory Medicine
Hallym University College of Medicine
Chuncheon 200-704, Korea
[2]Samsung Medical Center
Sungkyunkwan University School of Medicine
Seoul 135-710, Korea
[3]Corresponding author: Tel, 82-2-3410-2705;
Fax, 82-2-3410-2719; E-mail, jwonk@smc.samsung.co.kr

Accepted 24 October 2006

Abbreviations: BLD, background linkage disequilibrium; FET, Fisher's exact test

Abstract

Suitability of a specific population for linkage disequilibrium mapping studies of complex traits may be assessed by investigating the background linkage disequilibrium (BLD). We are unaware of studies for quantifying the degree of BLD in the Korean population, although the population may be a good candidate for mapping of complex trait genes through whole-genome association studies. It is useful to investigate the properties of genetic isolates in East Asia and to compare them to genetic isolates in Europe. We analyzed the extent of BLD in the Korean population using 735 microsatellite markers and compared the results with the Icelander population, which is one of the European expanded genetic isolates. The Korean population exhibited a level of BLD comparable with the Icelander population. The inference of population structure using the model with admixture showed that each individual has allele copies originating from K populations in equal proportions. Therefore, we believe that factors other than genetic distance, such as recent admixture, have not contributed to the level of BLD. Our results showed that the Korean population, which is an expanded population with no evidence of admixture, has a BLD level comparable with the Icelander population. Therefore, the Korean population can be used for fine mapping of either complex traits or monogenic diseases.

Keywords: Asian continental ancestry group; genetic markers; linkage disequilibrium; microsatellite repeats; population characteristics

Introduction

Suitability of a specific population for linkage disequilibrium mapping studies of complex traits may be assessed by investigating the background linkage disequilibrium (BLD) (Feimer *et al*., 1997; Gordon *et al*., 2000). Although many empirical studies have evaluated the BLD in European isolates, such as the Finnish and Sardinians (de la Chapelle *et al*., 1998; Eaves *et al*., 2000; Angius *et al*., 2001), few data from other region, such as Asia, have been reported (Katoh *et al*., 2002).

The Korean population probably originated from the Tungus branch of Mongolian ethnic groups who inhabited the general area of the Atlantic Mountains in Central Asia (Kim, 1970). There is evidence for Yayoi migration from their original places *via* China and the Korean peninsula to Japan, starting approximately 2,300 years ago (Hanihara, 1991). The population size in Korea about 2,000 yr ago is estimated to be approximately 3 million. After several gradual expansions, the modern South Korean population now consists of about 48 million people (Statistics Bureau/Statistical Research and Training Institute 2002) and the combined population of the South and North is estimated at 70 million.

Koreans are racially and linguistically homogenous with no sizable indigenous minorities. Throughout the history of Korea, there seems to be no obvious admixture. To date, we are unaware of any studies for quantifying the degree of BLD in the Korean population, although the population may be a strong candidate for mapping of complex trait genes through whole- genome association studies. It is useful to investigate the properties of genetic isolates in East Asia and to compare them to genetic isolates in Europe.

We report here the extent of BLD and heterozygosities of 735 microsatellite markers among 98 Korean subjects and we compare the results with the Icelander population (Kong *et al*., 2002), which is one of the expanded European genetic isolates.

Materials and Methods

Subjects

We collected blood samples from 98 healthy, un-

related Korean individuals (male-female ratio, 2.2:1; average age, 24.6 yr; range, 21-32 yr) after obtaining informed consent. In order to compare the LD patterns among microsatellite markers between the Korean and Icelandic populations, we obtained Icelandic genotype data from deCODE Genetics after submitting a complete agreement form.

Markers and genotyping

We genotyped 811 fluorescently-labeled microsatellite markers from the ABI PRISM Linkage mapping set HD5 v2.5 (AppliedBiosystems, Foster City). PCR was carried out in 96-well plate in a volume of 7.5 μl containing 30 ng of genomic DNA, 2.5 pmol of each primer, 250 μM dNTPs, 2 mM $MgCl_2$, and 0.3U of *Taq* DNA polymerase. The thermocycling conditions were pre-denaturation at 95°C for 12 s followed by 10 cycles of 94°C for 15 s, 55°C for 15 s and 72°C for 30 s, and 20 cycle of 89°C for 15 s, 55°C for 15 s, and 72°C 30 s. PCR products were detected using an ABI PRISM 3100 Genetic Analyzer and analyzed using GeneScan and Genotyper software. All genotypes were independently double checked. The results were tested for Hardy-Weinberg equilibrium using the SAS Genetics program.

