Genomic inbreeding and runs of homozygosity are associated with cardiovascular disease traits in the Norfolk Island genetic isolate

Miles C Benton[1], Donia Macartney-Coxson[2], David Eccles[1], Geoff Chambers[3], Lyn Griffiths[1], and Rodney A Lea[1]

[1]Queensland University of Technology
[2]Institute of Environmental Science and Research (ESR)
[3]Victoria University of Wellington

September 12, 2018

1 Abstract

Background: When designing genetic studies for any isolated human group it is crucial to understand it's population structure (i.e. founder effects, admixture, inbreeding) in order to fully harness the power of the target population in the detection of disease related genetic markers. The aim of this study was to characterise underlying patterns of genomic homozygosity in the Norfolk Island genetic isolate with the idea that once characterised these properties will be useful in understanding the genetics of complex traits in this cohort. This report hinges on information from a reconstructed multigenerational pedigree from Norfolk Island and available SNP genotype data for over 500 individuals comprising the pedigree. Pedigree and marker derived inbreeding coefficients were calculated using the IBDLD software, as well as runs of marker derived locus-specific homozygosity. Homozygosity by descent (HBD) was used to assess the locus-specific patterns of inbreeding. Statistical association testing was performed to explore relationships between cardiovascular disease related endophenotypes and the identified inbreeding and homozygosity patterns.

Results: Calculation of inbreeding based on the reconstructed pedigree structure revealed an average inbreeding coefficient of 0.011, a relationship that lies between second cousins and second cousins once removed. A combination of pedigree and SNP information was used to identify an overall prevalence of consanguinity in the genotyped core-pedigree individuals of 87%, with a mean marker-generated inbreeding coefficient of 0.011, which is consistent with the pedigree derived estimate. Numerous runs of HBD were identified across the genotyped core-pedigree individuals, with large regions of homozygosity observed on chr 6 (in the HLA region). One of the primary goals of identifying these multiple levels of genomic structure was to be able to make use of them in the search for potential markers of complex disease. Statistically significant correlations were observed between several CVD risk traits and inbreeding. The strongest correlation was observed between an aggregate risk score for CVD and marker derived inbreeding

coefficient (r=0.389, P<2.4x10-11). Moreover, locus-specific HBD associations were observed for several CVD risk traits and SNPs localising to chr 5, 6 and 11 (P<0.05).

Conclusions: This analysis has further characterised the unique population structure forged in the Norfolk population over the last 200 years. Estimates for inbreeding are ˜10% and a new set of measurable indices that represent both runs of homozygosity have been added. Exploratory analyses comparing homozygosity indices with CVD risk traits suggest that individuals with close genetic ties to the original founders may have a potential predisposition to CVD. These findings will be important for future genetic analysis of complex disorders and indicate that homozygosity mapping may shed light on CVD risk in this cohort.

32 Background

33 The Norfolk Island community is an isolated population of whom the majority are direct descen-
34 dants of 18th century European Bounty mutineers and Polynesian women who relocated to NI from
35 Pitcairn Island in 1856 [1]. The majority this population can trace their genetic heritage back to
36 a small number of families derived from the original Bounty mutineers and Polynesians. Genomic
37 structure refers to the proportion of an individuals genome that is contributed to by either one
38 or various ancestors; structure is therefore an individuals 'genomic pattern'. This study aimed to
39 explore the phenomena that shaped the genomic structure of a population, leading to the forma-
40 tion of a unique genetic architecture with properties that aid in the identification and mapping of
41 complex disease.

42 *Inbreeding*

43 Just as admixture is important in shaping the genomic architecture of individuals within a popu-
44 lation, the concept of inbreeding (consanguinity) is equally important. Inbreeding is a descriptive
45 term for the offspring of matings between genetically related individuals. This phenomenon leads
46 to the reduction of genetic variation in offspring from inbred matings over time. The degree of
47 this genetic reduction (an increase in homozygosity) is determined by: a) the closeness of the con-
48 sanguineous relationship and b) the number of past consanguineous relationships within the same
49 lineage. The inbreeding coefficient, F, represents the probability of 2 alleles being identical by de-
50 scent (IBD). **Table 1** represents F values for one generation of inbreeding, with no prior generations
51 of inbreeding being taken into account. The value of F will increase above these expected values
52 with increasing generations of inbreeding (as might be expected in population isolates).

