View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Griffith Research Online

Linkage Disequilibrium Analysis in the Genetically Isolated Norfolk Island Population

C. Bellis[1], R.A Lea[1,2], D. Burgner[3], M. Ovcaric[1], S.C Heath[4], J. Blangero[5] and L.R Griffiths[1]

[1] Genomics Research Centre School of Medical Science Griffith University Gold Coast

PMB 50, GCMC Bundall 9726.

e-mail: l.griffiths@griffith.edu.au

Tel.: +61 7 5552 8808, Fax: +61 7 5552 8908

[2] Institute of Environmental Science and Research Ltd, School of Biological Sciences, Victoria University of Wellington, New Zealand

[3] School of Paediatrics and Child Health, University of Western Australia, Perth, Australia

[4] Centre National de Génotypage, 2 Rue Gaston Cremieux, Evry, France.

[5] Southwest Foundation for Biomedical Research, San Antonio, Texas, USA

Running title: Norfolk Island LD analysis

Keywords: Norfolk Island, linkage disequilibrium, genetic isolate, iNOS, Xq13.3

Word count: 3544

ABSTRACT

Norfolk Island is a genetic isolate, possessing unique population characteristics that could be utilized for complex disease gene localization. Our intention was to evaluate the extent and strength of linkage disequilibrium (LD) in the Norfolk isolate by investigating markers within Xq13.3 and the *NOS2A* gene encoding the inducible nitric oxide synthase (iNOS). A total of six microsatellite markers spanning approximately 11-Mb were assessed on chromosome Xq13.3 in a group of 56 males from Norfolk Island. Additionally, three single nucleotide polymorphisms (SNPs) localizing to the *NOS2A* gene were analysed in a subset of the complex Norfolk pedigree. With the exception of two of the marker pairs, one of which is the most distantly spaced markers, all the Xq13.3 marker pairs were found to be in significant LD indicating that LD extends up to 9.5-11.5Mb in the Norfolk Island population. Also, all SNPs studied showed significant LD in both Norfolk Islanders and Australian Caucasians, with two of the marker pairs in complete LD in the Norfolk population only. The Norfolk Island study population possesses a unique set of characteristics including founder effect, geographical isolation, exhaustive genealogical information and phenotypic data of use to cardiovascular disease (CVD) risk traits. With LD extending up to 9.5-11Mb, the Norfolk isolate should be a powerful resource for the localization of complex disease genes.

INTRODUCTION

Genetic investigation of inherited diseases has matured from the mapping of Huntington's disease in the early 1980's to a future which will potentially provide information pertaining to many thousands of genetic variants underlying both simple and complex disease. There is no doubt that central to the success of this genomic era has been the coordinated efforts which have resulted in firstly, the complete sequencing of the human genome and its public availability to researchers globally and secondly, development of high-throughput technology platforms amenable to dense marker maps in large study populations. However, population choice is of primary concern in any gene mapping study. Genetic isolates provide a potentially powerful sample population for disease gene mapping of complex multifactorial traits, due to the combined effects of geographical isolation, limited variation in environmental influences and purported higher levels of linkage disequilibrium (LD). Additionally, such populations generally arise from a small number of founding members, possibly from quite diverse cultural backgrounds, therefore introducing genetic admixture affects. There are many population isolate studies being investigated - the study populations themselves as diverse as the genetic disorders they are being used to research. European researchers have focused their efforts on historically and culturally distinct populations from Scandinavia including Finland (Peltonen *et al*, 1999; Varilo *et al*, 2000; Varilo *et al*, 2003; Wessman *et al*, 2002) and Iceland (Gulcher and Stefansson, 1998; Helgason *et al*, 2003), Mediterranean regions of Italy, especially Sardinia (Angius *et al*, 2002a; Angius *et al*, 2001; Angius *et al*, 2002b; Eaves *et al*, 2000; Falchi *et al*, 2004; Pugliatti *et al*, 2003; Tenesa *et al*, 2004; Zavattari *et al*, 2000) and Corsica, France (Latini *et al*, 2004). North American isolates have mainly been concerned with large extended pedigrees from for example, Hutterites (Abney *et al*, 2000; Abney *et al*, 2001; Abney *et al*, 2002; Newman *et al*, 2003;

Newman *et al*, 2004; Ober *et al*, 2001; Weiss *et al*, 2006), while several remote Polynesian populations are presently involved in disease gene mapping studies (Bonnen *et al*, 2006; Han *et al*, 2002; Kayser *et al*, 2003; Murray-McIntosh *et al*, 1998; Redd *et al*, 1995; Shmulewitz *et al*, 2001; Tsai *et al*, 2004; Wijsman *et al*, 2003). These studies indicate that the population isolate approach to identifying disease loci presents an important framework as a means of identifying genes involved in complex multifactorial diseases.

This study focused on the genetic isolate of Norfolk Island. The Island of Norfolk is situated approximately 1,700 kilometers northeast of Sydney, on the Norfolk Ridge, which runs from New Zealand to New Caledonia. It was initially a penal colony of the British Empire until the last convict settlers were transported to Tasmania in the 1850s (Hoare, 1999). At this time, the descendants of the Bounty mutineers and Tahitian women who were previously inhabitants of Pitcairn Island relocated to Norfolk, with a total population 194 (40 men, 47 women, 54 boys, 53 girls). This small population originated from nine paternal ('*Bounty*' mutineers) and twelve maternal (Tahitian) lineages, although, due to the violent dynamics of the colony, only one of the Bounty mutineers (John Adams) survived to inhabit Norfolk (Edgecombe, 1999; Hoare, 1999). Interestingly, approximately 80 percent of the male participants included in the present Xq13.3 LD study are directly related to John Adams, four of these individuals possess an unbroken patrilineage to Adams.

