

ORIGINAL ARTICLE

Linkage disequilibrium analysis in the genetically isolated Norfolk Island population

C Bellis¹, HC Cox¹, M Ovcacic¹, KN Begley², RA Lea^{1,3}, S Quinlan¹, D Burgner⁴, SC Heath⁵, J Blangero⁶ and LR Griffiths¹

¹Genomics Research Centre, School of Medical Science, Griffith University, Gold Coast, Bundall, Australia; ²Division of Information Services, Griffith University, Gold Coast, Bundall, Australia; ³School of Biological Sciences, Institute of Environmental Science and Research Ltd, Victoria University of Wellington, Wellington, New Zealand; ⁴School of Paediatrics and Child Health, University of Western Australia, Perth, Australia; ⁵Statistical Genetics, Centre National de Génotypage, 2 Rue Gaston Cremieux, Evry, France and ⁶Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX, USA

Norfolk Island is a human 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. A total of six microsatellite markers spanning ~11 Mb were assessed on chromosome Xq13.3 in a group of 56 men from Norfolk Island. Additionally, three single nucleotide polymorphisms (SNPs) localizing to the *NOS2A* gene were analyzed 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 marker, all the Xq13.3 marker pairs

were found to be in significant LD indicating that LD extends up to 9.5–11.5 Mb 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 risk traits. With LD extending up to 9.5–11 Mb, the Norfolk isolate should be a powerful resource for the localization of complex disease genes. *Heredity* (2008) **100**, 366–373; doi:10.1038/sj.hdy.6801083; published online 19 December 2007

Keywords: genetic isolate; *NOS2A*; linkage disequilibrium; Norfolk Island; Xq13.3

Introduction

Genetic investigation of inherited diseases has matured from the mapping of Huntington's disease in the early 1980s to a future that will potentially provide information pertaining to many thousands of genetic variants underlying both simple and complex diseases. There is no doubt that central to the success of this genomic era has been the coordinated efforts which have resulted in, first, the complete sequencing of the human genome and its public availability to researchers globally and, second, 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, 2003; Wessman *et al.*, 2002) and Iceland (Gulcher and Stefansson, 1998; Helgason *et al.*, 2003), Mediterranean regions of Italy, especially Sardinia (Eaves *et al.*, 2000; Zavattari *et al.*, 2000; Angius *et al.*, 2001, 2002a, b; Pugliatti *et al.*, 2003; Falchi *et al.*, 2004; Tenesa *et al.*, 2004) 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, 2001, 2002; Ober *et al.*, 2001; Newman *et al.*, 2003, 2004; Weiss *et al.*, 2006), while several remote Polynesian populations are presently involved in disease gene-mapping studies (Redd *et al.*, 1995; Murray-McIntosh *et al.*, 1998; Shmulewitz *et al.*, 2001; Han *et al.*, 2002; Kayser *et al.*, 2003; Wijsman *et al.*, 2003; Tsai *et al.*, 2004; Bonnen *et al.*, 2006). 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 ~1700 km

Correspondence: Professor LR Griffiths, School of Medical Science, Genomics Research Centre, Griffith University, Gold Coast Campus, PMB 50, GCMC, Gold Coast, Bundall 9726, Australia.

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

Received 12 July 2007; revised 12 November 2007; accepted 16 November 2007; published online 19 December 2007

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 and 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, pedigree-based searches undertaken in the genealogical program Brother's Keeper (Version 6.0, Rockford, MI, USA) revealed that ~80% 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 composed 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 toward 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 (CVD) 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.

Materials and methods

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 that initially included 86 male samples. Since the Norfolk Island community possesses the unique characteristic of belonging to a single large complex family, the number of unrelated individuals available for LD analysis was limited. This was exemplified when the calculations concerning relatedness within the Norfolk Island population determined the mean inbreeding coefficient to be 0.0044. The maximum individual inbreeding coefficient observed was 0.0684, which is equivalent to an offspring of first cousins. The pair-wise coefficient of relationship values were interesting since, among the related pairs, most are less than third degree relatives ($\phi^2 = 0.125$) (Bellis *et al.*, 2005). However, to avoid upwardly biasing LD results, first and second degree relatives were identified and subsequently removed from the analysis, resulting in a final sample size of 56 men, which is comparable to other population sizes for LD estimates (Angius *et al.*, 2001, 2002a; Marroni *et al.*, 2006). Furthermore, aims of this investigation were to study the descendants of the initial European men and their Tahitian wives and to accurately capture LD patterns within the complex Norfolk Island pedigree. Pedigree analysis indicated that a total of 113 individuals were unrelated (first and second degree relatives were excluded), including 59 men. Results are based on 56 men 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, Sardinia and Brazilian populations (Laan and Paabo, 1997; Kaessmann *et al.*, 1999; Zavattari *et al.*, 2000; Angius *et al.*, 2001, 2002a; Pereira and Pena, 2006). 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 multiallelic 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 *et al.*, 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 short tandem repeat (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 Genetic Data Analysis (Lewis and Zaykin, 2002) as follows:

