

ARTICLE

An assessment of the Irish population for large-scale genetic mapping studies involving epilepsy and other complex diseases

Colm T O'Dushlaine¹, Ciara Dolan², Michael E Weale³, Alice Stanton², David T Croke⁴, Reetta Kalviainen^{5,6}, Kai Eriksson^{5,6}, Anne-Mari Kantanen^{5,6}, Rachel A Gibson^{7,8}, David Hosford^{7,8}, Sanjay M Sisodiya^{9,10}, Michael Gill¹, Aiden P Corvin¹, Derek W Morris¹, Norman Delanty^{11,12} and Gianpiero L Cavalleri^{*,11,12}

¹Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity College Dublin, Ireland; ²Molecular and Cellular Therapeutics, RCSI Research Institute, Royal College of Surgeons in Ireland, Dublin, Ireland; ³Institute for Genome Sciences and Policy, Center for Population Genomics and Pharmacogenetics, Duke University, Durham, NC;

⁴Department of Clinical Neurological Sciences and Molecular and Cellular Therapeutics, RCSI Research Institute Royal College of Surgeons in Ireland; ⁵Kuopio Epilepsy Center, Department of Neurology, University Hospital of Kuopio, Kuopio, Finland; ⁶Department of Medicine, University of Tampere; ⁷GlaxoSmithkline, Research Triangle Park, NC, USA;

⁸GlaxoSmithkline, New Frontiers Science Park, Harlow, UK; ⁹Department of Clinical and Experimental Epilepsy, Institute of Neurology, University College London, Queen Square, London; ¹⁰National Society for Epilepsy, Chalfont-St-Peter, Bucks, UK; ¹¹Division of Molecular and Cellular Therapeutics, Department of Clinical Neurological Sciences, RCSI Research Institute Royal College of Surgeons in Ireland, Ireland; ¹²Division of Neurology, Beaumont Hospital, Dublin, Ireland

The recent completion of the International HapMap Project has rapidly advanced our understanding of linkage disequilibrium (LD) in the human genome. Today, tagging SNPs (tSNPs) can be quickly and easily selected and consequently HapMap data are regularly applied to both small- and large-scale genetic mapping studies. However, to correctly interpret the application of HapMap-derived tSNPs in a genetic mapping study, an understanding of how well HapMap data represents LD in the study population is critical. The Irish population had not previously been characterised in this way. Here, we do so using a set of 4424 SNPs selected from 279 candidate genes for epilepsy genotyped across 1118 healthy individuals from the Irish, British, Finnish and Australian populations. By considering the Irish population alongside surrounding European populations, our results confirm that the HapMap European-derived population accurately estimates patterning of LD in European descent populations. The Irish population appears notably well matched to the European HapMap population, and is markedly similar to the neighbouring British population. Although we were unable to detect significant substructure within the Irish population (a favourable result for genetic mapping), methods for controlling stratification should always be incorporated. This analysis therefore confirms that the genetic architecture of the Irish population is well suited to the study of complex traits and that tSNPs selected using the HapMap data can be confidently applied to the Irish population.

European Journal of Human Genetics (2008) 16, 176–183; doi:10.1038/sj.ejhg.5201938; published online 31 October 2007

Keywords: genetic mapping; tagging SNPs; tag transferability; linkage disequilibrium; structure

*Correspondence: Dr GL Cavalleri, Division of Molecular and Cellular Therapeutics, Department of Clinical Neurological Sciences, RCSI Research Institute Royal College of Surgeons in Ireland, Ireland.

Tel: +3 531 402 2795; E-mail: gcavalleri@rcsi.ie

Received 4 August 2007; revised 6 September 2007; accepted 11 September 2007; published online 31 October 2007

Introduction

The recent completion of the HapMap^{1,2} project has rapidly advanced our understanding of linkage disequilibrium (LD) across the human genome. By analysing the

HapMap data, it is now possible to identify the set of variants that represents or effectively tags a particular gene or region, or, indeed, the entire set of common variants genotyped by the HapMap consortium. This information is used to maximise the power and efficiency of genetic association studies and influences SNP selection in candidate gene studies, as well as informing the design of large-scale whole-genome association (WGA) studies (see www.illumina.com for examples of WGA panels designed in this way). We have recently completed a large-scale candidate gene association study for epilepsy.³

One of the central questions surrounding the utility of HapMap project data in association study design is how well the pattern of LD in the four reference populations (ie, Caucasian, Han Chinese, Japanese and Yoruban) matches the study population, in other words, how transferable or portable are tagging SNPs (tSNPs) selected in HapMap to the particular population being examined. A number of analyses of haplotypic variation and portability of SNPs derived from HapMap have recently been carried out.^{4–9} These have highlighted how variation in European populations in particular is in close agreement with data from the HapMap Caucasian reference sample, consistent with expectation given the recent ancestral origin of European populations.