Norfolk's history is particularly well-documented especially since anthropologists from the Island have maintained an exhaustive genealogical history in the form of a single large family pedigree comprised of ~6500 individuals who have contributed to the present-day population. Of the ~1200 current permanent residents, up to 80% can trace their heritage back to the Island's initial founders. In addition, severe immigration and

quarantine legislation restrict new founders from migrating to Norfolk. This, together with its isolation from other populations, makes Norfolk a potentially valuable resource for the mapping of genes involved in the pathogenesis of complex, yet common disorders such as hypertension, diabetes and obesity, which are known to be prevalent in the Polynesian admixture populations of the South Pacific (Abbott *et al*, 2001). Norfolk Island is an isolated community with a strict quarantine and an unusual health care system. Inhabitants are not covered by Australian or New Zealand health systems consequently, health is administered by the Norfolk Government. Such health care is currently directed towards emergency situations with minimal public health or preventive care. Considering that to date, there has been virtually no public health screening of the Norfolk Island community, our recent health study conducted on Norfolk Island highlighted the severity and extent of cardiovascular disease risk factors within this isolated population (Bellis *et al*, 2005). Extensive analysis involving heritability estimates and power calculations based on the complex Norfolk Island pedigree suggested that this population possesses unique characteristics which could aid in facilitating identification of genes involved with complex multifactorial diseases such as CVD.

METHODS AND MATERIALS

Population samples

Collection and initial analysis of the Norfolk population has previously been described in detail (Bellis *et al*, 2005). Briefly, recruitment of individuals from Norfolk Island over the age of 18 was made possible through local media announcements via radio and newspaper. Interested participants were included in the study after providing a signed informed consent statement. Ethical clearance for the Health Study was granted by the

Griffith University Human Research Ethics Committee prior to collections of samples or phenotyping of participants.

Laboratory procedures

Genomic DNA was extracted from whole blood using a standard salting out method (Miller *et al*, 1988).

X Chromosomal DNA analysis

Levels of LD in the Xq13.3 region were assessed in a study group which initially included 86 male samples. Since the Norfolk Island community possesses the unique characteristics of belonging to a single large complex family, the constraints of the pedigree were such that there are limited numbers of participants who are unrelated to other participants considering calculations concerning relatedness within the Norfolk Island population determined the mean inbreeding coefficient to be 0.0044 (with a maximum observed inbreeding coefficient of 0.0684). The pair-wise coefficient of relationship values were interesting since, among the related pairs, most are less than 3rd degree relatives ($phi^2 = 0.125$) (Bellis *et al*, 2005). However, to avoid upwardly biasing LD results, 1st and 2nd degree relatives were identified and subsequently removed from the analysis, resulting in a final sample size of 56 males, which is comparable to other population sizes for LD estimates (Angius *et al*, 2002a; Angius *et al*, 2001; Marroni *et al*, 2006). Furthermore, aims of this investigation were to study the descendents of the initial European males and their Tahitian wives and to accurately capture LD patterns within the complex Norfolk Island pedigree. Pedigree analysis indicated a total of 113 individuals were unrelated (1st and 2nd degree relatives were excluded), including 59 males. Results are based on 56 males from this group.

To evaluate the extent of LD within the Norfolk isolate we chose six microsatellite markers (DXS983, DXS8092, DXS8082, DXS1225, DXS8037 and DXS986) located in the Xq13.3 region which has been regularly studied for LD estimation in genetic isolates, including in the Finnish, Saami, Sardinina and Brazillian populations (Angius *et al*, 2002a; Angius *et al*, 2001; Kaessmann *et al*, 1999; Laan and Paabo, 1997; Pereira and Pena, 2006; Zavattari *et al*, 2000). Primer sequences were obtained from the Genome Database. PCR cycling conditions were explained elsewhere (Angius *et al*, 2001). PCR product sizes were assessed via ABI PRISM 310 DNA analyzer (PE Biosystems) and data were processed using GENESCAN v3.1 and GENOTYPER v2.5 software.

For the six microsatellite markers, the normalized disequilibrium, D', was calculated by using a multi-allelic extension of the Lewontin standardized measure of disequilibrium(Lewontin, 1988) between the various marker loci pairs. Pair-wise LD was evaluated using a test analogous to Fisher's exact test, but extended to a contingency table of arbitrary size as implemented in the software ARLEQUIN (Schneider, 2000). Disequilibrium across each locus was plotted by the GOLD program (Abecasis and Cookson, 2000).

To minimize the risk of type 1 error associated with the multiple pair-wise LD testing, raw P values were corrected through multiple comparison procedures, namely the Holm-Sidak step down procedure. The adjusted P value is derived as follows:

$$P_2^{'} = 1 - (1 - P)^{m}$$

where m is the number of P values greater than or equal to that being corrected, P is the raw value resulting from the statistical procedure used to test the hypothesis and P'_s is the Holm-Sidak corrected P value. This is believed to be a more elegant approach than the

Bonferroni-based procedure, which can produce corrected values of *P* that exceed 1 (Ludbrook, 1998).

Expected gene diversity values and their corresponding sample variance were estimated for six STR markers on Xq13.3 in the Norfolk Island population. These data were compared to Sardinian isolates from Ogliastra and Talana (Angius *et al*, 2002a). Allele frequencies were determined by gene counting. The genetic diversity value is equivalent to heterozygosity for diploid data and is defined as the probability that two randomly chosen alleles are different in the sample. An unbiased estimate of gene diversity ($\hat{h}$) was calculated by GDA (Lewis and Zaykin, 2002) as follows:

$$\hat{h} = \frac{2n}{2n-1}\left(1-\sum_{i=1}^{k} p_i^2\right)$$

and its sample variance, $V(\hat{h})$,

$$V(\hat{h}) = \frac{2}{2n(2n-1)}\left\{2(2n-2)\times\left[\sum_{i=1}^{k} p_i^3 - \left(\sum_{i=1}^{k} p_i^2\right)^2\right] + \sum_{i=1}^{k} p_i^2 - \left(\sum_{i=1}^{k} p_i^2\right)^2\right\}$$

where n is the number of gene copies, k is the number of alleles, and p_i is the frequency of the *i*-th allele of the locus considered (Nei, 1987; Nei and Roychoudhury, 1974). The average gene diversity, $\hat{H}$, is estimated by sampling r loci from the genome. Namely,

$$\hat{H} = \sum_{j=1}^{r} \hat{h}_j / r$$

where $\hat{h}_j$ is the value of $\hat{h}$ for the *j*th locus. The sampling variance of $\hat{H}$ may be obtained by:

$$V(\hat{H}) = V\left(\hat{h}\right)/r$$

where $V\left(\hat{h}\right)$ is the variance of $\hat{h}$ and is given by

$$V(\hat{h}) = \sum_{j=1}^{r} (\hat{h}_j - \hat{H})^2 / (r-1)$$

Total sampling variance, $V(\hat{h})$, is comprised of intralocus variance, $V(h)$, and interlocus variance $V_s(h)$, such that:

$$V(\hat{h}) = V(h) + V_s(h)$$

Generally, interlocus variance is much larger than intralocus variance (Nei, 1987).

