nature genetics

Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies

Susan Service[1], Joseph DeYoung[2], Maria Karayiorgou[3], J Louw Roos[4], Herman Pretorious[4], Gabriel Bedoya[5], Jorge Ospina[6], Andres Ruiz-Linares[5,7], António Macedo[8], Joana Almeida Palha[9], Peter Heutink[10,11], Yurii Aulchenko[12], Ben Oostra[12], Cornelia van Duijn[12], Marjo-Riitta Jarvelin[13,14], Teppo Varilo[15,16], Lynette Peddle[17], Proton Rahman[18], Giovanna Piras[19], Maria Monne[19], Sarah Murray[20], Luana Galver[20], Leena Peltonen[15,16], Chiara Sabatti[21,22], Andrew Collins[23] & Nelson Freimer[1,24,25]

The genome-wide distribution of linkage disequilibrium (LD) determines the strategy for selecting markers for association studies, but it varies between populations. We assayed LD in large samples (200 individuals) from each of 11 well-described population isolates and an outbred European-derived sample, using SNP markers spaced across chromosome 22. Most isolates show substantially higher levels of LD than the outbred sample and many fewer regions of very low LD (termed 'holes'). Young isolates known to have had relatively few founders show particularly extensive LD with very few holes; these populations offer substantial advantages for genome-wide association mapping.

There remains a great deal of uncertainty about how to design and carry out genome-wide association studies of common diseases[1]. It is not yet feasible to directly assay the entire genome for association between diseases and candidate variants. Current genome-wide studies must therefore be indirect, identifying associations by assaying SNPs in LD with untyped variants that are causal for a given trait. The extent and distribution of LD determines the number of SNPs required for such studies. Variations in demographic histories between populations generate substantial differences in the extent and distribution of LD[2] and therefore cause variability between populations in the number of SNPs required for association studies. It has been suggested that population isolates, particularly those founded recently, have longer stretches of LD than outbred populations[3–5]. Hence, in comparison with outbred populations, such isolates may require fewer markers for genome-wide association studies or may achieve better genome-wide coverage with equivalent numbers of markers. However, these possible advantages of isolates have not yet been quantified.

We evaluated LD in 11 populations that have been proposed as isolates, together with an outbred European-derived sample. The demographic histories of these isolates are well documented (**Supplementary Note** and **Supplementary Table 1** online). Additionally, the histories of these isolates differ according to factors hypothesized to influence LD; for example, some populations experienced admixture at the time of founding, whereas others differed markedly in the size of the founding population or in the time since their founding. Notably, we used a large sample size in this study (200 persons in each population); small samples generally overestimate the strength of LD, because they may not include rare, recombinant haplotypes[6,7].

We analyzed LD in these populations across the entire length of chromosome 22. We chose a marker density that previous smaller-scale surveys of LD suggested would adequately assay LD in the

[1]Center for Neurobehavioral Genetics and [2]Southern California Genotyping Consortium, University of California, Los Angeles, California 90095, USA. [3]Rockefeller University, New York, New York 10021, USA. [4]University of Pretoria, Department of Psychiatry & Weskoppies Hospital, Pretoria, Republic of South Africa 0001. [5]Laboratory of Molecular Genetics and [6]Department of Psychiatry, University of Antioquia, Medellin, Colombia. [7]The Galton Laboratory, Department of Biology (Wolfson House), University College London, London NW1 2HE, UK. [8]Institute of Medical Psychology, Faculty of Medicine, University of Coimbra, 3000-033 Coimbra, Portugal. [9]Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, 4704-553 Braga, Portugal. [10]Department of Human Genetics, Section of Medical Genomics and [11]Center for Neurogenomics and Cognitive Research, Vrije University and Vrije University Medical Center, 1081 BT, Amsterdam, The Netherlands. [12]Genetic Epidemiology Unit, Departments of Epidemiology & Biostatistics and Clinical Genetics, Erasmus Medical Center, 3000 DR Rotterdam, The Netherlands. [13]Department of Public Health Science and General Practice, University of Oulu, 90220 Oulu, Finland. [14]Department of Epidemiology and Public Health, Imperial College London, London W2 1PG, UK. [15]Department of Medical Genetics, University of Helsinki, 00290 Helsinki, Finland. [16]Department of Molecular Medicine, National Public Health Institute, Biomedicum Helsinki, 00290 Helsinki, Finland. [17]Newfound Genomics, 187 Lemarchant Road, St. John's, Newfoundland, Canada A1C 2H5. [18]Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1C 5S7. [19]Division of Haematology, 'San Francesco' Hospital, ASL3 Nuoro, Italy. [20]Illumina, Inc., 9885 Towne Centre Dr., San Diego, California 92121, USA. [21]Department of Human Genetics and [22]Department of Statistics, University of California, Los Angeles, California 90095 USA. [23]Human Genetics Research Division, University of Southampton, Southampton General Hospital, Tremona Road, Southampton SO16 6YA, UK. [24]The Jane and Terry Semel Institute for Neuroscience and Human Behavior and [25]Department of Psychiatry, University of California, Los Angeles, California 90095 USA. Correspondence should be addressed to N.F. (nfreimer@mednet.ucla.edu).