As a relatively isolated island population on the North Western extreme of Europe, the genetic architecture of the Irish population should be well suited to the study of complex traits. There is evidence to suggest that the population history of Ireland has been relatively undisturbed by secondary migrations.^{10,11} Studies of the Y chromosome and mtDNA point to a common genetic legacy in Ireland that probably extends back to the repopulation of the island after the last glacial maximum, around 9000 years ago, from population centres in the Iberian peninsula and south western France.^{10,12–14} The same studies suggest a minimal genetic influence of recent human migrations to Ireland over the last three millennia. Therefore, it is possible that the current Irish population exhibits reduced genetic diversity across the autosomes, and decreased heterogeneity within disease and treatment response-related alleles, all of which are critical assets for mapping clinically functional genetic variation.

Here, we assess for the first time the allelic architecture of the Irish autosomal genome and compare its characteristics with two other European populations (UK and Finland), and a population of European descent (Australia). In addition, we have evaluated each population in turn against the European HapMap sample. We have carried out this analysis to (a) assess the suitability of each population for use in genetic association studies where, as is often the case, the European HapMap sample is the reference panel used for choosing tSNPs and (b) inform on population structure between populations used in a recent genetic mapping effort in epilepsy.³

With the exception of the HapMap population, this study is based solely on data generated in healthy control samples which formed part of a large-scale genetic association study of epilepsy.³ In total, 4424 SNPs were genotyped in 1118 healthy individuals from Ireland, the UK, Finland and Australia. We investigated (a) the degree to which allele frequency and LD values in each population correlated with those observed in the HapMap reference sample, (b) the extent of haplotype diversity in each population, (c) the degree of differentiation between the populations and (d) the transferability of tags across each of the populations.

Methods

Samples

This study was approved by the relevant local Ethics Committees of the participating institutions. Participating individuals were recruited after written informed consent was obtained.

DNA from 1118 healthy individuals, sampled from four 'test' populations, was included in this study (Ireland, $n=358$; UK, $n=359$; Finland, $n=309$; and Australia, $n=92$).

The UK cohort consisted of randomly selected participants of the British 1958 Birth Cohort Collection (all participants of which were unrelated and of British ancestry). The Irish cohort consisted of randomly selected participants of the Allied Irish Bank blood pressure study, a cohort of current and retired bank employees and their spouses¹⁵ (all participants were unrelated and of Irish ancestry). The Australian cohort consisted of healthy unrelated individuals recruited from Australian blood donation centres (all individuals were of Australian Caucasian ancestry). The Finnish cohort consisted of individuals of self-identified Finnish origin recruited from various sites around southern Finland.

We used HapMap data from European and Yoruban populations (60 unrelated individuals from each population, HapMap data Rel21a/phase II Jan 07).

Selection of candidate genes and SNPs

Candidate genes were selected for use in a large-scale genetic association study of sporadic forms of epilepsy. The genes included members of the ligand and voltage-gated ion channels and pharmacogenetic candidates for the major antiepileptic drugs. The genomic location for each candidate gene was determined using the 'ref seq' track from build 35 of the UCSC genome browser (<http://genome.ucsc.edu>). We added 10 kb of sequence upstream and 5 kb of sequence downstream to each UCSC 'ref seq' defined genomic location to capture regulatory regions surrounding each candidate gene. In total, 279 candidate genes were selected (details available in Supplementary Table S1).

tSNPs were selected using the pairwise r^2 method as implemented in the Tagger tool (<http://www.broad.mit.edu/mpg/tagger/server.html>). HapMap Phase II data (HapMapdata Rel21a/phase II Jan 07) were used, and the r^2 threshold was set at 0.7. Only variants with a minor allele frequency greater than 5% were considered for tagging. The regions selected for tagging correspond to genomic addresses listed in Supplementary Table S1. Variation across the 279 candidate genes that was determined as functional was genotyped directly, regardless of LD or presence in HapMap data set. For a description of 'functionality' as used here, and the full list of SNPs considered in this study see Cavalleri *et al.*³

Genotyping

Genotyping was performed at Duke University using the Illumina GoldenGate platform (1536 SNP arrays). We applied stringent quality control criteria to prevent poor genotyping calls (details available on request).

