Clinical case study meets population cohort: Identification of a *BRCA1* pathogenic founder variant in Orcadians

Shona M. Kerr*[1], Emma Cowan*[2], Lucija Klaric[1], Christine Bell[2], Dawn O'Sullivan[2], David Buchanan[1], Joseph J. Grzymski[3,4], Cristopher V. van Hout[5,6], Gannie Tzoneva[5], Alan R. Shuldiner[5], James F. Wilson[#1,7] and Zosia Miedzybrodzka†[#2,8]

[1]MRC Human Genetics Unit, University of Edinburgh, Institute of Genetics and Cancer, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK.

[2]Department of Medical Genetics, Ashgrove House, NHS Grampian, Aberdeen, AB25 2ZA, UK.

[3]Center for Genomic Medicine, Desert Research Institute, Reno, Nevada, United States.

[4]Renown Health, Reno, NV, USA.

[5]Regeneron Genetics Center, Tarrytown, NY, USA

[6]Laboratorio Internacional de Investigatión sobre el Genoma Humano, Campus Juriquilla de la Universidad Nacional Autónoma de México, Querétaro, Querétaro 76230, México.

[7]Centre for Global Health Research, Usher Institute, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK.

[8]Medical Genetics Group, School of Medicine, Medical Sciences, Nutrition and Dentistry, University of Aberdeen, Polwarth Building, Aberdeen, AB25 2ZD, UK.

* These authors contributed equally to this work

[#] Indicates joint senior authors

† To whom correspondence should be addressed: zosia@abdn.ac.uk

This work was funded by the MRC University Unit award to the MRC Human Genetics Unit, University of Edinburgh, MC_UU_00007/10. LK was supported by an RCUK Innovation Fellowship from the National

Productivity Investment Fund (MR/R026408/1). ORCADES was supported by the Chief Scientist Office of the Scottish Government (CZB/4/276 and CZB/4/710), a Royal Society URF to JFW and Arthritis Research UK.

Abstract

We multiply ascertained the *BRCA1* pathogenic missense variant c.5207T>C; p.Val1736Ala (V1736A) in clinical investigation of breast and ovarian cancer families from Orkney in the Northern Isles of Scotland, UK. We sought to investigate the frequency and clinical relevance of this variant in those of Orcadian ancestry as an exemplar of the value of population cohorts in clinical care, especially in isolated populations. Oral history and birth, marriage and death registrations indicated genealogical linkage of the clinical cases to ancestors from the Isle of Westray, Orkney. Further clinical cases were identified through targeted testing for V1736A in women of Orcadian ancestry attending National Health Service (NHS) genetic clinics for breast and ovarian cancer family risk assessments. The variant segregates with female breast and ovarian cancer in clinically ascertained cases. Separately, exome sequence data from 2,088 volunteer participants with three or more Orcadian grandparents, in the ORCADES research cohort, was interrogated to estimate the population prevalence of V1736A in Orcadians. The effects of the variant were assessed using Electronic Health Record (EHR) linkage. Twenty out of 2,088 ORCADES research volunteers (~1%) carry V1736A, with a common haplotype around the variant. This allele frequency is ~480-fold higher than in UK Biobank participants. Cost-effectiveness of population screening for *BRCA1* founder pathogenic variants has been demonstrated at a carrier frequency below the ~1% observed here. Thus we suggest that Orcadian women should be offered testing for the *BRCA1* V1736A founder pathogenic variant, starting with those with known Westray ancestry.

Keywords

Breast cancer, ovarian cancer, BRCA1, founder, population screening, Orkney

Introduction

Pathogenic variants in *BRCA1* confer a high lifetime risk of breast and ovarian cancer (1-3). Genetic testing for pathogenic variants in *BRCA1* and *BRCA2* is widely available in breast and ovarian cancer, to enable not only early detection and risk reduction, but also to guide cancer treatment, e.g. consideration of the use of olaparib in chemotherapy regimens (4). Predictive testing of unaffected family members is well established, with pre-symptomatic carriers of *BRCA1* pathogenic variants being offered risk-reducing prophylactic bilateral mastectomy, bilateral salpingo-oopherectomy and annual magnetic resonance breast imaging as standard care.