Received 8 November 2005; accepted 28 February 2006; published online 2 April 2006; doi:10.1038/ng1770

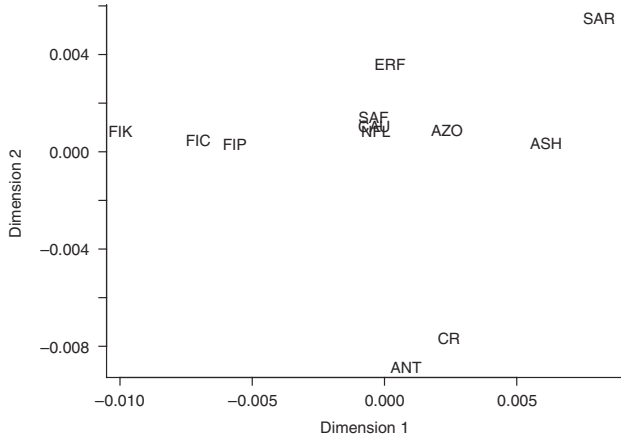

Figure 1 Result of non-metric multidimensional scaling applied to Nei's genetic distance between all pairs of populations. Population abbreviations are as in **Table 1**.

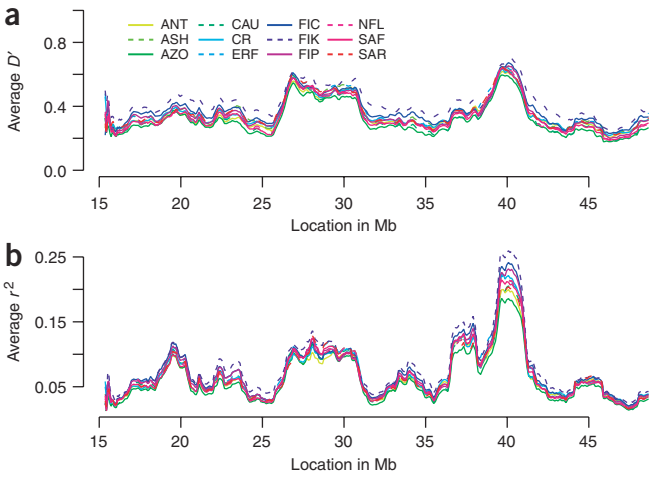

Figure 2 Distribution of linkage disequilibrium on chromosome 22. Average D' (**a**) and r^2 (**b**) coefficients plotted in sliding windows containing all markers separated by less than 500 kb in successive 1.7-Mb segments (1.6-Mb overlap). Population abbreviations are as in **Table 1**.

isolates[8–13] and that was similar to the density in currently available assays for genome-wide association studies[14]. The 2,486 SNPs that met the criteria for successful genotyping (see Methods) spanned 34.2 Mb of chromosome 22, at an average spacing of one marker every 13.8 kb (median, one marker per 8.5 kb). Seventy-eight percent of gaps between markers were under 20 kb, and only 3.5% of gaps were greater than 50 kb. These markers were in Hardy-Weinberg equilibrium in all populations (after correcting for multiple testing) and were not monomorphic in any population. Most markers (98%) had over 180 persons genotyped in each population.

The mean heterozygosity of the markers was similar in the different populations, ranging from 0.359 (Sardinia) to 0.373 (Antioquia). The 12 populations also had similar numbers of markers with minor allele frequency (MAF) $<10\%$, ranging from 10% of markers in Antioquia to 14% in Sardinia. We estimated the genetic distance between all pairs of populations with all marker data using Nei's D[15], and we applied nonmetric multidimensional scaling to this distance matrix to obtain a two-dimensional representation of interpopulation distances (**Fig. 1**). The outbred sample was most similar to the Newfoundland and Afrikaner populations and least similar to the populations of Antioquia, Kuusamo and Sardinia. The Costa Rican and Antioquian populations were as similar to each other as were the three Finnish samples.