53 An inbreeding calculation may be used to determine the general genetic distance among relatives

3

by multiplying the inbreeding coefficient by two, because any progeny would have a 1 in 2 risk of actually inheriting the identical alleles from both parents. For instance, the parent/child or sibling/sibling relationships have 50% identical genetics. Generally the sharing of identical genetic material is considered detrimental to the overall viability and strength of the total gene pool as it greatly reduces the overall genetic variation, which can lead to a decrease in the fitness of a population [2]. This generally leads to an increased chance of sequential offspring inheriting recessive or deleterious traits. This reduction in genetic diversity can be a useful tool in the detection and identification of disease related genes as the increased frequencies of homozygous alleles can be viewed as an 'amplification' of potential genetic markers of disease within the inbred cohort [3]. Thus, the presence of inbreeding and AIMs can provide increased power for the detection of disease markers as demonstrated by a number of studies to date [4]; [5]; [6].

Runs of homozygosity/homozygosity by descent

Individuals born of consanguineous (inbred) union have genomic segments which are homozygous as a result of inheriting identical ancestral genomic segments from both parents. Homozygosity by descent (HBD), also known as autozygosity, explains the presence of the same alleles at a given position that arise from inheritance of an identical ancestral allele through consanguineous mating events. Runs of HBD can be viewed as segments of the genome that span large regions of chromosomes, sometimes several Mb in size, that result from recombination (or lack of recombination) of chromosomal segments inherited from the same ancestral source. Numerous studies have highlighted increased prevalence of recessive diseases within consanguineous populations [7]; [8]; [9]. HBD-mapping involves the detection of large 'runs' of homozygous alleles, and their association with disease within small inbred families or population isolates. This method was originally developed using micro-satellites/STRs, but was adapted to denser SNP arrays [10]. SNP based HBD-mapping has been used successfully to identify rare recessive disease associated genes in numerous closely related populations [11]. HBD-mapping has also been utilised to detect disease 'signatures', a collective group of alleles spread across the genome, in more distantly related co-

horts [12]. Recent advances in next-generation sequencing technologies are providing easy access to an even greater depth of information. For example the use of exome sequencing to accurately identify runs of homozygosity in large cohorts of individuals [13].

History of the Norfolk Island population

Norfolk Island is a small volcanic island located in the South Pacific about 1600 km north-east of Sydney, Australia. The Island was initially populated by the seafaring Polynesians during the 14th/15th century, however their settlement was brief and unsuccessful and they moved on. It wasn't until 1774 that Captain James Cook rediscovered NI, and it was subsequently used as a British penal colony on and off from 1788 until 1855. The history of the current occupants of NI originates from the events and actions of the Bounty Mutineers. The mutiny of the HMS Bounty was led by acting Lieutenant Fletcher Christian, who along with 18 other men overthrew the captain and remaining crew members, and sailed to Tahiti. In an attempt to hide from British retribution, Christian lead 9 Bounty Mutineers (Isle of Mann and British Ancestry), 6 Tahitian men (Polynesian Ancestry) and 12 Tahitian women plus a baby girl (Polynesian Ancestry) to the Pitcairn Islands, where the Bounty was sunk to avoid detection as well as eliminate escape from the island. Conflict was rife amongst the small community during the first 3 years of the settlement of Pitcairn, with numerous murders as well as suicide, leaving all Tahitian men and 7 Mutineers dead. When Ned Young succumbed to respiratory failure in 1800, John Adams was the sole surviving male on Pitcairn. Slowly the island's population continued to grow, later bolstered by three further male European settlers – John Buffett and John Evans in 1823, and G.H. Nobbs in 1828 [14]. These three men were the only outsiders to settle permanently on the island and subsequently they married into the community. Due to population growth and the severe dwindling of resources on Pitcairn the British Government allowed the Pitcairn Islanders to settle NI, which had recently been abandoned by the British as a penal colony. On June 8 1856 the descendants of the Bounty Mutineers and Tahitian women who were previously inhabitants of Pitcairn Island were relocated, with a total population of 193 settlers making NI their new home.

Today NI has a population of 1576 permanent residents, with approximately 1200 of these being adults. A recent census documents approximately half (N=750) of the permanent population are descendants of the Pitcairn founders. Until recently, strict immigration and quarantine laws have severely restricted the settling of new founders to the island. These laws, along with the sheer isolation of the island have resulted in a lack of interaction with other populations.

Previously Identified Genomic Structure within the NI population

The concepts explained above outline the mechanisms that create and shape the unique genomic structure seen within admixed and genetic isolate populations. Previous work estimated global admixture within the NI population using a set of 128 AIMs derived from the HapMap database [15]. It was established that the ancestry proportions in the population were 88% European vs 12% Polynesian. Further work and initial validation of the broader pedigree structure was documented using a small set of microsatellite markers. This study included an overview of the presence of inbreeding with the population, estimating the average inbreeding coefficient as F=0.011 for the established pedigree [16]. To date these indices have not been estimated using the recently available SNP data, which should lead to increased accuracy of estimation. The SNP data can be used to facilitate the estimation of locus specific admixture as well as runs of homozygosity, two structural elements that have yet to be derived in the NI population. Doing this will expand knowledge of the NI population by building upon the previously identified unique structure. It is likely that this structure could influence genetic associations with disease in the NI population, and is an integral component of disease gene mapping when using population isolates. It is important to take factors such as genomic structure into account as they have been demonstrated to have the potential to artificially inflate the false discovery rate in association studies when not accounted for [17]. As such this cannot be ignored, and should be considered crucial to experimental design moving forward into future analyses.