In isolate populations, a pathogenic variant present in a founding or early member can become widespread in later generations, contributing significantly to the overall disease burden. Pathogenic variants in *BRCA1* and *BRCA2* have been described in several isolate and founder populations worldwide, notably Ashkenazi and Sephardi Jews, and Icelanders (5, 6). Genetic screening programmes focused on founder variants in such genes can be cost-effective (7-10).

The Northern Isles of Scotland – the Orkney and Shetland archipelagos- have the most divergent and isolated of all British and Irish populations, with the highest degree of kinship and Norse admixture in the British Isles and Ireland, evidenced in the extensive genealogies, and genome-wide analyses (11). Research by ourselves and others has demonstrated enrichment of rare and low frequency functional variants in isolated populations, including Orkney (12). Enriched rare variants of large effect are of most clinical relevance.

Viking Genes (www.ed.ac.uk/viking) comprises three Northern Isles cohort studies aiming to explore genetic causes of disease - Orkney Complex Disease Study (ORCADES), VIKING I and VIKING II. ORCADES contains a rich data resource of more than 2,000 deeply phenotyped and exome sequenced research subjects with three or four grandparents from the Orkney Islands, ideal for analyses of the frequency and penetrance of clinically relevant variants in the Orcadian population. UK Biobank is a large-scale cosmopolitan biomedical database containing genetic, lifestyle and health information from half a million participants in the UK (13). Linkage to NHS routine electronic health record (EHR) data adds a longitudinal

77 component to both the ORCADES and UK Biobank cohorts. These research cohorts, although not perfect
78 representations of the populations from which they sought to recruit, are sufficiently unbiased for
79 estimation of population frequency of genetic variants.

80 The NHS Grampian genetics team observed the *BRCA1* missense variant, c.5207T>C; p.Val1736Ala, in a
81 number of ovarian and breast cancer cases from Orkney. The *BRCA1* c.5207T>C; p.Val1736Ala variant is a
82 conservative amino acid substitution in the carboxyl-terminal domain, a region known to be important in
83 BRCA1 function. *In vitro* studies suggested that the variant disrupts BRCA1 activity (14, 15). Independent
84 evidence for pathogenicity comes from saturation genome editing of *BRCA1* exons in HAP1 cells, which
85 revealed V1736A to be non-functional in cultured cells (16). A report was published of a severe phenotype
86 patient with ovarian cancer at age 28, short stature, microcephaly and significant toxicity from
87 chemotherapy, with compound heterozygous *BRCA1* variants, c.2457delC, and c.5207T>C; as well as loss
88 of heterozygosity (LOH) in associated tumours (17). This, together with segregation data, led to
89 reclassification of V1736A from a variant of unknown significance to likely pathogenic (17). The
90 interpretation that V1736A is a pathogenic variant was corroborated by an expert panel in the Clinvar
91 database (18), accession VCV000037648, annotated as pathogenic by multiple sources. Our own co-
92 segregation studies in the Orcadian clinical super-kindred detailed below (Methods and data available on
93 request) support the pathogenic nature of the variant.

94 Here, we report for the first time the multiple ascertainment of the *BRCA1* pathogenic missense variant
95 c.5207T>C; p.Val1736Ala (V1736A) rs45553935 as part of routine clinical care in breast and ovarian cancer
96 families from Orkney. We then demonstrate relatedness of V1736A gene carriers using genealogy and
97 haplotyping, and use ORCADES to estimate the population based variant frequency, consider penetrance
98 and make the case for population screening of the variant in ancestral Orcadians. This is an exemplar of
99 the value of phenotyped population cohorts for informing genetic health policy.

100

101 **Materials (Subjects) and Methods**

Clinical case note review Women with breast and / or ovarian cancer with the *BRCA1* missense variant, c.5207T>C; p.Val1736Ala, were identified at the Orkney genetic clinic, and family history of cancer in consenting living and deceased family members was recorded and confirmed from medical records as part of routine clinical genetics care. These oral histories from multiple consultands from multiple nuclear families was supplemented with genealogical information from the Scottish Register of Births, Marriages and Deaths to link family members genealogically.