We calculated D' and r^2 between all pairs of markers. The profile of LD for both of these measures was similar to that previously observed on chromosome 22 in European-derived samples[16]. Interpopulation differences were apparent from the moving average of D' and r^2 in sliding windows along the chromosome (**Fig. 2**), with the highest LD in the three Finnish samples and the lowest LD in the Azores sample.

We constructed LD maps[17] for all populations. In this method, distances are estimated in linkage disequilibrium units (LDUs) between adjacent pairs of SNPs. The LDU represents the product of the physical distance between markers and a parameter that reflects the exponential decline in the probability of association between markers according to physical distance. The LDU scale has additive distances, and these maps are analogous to recombination maps. The LD map is more robust than haplotype block methods to variation in marker density[18], and previous studies suggest that spacing markers evenly on the LDU map is a simple but effective strategy for determining appropriate marker density for association studies.

The profiles of the LD maps were very similar in all 12 populations examined, as has been observed previously[19,20] (**Fig. 3**). This pattern corresponds to regions of elevated and diminished recombination that are common to all populations, as evidenced by comparison with the linkage map[21]. Although the LD maps for the 12 populations had similar shapes, they differed substantially in their overall length (**Fig. 3** and **Table 1**). The shortest map (and hence, the population with the most extensive LD) was in the Finnish isolate from Kuusamo, and the longest maps were in the Azorean, Newfoundland, Afrikaner and outbred samples.

It has been previously observed that LD between markers is indeterminate when the LDU distance between them is ≥ 3 (ref. 19). Therefore, in common with previous studies using this methodology, we defined an LD 'hole' to be a chromosomal segment with a gap of ≥ 2.5 LDU between adjacent SNPs[22]. In our data set, the

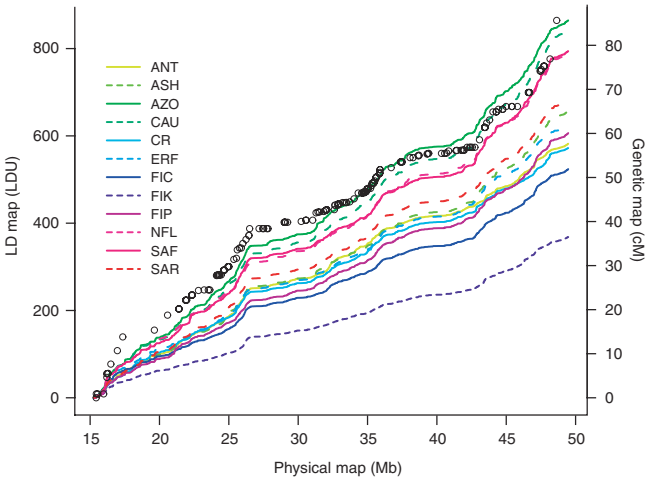

Figure 3 Quantitative comparison of the LD maps of the 12 populations and physical and genetic maps. Open circles represent the genetic map plotted against the physical map. Note the second *y*-axis showing the scale of the genetic map. Population abbreviations are as in **Table 1**.

Table 1 LD metric map length

Population	Length of LD map in LDU	s.e.m.	LDU/Mb ratio	Number of LD holes	Total size LD holes (kb)	Swept radius
ANT	581.9	34.2	17.01	31	1,092	164.62
ASH	656.5	26.1	19.19	26	975	155.84
AZO	864.5	26.2	25.27	84	2,709	130.51
CAU	845.1	29.6	24.70	84	2,574	129.55
CR	572.1	31.1	16.72	23	821	161.95
ERF	620.8	30.4	18.15	29	1,166	156.11
FIC	523.9	15.2	15.31	21	821	161.52
FIK	368.3	19.4	10.77	5	146	202.75
FIP	606.5	20.9	17.73	26	731	149.88
NFL	790.4	23.5	23.10	69	2,014	138.93
SAF	794.1	23.7	23.21	76	2,633	141.29
SAR	681.4	26.0	19.92	38	1,241	145.51