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

Since both the number of alleles sampled and their evenness define diversity, it should be clear that large values of \hat{h} represent very diverse samples. The sampling variance of this measure, $V(\hat{h})$, was formulated by Nei (1987) and is given by the following formula:

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

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 and Roychoudhury, 1974; Nei, 1987). The average gene diversity, $\text{avg}(\hat{h})$, is estimated by sampling r loci from the genome. Namely,

$$\text{avg}(\hat{h}) = \frac{\sum (2n_i)/(2n_i-1) \left(1 - \sum_{i=1}^k p_i^2 \right)}{r}$$

where r is the number of sampled loci and n_i is the sample size (gene copies sampled) per locus.

The sampling variance of this estimator is given by:

$$V(\text{avg}(\hat{h})) = \frac{\sum_{j=1}^r \left((2n_j)/(2n_j-1) \left(1 - \sum_{i=1}^k p_i^2 \right) - \text{avg}(\hat{h}) \right)^2}{r-1}$$

This sampling variance $V(\text{avg}(\hat{h}))$, is composed of intralocus variance, $V(h)$, and interlocus variance $V_s(h)$, such that:

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

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, first, between Norfolk Island and Talana, and second 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 single nucleotide polymorphisms (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; 1.75 mmol l⁻¹ MgCl₂, 1 unit of Taq polymerase, 200 μmol l⁻¹ deoxyribonucleotide triphosphates, 1 × standard PCR buffer, 0.2 μmol l⁻¹ each of forward and reverse primer, 40 ng of genomic DNA, made up to a final volume of 25 μl 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.

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 0 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.

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

Table 1 Primer sequence data for *NOS2A* SNPs (–1026G/T, –1659C/T and –2447C/G) including expected restriction digest fragment sizes

Marker	Forward and reverse primers	Restriction enzyme	Digest products (genotype)
1	5'-TGGGAAAGACAAGAAGGAAATGAGTGGAC-3' 5'-GTAGAGACTGGGTTTCACCATGTTGTCCA-3'	<i>Nla</i> III	281+18+9 (T/T) 200+81+18+9 (G/G)
2	5'-AAGTAGATGGGATTCTTGAGGATTG-3' 5'-CAAGGCCTAGAAAAATAAATGGATAATT-3'	<i>Bfa</i> I	303+9 (C/C) 228+75+9 (T/T)
3	5'-CCTCCTGAGTAGCTAGGACTACAGTAGCTGGTCATG-3' 5'-GTTAATGTCTGTGATGCACACCACGCAG-3'	<i>Cac</i> 81	317 (C/C) 279+38 (G/G)

Marker 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 **T**.

where,

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

and N refers to the 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 11.5 Mb. Table 2 illustrates 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 ~11.5 Mb), it is evident that all of the marker pairs are in significant LD ($P < 0.05$) both before and following correction.

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). For comparison purposes, gene diversity values 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).

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

Comparison between Norfolk Island, Talana and Ogliastra indicated that Talana possesses the lowest average genetic diversity (0.77, 0.67 and 0.75, respectively) as well as the lowest individuals locus diversity (0.617 at DXS8082). This difference is also evident when

Table 2 Pair-wise LD results between Xq13.3 markers

M1	M2	Distance		D'	P -value
		cM ^a	Mb ^b		
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

P 's 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).

^aGenetic distance in cM and physical distance in Mb as presented by Laan and Paabo (1997).