Genotype check

Five samples on each 96-well DNA plate were replicated to ensure consistency in genotyping. The duplicate concordance rate was 99.99%. We ensured the accuracy of our DNA management and genotyping systems by comparing, across 12 SNPs, genotypes generated on the Illumina platform with those generated using a different technology (ABI Taqman or sequencing). In each comparison, different DNA aliquots of the same samples were used. All 12 SNPs matched exactly across the two different platforms.

Comparison of allele frequencies

SNP minor allele frequencies were compared between the reference European HapMap population and each of our four 'test' cohorts. Minor allele frequencies were calculated for each population using PowerMarker (v3.25).¹⁶ Symmetric SNPs (alleles = C,G or A,T) were removed because of known discordance between the results of allele assignment methodology used by Illumina and HapMap. Fortunately, only 9.5% of SNPs in our panel were symmetric, due to the more frequent occurrence of transitions (alleles = A,G or C,T) over transversions. For each pair of populations, the minor allele frequencies of SNPs common to both populations were compared. Significant differences between pairs of populations were assessed by calculating 95% confidence intervals on the differences in correlations via a Fisher's z transformation.

Calculation of linkage disequilibrium

PLINK¹⁷ was used to calculate LD (measured by pairwise r^2) between all pairs of SNPs for each of 279 genes for each of the four test populations (Ireland, UK, Finland, Australia) from the Illumina Golden Gate platform. The corresponding r^2 value was obtained for SNP pairs also present in the reference European HapMap population. This enabled a

one-to-one comparison of r^2 values between pairs of SNPs in each test population and the European HapMap population. Results from all genes were pooled to provide a summary estimate of the extent of correlation between the r^2 value for a given test population and that of the HapMap group. As with the comparisons of allele frequencies, a Fisher's z transformation was used to check for significant differences between populations in terms of their correlation with HapMap. The relationship between LD and distance was also investigated for each chromosome and each population using Haploview¹⁸ under default settings. The results were pooled to provide a summary over all chromosomes for each population.

Analysis of haplotype diversity

Haplotype blocks were defined using the method described by Gabriel *et al.*¹⁹ as implemented in HaploView. Blocks were first estimated on a per gene basis in the European HapMap sample. Haplotype diversity was then calculated for each haplotype block as $h = (1 - \sum x_i^2)n/(n-1)$, where x_i is the haplotype frequency and n is the number of genotyped individuals.²⁰ The same haplotype block structure identified in the HapMap reference populations was applied to each of the four test populations. Haplotype diversity was then calculated for these blocks, allowing a direct comparison of haplotype diversity to be made between all populations.

Analysis of F_{st}

The extent of population isolation was estimated, using PowerMarker (v3.25),¹⁶ by calculating pairwise F_{st} values between all populations examined in this study. For both this analysis and the principal components analysis (PCA) (see below), only 'independent' SNPs were considered. Operationally, we define independence as a set of SNPs within which all sample estimates of pairwise r^2 are less than 0.2. We further reduced this set by removing symmetric SNPs (see 'Comparison of allele frequencies' section) as due to differing methods of allele assignment, symmetric SNPs show high levels of discordance between HapMap and Illumina genotyping platforms. A final total of 4020 SNPs were included in the F_{st} and PCA.

Principal components analysis

PCA was applied to the pairwise genetic distance matrix for all individuals to infer continuous axes of variation using the ADE4 software for the R statistical package.²¹ This was used to summarise the variance across populations used in our analysis into trends of maximum relevance known as principal components. The pairwise genetic similarity matrix was created using the identity-by-state clustering function in PLINK.¹⁷

Genomic inflation factors

Genomic control is a statistical approach that can be used to control for population heterogeneity.²² We estimated genomic inflation factors for Irish, UK, Finnish and Australian populations using the full set of 4020 SNPs. Case and control groups of equal size were created for each population by random assignment. The genomic inflation factor was calculated as described in Devlin *et al.*²² This process was repeated 100 times. Reported lambda values represent the mean from these 100 runs.

Results

Correlation of SNP frequencies

Using the HapMap data as the reference, allele frequencies were compared between the European HapMap reference population and each of our test populations for all relevant SNPs (Supplementary Figure 1). The Pearson's correlations for the European reference vs each of our test populations was: 0.957 (Ireland), 0.960 (UK), 0.908 (Finland) and 0.942 (Australia) ($P < 0.0005$ in all cases). To test whether any one of these correlation coefficients represented significant improvements over others (eg, is UK/HapMap significantly better correlated than Finland/HapMap), we calculated 95% confidence intervals on the differences in correlations via a Fisher's z transformation. We found that, while all correlations were themselves significant, none represented significant improvements over other correlations.