Third column shows the delete-d jackknife estimate of the s.e.m. of the length of the LD map. A hole is defined as a gap in the LD map of >2.5 LDU. Swept radius indicates the mean useful extent of LD for association mapping on chromosome 22 (in kb). ANT: Antioquia, Colombia; ASH: Ashkenazi; AZO: Azores; CAU: outbred European-derived sample; CR: Central Valley of Costa Rica; ERF: a village in southwestern Netherlands; FIC: early-settlement Finland; FIK: Finnish subisolate of Kuusamo; FIP: Finland nationwide; NFL: Newfoundland; SAF: Afrikaner; SAR: province of Nuoro in Sardinia.

median physical distance between markers defining a hole is larger than the median distance between non-hole markers (28.1 kb versus 8.4 kb), even though most of the larger gaps between markers do not contain LD holes (only 8.3% of intermarker gaps >50 kb contain LD holes). The correlation between the total number of LD holes in a 3-Mb window and the recombination rate for that window (downloaded from HapMap) is 0.83 ($n = 21$ comparison points). The number of holes is not uniformly distributed across the 12 populations ($\chi^2 = 197$, 11 degrees of freedom (d.f.), $P = 0$) and is positively related to the length of the LD maps (**Table 1**, **Fig. 4**).

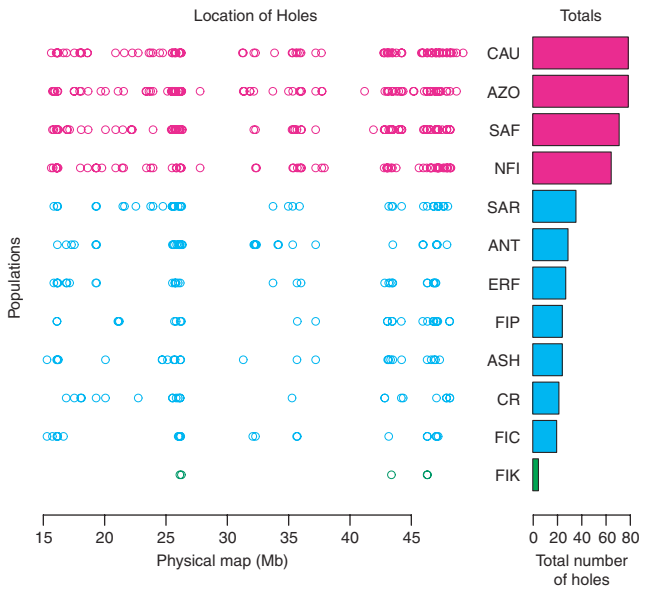

Figure 4 Distribution of LD holes (defined as a gap in the LD map of ≥2.5 LDU) by population. The null hypothesis of uniformity of the distribution of the total number of holes cannot be rejected for populations with the same color coding. Within each group, one cannot reject the hypothesis of a uniform distribution of LD holes (blue group: $\chi^2 = 6.9$, 6 d.f., $P = 0.33$; pink group $\chi^2 = 2$, 3 d.f., $P = 0.57$). Population abbreviations are as in **Table 1**.

As r^2 is the standard measure of whether the LD between two markers is sufficient for detecting associations, for each population, we calculated r^2 between pairs of markers that defined a hole. We compared this value to the r^2 value for a randomly selected pair of markers (matched to be approximately the same physical distance apart). The r^2 values are much higher between marker pairs that do not flank holes than between marker pairs that flank a hole. The results show that the holes are not simply segments with a large physical distance between markers (**Fig. 5a**). In fact, marker pairs within 100 kb of either side of an LD hole also have lower estimated r^2 than do non-hole areas of a similar size, indicating reduced power to detect association in a larger region around the hole-defining markers (**Fig. 5b**).

To assess the impact of increasing marker density of the number of LD holes, we constructed and evaluated chromosome 22 LD maps for the Centre d'Etude du Polymorphisme Humain (CEPH) samples from successive releases of HapMap data. The LD map constructed from release #19 of the HapMap data used nearly three times as many markers as the LD map constructed from release #13 of the HapMap data (27,060 SNPs versus 9,658 SNPs), yet the number of LD holes decreased only by one-half (30 LD holes versus 64 LD holes).

There are substantial differences between the populations in the overall magnitude of LD; the length of the LD map in the majority of the isolates was ~20–45% lower than in the outbred sample. The length of the LD map provides a guide to the minimum number of markers needed for association analyses, which require equal marker spacing on the LDU scale[19,20]. Considering the length of the LD maps only, our data suggest that association analyses in samples from Finland, the Dutch isolate, Costa Rica, Antioquia, Sardinia and the Ashkenazim would require at least 30% fewer markers than studies in outbred populations. The advantages of working in such isolates, which display several-fold fewer LD holes than the outbred sample, are also apparent in considering the additional markers needed to fill such holes. Marker sets currently being proposed for genome-wide LD mapping studies are spaced at similar densities to that used in the current analysis, suggesting that, in these isolates, such mapping sets may provide much better genome coverage for detecting LD than in outbred populations.