Table 3 Gene diversity data comparisons for six microsatellite markers on Xq13.3 in known genetically isolated populations

Population	Sample size	Number of haplotypes	Number of alleles (diversity), based on marker						Mean (\pm s.d.) gene diversity	Total variance	Interlocus variance	Intralocus variance
			DXS983	DXS986	DXS8092	DXS8082	DXS1225	DXS8037				
Norfolk Island	56	43	5 (0.64)	9 (0.79)	9 (0.84)	10 (0.79)	8 (0.79)	6 (0.79)	0.77 (\pm 0.43)	0.004616	0.000563	
Ogliastra ^a	91	86	7 (0.79)	10 (0.76)	11 (0.85)	11 (0.63)	12 (0.69)	9 (0.76)	0.75 (\pm 0.41)	0.005690	0.000606	
Talana ^a	50	35	5 (0.63)	6 (0.75)	8 (0.70)	7 (0.62)	5 (0.62)	7 (0.67)	0.67 (\pm 0.37)	0.002682	0.001108	
Finnish ^b	80	75	6 (0.70)	11 (0.80)	12 (0.85)	9 (0.75)	9 (0.73)	8 (0.71)	0.74			
Saami ^b	54	32	4 (0.58)	10 (0.80)	8 (0.83)	6 (0.76)	5 (0.50)	3 (0.44)	0.67			
CEPH ^b	20		6 (0.81)	11 (0.73)	12 (0.94)	10 (0.73)	12 (0.75)	9 (0.78)	0.79			

Total variance data in bold indicate significant difference between Norfolk Island and Talana populations.

^aAngius *et al.* (2002a).

^bLaan and Paabo (1997).

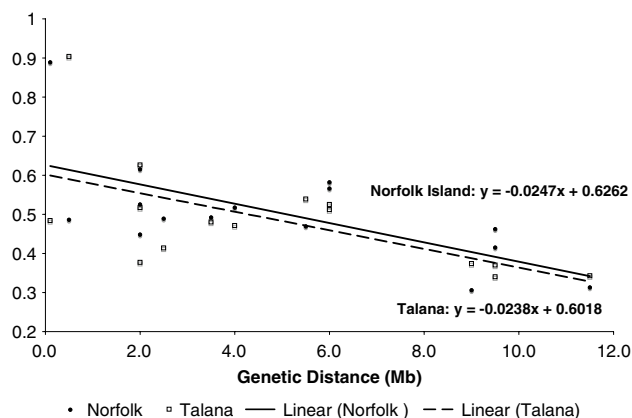


Figure 1 Strength of linkage disequilibrium (LD) evaluated as multiallelic D' values in the Norfolk Island population versus Talana population stratified physical distances for the chromosome Xq13.3 region.

comparing the total variance of genetic diversity. The difference was significant between the Norfolk Island and Talana populations ($P = 0.01$). However, considering the Norfolk Island sample possessed higher levels of interlocus variance (Table 3), Kalinowski (2002) suggests that if more markers were sampled in the Norfolk population, this variance could be minimized.

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 1 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 similar rate over genetic distance to the Norfolk Island population.

LD analysis *NOS2A* SNPs

This study investigated three SNPs within the *NOS2A* gene located on chromosome 17q11.2–q12 that were previously identified and analyzed 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 samples, 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. Table 4 illustrates the minor allele frequency (MAF) comparison between different study populations.

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 4, 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 MAFs for the three SNPs in the Norfolk population are indicative of a population

Table 4 Comparison of minor allele frequencies (MAF) 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 ^a	<0.00001 ^a	0.00297 ^a

^aNot significant if Gambian data are removed from analysis.

with reduced genetic heterogeneity when compared to outbred populations.

Highly significant LD was found to exist between alleles of all three SNPs in the Norfolk Islander population ($P < 0.000001$). 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 the –1026/–1659 marker pair, all the SNPs studied in the Norfolk population were found to be in complete and significant LD with each other.

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 5).

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 (Kaessmann *et al.*, 1999; Zavattari *et al.*, 2000; Angius *et al.*, 2001, 2002a; Pereira and Pena, 2006) allowing comparison between different isolates. The Norfolk Island population study found that LD extends up to 9.5–11 Mb 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 MAF of all three SNPs (when compared to outbred Australian, UK and Gambian populations) and increased frequency of the most common haplotype, both are indicators of

Table 5 NOS2A SNP haplotype frequencies in Norfolk Islander, Caucasian and Gambians

Identifier	Haplotype			Population frequency			
	−1026g/t	−1659c/t	−2447g/c	Gambian ^a	UK Caucasian ^a	Australian Caucasian	Norfolk Islander
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

^aBurgner *et al.* (2003).