Correlation of LD between test populations and the reference HapMap European population

The pairwise disequilibrium coefficient (r^2) was calculated between all pairs of SNPs within each gene, for each test population separately. Similarly, r^2 was calculated for each of the corresponding SNP pairs in the European HapMap reference sample. All pairwise r^2 were pooled across genes within each test population and correlated with the European HapMap r^2 data (Figure 1). These correlations are summarised in Table 1. The Irish group showed the highest mean correlation with the HapMap data followed by UK, Finland and Australia.

Confidence intervals calculated using Fisher's z transformation on correlation values failed to show significant differences between pairs of populations.

The range of LD across each population

D' and r^2 measures of LD were calculated for pairs of SNPs with a minimum LOD score of 3 (to guard against spurious

Table 1 Summary statistics for r^2 LD comparison between HapMap reference and test populations

Population	N	Pearson	P
Ireland	25 325	0.93	<0.0005
United Kingdom	25 165	0.92	<0.0005
Finland	19 736	0.88	<0.0005
Australia	25 191	0.90	<0.0005

N is the number of SNP pairs overlapping between data sets (test population and HapMap European reference), P is the significance of the Pearson correlation.

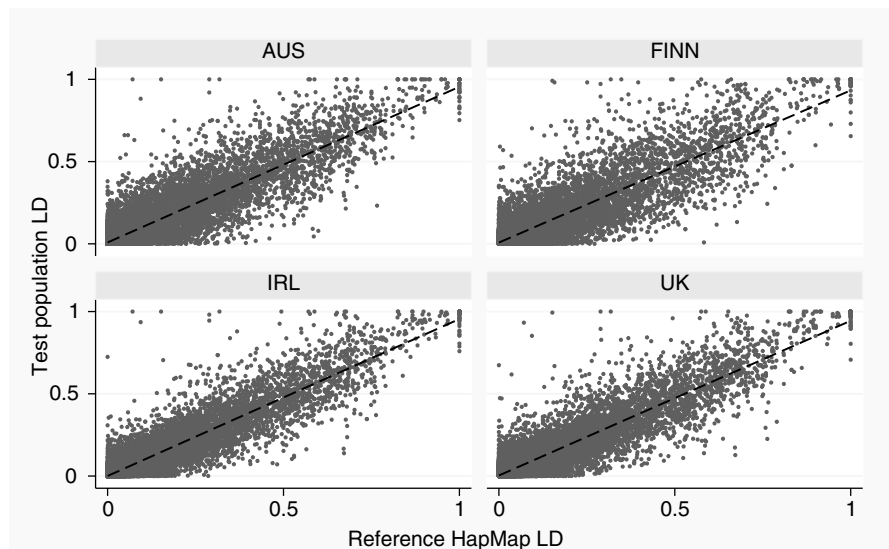


Figure 1 The correlation in r^2 values between each test population and the HapMap European reference. All values for all genes were pooled before estimating the overall correlation between a given population and the HapMap European reference panel. Each point represents the LD values for a pair of SNPs existing in both populations. The regression line is also shown. AUS, Australia; FINN, Finland; IRL, Ireland; UK, United Kingdom. Correlation coefficients were as follows: IRL: 0.93, UK: 0.92, AUS: 0.90, FINN: 0.88.

LD values) and within 100 kb of each other. As the estimation of both r^2 and D' are influenced by sample size, we restricted our analysis to a fixed number of individuals in each of our populations (with the exception of the HapMap populations). This number was set at 92, the number of individuals in our smallest population (Australia). Individuals were selected at random. A threshold of 100 kb was assigned on the basis that very few observations were made beyond 100 kb, owing to the candidate gene nature of the data set. These data were pooled across all pairs of SNPs and all chromosomes (see Table 2). As the range of LD is largely determined by the patterning of recombination, we focused on D' (as opposed to r^2 which is influenced by allele frequency), for which following mean values were observed: Yoruban HapMap (0.863), European HapMap (0.880), UK (0.893), Ireland (0.896), Australia (0.896) and Finland (0.902).