Two of the most important determinants of the extent of LD in a population are the effective population size (influenced strongly by the number of founders and expansion rate of the population) and the

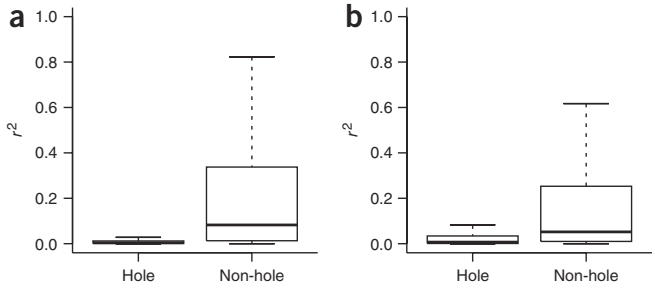

Figure 5 Box plots of r^2 values between markers surrounding LD holes and in non-hole regions. The dark line in the center of the box is the median, the edges of the box are located at the quartiles and the range of non-outlier points is indicated by the dashed lines. (**a**) r^2 between pairs of markers that define a hole and between randomly selected marker pairs a similar physical distance apart. (**b**) r^2 between pairs of markers within 100 kb of either marker that defined a hole and between pairs of markers in randomly selected regions.

time since founding. These factors affect, respectively, the number of different haplotypes present in the population and the number of opportunities for recombination between markers. In the Kuusamo isolate, the precise historical evidence documenting recent founding by a few individuals, subsequent population bottleneck and then rapid expansion is consistent with the observation that LD in this sample is much more extensive than in any of the other samples. Conversely, the relatively large number of founders of the Newfoundland and Azorean populations and low rates of expansion provide reasonable explanations for why less LD is seen in these samples. Admixture and inbreeding are two other possible explanations for the variation in length of LD map and number of LD holes among the isolates. Admixture at the time of founding is a possible source of enhanced LD in a population. Two isolates with well-documented initial admixture (Costa Rica and Antioquia[23]) display a greater extent of LD and fewer LD holes than the outbred sample. It is difficult to untangle, retrospectively, the relative contributions of admixture from other aspects of population history. Investigation of the extent and distribution of LD in Native American groups from the vicinities of Costa Rica and Antioquia may provide useful background information for the design of association studies in these two isolates. Inbreeding is likely to be higher in isolates than in outbred populations; however, in our study, marker heterozygosities were similar in the different populations, indicating that this was probably not a large factor in creating variability among populations in the extent of LD (**Supplementary Note**).

In summary, our comparison of 12 populations demonstrates that population isolates that have experienced marked recent population growth, starting from a small founding population, have higher overall levels of LD than do outbred populations, and they have far fewer regions of very low LD. Therefore, in such isolates it will be possible to conduct genome-wide association studies using fewer SNPs than in outbred samples. Similarly, these isolates will have fewer regions in which associations are undetectable by off-the-shelf marker sets now available for genome-wide studies.

METHODS

Samples. We genotyped 200 independent persons from each population. The samples from the population isolates were collected as part of genetic studies of common diseases. Subjects were chosen mostly from control groups, except in four populations, where they were drawn from the unrelated parents of affected individuals (**Supplementary Note**). The outbred sample consisted of 60 parents from CEPH trios (the same samples used in ref. 24) and 140 apparently healthy, unrelated self-declared European-Americans from the Coriell Institute Human Variation Panel. Informed consent was obtained from all study subjects, and the University of California, Los Angeles Institutional Review Board approved this study.

Before we began genotyping, we determined concentrations of genomic DNA samples by a Quant-iT PicoGreen (Invitrogen) dsDNA assay. Samples were then normalized to a concentration of 40–70 ng/μl. Samples containing <40 ng/μl genomic DNA were concentrated by partial dehydration to 90 °C using a 96-well heat block. Samples containing more than 70 ng/μl genomic DNA were diluted using 1× TE.