Estimates of genetic diversity and its variance were calculated by formula described above. Independent samples *t*-test was used to determine whether the difference was significant for the following comparisons firstly, between Norfolk Island and Talana, and secondly between Norfolk Island and Ogliastra.

NOS2A Genotyping and LD analysis

This study aimed to investigate three polymorphic sites within the *NOS2A* gene located on chromosome 17q11.2-q12. The three markers were SNPs and were investigated in a subset of the Norfolk Island sample population (n=227). The PCR primer sequences for three SNPs are presented in Table 1. The following simplex reaction was used to amplify each of the SNPs; 1.75mmol/L $MgCl_2$, 1 unit of *Taq* polymerase, 200μmol/L dNTPs, 1 X standard PCR buffer, 0.2μmol/L each of forward and reverse primer, 40ng of genomic DNA, made up to a final volume of 25μL with sterile distilled water. Thermal cycling conditions were 1 cycle at 94°C for 4 min, 35 cycles of 94°C for 1 min, 60°C for 1 min and 1 cycle of 72°C for 2 min.

Table 1 here

LD between the three SNPs (-1026/-1659/-2447) was calculated by the formula D=p_{11}-p_1-q_1, where p_{11} is the observed frequency of the haplotype and p_1 and q_1 are the individual allele frequencies. D' is the normalized value of D and has a value between + 1 and – 1. A value of zero therefore represents that the markers are in linkage equilibrium, whereas +1 indicates that LD is at its theoretical maximum and a value of <0 that the rare alleles are in the repulsion phase. Statistical significance of D' was calculated using a χ^2 statistic with one degree of freedom.

$$\chi^2 = \rho^2 N$$

where,

$$\rho = \frac{D}{\sqrt{p_1 p_2 q_1 q_2}}$$

and N=number of chromosomes observed (Hartl and Clark, 1997).

RESULTS

LD Analysis of Xq13.3 STR markers

A total of six microsatellite markers were used to assess the levels of LD present in the genetically isolated population of Norfolk Island. The specific markers spanned a region of up to 11.5Mb. Table 2 and Figure 1 illustrate the extent and strength of LD over this region in a sample of male Norfolk Islanders. With the exception of two marker pairs (one of which is the most distantly spaced markers, DXS983 and DXS986 ~ 11.5Mb), it is evident that all of the marker pairs are in significant LD (*P*<0.05) both before and following correction.

Table 2 here

Figure 1 here

Average gene diversity estimates for this subset of Norfolk Island individuals indicated that the study population possesses a similar homogeneous genetic architecture when compared to other genetic isolates (Angius *et al*, 2002a). As a comparison data from GDV reported in Saami and Finnish samples were included (Laan and Paabo, 1997). However, sample variances were not available in these additional samples, therefore preventing calculation of confidence intervals (Table 3).

Table 3 here

Heterozygosity is estimated in haploid data by calculating gene diversity. The frequency of heterozygotes is important in population comparisons since each heterozygote carriers different alleles and represents the existence of variance.

The variance of the distribution of genetic distance across loci is called the interlocus variance. This variance increases as populations are isolated for longer periods of time. Interlocus variance, $V(h)$, is one of the two contributors to the sampling variance, $V(\hat{h})$. A second source of sampling variance is intralocus variance, $V_s(h)$, which is the sampling variance attributable to sampling a limited number of individuals at the loci being examined. Each of these variances are related by:

$$V(\hat{h}) = V(h) + V_s(h)$$

Expressions for the total variance of sample gene diversity and heterozygosity must account for both within and between population variation for the case of more than one population sampled. The total variance of sample heterozygosity is minimized as sample sizes increase (Weir *et al*, 1990). For a mixed-mating or random mating system, this variance results mostly from associations of genes between individuals. Because of this, the total variance of sample heterozygosity in minimized most efficiently by sampling

more individuals, rather than increasing the number of loci sampled. In contrast, increasing the number of loci sampled, rather than the number of individuals sampled, has the strongest minimization of variance effect for unlinked loci in populations at migration-drift equilibrium (Kalinowski, 2002).

Comparison between Norfolk Island, Talana and Ogliastra indicated that Talana possesses the lowest average genetic diversity (0.77, 0.66 and 0.75 respectively) as well as the lowest individuals locus diversity (0.617 at DXS8082). This difference is also evident when comparing the total variance of genetic diversity (Figure 2). The difference was significant between the Norfolk Island and Talana population (p=0.01). However, considering the Norfolk Island sample possess higher levels interlocus variance (Figure 2), more markers were sampled, this variance could be reduced.

Figure 2 here

Figure 3 here

Decay of D' in the Xq13.3 genomic region of sampled Norfolk Islanders is plotted against the same data from the genetic isolate of Talana (Angius *et al*, 2002a). Figure 3 illustrates that although the gene diversity estimates are higher in the Norfolk Island population when compared to Talana, the levels of D' are higher and LD decays at a slower rate over genetic distance for the Norfolk Island population.

LD Analysis NOS2A SNPs

This study investigated three SNPs within the *NOS2A* gene located on chromosome 17q11.2-q12 which were previously identified and analysed in a Gambian and UK Caucasian population (Burgner *et al*, 2003). The SNPs (-1026g/t, -1659c/t, -2447c/g) are clustered closely together in the proximal promoter region of the *NOS2A* gene and have

been shown to be in complete and near complete LD in the Gambian and UK sample, respectively (Burgner *et al*, 2003). The initial objective of this study was to determine the frequency, and assess LD, of these SNP alleles in an isolated population from Norfolk Island. Genotype and allele frequencies for all SNPs are presented in table 4, while Table 5 illustrates the minor allele frequency (MAF) comparison between different study populations.