Genotypic data derived from CEPH families for each marker were obtained from the CEPH database (http://www.ceph.fr/). The order and sex-averaged distance of the markers were based on the deCODE map (Kong et al., 2002). The average spacing of the markers in our data set was $<$ 5 cM.

Statistical methods

Fisher's exact test (FET) was used to detect significant LD between every pair of the microsatellite markers. BLD tests were performed between all pairs of adjacent markers on all 23 chromosomes using 735 informative markers (average heterozygosity 0.71 $\pm$ 0.13) that met the criteria of Hardy-Weinberg equilibrium. All of our pairwise tests used the same sample size. In order to compare the BLD strength between Korean and Icelander populations, it was necessary to equilibrate the sample size used in order to avoid concluding that the larger population had more LD simply because there was more power to detect LD in a larger sample. We analyzed the extent of BLD among the Icelandic population through use of 41 locus pairs matched to Korean data with a 100 $\pm$ 10 sample size. For the X chromosome, we calculated the LD value based on combined data composed of phased male genotype data and we assigned female haplotype data using a haplotype frequency based method.

We used logistic regression analysis to model the probability of being in BLD as a function of distance between loci and the number of alleles at each of the loci (Service *et al*., 2001). This analysis was performed using 156 locus pairs spread throughout the genome, with the requirement that pairs were separated by 20 cM so that they would be roughly independent. We chose 500 locus pairs at random and examined the BLD between unlinked loci to determine a baseline level of FET significance in our sample.

To evaluate the presence of population structure and identify admixed individuals, we used the model-based clustering methods of the STRUCTURE program (50,000 burn-in and 500,000 iteration, K = 1-4) (Pritchard *et al*., 2000).

Results

Overall heterozygosities of the markers from Koreans exhibited a higher similarity to those from Japanese subjects (Ikari *et al*., 2002) compared to Caucasians (Figure 1 and supplementary Table 1). Of 735

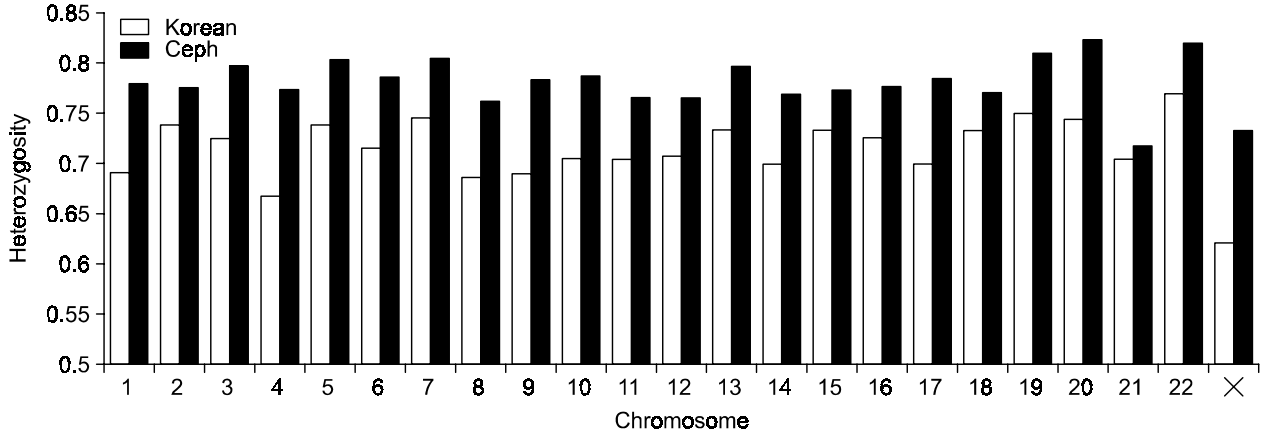

Figure 1. Comparison of heterozygosities of 735 microsatellite markers between Koreans and CEPH familes.