This study aimed to expand upon previous work that identified estimates of ancestry and pedigree-based inbreeding within the NI population. This was achieved using dense genotype data, which resulted in the improvement in accuracy of various structure parameter estimates (admixture and inbreeding). By incorporating a reconstructed pedigree, recently available genome-wide SNP data and new software methods, extensive runs of both homozygosity as well as locus-specific admixture were observed in the genotyped core-pedigree individuals. These indices were previously unidentified in the NI population and will aid in future studies such as 'admixture-mapping' and disease association studies. This work also reinforces previous findings, and documents novel observations, demonstrating that the unique genomic structure in the NI pedigree has made it an ideal tool for disease gene/marker mapping, leading to the identification of genomic regions and structure correlated to a variety of MetS and CVD related traits for which the population is at a known increased prevalence.

Methods

Cohort Collection and Ethics

Accurate and detailed historical accounts have been used by genealogists to create and maintain a well-documented database of the entire Norfolk Island population, spanning all the way back to the original founders. This pedigree has been drawn up and is maintained in a genealogy program known as Brother's Keeper. The pedigree includes ˜5700 individuals coalescing over 11 generations or 200 years back to the original 9 European sailors and 12 Tahitian women. The Norfolk Island Health Study, which has already been well established in previous research [18]; [19], sampled individuals from the lower four generations of the pedigree and included 386 (64 %) individuals possessing lineages back to the founders and 216 individuals (36 %) who were considered to be new founders and did not show direct ancestral links . An updated core pedigree was constructed using this information and genetic information as it became available through genetic studies. Currently the core pedigree structure contains those individuals that are most closely related and coalesce directly back to the original founders. The Norfolk Island Health Study (NIHS) has already been

well established in previous research . In this study, we used a representative group of Norfolk Island individuals selected from the pedigree, meaning that they relate back to the original founders, and we have phenotype and genotype information for them. The total number of core pedigree members selected was 330 (this was adjusted to exclude individuals under the age of 18 years), which consisted of 152 males and 178 females. All individuals gave written informed consent. Ethical approval was granted prior to the commencement of the study by the Griffith University Human Research Ethics Committee (ethical approval no: 1300000485) and the project was carried out in accordance with the relevant guidelines, which complied with the Helsinki Declaration for human research.

Genome-wide Genotyping

EDTA anticoagulated venous blood samples were collected from all participants. Genomic DNA was extracted from blood buffy coats using standard phenol-chloroform procedures (Qiagen). Genome-wide genotyping was carried out using the Illumina Human610-Quad v1.0 beadchip. Raw data from Illumina idat files was SNP genotyped in R using the CRLMM package [20]. Genotype data then underwent QC routines using PLINK [21]. Briefly, SNP analysis was restricted to autosomal SNPs with minor allele frequency >0.01, call rate >0.95 and Hardy-Weinberg equilibrium testing p-value >10-5. After quality control, 590,603 SNPs were used for further analyses.

Inbreeding and runs of homozygosity in Norfolk Island

Estimation of pedigree-based inbreeding (FPED) was done in R using the package 'pedigree' [**21**], calculations were based upon the reconstructed 1388 core-pedigree (see [22] for details). Marker derived inbreeding (FMARKER) was calculated using the software IBDLD [23]. IBDLD is implemented in a way that it overcomes the issues around exact multipoint estimation of IBD in large pedigrees, and also eliminates the difficulty of accommodating the background linkage disequilibrium (LD) that is present in high-density genotype data [23]. While generating IBD matrices is the primary function of IBDLD, it also has several other important applications including: calcu-

lation of inbreeding coefficients for each genotyped individual per chromosome; per chromosome estimation of allelic homozygosity (homozygous by descent probability [HBD]), and segmental sharing analysis. Unless otherwise stated, all analysis runs of IBDLD were set at 10000 simulations. Inbreeding coefficients, calculated on a per chromosome basis using the *-ibc* function of IBDLD, were averaged for each individual to obtain an overall genome-wide FMARKER. To summarise the number of consanguineous relationships a cut-off was implemented at 4 decimal places to determine final FMARKER values. This cut-off was introduced as a number of individuals were observed to have F values in the range of 1x10-6- 1x10-8; these values are so close to approaching 0 that it was deemed more accurate to refer to them as not showing any inbreeding in their ancestry.