Table 4 here

Genotype frequencies for all markers were found to comply with Hardy-Weinberg equilibrium ratios (P>0.05) in both the Australian Caucasian control group and the Norfolk Island population. There were no significant differences detected between any of the three *NOS2A* SNPs in the Norfolk Island population and the Australian Caucasian control group (P>0.05).

As can be seen in Table 5, the frequencies of the rarer allele for these SNPs differ substantially between the two ethnic groups from Gambia and the United Kingdom and for SNPs -1026g/t and -1659c/t, the difference is significant (Burgner *et al*, 2003). Although not statistically significant, reduced minor allele frequencies for the three SNPs in the Norfolk population are indicative of a population with reduced genetic heterogeneity when compared to outbred populations.

Table 5 here

Highly significant LD was found to exist between alleles of all three SNPs in Norfolk Islander population (Table 6). This is consistent with previous work (Bugeja *et al*, 2005) and is to be expected given the close proximity of the SNPs, the relatedness of the Norfolk population and the prevalence of the minor SNP allele (>10%). With the

exception of -1026/-1659 marker pair, all the SNPs studied in the Norfolk population were found to be in complete and significant LD with each other.

Table 6 here

Haplotype analysis of NOS2A SNPs

Haplotype analysis suggested that the Norfolk Island population displays reduced heterogeneity at the three SNPs investigated in the *NOS2A* proximal promoter region. Although comparison to other Caucasian study groups suggests that the same common haplotype is shared, the Norfolk Island population has higher frequency of this haplotype when compared to other presented population data (Table 7).

Table 7 here

DISCUSSION

The primary focus of the current study was to determine the extent of LD across the Xq13.3 region in a genetically isolated population from Norfolk Island. This particular region has been routinely utilized for LD estimation in numerous populations (Angius *et al*, 2002a; Angius *et al*, 2001; Kaessmann *et al*, 1999; Pereira and Pena, 2006; Zavattari *et al*, 2000) allowing comparison between different isolates. The Norfolk Island population study found that LD extends up to 9.5-11Mb in this particular region, with 13 out of the 15 marker pairs in LD, which is at least comparable to or exceeding that found in other genetic isolates, such as the secluded population of Talana on the island of Sardinia (Angius *et al*, 2002a). Additionally, gene diversity calculations based on the six microsatellite markers in the same Xq13.3 region suggest that the sampled population of Norfolk Island shows reduced genetic expected heterogeneity, again similar to other isolate populations (Angius *et al*, 2002a).

Analysis of three SNPs within the *NOS2A* gene showed interesting results in terms of Norfolk Island's genetic architecture, including a reduced minor allele frequency of all three SNPs (when compared to outbred Australian, United Kingdom and Gambian populations) and increased frequency of the most common haplotype, both are indicators of increased genetic homogeneity, which is expected considering Norfolk's limited number of initial founders, and stable population growth in isolation from other populations. Although no significant differences were observed regarding MAF and haplotype frequency between Norfolk and an outbred Australian population, it should be noted that the three SNPs in a 2kb region provide very few data points for comparison and substantial LD was expected in this limited region. Further studies would obviously increase SNP saturation over a greater genetic distance providing significantly more points for comparison. Additionally, introduction of phenotype correlation analysis may be worth investigating, particularly studying ther markers within the proximal promoter region of this gene (Hobbs *et al*, 2002). In relation to this, genome wide linkage analysis for a number of CVD-related quantitative traits in the Norfolk Island population is currently in progress. This present study provides preliminary comparisons between marker types and the Norfolk population, which should be useful for follow-up analysis stemming from the STR genome scan data.

Founder population isolates offer several significant advantages for disease gene mapping over mainstream populations. Firstly, the limited number of ancestors minimizes genetic heterogeneity and therefore it is expected that there will be fewer susceptibility genes with greater overall effect. Also, in geographically and culturally isolated populations, environmental 'noise' is reduced, minimising the confounding effects of non-genetic variables. These characteristics have been exploited in the study of rare genetic disorders

caused by single defective genes (Newman *et al*, 2003; Nikali *et al*, 1995; Puffenberger *et al*, 1994) and can also be advantageous in complex disease gene mapping (Bourgain *et al*, 2000; Peltonen *et al*, 2000; Sheffield *et al*, 1998; Shifman and Darvasi, 2001). Furthermore, the kinship (relatedness) coefficient is much higher in isolated founder populations. This is important because very large extended pedigrees, suitable for powerful linkage analysis of complex traits, can often be identified (Bourgain *et al*, 2000). A recent study investigating genome scan information in a Hutterite isolate, stressed the importance of good genealogical information (Newman *et al*, 2001). This study noted that a failure to take full pedigree information into account can reduce the power to detect linkage, or inflate LOD scores and also the failure to account for relatedness can affect association studies. Hence it is important that good genealogical information with defined pedigree information is available. This allows the analysis of large extended pedigrees and greatly increases the power and accuracy of a gene mapping study. In addition, a good understanding of local population history is important for evaluating factors such as the number of founders, population size, consanguinity, immigration, population expansion rate and genetic drift. The best candidate populations for detecting associations with common genetic variants are believed to be isolates with a small effective number of unrelated founders (10-100), as this offers the advantage of a smaller number of disease susceptibility variants within the test populations compared with out-bred populations (Bourgain *et al*, 2000; Peltonen *et al*, 2000; Sheffield *et al*, 1998).

REFERENCE LIST

Abbott W, Scragg R, Marbrook J (2001). Differences in disease frequency between Europeans and Polynesians: directions for future research into genetic risk factors. *Pac Health Dialog* **8**(1)**:** 129-156.

Abecasis GR, Cookson WO (2000). GOLD--graphical overview of linkage disequilibrium. *Bioinformatics* **16**(2)**:** 182-183.

Abney M, McPeek MS, Ober C (2000). Estimation of variance components of quantitative traits in inbred populations. *Am J Hum Genet* **66**(2)**:** 629-650.

Abney M, McPeek MS, Ober C (2001). Broad and narrow heritabilities of quantitative traits in a founder population. *Am J Hum Genet* **68**(5)**:** 1302-1307.

Abney M, Ober C, McPeek MS (2002). Quantitative-trait homozygosity and association mapping and empirical genomewide significance in large, complex pedigrees: fasting serum-insulin level in the Hutterites. *Am J Hum Genet* **70**(4)**:** 920-934.