Haplotype diversity

Correlations between haplotype diversity estimates for the European HapMap reference population and each of our test populations are shown in Supplementary Figure 2. The Spearman's correlations for these comparisons are presented in Table 3. The overall picture shows that test populations are highly concordant with the European reference in terms of haplotype diversity. Haplotype

Table 2 Summary statistics for analysis of range of LD

Population	N	Mean r^2	SD r^2	Mean D'	SD D'
Ireland	3532	0.358	0.232	0.896	0.150
United Kingdom	3605	0.355	0.231	0.893	0.152
Finland	3510	0.359	0.235	0.902	0.148
Australia	3642	0.368	0.230	0.896	0.149
<i>HapMap reference populations</i>					
CEU	10133	0.335	0.213	0.880	0.157
YRI	4347	0.330	0.219	0.863	0.168

CEU, HapMap European population; YRI, HapMap Yoruban population. Values presented reflect means calculated using a maximum distance between SNPs of 100 kb and a LOD score cutoff of 3.

Table 3 Summary statistics for analysis of haplotype diversity

Population	Mean diversity	N	Spearman Rho vs HapMap	P
Ireland	0.579	283	0.955	<0.0005
United Kingdom	0.572	255	0.956	<0.0005
Finland	0.567	235	0.927	<0.0005
Australia	0.579	275	0.937	<0.0005

Mean diversity refers to the mean haplotype diversity in the population in question (see Methods for details). The Spearman correlation refers to the correlation between haplotype diversity in the population in question and that in the HapMap CEPH reference sample, the significance of which is given by P .

frequencies from the Irish and UK populations are in closest agreement with the frequencies estimated in the European reference population. The Finnish group is least correlated with the European reference. Mean diversity values within each population are also presented, highlighting a minimum diversity in the Finnish population, followed by the British population. The Irish and Australian populations had similar diversity estimates.

Analysis of population structure

The extent to which each of our test populations differed from each other, as measured by F_{st} , was also investigated. Each test population was compared individually to the reference HapMap European and Yoruban populations and also in a pairwise fashion to each other (Figure 2). As expected, highest F_{st} values were observed when compared Yoruban with the Irish, British, Finnish and Australian populations. The European HapMap and Finnish populations were slightly more distant from each other than was the European HapMap population from any of the other test populations (Ireland, UK and Australia).

Figure 3 displays axes 1 and 2 of the PCA for all individuals. Axis 1, accounting for 58% of the variance, is driven by the Yoruban group and differentiates African from European populations. Axis 2, accounting for 1.2% of the variance appears to be driven by a North–South axis of European genetic structure and differentiates the Finnish population from the highly overlapping Irish, British and Australian populations. Axis 3 (not illustrated in Figure 3) accounts for just 0.8% of the total variance and is driven by a separation of the European HapMap population from the Irish, British and Australian populations.

Population structure was also examined within the Irish population using data on grandparental province of origin. Individuals where all four grandparents originated from the same province ($n=167$) were assigned to one of the four provinces of Ireland (Ulster (North), $n=12$; Munster (South), $n=51$; Leinster (East), $n=86$; Connaught (West), $n=18$). Using PCA for these individuals, no evidence was found supporting the existence of distinct clusters by province (Supplementary Figure 3).

Despite being unable to clearly illustrate genetic clusters within the Irish population or indeed between the Irish, British and Australian populations, all human populations are to some degree substructured genetically. Methods for estimating and controlling this substructure should always be incorporated into population-based genetic mapping studies. Genomic inflation factors²² for the Irish, British, Finnish and Australian populations were calculated at 1.024, 1.023, 1.023 and 1.037, respectively.

Tag transferability

The portability of tSNPs was also assessed. For all SNPs genotyped in the test populations, the European HapMap reference sample was screened for pairwise r^2 values >0.7 .