Runs of homozygosity were estimated using the *-hbd* function of IBDLD, again on a per chromosome basis. IBDLD refers to marker-derived homozygosity as homozygous by descent (HBD), which is the probability that two alleles are inherited from a single source (ancestor). After running IBDLD a file containing HBD results is generated, in which each SNP is ranked between 0 and 1 (0 being no probability of HBD, 1 being 100% probability of HBD) for each genotyped individual. To visualise these results both genome-wide and chromosome wide plots of average HBD were generated in R [24]. 'Peaks', or extended runs of HBD (homozygosity), were inferred as a set of SNPs continuously exhibiting an average HBD probability greater than the population average of 0.011.

Correlations between Inbreeding and Endophenotypes

Isolated populations have been repeatedly shown to have increased power to detect disease associated markers [24]; [25]; this is due to the unique properties such as population inbreeding that arise due to founder effects and genetic bottlenecks. In order to explore the potential importance and power of the NI pedigree for disease association marker detection correlation analyses between estimates of inbreeding with endophenotypes for CVD were carried out. All correlations were generated in R [24] using the Pearson's correlation function (2-tailed P-values were generated). **The endophenotypes examined here included ....**

Results

Homozygosity and Inbreeding in Norfolk Island Pedigree

Inbreeding reduces genetic diversity within a population. The measure for inbreeding (also known as consanguinity) within a population is measured by the inbreeding coefficient (F). Pedigree-based inbreeding (FPED) was estimated using the reconstructed core-pedigree (N=1388). The mean FPED was 0.011 with a maximum FPED=0.28. Of the 1388 individuals in this pedigree, 400 were estimated to have an inbreeding coefficient greater than zero indicating inbreeding has occurred sometime in the past. Genetic markers can also be used to estimate inbreeding. Using data from 502 SNP-genotyped individuals from the core NI pedigree we calculated FMARKER values. **Figure 1** shows average FMARKER values per chromosome. It was observed that 87% (N=439) of all 502 genotyped individuals exhibited an F value greater than 0 (using a cut-off at 4 decimal places, see methods for detail). Using this information the global average F was calculated (F=0.011), with a maximum F=0.215 being observed. It is interesting to note that both the FPED and FMARKER are identical (0.011); this validates the accuracy of the updated core-pedigree and the approach used in the cleaning and reconstruction process.

The next step was to characterise the patterns of inbreeding across the genome of the group of core pedigree individuals. This was performed by calculating runs of homozygosity by descent (HBD) ie. areas of the genome that show a reduction of genetic diversity due to inheritance of analogous alleles. **Figure 2** shows a genome-wide profile of HBD probability across genotyped core-pedigree members. Visualisation of the HBD data was split into 2 separate plots. **Figure 2 A** displays an average locus-specific homozygosity profile for those individuals exhibiting high HBD probability (determined as at least one locus HBD > 0.75). The average HBD for this subset of individuals was 0.042. The visualisation of peaks arising from individuals with higher than normal levels of HBD is informative of potentially smaller groups of closer related individuals within the wider NI pedigree structure. These could potentially be interesting for investigating disease associations, or could

identify smaller sub-pedigrees to further facilitate the tracking of complex traits such as migraine and ocular disorders (glaucoma), both of which show increased prevalence in the NI population [26]; [27]; [28]. **Figure 2 B** shows the average genome-wide locus-specific homozygosity profile for all 502 genotyped core-pedigree individuals. The average genome-wide HBD for all genotyped individuals was 0.011; it should be noted that this is the exact same value as the estimated FMARKER for these individuals. This is due to the fact that both methods are calculating inbreeding with FMARKER being a genome average and HBD being locus-specific. This broader genome-wide profile of all genotyped individuals shows numerous areas of greatly increased HBD, several of the longer genomic regions (HBD > 0.011) are detailed in **Table 2.** The largest observed 'peak' of HBD probability on chromosome 6 was 0.13, which spans 18Mb showing an average HBD of 0.028 across the span. This peak on chromosome 6 was unique when compared to other areas of increased HBD showing multiple peaks within the same determined run of HBD (**Figure 3**). Interestingly the area of highest HBD resides on top of the well-defined human leukocyte antigen (HLA) region; a highly variable area of the genome well studied and known for its role in the immune system/response and disease. Another region of high HBD was observed as 2 peaks on chromosome 11. The second peak lies on a large family of olfactory receptor genes. These genes are important in the detection and interpretation of odours [29], and are reported to show increased genetic variation in order to account for the potential limitless amount of detectable odours [30]. **Additional File 1** shows exactly the same data as the genome-wide figures, but have been visualised in smaller blocks of chromosomes in order to better display the location and extent of HBD across a given chromosomal region.