Angius A, Bebbere D, Petretto E, Falchi M, Forabosco P, Maestrale B *et al* (2002a). Not all isolates are equal: linkage disequilibrium analysis on Xq13.3 reveals different patterns in Sardinian sub-populations. *Hum Genet* **111**(1)**:** 9-15.

Angius A, Melis PM, Morelli L, Petretto E, Casu G, Maestrale GB *et al* (2001). Archival, demographic and genetic studies define a Sardinian sub-isolate as a suitable model for mapping complex traits. *Hum Genet* **109**(2)**:** 198-209.

Angius A, Petretto E, Maestrale GB, Forabosco P, Casu G, Piras D *et al* (2002b). A new essential hypertension susceptibility locus on chromosome 2p24-p25, detected by genomewide search. *Am J Hum Genet* **71**(4)**:** 893-905.

Bellis C, Hughes RM, Begley KN, Quinlan S, Lea RA, Heath SC *et al* (2005). Phenotypical characterisation of the isolated norfolk island population focusing on epidemiological indicators of cardiovascular disease. *Hum Hered* **60**(4)**:** 211-219.

Bonnen PE, Pe'er I, Plenge RM, Salit J, Lowe JK, Shapero MH *et al* (2006). Evaluating potential for whole-genome studies in Kosrae, an isolated population in Micronesia. *Nat Genet* **38**(2)**:** 214-217.

Bourgain C, Genin E, Quesneville H, Clerget-Darpoux F (2000). Search for multifactorial disease susceptibility genes in founder populations. *Ann Hum Genet* **64**(Pt 3)**:** 255-265.

Bugeja MJ, Booth DR, Bennetts BH, Heard RN, Burgner D, Stewart GJ (2005). An investigation of NOS2A promoter polymorphisms in Australian multiple sclerosis patients. *Eur J Hum Genet* **13**(7)**:** 815-822.

Burgner D, Rockett K, Ackerman H, Hull J, Usen S, Pinder M *et al* (2003). Haplotypic relationship between SNP and microsatellite markers at the NOS2A locus in two populations. *Genes Immun* **4**(7)**:** 506-514.

Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J *et al* (2000). The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nat Genet* **25**(3)**:** 320-323.

Edgecombe J (1999). *Norfolk Island - South Pacific : island of history and many delights*, 2nd edn. J.M. Edgecombe: Thornleigh, Australia, 179pp.

Falchi M, Forabosco P, Mocci E, Borlino CC, Picciau A, Virdis E *et al* (2004). A genomewide search using an original pairwise sampling approach for large genealogies identifies a new locus for total and low-density lipoprotein cholesterol in two genetically differentiated isolates of Sardinia. *Am J Hum Genet* **75**(6)**:** 1015-1031.

Gulcher J, Stefansson K (1998). Population genomics: laying the groundwork for genetic disease modeling and targeting. *Clin Chem Lab Med* **36**(8)**:** 523-527.

Han Z, Heath SC, Shmulewitz D, Li W, Auerbach SB, Blundell ML *et al* (2002). Candidate genes involved in cardiovascular risk factors by a family-based association study on the island of Kosrae, Federated States of Micronesia. *Am J Med Genet* **110**(3)**:** 234-242.

Hartl DL, Clark AG (1997). *Principles of population genetics*, 3rd edn. Sinauer Associates: Sunderland, MA, xiii, 542pp.

Helgason A, Nicholson G, Stefansson K, Donnelly P (2003). A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann Hum Genet* **67**(Pt 4)**:** 281-297.

Hoare M (1999). *Norfolk Island : a revised and enlarged history 1774-1998*, 5th Edition edn. Central Queensland University Press: Rockhampton, 228pp.

Hobbs MR, Udhayakumar V, Levesque MC, Booth J, Roberts JM, Tkachuk AN *et al* (2002). A new NOS2 promoter polymorphism associated with increased nitric oxide production and protection from severe malaria in Tanzanian and Kenyan children. *Lancet* **360**(9344)**:** 1468-1475.

Kaessmann H, Heissig F, von Haeseler A, Paabo S (1999). DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat Genet* **22**(1)**:** 78-81.

Kalinowski ST (2002). Evolutionary and statistical properties of three genetic distances. *Mol Ecol* **11**(8)**:** 1263-1273.

Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill P, Shen P *et al* (2003). Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet* **72**(2)**:** 281-302.

Laan M, Paabo S (1997). Demographic history and linkage disequilibrium in human populations. *Nat Genet* **17**(4)**:** 435-438.

Latini V, Sole G, Doratiotto S, Poddie D, Memmi M, Varesi L *et al* (2004). Genetic isolates in Corsica (France): linkage disequilibrium extension analysis on the Xq13 region. *Eur J Hum Genet* **12**(8)**:** 613-619.

Lewis PO, Zaykin D. (2002).

Lewontin RC (1988). On measures of gametic disequilibrium. *Genetics* **120**(3)**:** 849-852.

Ludbrook J (1998). Multiple comparison procedures updated. *Clin Exp Pharmacol Physiol* **25**(12)**:** 1032-1037.

Marroni F, Pichler I, De Grandi A, Beu Volpato C, Vogl FD, Pinggera GK *et al* (2006). Population isolates in South Tyrol and their value for genetic dissection of complex diseases. *Ann Hum Genet* **70**(Pt 6)**:** 812-821.

Miller SA, Dykes DD, Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**(3)**:** 1215.

Murray-McIntosh RP, Scrimshaw BJ, Hatfield PJ, Penny D (1998). Testing migration patterns and estimating founding population size in Polynesia by using human mtDNA sequences. *Proc Natl Acad Sci U S A* **95**(15)**:** 9047-9052.

Nei M (1987). *Molecular evolutionary genetics*. Columbia University Press: New York, x, 512pp.

Nei M, Roychoudhury AK (1974). Sampling variance of heterozygosity and genetic distance. *Genetics* **76:** 379-390.

Newman DL, Abney M, Dytch H, Parry R, McPeek MS, Ober C (2003). Major loci influencing serum triglyceride levels on 2q14 and 9p21 localized by homozygosity-by-descent mapping in a large Hutterite pedigree. *Hum Mol Genet* **12**(2)**:** 137-144.