SNP markers, genotyping and scoring. Genotyping reactions were performed using Illumina BeadLab kit reagents and protocols[25]. We performed genotyping, using 250 ng of normalized genomic DNA from each sample as template for the Illumina GoldenGate genotype assay, on the Illumina BeadLab 1000 platform. Samples were biotinylated and then captured by avidin-coated paramagnetic particles. Pooled oligonucleotides were then hybridized to the bound genomic DNA. After removal of unbound or mismatched oligonucleotides by post-hybridization washes, a polymerase-mediated extension of the allele-specific oligonucleotides (ASOs) was performed to fill in the gap between the ASOs and locus-specific oligonucleotides (LSOs). A separate ligation step sealed the nick between the end of the extension and the LSO(s). These single-stranded templates were then used in a multiplex PCR reaction. Simultaneous PCR amplification of 1,536 loci per sample per oligonucleotide pool was performed using Titanium Taq (BD Biosciences) and 0.5 units UDG (Invitrogen) per 60 μl reaction.

After post-PCR washes, the amplification products were hybridized to derivatized beads arranged in a Sentrix Array Matrix (SAM). Each array contains 96 bundles of 50,000 fiber-optic filaments. Each filament is chemically etched to create a 3-μm well. Each well houses a single bead, and each bead contains the complementary address sequence of one of the oligonucleotide pool loci. With 50,000 beads and 1,536 bead types, each bead type is represented in the array approximately 30 times. Each genotype is, therefore, an average of 30 independent measurements of fluorescence intensities. Collection of fluorescence intensities was performed on a BeadArray Reader (Illumina) confocal fluorescence scanner. Data collection was accomplished using the BeadScan (Illumina) software package.

GenCall version 6.0.7 (Illumina) was used to cluster fluorescence intensities. Automated clustering was reviewed by hand and corrected where appropriate. The clustering for each population was reviewed and adjusted separately. After clustering, genotypes were called and quality scores assigned by GenCall. GTS reports (Illumina) were used to remove failed markers, failed samples and individual genotypes with quality scores below the 0.25 threshold from the final data set.

Error checking and selection of SNPs for LD analysis. Only the 2,589 markers scored in all 12 populations were considered for analysis. We then evaluated these markers by testing for Hardy-Weinberg equilibrium, estimating homozygosity and checking for mendelian transmission (**Supplementary Methods** online); the 2,486 markers that met all criteria were used in subsequent analyses.

Comparison of allele frequencies and calculation of pairwise LD measurements. Allele frequency distributions in each population were compared with the outbred sample using a χ^2 test. A marker was considered to have a significantly different allele frequency distribution from the outbred sample if $\chi^2 > 18$, to correct for multiple comparisons. The genetic distance between all pairs of populations was estimated using Nei's D[15], using all marker data and the GENDIST program of the PHYLIP package. Nei's D is estimated by

$$D = -\ln\left(\frac{\sum_m \sum_i p_{1mi} p_{2mi}}{\left(\sum_m \sum_i p_{1mi}^2\right)^{0.5} \left(\sum_m \sum_i p_{2mi}^2\right)^{0.5}}\right) \quad (1)$$

where m indexes the number of markers, i indexes the alleles at each marker and p_{1mi} is the frequency of allele i of marker m in population 1 (similarly for population 2). The matrix of genetic distances was used as a difference matrix in Kruskal's non-metric multidimensional scaling to obtain a two-dimensional representation of the proximity of populations to one another. Estimates of D' and r^2 between all marker pairs in each population were calculated using GOLD[26].

Construction of metric LD maps. For LD map construction[17] we computed the pairwise association, ρ, which equates to the absolute value of D', for pairs of SNPs and the corresponding information K_ρ under the null hypothesis that $\rho = 0$, such that $\rho^2 K_\rho = \chi^2_1$ (ref. 27). The predicted association between two SNPs separated by distance $\sum d_i$ (in kb) is modeled as

$$\rho = (1 - \mathrm{L})\mathrm{M}e^{-\sum \varepsilon_i d_i} + \mathrm{L}.$$

Here, d_i is the length in kb of the ith map interval between adjacent SNPs, ε_i is the exponential decline of association with distance and the intercept M is <1 if founding haplotypes have polyphyletic origin. The asymptote L reflects association at large distance that is not due to linkage and is predicted as the weighted mean deviation for a normal distribution[28]. The estimation of ε_i is achieved by composite likelihood using the informative pairwise data that include the ith interval. SNP pairs separated by more than 500 kb and pairs separated by more than 100 map intervals were excluded as uninformative. The

M parameter was re-estimated at regular intervals during map construction, which proceeds by iteratively computing the additive LD unit distances $\sum \varepsilon_i d_i$. The model is fitted from the composite likelihood

$$\exp[-\sum K_\rho(\hat{\rho} - \rho)^2/2],$$

where $\hat{\rho}$ is an estimate with prediction ρ and the summation is over all SNP pairs used for analysis. LDU maps were assembled in segments of ~500 markers, with a 25-marker overlap between adjacent segments and the overlap LDU distance averaged. These defaults give rapid construction of LD maps with the flanking intervals contributing efficiently to each interval estimate.