Correlation between inbreeding/HBD and CVD endophenotypes

An exploratory correlation analysis was conducted to investigate relationships between genomic inbreeding and 10 CVD risk traits (endophenotypes). **Table 3** shows that all 10 traits exhibited some evidence of association with inbreeding (P<0.05) The strongest correlation was between CVD risk and FMARKER (Pearson's r=0.389, P<2.4x10-11). These new results are consistent with

previously reported between inbreeding and CVD related traits in the NI population [16]. The current study therefore supports these findings, and builds upon them with previously unidentified trait relationships which could indicate important areas for future research.

New stuff here.

Discussion

This section explored the unique genomic structure that underlies the NI pedigree. This structure has resulted from the rich history of the original Bounty Mutineers and Polynesian founders, being shaped by genetic bottle necks, founder effects and population admixture over the span of approximately 200 years. Previous calculations for both admixture and inbreeding coefficients have been estimated in the NI population and using these metrics it has been established that there are correlations between ancestry and CVD risk in NI [31]; [32]. More specific population effects upon genomic structure, such as locus-specific admixture and runs of homozygosity, have not previously been explored in the NI population. These indices have potential implications in terms of disease association and will provide important foundations for future studies in NI, especially in the investigation of disease phenotypes that differ in frequency between the European and Polynesian ancestral populations.

Using a dense set of SNPs an estimate of average inbreeding coefficient of F=0.011 was calculated for the NI population. An F of this value indicates an average relationship level somewhere between second cousins (0.0156) and second cousins once removed (0.0078). This estimate is substantially smaller than that calculated previously by Macgregor et al., F=0.044; there are several potential explanations for this [16]. Firstly, this study was able to use a more accurate, updated pedigree. Secondly, a greater depth of data in terms of a dense set of genotype data (SNPs) was available, as opposed to a small set of microsatellites (STRs). This genomic data and the updated genealogical

12

information facilitated the reconstruction of a more accurate representation of the core NI pedigree, which should aid in more accuracy in such calculations.

As HBD is related to inbreeding, genomic regions which show increased HBD probabilities could potentially identify loci that show a lack of genetic diversity. There are several such areas with above average HBD in the genotyped core-pedigree NI population. There are major implications for this drop in genetic diversity, especially in areas such as the HLA region on chromosome 6, which was identified as exhibiting the largest average HBD. It is well established that the HLA region contains a large amount of immune function related genes, many coding for immune cell receptors that will potentially bind and recognise antigens (foreign peptides). The HLA region requires increased genetic variation to allow near limitless increased receptor specificity from a limited number of genes. High genetic variation is critical to the function of the HLA region as it allows near limitless variation to be introduced to the receptors antigen 'recognition site', which in turn increases the potential number of foreign antigens that can be detected. Decreased variation in this region could potentially be detrimental.

Chromosome 11 also had a larger degree of concentrated increased HBD, one particular peak was observed upon a large region of olfactory receptor genes. Olfactory receptors determine the way in which an organism interprets odours [30]; [33]. As with the HLA region, the olfactory receptor genes are limited in number, therefore increased variation within the gene family is required to enable detection and interpretation of near limitless possible odours [34]; [35]; [36]. Thus decreased genetic variation across this region could lead to a potential reduction in odour detecting abilities. Interestingly HLA may also be related to people's detection and perception of the odour; with several studies observing association between HLA variation and preference to odour; this may be involved in mate selection [37]; [38], as at least one study found a lower than expected rate of HLA similarity between spouses in an isolated community [39]. Additionally, research has shown that more married couples have distinctly different HLA (MHC) genomic backgrounds/variation than would be expected by chance alone, suggesting that selection is potentially driving for composition

and differentiation within the immune systems of offspring so they are able to adapt to the threat of new diseases. Another reason for this could also an avoidance of inbreeding in an attempt to maintain a higher amount of genetic diversity within a population. In this context it is interesting that in this study has identified decreased variation in the form of increased HBD at both the HLA and olfactory receptors gene loci which warrants further exploration.