Newman DL, Abney M, McPeek MS, Ober C, Cox NJ (2001). The importance of genealogy in determining genetic associations with complex traits. *Am J Hum Genet* **69**(5)**:** 1146-1148.

Newman DL, Hoffjan S, Bourgain C, Abney M, Nicolae RI, Profits ET *et al* (2004). Are common disease susceptibility alleles the same in outbred and founder populations? *Eur J Hum Genet* **12**(7)**:** 584-590.

Nikali K, Suomalainen A, Terwilliger J, Koskinen T, Weissenbach J, Peltonen L (1995). Random search for shared chromosomal regions in four affected individuals: the assignment of a new hereditary ataxia locus. *Am J Hum Genet* **56**(5)**:** 1088-1095.

Ober C, Abney M, McPeek MS (2001). The genetic dissection of complex traits in a founder population. *Am J Hum Genet* **69**(5)**:** 1068-1079.

Peltonen L, Jalanko A, Varilo T (1999). Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* **8**(10)**:** 1913-1923.

Peltonen L, Palotie A, Lange K (2000). Use of population isolates for mapping complex traits. *Nat Rev Genet* **1**(3)**:** 182-190.

Pereira RW, Pena SD (2006). Phylogeography of haplotypes of five microsatellites located in a low-recombination region of the X chromosome: studies worldwide and in Brazilian populations. *Genetica* **126**(1-2)**:** 243-250.

Puffenberger EG, Hosoda K, Washington SS, Nakao K, deWit D, Yanagisawa M *et al* (1994). A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. *Cell* **79**(7)**:** 1257-1266.

Pugliatti M, Sotgiu S, Sadovnick AD, Yee IM, Sotgiu MA, Poser CM *et al* (2003). Multiple sclerosis in Northern Sardinia, Italy: a methodological approach for genetic epidemiological studies. *Neuroepidemiology* **22**(6)**:** 362-365.

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, Stoneking M (1995). Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* **12**(4)**:** 604-615.

Schneider S, Roessli, D. and Excoffier. L. (2000). Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.

Sheffield VC, Stone EM, Carmi R (1998). Use of isolated inbred human populations for identification of disease genes. *Trends Genet* **14**(10)**:** 391-396.

Shifman S, Darvasi A (2001). The value of isolated populations. *Nat Genet* **28**(4)**:** 309-310.

Shmulewitz D, Auerbach SB, Lehner T, Blundell ML, Winick JD, Youngman LD *et al* (2001). Epidemiology and factor analysis of obesity, type II diabetes, hypertension, and dyslipidemia (syndrome X) on the Island of Kosrae, Federated States of Micronesia. *Hum Hered* **51**(1-2)**:** 8-19.

Tenesa A, Wright AF, Knott SA, Carothers AD, Hayward C, Angius A *et al* (2004). Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. *Hum Mol Genet* **13**(1)**:** 25-33.

Tsai HJ, Sun G, Smelser D, Viali S, Tufa J, Jin L *et al* (2004). Distribution of genome-wide linkage disequilibrium based on microsatellite loci in the Samoan population. *Hum Genomics* **1**(5)**:** 327-334.

Varilo T, Laan M, Hovatta I, Wiebe V, Terwilliger JD, Peltonen L (2000). Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo. *Eur J Hum Genet* **8**(8)**:** 604-612.

Varilo T, Paunio T, Parker A, Perola M, Meyer J, Terwilliger JD *et al* (2003). The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum Mol Genet* **12**(1)**:** 51-59.

Weir BS, Reynolds J, Dodds KG (1990). The variance of sample heterozygosity. *Theor Popul Biol* **37**(1)**:** 235-253.

Weiss LA, Pan L, Abney M, Ober C (2006). The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet* **38**(2)**:** 218-222.

Wessman M, Kallela M, Kaunisto MA, Marttila P, Sobel E, Hartiala J *et al* (2002). A susceptibility locus for migraine with aura, on chromosome 4q24. *Am J Hum Genet* **70**(3)**:** 652-662.

Wijsman EM, Rosenthal EA, Hall D, Blundell ML, Sobin C, Heath SC *et al* (2003). Genome-wide scan in a large complex pedigree with predominantly male schizophrenics from the island of Kosrae: evidence for linkage to chromosome 2q. *Mol Psychiatry* **8**(7)**:** 695-705, 643.

Zavattari P, Deidda E, Whalen M, Lampis R, Mulargia A, Loddo M *et al* (2000). Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum Mol Genet* **9**(20)**:** 2947-2957.

Table 1. Primer sequence data including expected restriction digest fragment sizes.

Marker	Forward and Reverse Primers	Restriction enzyme	Digest Products (genotype)
1	5'-TGGGAAAGACAAGAAGGAAATGAGTGGAC-3'	*Nla*III	281+18+9 (T/T)
	5'-GTAGAGACTGGGTTTCACCATGTTGTCCA-3'		200+81+18+9 (G/G)
2	5'-AAGTAGATGGGATTTCTTGAGGATTG-3'	*Bfa*I	303+9 (C/C)
	5'-CAAGGCCTAGAAAAATAAATGGATAATT-3'		228+75+9 (T/T)
3	5'-CCTCCTGAGTAGCTAGGACTACAGTAGCTGG**<u>T</u>**CATG-3'	*Cac*81	317 (C/C)
	5'-GTTAATGTCTGTGATGCACACCACGCAG-3'		279+38 (G/G)

Maker information 1. -1026G/T, 2. -1659C/T, 3. -2447C/G.

Deliberate mismatch introduced in forward primer of -2447C/G removes unwanted *Cac*81 restriction site. Identified as underlined and bold **<u>T</u>**.

Table 2. Pair-wise LD results between Xq13.3 markers. *P* shown in *bold* are lower than 0.05 after correction for multiple testing by using the Holm-Sidak step down procedure (Ludbrook, 1998, Lautenberger et al., 2000).