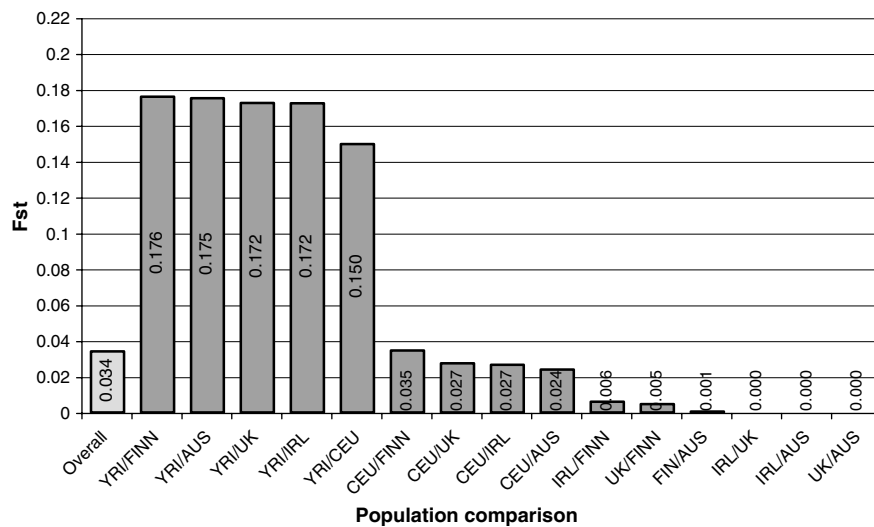


Figure 2 Population differentiation as measured by F_{st} . F_{st} values were calculated between pairs of populations (dark grey) and across all populations considered together (light grey). AUS, Australia; FINN, Finland; IRL, Ireland; UK, United Kingdom; CEU, HapMap European population, YRI, HapMap Yoruban population.

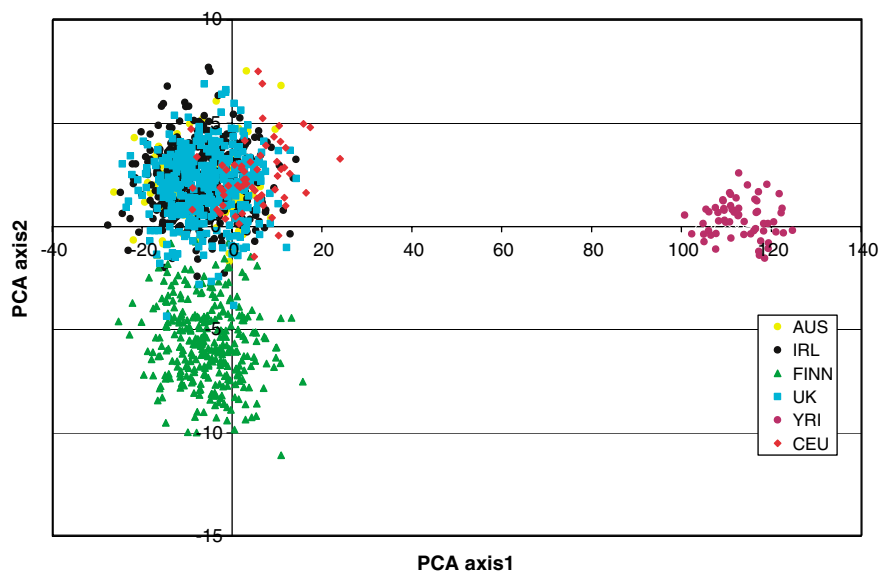


Figure 3 Principal components analysis. AUS, Australia; FINN, Finland; IRL, Ireland; UK, United Kingdom; CEU, HapMap European population; YRI, HapMap Yoruban population.

Depending on the population, between 32 and 37 SNP pairs were identified that reached this threshold. Each of these SNP pairs represents a SNP/proxy tSNP combination in both the European HapMap sample and in our test populations. The relatively small number of SNP pairs that reach this r^2 threshold reflects the fact that SNPs were originally selected in this study as tag SNPs in epilepsy candidate genes³ and would not, therefore, be in high LD with each other. The reason a number of high LD SNP pairs do occur is that in addition to selected tag SNPs, a number of functional SNPs were selected for analysis at these genes, irrespective of their LD relationship with neighbouring

SNPs.³ The r^2 values for these SNP pairs in the European HapMap population were correlated to the r^2 values calculated for the same SNP pairs in each of the four test populations (Figure 4). Overall, the Irish ($P=0.0004$) and UK ($P=0.0002$) populations had the highest correlations for portable tags, with the Finnish population showing the least portability.

Discussion

This study represents the first attempt to explore and describe the genetic architecture of the Irish autosomal