The initial aim was to identify relationships between disease related endophenotypes and indices of the unique structural properties measured in the NI population. Using updated estimates of both marker-derived admixture and inbreeding numerous significant relationships between both measures and a range of Metabolic and CVD related traits were observed, which included a robustly calculated risk score for CVD and clinical diagnosis of Metabolic Syndrome. This is not the first study to identify these factors showing increased prevalence within the NI cohort. An initial study by Bellis et al., (2006) established the basis of increased risk for CVD related disorders and outlined baseline phenotype data . This was followed by several linkage analyses using STR data which established initial genomic maps and identified loci showing significant associations with CVD traits [40]; [19]. This study builds upon the previous work and identifies novel findings. This is the first study to examine the high-density SNP data in association with CVD/metabolic related traits, and to use an integrative genomics approach. Showing these strong structure vs risk trait relationships provides evidence that the reconstruction of the NI core-pedigree is robust and that the genomic data (SNPs) are concordant with this. This provides confidence for future disease gene mapping studies including: association; linkage and 'admixture-mapping' in this population.

Conclusion

Study of isolated populations require an understanding of the unique population history and admixture, which has led to unique genomic structure. Genomic structure in populations has the potential to influence genetic associations with disease, and is therefore important to consider in future study design. This knowledge can then be used appropriately as a valuable tool in disease

mapping and association studies. This work increases the accuracy of previous estimates of inbreeding and documents for the first time runs of HBD in the NI population. Additionally, this study identified significant correlations between these unique structural components and disease risk traits for Metabolic Syndrome and CVD. Importantly both increased prevalence and underlying population/genetic based association with Metabolic Syndrome in the NI pedigree has been identified. This provides strong justification for further examination of the NI population in the context of Metabolic Syndrome risk and prevalence. Future research should focus in on the identified area's of locus-specific admixture and HBD in light of the correlations with MetS and related traits.

Competing interests

The authors have no competing interests to declare.

Authors' Contributions

Contributions here. . . .

Funding

References

1. Edgecombe J. Norfolk Island - South Pacific: Island of History and Many Delights [Internet]. J.M. Edgecombe; 1991. Retrieved from: https://books.google.com/books?id=HbfiAAAAMAAJ

2. Bittles AH, Neel JV. The costs of human inbreeding and their implications for variations at the DNA level. Nature Genetics [Internet]. Springer Nature; 1994;8:117–21. Retrieved from: https://doi.org/10.1038%2Fng1094-117

3. Bulayev OA, Pavlova TA, Bulayeva KB. The effect of inbreeding on aggregation of complex diseases in genetic isolates. Russian Journal of Genetics [Internet]. Pleiades Publishing Ltd; 2009;45:961–8. Retrieved from: https://doi.org/10.1134%2Fs1022795409080109

4. Latini V, Sole G, Doratiotto S, Poddie D, Memmi M, Varesi L, et al. Genetic isolates in Corsica (France): linkage disequilibrium extension analysis on the Xq13 region. European Journal of Human Genetics [Internet]. Springer Nature; 2004;12:613–9. Retrieved from: https://doi.org/10.1038%2Fsj.ejhg.5201205

5. Thompson EE, Sun Y, Nicolae D, Ober C. Shades of gray: a comparison of linkage disequilibrium between Hutterites and Europeans. Genetic Epidemiology [Internet]. Wiley; 2009;34:133–9. Retrieved from: https://doi.org/10.1002%2Fgepi.20442

6. Kenny EE, Kim M, Gusev A, Lowe JK, Salit J, Smith JG, et al. Increased power of mixed models facilitates association mapping of 10 loci for metabolic traits in an isolated population. Human Molecular Genetics [Internet]. Oxford University Press (OUP); 2010;20:827–39. Retrieved from: https://doi.org/10.1093%2Fhmg%2Fddq510

16

7. el-Hazmi MA, al-Swailem AR, Warsy AS, al-Swailem AM, Sulaimani R, al-Meshari AA. Consanguinity among the Saudi Arabian population.. Journal of Medical Genetics [Internet]. BMJ; 1995;32:623–6. Retrieved from: https://doi.org/10.1136%2Fjmg.32.8.623

8. Zlotogora J. Autosomal recessive diseases among Palestinian Arabs.. Journal of Medical Genetics [Internet]. BMJ; 1997;34:765–6. Retrieved from: https://doi.org/10.1136%2Fjmg.34.9.765

9. Farrer LA. Identification of multiple loci for Alzheimer disease in a consanguineous Israeli-Arab community. Human Molecular Genetics [Internet]. Oxford University Press (OUP); 2003;12:415–22. Retrieved from: https://doi.org/10.1093%2Fhmg%2Fddg037

10. Woods CG. A new method for autozygosity mapping using single nucleotide polymorphisms (SNPs) and EXCLUDEAR. Journal of Medical Genetics [Internet]. BMJ; 2004;41:e101–e101. Retrieved from: https://doi.org/10.1136%2Fjmg.2003.016873