A delete-d jackknife estimate[29] of the s.e.m. around the length of each population's LD map was found by taking 100 samples of 180 persons and recomputing the LD map for each jackknife sample. The s.e.m. for *m* draws of size $n - d$ from a sample of *n* subjects is found by

$$\left\{\frac{n-d}{dm}\sum_{i=1}^{m}\left(L_{n-d,i} - \frac{1}{m}\sum_{j=1}^{m} L_{n-d,j}\right)^2\right\}^{0.5} \quad (2)$$

In our application, $m = 100$, $d = 20$, $n = 200$ and $L_{n-d,i}$ represents the length of the LD map from the *i*th sample of $n - d$ persons. We chose to use a jackknife estimate of the s.e.m. rather than a bootstrap estimate, because resampling persons with replacement (as in bootstrap sampling) may result in more LD than actually observed and may not accurately reflect the distribution of the length of the LD map.

LD hole distribution. To assess the significance of different total number of LD holes (defined as gaps between adjacent SNPs in the LD map of ≥2.5 LDU) across populations, we considered a model where each hole is an observation from a random element that can take on 12 possible values corresponding to the 12 populations under study. Fixing the number of holes to the observed 512, we considered the expected counts for each of the populations under the hypothesis of equal probability of all the populations and contrasted them with the observed totals for each population using a χ^2 test. We then analyzed population subgroups easily identifiable from the data to determine if the hypothesis of uniform probability would hold within these groups.

Recombination and LD holes. We downloaded the recombination data (calculated using the method of ref. 30) generated for the CEPH samples from release #16 of the HapMap data and compared the recombination rate from these data (averaged over windows of 3 Mb, with overlap of 1.5 Mb) to the total number of LD holes (in the same 3-Mb windows, with 1.5-Mb overlap).

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all of the study participants; the Center of Medical Systems Biology (CMSB); D. Ruano (Life and Health Sciences Research Institute (ICVS), University of Minho, Braga, Portugal); M.J. Soares, J. Valente and M.H. Azevedo (Instituto de Psicologia Médica, Faculdade de Medicina, Coimbra, Portugal); C. Pato and M.T. Pato (Center for Psychiatric and Molecular Genetics, and Department of Psychiatry, State University of New York, Syracuse, New York, and the Veterans Administration Medical Center, Washington, D.C.); A. Gabbas (Haematology Division and Bone Marrow Transplantation Unit, San Francesco Hospital, Nuoro, Italy); R. van Wyk, C. Botha and G. Valencia (Universidad de Antioquia) and P. Snijders for recruitment of subjects; M. Almonte and E. Slaten for genotyping assistance; R. Ophoff for comments on the manuscript and M. Levinson for assistance with graphics. We acknowledge funding from the Biotechnology and Biological Sciences Research Council (A.C.), the US National Institutes of Health (N.F., M.K., A.R.-L.), Newfound Genomics (P.R.), Colciencias (A.R.-L.), the Universidad de Antioquia (A.R.-L.), the National Alliance for Research in Schizophrenia and Depression (A.R.-L.), the Wellcome Trust (A.R.-L.), the Netherlands Organisation for Scientific Research (C. van D.), the Netherlands Diabetes Fund (C. van D.), the Netherlands Kidney Fund (C. van D.), the Netherlands Heart Foundation (C. van D.), the International Alzheimer Organisation (C. van D.), the Netherlands Brain Fund (C. van D.), the Center of Excellence of the Academy of Finland (L.P.), Nordic Center of Excellence in Disease Genetics (L.P.) and Biocentrum Helsinki, Finland (L.P.).