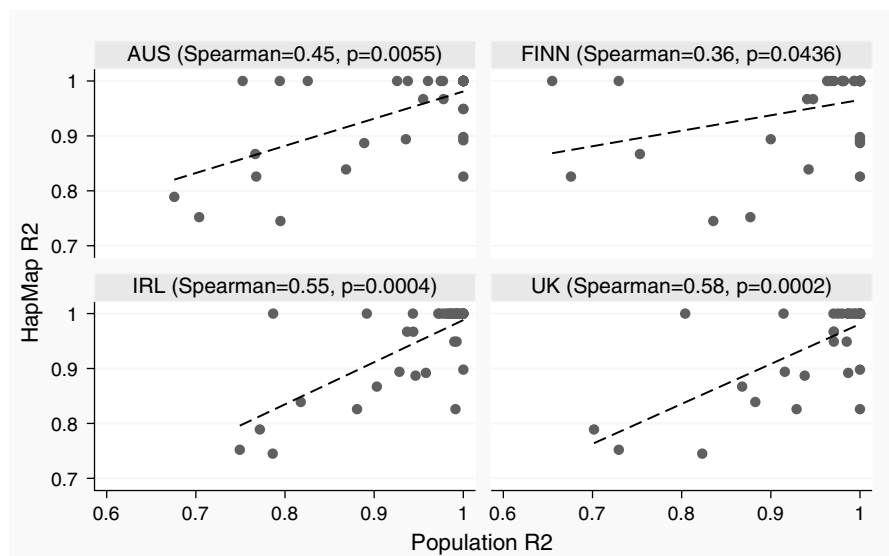


Figure 4 Comparisons of tag portability between HapMap and each of our populations. The regression line is also highlighted. AUS, Australia; FINN, Finland; IRL, Ireland; UK, United Kingdom.

genome. By considering the Irish population in the context of the European HapMap and other neighbouring European populations, this study reveals a number of important insights that are relevant to conducting genetic mapping in Irish ancestry populations.

We observed tight correlations between the Irish population and the European HapMap population in terms of minor allele frequency, LD values and haplotype diversity. Similar tight correlations were observed in the other test populations we considered. We also observed a close genetic relationship between the Irish population and neighbouring British and Australian populations. Furthermore, results presented here confirm previous observations that the patterning of LD observed in the European HapMap population represents to a large degree that observed in other European populations.^{4,6,8} Interestingly, among the populations we studied, the Irish population appeared the most highly correlated with European HapMap population in terms of LD (as measured by r^2), suggesting high portability of tSNPs, an observation confirmed by our tSNP transferability assessment. We note, however, that reported differences in correlations were not statistically significant.

Our results suggest the Irish population is genetically indistinguishable from the neighbouring British population using 4424 SNPs. The Irish population does not appear to show the genetic characteristics of a population isolate. However, this does not rule out the possibility of localised structuring of LD in Ireland, in particular in remote rural regions or islands off the mainland. It is also probable that some degree of genetic structure between Britain and Ireland will be resolved as more ancestral informative

markers are identified as has been the case within British populations.^{23,24}

Contrary to studies of the Y chromosome, and in agreement with a previous report²⁵ our assessment failed to detect evidence of population structure within Ireland even when only individuals with four grandparents from the same province were considered.^{10,25,26} It is reasonable to assume that by considering a larger number of SNPs (such as those currently available on WGA platforms) this picture would change as Irish specific ancestry informative SNPs begin to emerge. However, it is likely that any substructuring within the Irish population, although clear, would be slight in terms of magnitude. That this observation differs somewhat from results reported from Y chromosome would reflect the differing natures of the autosomal and Y chromosome systems, with the recombining autosomes being much less sensitive to substructure.

PCA using 4424 SNPs revealed the Irish, British and Australian populations to be indistinguishable from each other. However, in agreement with patterns observed with similar and other genetic systems, the Finnish population showed clear separation consistent with a general North–South axis of European genetic structure.^{27,28} The closeness of the Australian population to the British and Irish populations would reflect the history of emigration from the British and Irish Isles to Australia. Regardless of the ability to resolve structure within and between populations, methods for controlling substructure should always be incorporated into population-based genetic mapping studies. Our estimate of the genomic inflation factor²² for the Irish population ($\lambda = 1.024$) is similar to that observed in the other European populations studied here.

Our previous multi-centre study of epilepsy employing the same four 'test' populations indicated the possibility of numerous SNPs contributing to epilepsy in an apparent population-specific manner.³ The degree of similarity observed here between the Irish, British and Australian populations would suggest that apparent population-specific effects are probably related to phenotype (insufficient resolution) as opposed to genes. However, the structure observed between Finland and the other three 'test' populations highlights the importance of replicating association results in populations of similar descent, in particular in cases where structure has been characterised between the population in which the association was first reported and that in which replication is being considered.

In summary, our results suggest that the Irish population is very well matched to the European HapMap population and markedly similar to the neighbouring British population. Although we were unable to detect significant substructure within the Irish population (a favourable result for genetic mapping), methods for controlling stratification should always be incorporated. tSNPs selected using the HapMap data can be confidently applied to the Irish population.