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 2 kb 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 the 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. First, 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, minimizing the confounding effects of non-genetic variables. These characteristics have been exploited in the study of rare genetic disorders caused by single defective genes (Puffenberger *et al.*, 1994; Nikali *et al.*, 1995; Newman *et al.*, 2003) and can also be advantageous in complex disease gene mapping (Sheffield *et al.*, 1998; Bourgain *et al.*, 2000; Peltonen *et al.*, 2000; 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 under-

standing 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 outbred populations (Sheffield *et al.*, 1998; Bourgain *et al.*, 2000; Peltonen *et al.*, 2000).

Acknowledgements

We thank Paul Jardine (GU) for computational support and Dr Peter Visscher (QIMR) for his suggestions and continued assistance. We also like to acknowledge the input and suggestions of the two reviewers of this manuscript. Lastly, our appreciation for those from Norfolk Island who volunteered for this study must be noted as the project would not have been possible without participation by these individuals.

References

- 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: 129–156.
- Abecasis GR, Cookson WO (2000). GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* 16: 182–183.
- Abney M, McPeck MS, Ober C (2000). Estimation of variance components of quantitative traits in inbred populations. *Am J Hum Genet* 66: 629–650.
- Abney M, McPeck MS, Ober C (2001). Broad and narrow heritabilities of quantitative traits in a founder population. *Am J Hum Genet* 68: 1302–1307.
- Abney M, Ober C, McPeck MS (2002). Quantitative-trait homozygosity and association mapping and empirical genome-wide significance in large, complex pedigrees: fasting serum-insulin level in the Hutterites. *Am J Hum Genet* 70: 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: 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: 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: 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**: 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**: 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** (Part 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**: 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**: 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**: 320–323.
- Edgecombe J (1999). *Norfolk Island—South Pacific: Island of History and Many Delights*, 2nd edn. JM Edgecombe: Thornleigh, Australia, 179pp.
- Falchi M, Forabosco P, Mocchi 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**: 1015–1031.
- Gulcher J, Stefansson K (1998). Population genomics: laying the groundwork for genetic disease modeling and targeting. *Clin Chem Lab Med* **36**: 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**: 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** (Part 4): 281–297.
- Hoare M (1999). *Norfolk Island: A Revised and Enlarged History 1774–1998*, 5th 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**: 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**: 78–81.
- Kalinowski ST (2002). Evolutionary and statistical properties of three genetic distances. *Mol Ecol* **11**: 1263–1273.
- Kayser M, Brauer S, Weiss G, Schiefenhover 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**: 281–302.
- Laan M, Paabo S (1997). Demographic history and linkage disequilibrium in human populations. *Nat Genet* **17**: 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**: 613–619.
- Lautenberger JA, Stephens JC, O'Brien SJ, Smith MW (2000). Significant admixture linkage disequilibrium across 30 cM around the FY locus in African Americans. *Am J Hum Genet* **66**: 969–978.
- Lewis PO, Zaykin D (2002). Genetic Data Analysis (GDA): computer program for the analysis of genetic data. Free program distributed by the authors of the Internet from the GDA Homepage: <http://chee.unm.edu/gda/>.
- Lewontin RC (1988). On measures of gametic disequilibrium. *Genetics* **120**: 849–852.
- Ludbrook J (1998). Multiple comparison procedures updated. *Clin Exp Pharmacol Physiol* **25**: 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** (Part 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**: 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 USA* **95**: 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, McPeck 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**: 137–144.
- Newman DL, Abney M, McPeck MS, Ober C, Cox NJ (2001). The importance of genealogy in determining genetic associations with complex traits. *Am J Hum Genet* **69**: 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**: 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**: 1088–1095.
- Ober C, Abney M, McPeck MS (2001). The genetic dissection of complex traits in a founder population. *Am J Hum Genet* **69**: 1068–1079.
- Peltonen L, Jalanko A, Varilo T (1999). Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* **8**: 1913–1923.
- Peltonen L, Palotie A, Lange K (2000). Use of population isolates for mapping complex traits. *Nat Rev Genet* **1**: 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**: 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**: 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**: 362–365.
- Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, Stoneking M (1995). Evolutionary history of the COII/tRNA^{Lys} intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* **12**: 604–615.
- Schneider S, Roessli D, Excoffier L (2000). *Arlequin ver 2.000: A Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, 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**: 391–396.

- Shifman S, Darvasi A (2001). The value of isolated populations. *Nat Genet* **28**: 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**: 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**: 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**: 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**: 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**: 51–59.
- Weiss LA, Pan L, Abney M, Ober C (2006). The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet* **38**: 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**: 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**: 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**: 2947–2957.