M1	M2	Distance		D'	P Value
		cM[a]	Mb[a]		Norfolk Island
DXS983	DXS8092	1.6	5.5	0.469	0.31156
DXS983	DXS8082	1.7	9.0	0.306	**0.00219**
DXS983	DXS1225	2.0	9.5	0.415	**0.00312**
DXS983	DXS8037	2.0	9.5-11.5	0.462	**0.03795**
DXS983	DXS986	1.5	11.5	0.313	0.40950
DXS8092	DXS8082	0.1	3.5	0.492	**0.00748**
DXS8092	DXS1225	0.4	4.0	0.517	**0.00000**
DXS8092	DXS8037	0.4	4.0-6.0	0.582	**0.00219**
DXS8092	DXS986	0.1	6.0	0.566	**0.00000**
DXS8082	DXS1225	0.3	<0.5	0.486	**0.00000**
DXS8082	DXS8037	0.3	0.5-2.5	0.889	**0.00000**
DXS8082	DXS986	0.2	2.5	0.489	**0.00000**
DXS1225	DXS8037	0.0	<2.0	0.615	**0.00000**
DXS1225	DXS986	0.5	2.0	0.448	**0.00312**
DXS8037	DXS986	0.0	<2.0	0.525	**0.00000**

[a]Genetic distance in centiMorgans and physical distance in megabases as presented by Laan and Paabo (1997).

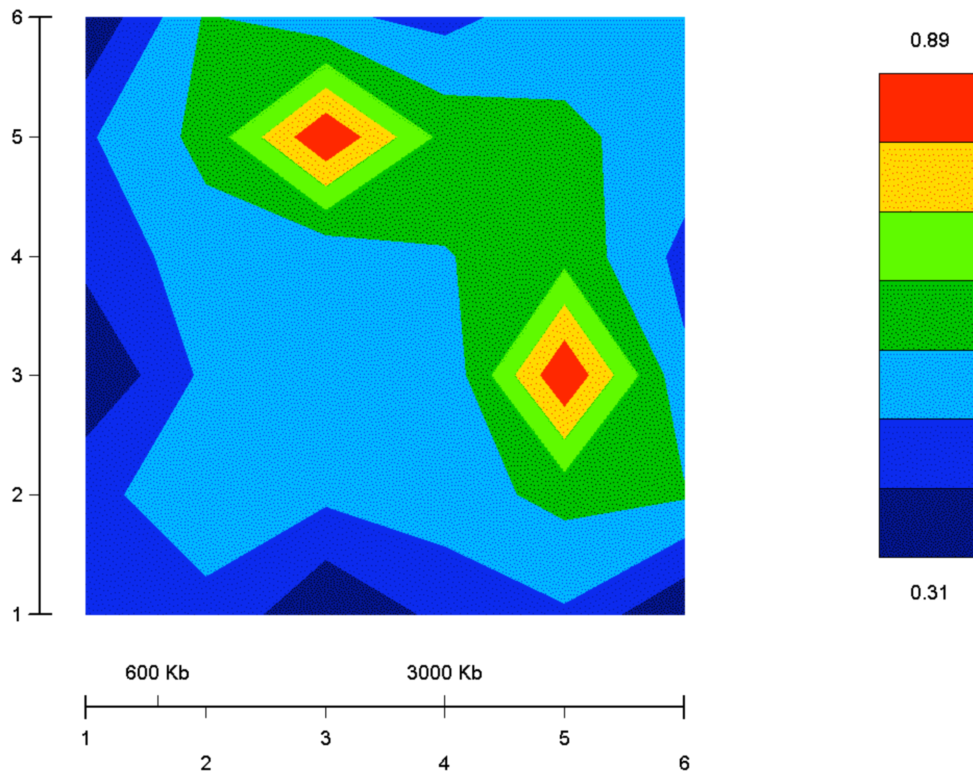

Figure 1. GOLD output of LD distributions present on Xq13.3 from the Norfolk Island population.

Table 3. Gene Diversity and Haplotype Data Comparisons for Six Microsatellite Markers on Xq13.3

Population	Sample size	No. of Haplotypes	No. of Alleles (Diversity), Based on Marker						Mean SD Gene Diversity
			DXS983	DXS986	DXS8092	DXS8082	DXS1225	DXS8037	
Norfolk Island	56	43	5 (.64)	9 (.79)	9 (.84)	10 (.79)	8 (.79)	6 (.79)	0.78 ± 0.43
Ogliastra[a]	91	86	7 (.79)	10 (.76)	11 (.85)	11 (.63)	12 (.69)	9 (.76)	0.75 ± 0.41
Talana[a]	50	35	5 (.63)	6 (.75)	8 (.70)	7 (.62)	5 (.62)	7 (.67)	0.67 ± 0.37
Finnish[b]	80	75	6 (.70)	11 (.80)	12 (.85)	9 (.75)	9 (.73)	8 (.71)	0.74
Saami[b]	54	32	4 (.58)	10 (.80)	8 (.83)	6 (.76)	5 (.50)	3 (.44)	0.67

[a](Angius et al., 2002)
[b](Laan and Paabo, 1997)