Table 1. Pairwise linkage disequilibrium (LD) based on Fisher's exact test (FET) using 41 matched locus pairs in the Korean and Icelandic populations.

Locus pairs	Chr	Distance (cM)	No. of alleles in locus 1		No. of alleles in locus 2		FET *P* values[a] in	
			Korean	Icelander	Korean	Icelander	Korean ($n = 98$)	Icelander ($n = 100 \pm 10$)[b]
D1S206-D1S495	1	1.4	9	9	12	13	0.849	0.492
D1S2726-D1S498	1	12.8	7	11	12	11	0.478	0.292
D1S498-D1S2635	1	9.3	12	11	8	13	0.325	0.636
D2S352-D2S367	2	2.7	9	10	13	15	0.297	0.480
D3S1266-D3S1609	3	3.4	6	5	5	8	0.893	0.449
D3S3725-D3S1565	3	3.5	11	11	7	8	0.329	0.930
D3S1565-D3S3715	3	5.7	7	8	5	6	0.814	**0.086**
D3S3686-D3S1580	3	3.9	10	13	16	12	0.653	**0.054**
D4S1597-D4S1595	4	4.4	7	10	7	8	0.530	0.670
D6S1721-D6S259	6	3.1	7	9	11	9	0.835	**0.023**
D6S257-D6S460	6	11.1	14	13	10	13	0.884	0.259
D7S641-D7S2464	7	3.0	7	5	5	7	0.804	**0.063**
D7S2464-D7S513	7	2.2	5	7	16	14	0.822	0.447
D7S2513-D7S661	7	1.6	12	11	8	11	**0.039**	**0.003**
D8S284-D8S256	8	5.5	16	13	10	12	0.865	0.304
D9S290-D9S164	9	10.5	7	7	10	11	0.997	0.149
D9S164-D9S1818	9	3.5	10	11	6	5	0.467	**0.079**
D10S547-D10S570	10	3.9	9	9	9	10	0.756	0.404
D10S196-D10S1652	10	10.5	6	8	8	12	0.404	0.784
D10S1652-D10S537	10	8.6	8	12	10	11	0.696	**0.041**
D10S1765-D10S185	10	7.0	10	11	10	10	1.000	0.972
D10S217-D10S1655	10	6.4	10	10	4	8	0.922	0.553
D11S4190-D11S915	11	4.5	11	10	11	9	0.106	0.253
D11S4102-D11S905	11	3.0	9	10	12	12	0.498	**0.052**
D11S905-D11S4191	11	7.6	12	12	13	15	**0.013**	0.604
D11S912-D11S4126	11	8.0	8	10	6	5	0.970	0.145
D12S364-D12S310	12	5.1	14	14	7	8	0.840	**0.064**
D12S310-D12S1682	12	3.0	7	8	11	8	0.899	0.518
D12S1718-D12S86	12	5.6	6	9	12	7	0.494	0.760
D12S1675-D12S1659	12	6.2	9	11	6	10	0.580	0.803
D13S1296-D13S156	13	6.9	14	13	10	8	**0.009**	**0.042**
D13S170-D13S265	13	5.1	14	14	8	12	0.185	0.240
D14S65-D14S985	14	9.3	12	11	9	7	0.953	0.869
D16S3049-D16S516	16	2.0	7	12	8	10	0.360	0.216
D18S59-D18S476	18	4.3	10	9	6	8	0.867	0.980
D18S476-D18S63	18	3.9	6	8	9	14	0.252	0.674
D19S931-D19S414	19	4.8	9	10	9	9	0.253	0.429
D19S220-D19S420	19	5.4	12	14	7	10	0.480	0.308
D22S539-D22S1174	22	4.1	6	9	10	9	0.858	0.890
D22S274-D22S1169	22	12.4	10	9	8	9	0.559	0.471
DXS1060-DXS1223	X	3.3	9	8	8	8	**0.050**	**0.000**

[a]The values with significant LD ($P \leq 0.05$) are underlined. Suggestive LD ($0.05 < P < 0.10$) are indicated in boldface. [b]Data are from Kong *et al.*, (2002). Chr, chromosome.