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Thomas, D.C., Haile, R.W. & Duggan, D. Recent developments in genomewide association scans: A workshop summary and review. *Am. J. Hum. Genet.* **77**, 337–345 (2005).
2. Stumpf, M.P.H. & Goldstein, D.B. Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium. *Curr. Biol.* **13**, 1–8 (2003).
3. Peltonen, L., Palotie, A. & Lange, K. Use of population isolates for mapping complex traits. *Nat. Rev. Genet.* **1**, 182–190 (2000).
4. Wright, A.F., Carothers, A.D. & Piratsu, M. Population choice in mapping genes for complex disease. *Nat. Genet.* **23**, 397–404 (1999).
5. Pritchard, J.K. & Przeworski, M. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* **69**, 1–14 (2001).
6. Teare, M.D., Dunning, A.M., Durocher, F., Rennart, G. & Easton, D.F. Sampling distribution of summary linkage disequilibrium measures. *Ann. Hum. Genet.* **66**, 223–233 (2002).
7. Tenesa, A. *et al.* Extent of linkage disequilbirum in a Sardinian sub-isolate: sampling and methodological considerations. *Hum. Mol. Genet.* **13**, 25–33 (2004).
8. Groenewald, J.Z., Liebenberg, J., Groenewald, I.M. & Warnich, L. Linkage disequilibrium analysis in a recently founded population: evaluation of the variegate porphyria founder in South African Afrikaners. *Am. J. Hum. Genet.* **62**, 1254–1258 (1998).
9. Gordon, D., Simonic, I. & Ott, J. Significant evidence for linkage disequilibrium over a 5-cM region among Afrikaners. *Genomics* **66**, 87–92 (2000).
10. Service, S.K., Ophoff, R. & Freimer, N.B. The genomewide distribution of background linkage disequilbrium in a population isolate. *Hum. Mol. Genet.* **10**, 545–551 (2001).
11. Hall, D., Wijsman, E.M., Roos, J.L., Gogos, J.A. & Karayiorgou, M. Extended inter-marker linkage disequilibrium in the Afrikaners. *Genome Res.* **12**, 956–961 (2002).
12. Varilo, T. *et al.* The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum. Mol. Genet.* **12**, 51–59 (2003).
13. Aulchenko, Y.S. *et al.* Linkage disequilibrium in young genetically isolated Dutch population. *Eur. J. Hum. Genet.* **12**, 527–534 (2004).
14. Maraganore, D.M. *et al.* High-resolution whole-genome association study of Parkinson disease. *Am. J. Hum. Genet.* **77**, 685–693 (2005).
15. Nei, M. Genetic distance between populations. *Am. Nat.* **106**, 283–292 (1972).
16. Dawson, E. *et al.* A first-generation linkage disequilibrium map of human chromosome 22. *Nature* **418**, 544–548 (2002).
17. Maniatis, N. *et al.* The first linkage disequilibrium (LD) maps: Delineation of hot and cold blocks by diplotype analysis. *Proc. Natl. Acad. Sci. USA* **99**, 2228–2233 (2002).
18. Ke, X. *et al.* The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Hum. Mol. Genet.* **13**, 577–588 (2004).
19. Tapper, W.J., Maniatis, N., Morton, N.E. & Collins, A. A metric linkage disequilibrium map of a human chromosome. *Ann. Hum. Genet.* **67**, 487–494 (2003).
20. De La Vega, F.M. *et al.* The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. *Genome Res.* **15**, 454–462 (2005).
21. Kong, X. *et al.* A combined linkage-physical map of the human genome. *Am. J. Hum. Genet.* **75**, 1143–1148 (2004).
22. Maniatis, N. *et al.* Positional cloning by linkage disequilibrium. *Am. J. Hum. Genet.* **74**, 846–855 (2004).
23. Carvajal-Carmona, L.G. *et al.* Genetic demography of Antioquia (Colombia) and the central valley of Costa Rica. *Hum. Genet.* **112**, 534–541 (2003).
24. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
25. Fan, J.-B. *et al.* Highly parallel SNP genotyping. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 69–78 (2004).
26. Abecasis, G.R. & Cookson, W.O. GOLD–graphical overview of linkage disequilibrium. *Bioinformatics* **16**, 182–183 (2000).
27. Collins, A. & Morton, N.E. Mapping a disease locus by allelic association. *Proc. Natl. Acad. Sci. USA* **95**, 1741–1745 (1998).
28. Morton, N.E. *et al.* The optimal measure of allelic association. *Proc. Natl. Acad. Sci. USA* **98**, 5217–5221 (2001).
29. Efron, B. & Tibshirani, R.J. *An Introduction to the Bootstrap* (Chapman and Hall, New York, 1993).
30. McVean, G.A. *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584 (2004).