11. Woods CG, Cox J, Springell K, Hampshire DJ, Mohamed MD, McKibbin M, et al. Quantification of Homozygosity in Consanguineous Individuals with Autosomal Recessive Disease. The American Journal of Human Genetics [Internet]. Elsevier BV; 2006;78:889–96. Retrieved from: https://doi.org/10.1086%2F503875

12. Keller MC, Simonson MA, Ripke S, Neale BM, Gejman PV, Howrigan DP, et al. Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor. Gibson G, editor. PLoS Genetics [Internet]. Public Library of Science (PLoS); 2012;8:e1002656. Retrieved from: https://doi.org/10.1371%2Fjournal.pgen.1002656

13. Carr IM, Bhaskar S, Sullivan JO, Aldahmesh MA, Shamseldin HE, Markham AF, et al. Autozygosity Mapping with Exome Sequence Data. Human Mutation [Internet]. Wiley; 2012;34:50–6. Retrieved from: https://doi.org/10.1002%2Fhumu.22220

14. Refshauge WF, Walsh RJ. Pitcairn Island: fertility and population growth 1790–1856. Annals of Human Biology [Internet]. Informa UK Limited; 1981;8:303–12. Retrieved from: https://doi.org/10.1080%2F03014468100005101

15. McEvoy BP, Zhao ZZ, Macgregor S, Bellis C, Lea RA, Cox H, et al. European and Polynesian

17

admixture in the Norfolk Island population. Heredity [Internet]. Springer Nature; 2009;105:229–34. Retrieved from: https://doi.org/10.1038%2Fhdy.2009.175

16. Macgregor S, Bellis C, Lea RA, Cox H, Dyer T, Blangero J, et al. Legacy of mutiny on the Bounty: founder effect and admixture on Norfolk Island. European Journal of Human Genetics [Internet]. Springer Nature; 2009;18:67–72. Retrieved from: https://doi.org/10.1038%2Fejhg.2009.111

17. Jiang Y, Epstein MP, Conneely KN. Assessing the Impact of Population Stratification on Association Studies of Rare Variation. Human Heredity [Internet]. S. Karger AG; 2013;76:28–35. Retrieved from: https://doi.org/10.1159%2F000353270

18. Bellis C, Hughes RM, Begley KN, Quinlan S, Lea RA, Heath SC, et al. Phenotypical Characterisation of the Isolated Norfolk Island Population Focusing on Epidemiological Indicators of Cardiovascular Disease. Human Heredity [Internet]. S. Karger AG; 2005;60:211–9. Retrieved from: https://doi.org/10.1159%2F000090545

19. Bellis C, Cox HC, Dyer TD, Charlesworth JC, Begley KN, Quinlan S, et al. Linkage mapping of CVD risk traits in the isolated Norfolk Island population. Human Genetics [Internet]. Springer Nature; 2008;124:543–52. Retrieved from: https://doi.org/10.1007%2Fs00439-008-0580-y

20. Scharpf RB, Irizarry RA, Ritchie ME, Carvalho B, Ruczinski I. Using the R Package crlmm for Genotyping and Copy Number Estimation. Journal of Statistical Software [Internet]. Foundation for Open Access Statistic; 2011;40. Retrieved from: https://doi.org/10.18637%2Fjss.v040.i12

21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics [Internet]. Elsevier BV; 2007;81:559–75. Retrieved from: https://doi.org/10.1086%2F519795

22. Benton MC, Stuart S, Bellis C, Macartney-Coxson D, Eccles D, Curran JE, et al. 'Mutiny on the Bounty': the genetic history of Norfolk Island reveals extreme gender-biased admixture. Investigative Genetics [Internet]. Springer Nature; 2015;6. Retrieved from: https://doi.org/10.

1186%2Fs13323-015-0028-9

23. Han L, Abney M. Identity by descent estimation with dense genome-wide genotype data. Genetic Epidemiology [Internet]. Wiley; 2011;:n/a–n/a. Retrieved from: https://doi.org/10.1002%2Fgepi.20606

24. Heutink P. Gene finding in genetically isolated populations. Human Molecular Genetics [Internet]. Oxford University Press (OUP); 2002;11:2507–15. Retrieved from: https://doi.org/10.1093%2Fhmg%2F11.20.2507

25. Kristiansson K, Naukkarinen J, Peltonen L. Isolated populations and complex disease gene identification. Genome Biology [Internet]. Springer Nature; 2008;9:109. Retrieved from: https://doi.org/10.1186%2Fgb-2008-9-8-109

26. Cox HC, Lea RA, Bellis C, Carless M, Dyer TD, Curran J, et al. A genome-wide analysis of 'Bounty' descendants implicates several novel variants in migraine susceptibility. neurogenetics [Internet]. Springer Nature; 2012;13:261–6. Retrieved from: https://doi.org/10.1007%2Fs10048-012-0325-x