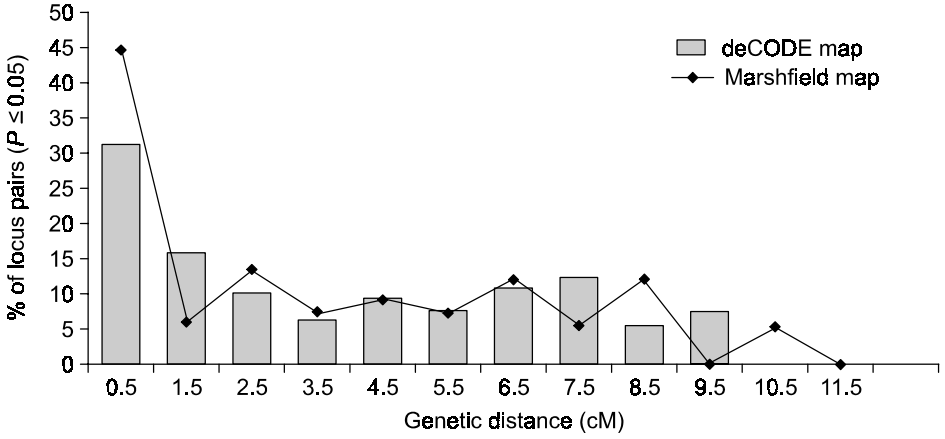

Figure 2. Percentages of *P*-values $\leq$ 0.05 from the FET of LD between 736 pairs of adjacent markers.

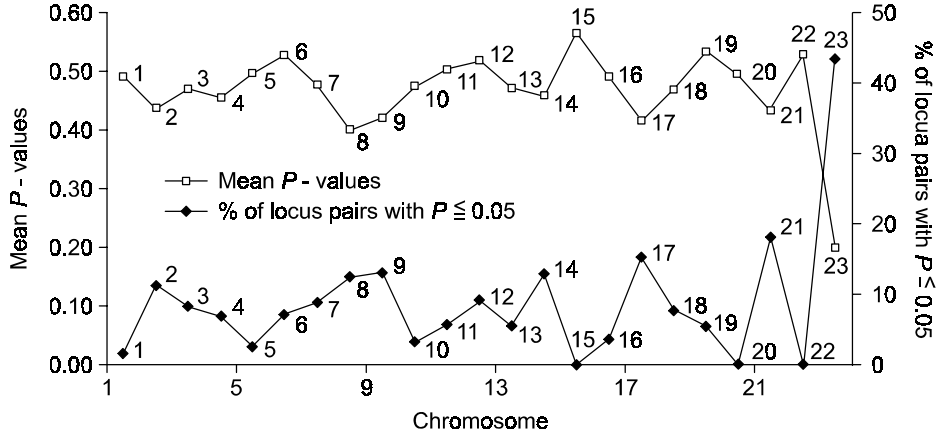

Figure 3. Mean *P*-values and percentages of *P*-values $\leq$ 0.05 from the FET by chromosome.

markers, 34 marker loci showed heterozygosity lower than 0.50.

In the 736 tests of adjacent markers, 69 (9%) and 26 (4%) were found to have *P* values from FET of $\leq$ 0.05 and $\leq$ 0.01, respectively. The FET *P* values from 500 locus pairs on different chromosomes showed 5% to be significant at the 0.05 level and 1% to be significant at the 1% level. Using 156 independent locus pairs, the distance between markers was found to significantly predict the probability of being in BLD, with markers $\leq$ 2 cM apart having a 3.2 times greater probability of an FET *P* value of $\leq$ 0.05 than markers $>$ 2 cM apart ($P = 0.04$, Confidence intervals 1.1-9.8). Sixteen percent of markers within 2 cM and 8% of markers $\geq$ 2 cM apart had FET *P* values of $\leq$ 0.05. However, the relationship was not close (correlation coefficient r = 0.12) as significant BLD was observed in numerous markers as far apart as 9 cM. The pattern of BLD according to the genetic distance was similar to the pattern of the Marshfield map, although there was a slight difference in the percentage of locus pairs with a significant LD at 1.5 and above 8.5 cM bins (Figure 2).