ORCADES Research Volunteer Recruitment Recruitment to ORCADES took place from 2005-2011, through advertisement and word of mouth. Volunteers were required to be aged 18 or over, with two or more grandparents born in Orkney. More than 90% had three or four Orcadian grandparents. The response rate was excellent, with the final cohort size comprising more than 10% of the total Orkney adult population. Participants attended at least two clinics in Orkney, one for fasting venepuncture and one for physical measurements, and provided broad-ranging consent for research, including for whole genome sequencing, and for their research data to be linkable to their NHS electronic health records. Blood (or very occasionally, saliva) samples from participants were collected, processed and stored using standard operating procedures and managed through a laboratory information management system at the Edinburgh Clinical Research Facility, University of Edinburgh.

ORCADES Cohort Pedigree Information Records of the births, marriages and deaths in Orkney are held at the General Register Office for Scotland (New Register House, Edinburgh). These records, along with relationship information obtained from study participants and genealogies available online, were used to construct a pedigree of ORCADES study participants using RootsMagic software (S&N Genealogy Supplies), which was then amended to reflect the genetic kinship between individuals using genotype data. The complete pedigree dates back to ~900 AD and comprises ~59,000 individuals.

Genotyping DNA from all ORCADES participants was used for genome-wide genotyping on the GSA BeadChip (Illumina) at the Regeneron Genetics Center. Monomorphic genotypes and genotypes with more than 2% of missingness and Hardy-Weinberg equilibrium (HWE) $p<10^{-6}$ were removed, as well as

individuals with more than 3% of missingness. Details of genotyping, sample and variant quality control of UK Biobank genotyping data are described in Bycroft *et al* (13).

Sequencing The fully quality controlled exome sequence data set was prepared at the Regeneron Genetic Centre, following the process detailed for UK Biobank by van Hout et al (19). Details of the quality control of whole exome sequencing on the 200,000 participants from the UK Biobank are described in Backman *et al* (20). The rs45553935 variant was validated by Sanger sequencing (further details in supplementary methods).

Haplotype analysis The ORCADES array genotype data were phased using Shapeit2 v2r837 (21), with the duoHMM option that uses the family-based nature of the data (22). Prior to phasing, the array genotype data were lifted over from the genome build GRCh38 to GRCh37 using liftOver, followed by quality control against the HRC reference panel with Rayner's HRC-1000G-check-bim (v4.2.13) script that was downloaded from https://www.well.ox.ac.uk/~wrayner/tools/. Details of phasing of UK Biobank genotyping data have been described (13). Then, the phased genotypes were used to determine a shared haplotype around rs45553935 using the coarse and fine methods described earlier (23).All methods were performed using R 4.0.2 (R Core Team 2020 https://www.R-project.org/). Haplotypes were defined with custom-built in-house scripts in R. Data handling was performed using data.table and tidyverse R packages. Plots were generated using ggplot2.

A single variant-based haplotype search was performed to determine the haplotype length between the different ORCADES carrier kindreds, and also with the UK Biobank carrier individuals, using a stepwise approach. Using phased genotype data, starting from the rs45553935 rare variant, one SNP at a time was added to define a haplotype. The procedure was repeated until haplotypes of two individuals (both known carriers) no longer matched, providing variant-level resolution of the haplotype length. The procedure was repeated for all pairs of individuals identified as carriers based on the exome sequencing data, both in the ORCADES and UK Biobank datasets. The shortest shared haplotype from ORCADES was then merged with the shortest shared haplotype from the UK Biobank to compare whether the haplotypes match across the 51 variants shared across genotyping chips. Two megabases of exome sequence around rs45553935 in a

carrier from the Healthy Nevada Project (24) were merged with the corresponding region in a carrier from the ORCADES study. 1974 variants overlapped between the two exomes. A similar approach was then taken to assess potential haplotype sharing. Beginning with the rs45553935 variant, moving one SNP at a time, we compared the two genotypes, repeating the procedure until we came to opposing homozygotes, beyond which haplotypes cannot be shared. Identity-by-descent (IBD) analysis was performed using KING 2.1.5 (25).

EHR Data Linkage in ORCADES NHS routine datasets linked to ORCADES participants in July 2021, including the Scottish Cancer Registry SMR06, were accessed using a secure process as for the Generation Scotland cohort (26).