27. Maher BH, Lea RA, Benton M, Cox HC, Bellis C, Carless M, et al. An X Chromosome Association Scan of the Norfolk Island Genetic Isolate Provides Evidence for a Novel Migraine Susceptibility Locus at Xq12. Zwick ME, editor. PLoS ONE [Internet]. Public Library of Science (PLoS); 2012;7:e37903. Retrieved from: https://doi.org/10.1371%2Fjournal.pone.0037903

28. Sherwin JC, Kearns LS, Hewitt AW, Ma Y, Kelly J, Griffiths LR, et al. Prevalence of Chronic Ocular Diseases in a Genetic Isolate: The Norfolk Island Eye Study (NIES). Ophthalmic Epidemiology [Internet]. Informa UK Limited; 2011;18:61–71. Retrieved from: https://doi.org/10.3109%2F09286586.2010.545933

29. Farley S. Regulating olfactory receptors. Nature Reviews Neuroscience [Internet]. Springer Nature; 2004;5:171–. Retrieved from: https://doi.org/10.1038%2Fnrn1361

30. Buck LB. Olfactory Receptors and Odor Coding in Mammals. Nutrition Reviews [Internet]. Oxford University Press (OUP); 2004;62:S184–S188. Retrieved from: https://doi.org/10.1111%

454 2Fj.1753-4887.2004.tb00097.x

455 31. Macgregor S, Bellis C, Lea RA, Cox H, Dyer T, Blangero J, et al. Legacy of mutiny on the
456 Bounty: founder effect and admixture on Norfolk Island. European Journal of Human Genetics
457 [Internet]. Springer Nature; 2009;18:67–72. Retrieved from: https://doi.org/10.1038%2Fejhg.
458 2009.111

459 32. Benton MC, Stuart S, Bellis C, Macartney-Coxson D, Eccles D, Curran JE, et al. 'Mutiny
460 on the Bounty': the genetic history of Norfolk Island reveals extreme gender-biased admixture.
461 Investigative Genetics [Internet]. Springer Nature; 2015;6. Retrieved from: https://doi.org/10.
462 1186%2Fs13323-015-0028-9

463 33. Malnic B, Godfrey PA, Buck LB. The human olfactory receptor gene family. Proceedings of
464 the National Academy of Sciences [Internet]. Proceedings of the National Academy of Sciences;
465 2004;101:2584–9. Retrieved from: https://doi.org/10.1073%2Fpnas.0307882100

466 34. Malnic B, Hirono J, Sato T, Buck LB. Combinatorial Receptor Codes for Odors. Cell [Internet].
467 Elsevier BV; 1999;96:713–23. Retrieved from: https://doi.org/10.1016%2Fs0092-8674%2800%
468 2980581-4

469 35. Niimura Y, Nei M. Evolutionary dynamics of olfactory and other chemosensory receptor genes
470 in vertebrates. Journal of Human Genetics [Internet]. Springer Nature; 2006;51:505–17. Retrieved
471 from: https://doi.org/10.1007%2Fs10038-006-0391-8

472 36. Khafizov K, Anselmi C, Menini A, Carloni P. Ligand specificity of odorant receptors. Journal
473 of Molecular Modeling [Internet]. Springer Nature; 2006;13:401–9. Retrieved from: https://doi.
474 org/10.1007%2Fs00894-006-0160-9

475 37. Ober C, Weitkamp LR, Cox N, Dytch H, Kostyu D, Elias S. HLA and Mate Choice in Humans.
476 The American Journal of Human Genetics [Internet]. Elsevier BV; 1997;61:497–504. Retrieved
477 from: https://doi.org/10.1086%2F515511

478 38. Jacob S, McClintock MK, Zelano B, Ober C. Paternally inherited HLA alleles are associated
479 with women's choice of male odor. Nature Genetics [Internet]. Springer Nature; 2002;30:175–9.

480 Retrieved from: https://doi.org/10.1038%2Fng830

481 39. Ober C, Weitkamp LR, Cox N, Dytch H, Kostyu D, Elias S. HLA and Mate Choice in Humans.

482 The American Journal of Human Genetics [Internet]. Elsevier BV; 1997;61:497–504. Retrieved

483 from: https://doi.org/10.1086%2F515511

484 40. Bellis C, Cox HC, Ovcaric M, Begley KN, Lea RA, Quinlan S, et al. Linkage disequilibrium

485 analysis in the genetically isolated Norfolk Island population. Heredity [Internet]. Springer Nature;

486 2007;100:366–73. Retrieved from: https://doi.org/10.1038%2Fsj.hdy.6801083