All of the FET *P* values in 41 pairs that matched the Icelanders are listed in Table 1. The results showed that the Korean population exhibited a BLD level comparable with the Icelander population. Koreans exhibited 4 of 41 pairs with a significant LD ($P \leq 0.05$) and 3 (D7S2513-D7S661, D13S1296-D13S156, DXS1060-DXS1223) of these corresponded to the Icelanders. The Icelanders exhibited 7 of 41 pairs with a significant LD ($P \leq 0.05$) and another 4 pairs exhibited a suggestive LD ($0.05 < P < 0.10$). The largest genetic distances for a significant LD were 7.6 and 8.6 cM in the Korean and Icelander populations, respectively. The BLD was not uniformly distributed as chromosomes 1, 5, 10, 15, 20, and 22 showed less BLD (Figure 3). Summaries for significant LD regions can be found in Table 2.

The inference of population structure using the model with admixture showed that each individual had allele copies originating from K populations in equal proportions. Therefore, factors other than genetic distance, such as recent admixture, did not contribute to the level of BLD.

Table 2. Summary of significant linkage disequilibrium regions.

Chr loc	cM loc	Locus pairs	Distance (CM)	No. of alleles in locus 1	No. of alleles in locus 2	FET *P*-values
2q14.3	135.3	D2S347-D2S2271	5.2	11	12	0.014
2q33.3	204.9	D2S325-D2S2321	0.0	12	4	0.000
3p24.3	43.9	D3S3659-D3S1266	6.7	10	6	0.013
5q21.1	111.4	D5S495-D5S433	1.2	14	11	0.013
6q26	174.5	D6S1599-D6S1719	4.1	18	8	0.015
7q11.22	79.8	D7S502-D7S2476	5.6	13	10	0.007
8p23.1	22.4	D8S503-D8S552	1.2	9	7	0.003
9p23	23.2	D9S168-D9S269	1.1	8	11	0.003
9p21.2	51.5	D9S161-D9S1853	2.0	6	8	0.008
10p11.21	62.3	D10S1780-D10S578	2.6	9	4	0.012
11p12	57.4	D11S905-D11S4191	7.6	12	13	0.013
13q21.33	62.8	D13S1296-D13S156	6.9	14	10	0.009
14q22.3	47.8	D14S276-D14S980	4.0	9	18	0.014
17p11.2	46.4	D17S1857-D17S1824	5.5	9	15	0.012
17q25.3	125.0	D17S836-D17S784	4.6	8	8	0.009

Discussion

We found that Koreans had the same or a relatively lower level of BLD than the Icelander population. Along with Finland, Sardinia, and Japanese, the Icelander population is an expanded genetic isolate with modestly higher levels of BLD on many chromosomal regions than outbred populations (Dunning *et al*., 2000; Eaves *et al*., 2000; Taillon-Miller *et al*., 2000). Also, the level of BLD among Koreans was similar to levels of the CEPH families despite differences in study design, including the marker set and the sample size. We observed that ~3% of locus pairs within 4 cM displayed an FET *P* value < 0.01, compared with ~4% of markers identified by Huttley *et al*. (1999). We also observed non-uniformity of BLD between chromosomes. Our results and the genome-wide study of Service *et al*., (2001) of a recently identified genetic isolate, both found less BLD on chromosomes 1, 10, 15, and 22 than on other chromosomes. We assessed BLD with a recently constructed high-resolution recombination map. Thus, the effects of artifacts due to inaccurate designation of genetic mapping on the non-uniformity of BLD are probably smaller than in previous studies.