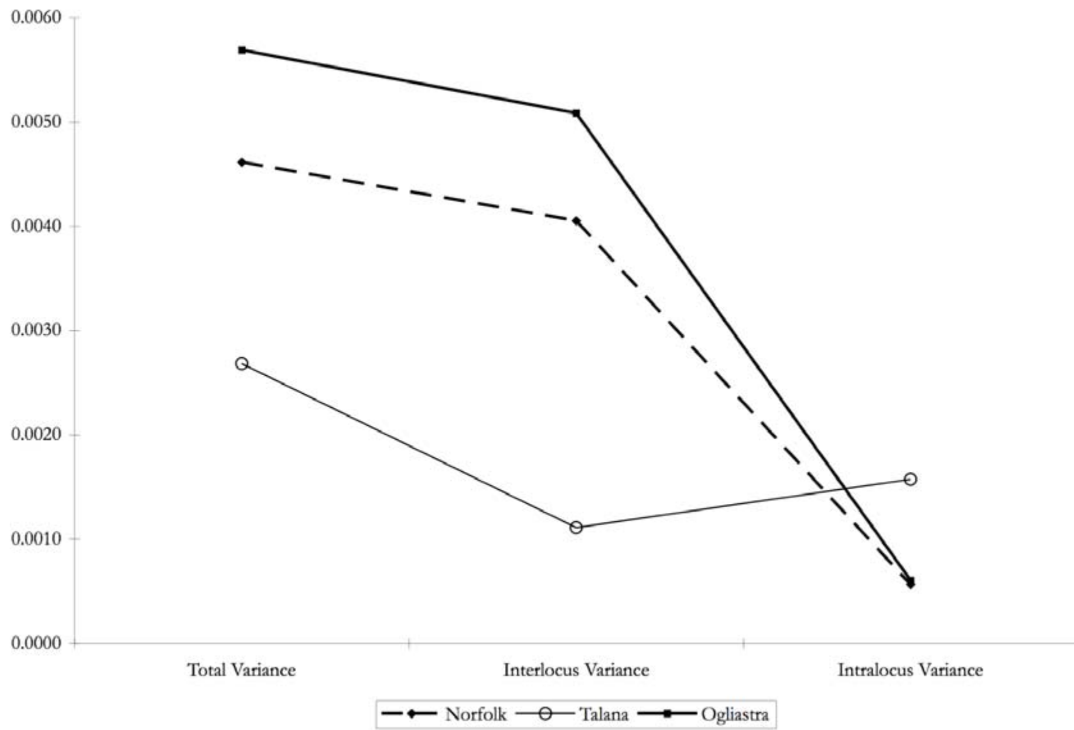

Figure 2. Comparison between variance of genetic diversity estimate in Norfolk Island, Talana and Ogliastra populations.

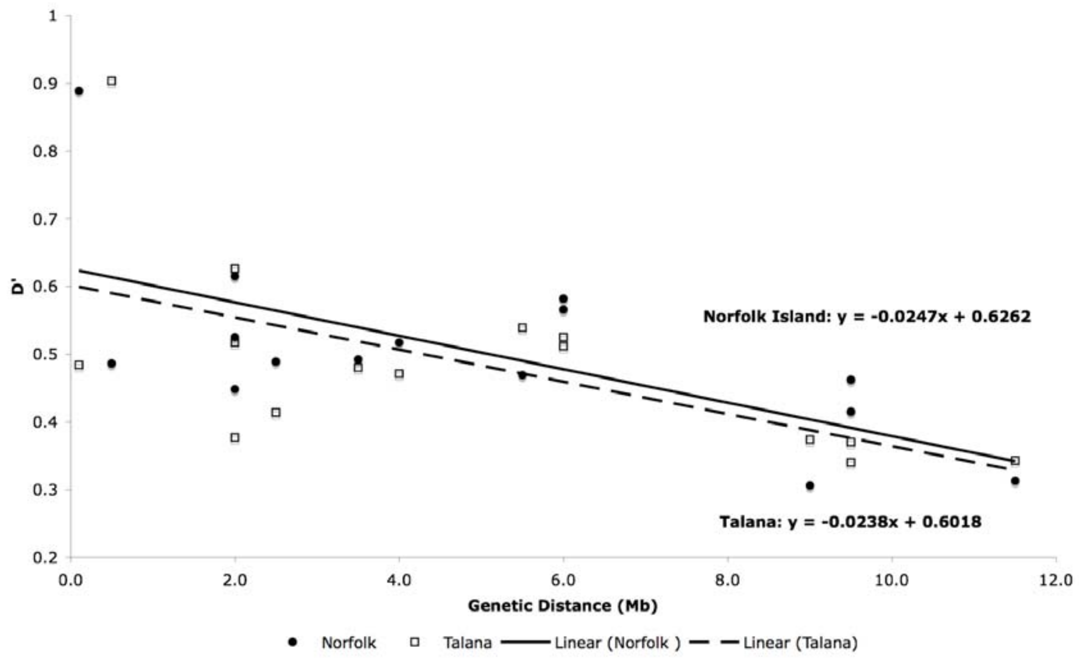

Figure 3. Strength of LD evaluated as multiallelic D' values in the Norfolk Island population versus Talana population stratified physical distances for the chromosome Xq13.3 region.

Table 4. Distribution of genotype and allele frequencies for the -1026g/t, -1659c/t and -2447c/g *NOS2A* promoter SNPs in Norfolk Island population compared to an Australian cohort.

		Genotypes				Alleles		HWE	
SNP	Group	1 1	1 2	2 2	N	1	2	χ^2	*P*
-1026	NI	101	55	11	167	77%	23%	0.859	0.354
	Control	128	84	22	234	73%	27%	2.187	0.139
-1659	NI	190	33	1	224	92%	8%	0.116	0.733
	Control	200	48	4	252	89%	11%	0.321	0.571
-2447	NI	123	76	20	219	74%	26%	2.593	0.107
	Control	110	101	32	243	66%	34%	1.303	0.254

Table 5. Comparison of minor allele frequencies for proximal *NOS2A* promoter region SNPs in Caucasian, Gambian and Norfolk Islander samples.

	Polymorphism		
	-1026t	*-1659t*	*-2447g*
Gambian	0.46	0.25	0.4
UK Caucasians	0.227	0.097	0.278
Australian Caucasians	0.27	0.11	0.34
Norfolk Islanders	**0.23**	**0.08**	**0.26**
χ^2	59.31	54.74	19.83
P-value	<0.00001*	<0.00001*	0.00297*

* Not significant if Gambian data are removed from analysis

Table 6. Linkage disequilibrium (D') comparison between SNPs of the proximal *NOS2A* promoter region in Norfolk Islanders and Australian Caucasians.

	Norfolk Island			Australian Caucasian case/control		
	-1026g/t	-1659c/t	-2447c/g	-1026g/t	-1659c/t	-2447c/g
-1026g/t	1	0.924	1	1	0.999	0.959
-1659c/t	—	1	1	—	1	0.981
-2447c/g	—	—	1	—	—	1

* all D' values are highly significant $P<0.000001$

Table 7. *NOS2A* SNP haplotype frequencies in Norfolk Islander, Caucasian and Gambians.

	Haplotype			Population Frequency			
Identifier	-1026 g/t	-1659 c/t	-2447 g/c	Gambian[a]	Uk Cauc.[a]	Au Cauc.	NI
1	g	c	c	0.619	0.663	0.647	0.736
2	t	c	c	0.070	—	0.006	0.011
3	t	t	g	0.243	0.102	0.105	0.051
4	t	c	g	0.067	0.130	0.167	0.144
5	g	c	g	—	0.105	0.073	0.050

[a]45