Results

Clinical ascertainment of the kindred The rs45553935 (V1736A) variant was ascertained independently in nine diagnostic tests of breast and ovarian cancer patients. Six female obligate carriers were also identified. Fourteen V1736A carriers (eight females and six males) were identified from predictive tests in unaffected relatives of NHS patients. Five of these eight females have undergone risk reducing surgery and none have yet developed cancer. This gives a total of 23 positive NHS results in females.

Population frequencies of the BRCA1 *variant* Data from 2,088 ORCADES participants (819 male and 1,269 female) passed all exome sequence and genotype quality control thresholds. There are twenty heterozygous carriers of the V1736A variant in the ORCADES exome dataset, of whom seven are female. No other *BRCA1* variants reported as pathogenic or likely pathogenic in ClinVar were present in the ORCADES exome dataset, including the Scottish pathogenic founder variant 2800delAA (p.Lys894fs) (27). None of the female carriers in ORCADES are compound heterozygotes for pathogenic or likely pathogenic exonic *BRCA1* alleles, for which there are multiple submitters and no conflicts in ClinVar, neither do they carry known pathogenic exonic variants in the cancer susceptibility genes *APC, BRCA2, RET, PALB2, MAX, TMEM127, BMPR1A, SMAD4, TP53, MLH1, MSH2, MSH6, PMS2, MEN1, MUTYH, NF2, SDHD, SDHAF2, SDHC, SDHB, PTEN, RB1, VHL* or *WT1*.

Commented [MPZH1]: Ooops- amazing what falls out

178 Information on allele frequencies in populations can be obtained from the Genome Aggregation Database
179 (gnomAD) (28). GnomAD v2.1.1, containing 125,748 exomes and 15,708 genomes from unrelated
180 individuals, has no instances of the c.5207T>C; p.Val1736Ala variant, emphasising its rarity. Consistent
181 with this, the variant is not observed at significant frequency in several population research cohorts
182 including the Viking Health Study Shetland (Table 1). The DiscovEHR study (29) browser also indicates no
183 V1736A carriers were recorded in 92,453 individuals in the Pennsylvanian MyCode population cohort (30).
184 In contrast, the first 200,000 exome sequences from the UK Biobank (19, 31) contain four instances (Table
185 1). This corresponds to a UK allele frequency of 0.00001, ~480-fold lower than we found in Orkney. None
186 of the four UK Biobank subjects was born in the Northern Isles; two live in Scotland and two in England.
187 Three of the four UK Biobank research participant V1736A carriers are female. One, aged in her late fifties
188 at assessment, has ovarian cancer ICD-10 codes in the EHR dataset (Table 1). The two other female variant
189 carriers, in their 50s and 60s at assessment, and the single male, have no reported ICD-10 codes relating
190 to hereditary breast–ovarian cancer (HBOC). Small numbers of variant carriers are also reported in two
191 databases of genomic data from cancer cases, CanVar (32) and BCAC, the Breast Cancer Association
192 Consortium (Table 1). However, the V1736A variant was not observed in sufficient numbers of cases and
193 controls to allow for estimation of cancer risks in BCAC (33).

194 *Origin of the V1736A variant* Oral histories and data from the Scottish register of births, marriages and
195 deaths clinical genealogy service traced the clinical cases to two lineages with founders born c. 1800, on
196 the small island of Westray, in the North Isles of Orkney. Of the ORCADES variant carriers, eight out of
197 twenty had four grandparents born in Westray, and all but one of them had at least one Westray
198 grandparent. Of all 80 grandparents of the carriers, 60% were from Westray, with most of the remainder
199 coming from other parishes or isles of Orkney (Figure 1).

200 The pathogenic variant carriers ascertained clinically and in the ORCADES study fit into five kindreds
201 descending from five separate couples. There are four kindreds from the ORCADES, and two from the
202 Clinical Genetics dataset, of which two overlap, giving a total of five kindreds. While most of these
203 kindreds show distant kinship with one another, e.g. being fifth or sixth cousins, this is not uncommon

among individuals with Westray heritage, and so it was not possible to be certain of the path of segregation of the variant to each of the carriers living today. Ancestral non-paternity and adoption events may have also influenced the path of segregation down the pedigree. Some of the connections are likely to date prior to *c.*1750, before which few paper genealogy records are available. Given the population structure of the island, and limited contribution of ancestors to descendants, it is likely that the kindreds which cannot be linked together genealogically do in fact connect in the preceding few generations at some point before *c.*1700, as demonstrated by their shared haplotype (see below). What is clear is that the ancestry goes back over 250 years in the island of Westray (Figure 2).