The migration of the Yayoi people leads us to expect a common genetic affinity between contemporary populations in Korea and Japan. Several previous genetic studies among East Asian populations, including a phylogenic study using polymorphic loci on the Y chromosome (Kim *et al*., 2000) and comparison studies of the distribution of HLA haplotypes (Park *et al*., 1998), revealed a closer genetic relationship between Japanese and Koreans than relationships between other surveyed Asian populations. These studies showed a higher degree of polymorphism in the distribution of HLA haplotypes among Koreans than in Japanese. Recently, it has been reported that the Japanese people have a relatively higher level of BLD than the Finnish, European American, and Sardinian populations, although the results were restricted to the X chromosome and mitochondirial DNA (Katoh *et al*., 2002).

We found that BLD was negatively related to the distance between the loci. Also, BLD in Koreans over large genetic distances is in agreement with results from previous empirical works. We used marker order and genetic distance based on the deCODE map and observed significant LD in locus pairs ordered in reverse on the Marshfield map. The LD based on the Marshfield map is more extensive than the LD based on the deCODE map.

The average heterozygosity in Koreans was slightly lower than in Caucasians, which probably indicates the genetic homogeneity of the Korean population. Our results provide a useful database for mapping studies among Koreans by showing that the Korean population, which is an expanded population with no evidence of admixture, has a BLD level comparable with the Icelander population. Therefore, the Korean population can be used for fine mapping of either complex traits or monogenic diseases.

Acknowledgement

This work was supported by The Center for Genome Research, Samsung Biomedical Research Institute.

References

Angius A, Melis PM, Morelli L, Petretto E, Casu G, Maestrale GB, Fraumene C, Bebbere D, Forabosco P, Pirastu M. Archival, demographic and genetic studies define a Sardinian sub-isolate as a suitable model for mapping complex traits. Hum Genet 2001;108:198-209

de la Chapelle A, Wright FA. Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. Proc Natl Acad Sci 1998;95:12416-23

Dunning AM, Durocher F, Healey CS, Teare MD, McBride SE, Carlomagno F, Xu CF, Dawson E, Rhodes S, Ueda S, Lai E, Luben RN, Van Rensburg EJ, Mannermaa A, Kataja V, Rennart G, Dunham I, Purvis I, Easton D, Ponder BA. The extent of linkage disequilibrium in four populations with distinct demographic histories. Am J Hum Genet 2000;67:1544-54

Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA. The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nat Genet 2000;25:320-3

Feimer NB, Service SK, Slatkin. Expanding on population studies. Nat Genet 1997;17:371-3

Gordon D, Simonic I, Ott J. Significant evidence for linkage disequilibrium over a 5-cM region among Afrikaners. Genomics 2000;66:87-92

Hanihara K. Dual structure model for the population history of the Japanese. Japan Review 1991;2:1-33

Huttley GA, Smith MW, Carrington M, O'Brien S. A scan for linkage disequilibrium across the human genome. Genetics 1999;152:1711-22

Ikari K, Onda H, Furushima K, Maeda S, Harata S, Takeda J. Establishment of an optimized set of 406 microsatellite markers covering the whole genome for the Japanese population. J Hum Genet 2002;46:207-10

Katoh T, Mano S, Ikuta T, Munkhbat B, Tounai K, Ando H, Munkhtuvshin N, Imanish T, Inoko H, Tamiya G. Genetic isolated in east Asia: A study of linkage disequilibrium in the X chromosome. Am J Hum Genet 2002;71:395-400

Kim N. People and language. In *Korea-its People and Cultures*, 1970, Hakwon-sa, Seoul

Kim W, Shin DJ, Harihara S, Kim YJ. Y chromosomal DNA variation in east Asian populations and its potential for inferring the peopling of Korea. J Hum Genet 2000;45:76-83

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeisson TE, Gulcher JR, Stefansson K. A high-resolution recombination map of the human genome. Nat Genet 2002;45:592-9

Park MH, Hwang YS, Park KS, Tokunaga K, Akaza T, Kim SI. HLA haplotypes in Koreans based on 107 families. Tissue Antigens 1998;51:347-55

Pritchard JK, Stephens M and Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945-59

Service SK, Ophoff RA and Freimer NB. The genome-wide distribution of background linkage disequilibrium in a population isolate. Hum Mol Genet 2001;10: 545-51

Taillon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen T, Cao A, Kere J, Pilia G, Rice JP, Kwok PY. Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nat Genet 2000;25:324-8