Acknowledgements

This work was partly supported by PRTL3 (ND, CD, GLC) and a MRC grant (SMS, DBG). The work was partly undertaken at UCLH/UCL who received a proportion of funding from the UK Department of Health's NIHR Biomedical Research Centre's funding scheme. GSK Research and Development provided funding for the recruitment of the Finnish subjects. We thank John C Mulley of the Department of Medicine, University of Adelaide, Australia, for contributing Australian DNA samples. We thank two anonymous reviewers for their constructive comments.

References

- 1 The International HapMap Consortium: A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; **449**: 851–861.
- 2 The International HapMap Consortium: A haplotype map of the human genome. *Nature* 2005; **437**: 1299–1320.
- 3 Cavalleri GL, Weale ME, Shianna KV *et al*: Multicentre search for genetic susceptibility loci in sporadic epilepsy syndrome and seizure types: a case-control study. *Lancet Neurol* 2007; **6**: 970–980.
- 4 Conrad DF, Jakobsson M, Coop G *et al*: A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* 2006; **38**: 1251–1260.
- 5 Gonzalez-Neira A, Ke X, Lao O *et al*: The portability of tagSNPs across populations: a worldwide survey. *Genome Res* 2006; **16**: 323–330.
- 6 Ribas G, Gonzalez-Neira A, Salas A *et al*: Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. *Hum Genet* 2006; **118**: 669–679.
- 7 Smith EM, Wang X, Littrell J *et al*: Comparison of linkage disequilibrium patterns between the HapMap CEPH samples and a family-based cohort of Northern European descent. *Genomics* 2006; **88**: 407–414.
- 8 Mueller JC, Lohmussaar E, Magi R *et al*: Linkage disequilibrium patterns and tagSNP transferability among European populations. *Am J Hum Genet* 2005; **76**: 387–398.
- 9 de Bakker PI, Burtt NP, Graham RR *et al*: Transferability of tag SNPs in genetic association studies in multiple populations. *Nat Genet* 2006; **38**: 1298–1303.
- 10 Hill EW, Jobling MA, Bradley DG: Y-chromosome variation and Irish origins. *Nature* 2000; **404**: 351–352.
- 11 McEvoy B, Richards M, Forster P, Bradley DG: The Longue Duée of genetic ancestry: multiple genetic marker systems and celtic origins on the Atlantic facade of Europe. *Am J Hum Genet* 2004; **75**: 693–702.
- 12 Helgason A, Hickey E, Goodacre S *et al*: mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 2001; **68**: 723–737.
- 13 Semino O, Passarino G, Oefner PJ *et al*: The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 2000; **290**: 1155–1159.
- 14 Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB: Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci USA* 2001; **98**: 5078–5083.
- 15 O'Brien E, Murphy J, Tyndall A *et al*: Twenty-four-hour ambulatory blood pressure in men and women aged 17–80 years: the Allied Irish Bank Study. *J Hypertens* 1991; **9**: 355–360.
- 16 Liu K, Muse SV: PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 2005; **21**: 2128–2129.
- 17 PLINK. [<http://pngu.mgh.harvard.edu/~purcell/plink/>].
- 18 Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- 19 Gabriel SB, Schaffner SE, Nguyen H *et al*: The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–2229.
- 20 Nei M, Tajima F: DNA polymorphism detectable by restriction endonucleases. *Genetics* 1981; **97**: 145–163.
- 21 The R Project for Statistical Computing. [<http://www.r-project.org/>].
- 22 Devlin B, Roeder K: Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.
- 23 Vitart V, Carothers AD, Hayward C *et al*: Increased level of linkage disequilibrium in rural compared with urban communities: a factor to consider in association-study design. *Am J Hum Genet* 2005; **76**: 763–772.
- 24 The Wellcome Trust Case-Control Consortium: Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661–678.
- 25 Dolan C, O'Halloran A, Bradley DG *et al*: Genetic stratification of pathogen-response-related and other variants within a homogeneous Caucasian Irish population. *Eur J Hum Genet* 2005; **13**: 798–806.
- 26 Moore LT, McEvoy B, Cape E, Simms K, Bradley DG: A y-chromosome signature of hegemony in gaelic ireland. *Am J Hum Genet* 2006; **78**: 334–338.
- 27 Bauchet M, McEvoy B, Pearson LN *et al*: Measuring European population stratification with microarray genotype data. *Am J Hum Genet* 2007; **80**: 948–956.
- 28 Cavalli-Sforza LL: *The History and Geography of Human Genes*. Princeton University Press, 1994.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)