All V1736A carriers tested share a common haplotype All twenty heterozygous V1736A carriers in ORCADES share the same haplotype at the variant locus, with a minimum length of ~2 Mb (Figure 3). Access was given to exome sequence data surrounding the same pathogenic variant in a breast cancer patient (Table 1) described by Grzymski *et al* (24), for comparative haplotype analysis. Analysis of exome data from this carrier participant in the Healthy Nevada Project (24) revealed that there were no opposing homozygote genotypes versus an ORCADES carrier across 676 SNPs, totalling 407 kb, consistent with them sharing one haplotype identical-by-descent in this region.

Analysis of haplotypes in the genotype data from the four UK Biobank participants carrying the variant showed that they all shared a ~1.1 Mb long haplotype, which was identical to the Westray haplotype from ORCADES. DNA was not available for haplotype analysis from the family described in Domchek *et al* (17).

Penetrance of V1736A in Orcadians In addition to the exome sequence information, and the detailed study data collected in the recruitment clinics, linkage to routine NHS data in the electronic health record (EHR) provides a longitudinal component to many research cohorts, including ORCADES and the UK Biobank. The morbidity (hospital admissions, SMR01) and cancer registry (SMR06) datasets are particularly useful for research on people living in Scotland. These data have been obtained for almost all participants in the ORCADES cohort.

The mean age of the seven female carriers at time of recruitment to ORCADES (baseline) was 54, and six gave permission for EHR linkage. Two V1736A carrier participants died over 80 years of age, one of whom

had ovarian cancer recorded as cause of death. None of the remaining carriers had a diagnosis of breast or ovarian cancer recorded, and four with EHR linkage survive (Table 2). Nine female obligate carriers linking two branches of the ORCADES pedigrees together were also ascertained (Table 2). Research study family history questionnaires reported that two died of breast cancer, three died of other causes and three remained cancer-free (for one there was no information). Scottish death registration certificates reveal that the great-grandmother of four of the ORCADES carriers died of breast cancer in the mid-1930s, and their close relative who died of breast cancer in her late fifties.

Together with the clinically-ascertained cases, we have thus identified a total of 37 women of Orcadian heritage with the variant, only two of whom overlap between the clinically-ascertained cases and ORCADES/obligate carriers. Importantly, comparison of common ancestors between the clinically ascertained and population-based pedigrees demonstrated that it is likely that only two out of the seven female variant carriers in the ORCADES population cohort have already been offered genetic testing as part of the cascade testing of the index family.

Discussion

Our combined approach of a family-based case study and systematic analysis within a population cohort has identified the pathogenic variant p.Val1736Ala *BRCA1* in 1% of Orcadians. All carriers share a novel rare long haplotype background. The variant is likely to have arisen in a founder individual from Westray, Orkney, at least 250 years ago. Individuals not ascertained by cascade testing of relatives from the clinical pedigree have been identified by the cohort study. In a previous work on the rare *KCNH2* variant, p.Gly584Ser (23), which causes long QT syndrome, a form of familial cardiac arrythmia, we similarly identified carriers in our Northern Isles research populations who could not have been ascertained from oral history based cascade testing. Indeed, for both actionable variants, up to seventh degree affected relatives were ascertained. This highlights the value of research cohorts in describing the burden of rare clinical variants, which may in turn help planning genetic services.

Although it possible that individuals with a family history of disease might be more likely to participate in genetic research studies, in ORCADES bias with respect to the breast or ovarian cancer phenotypes is mitigated by the large size of the cohort as a proportion of the Orcadian population. Furthermore, the recruitment information referred to "common diseases such as heart disease, eye disease, stroke and diabetes" and not cancers specifically.

Pathogenic variants in actionable genes like *BRCA1* are often considered to be more penetrant in the clinical context of a family history of the relevant condition than in population-based cohorts, due to co-inheritance of multiple lower penetrance modifiers, and ascertainment bias contributes to risk over-estimation (34). However, the penetrance of the predominant Ashkenazi pathogenic sequence variants is demonstrated as largely related to the variants themselves, with minor contribution of the specific family history (35). In the clinical genetics setting, *BRCA1* penetrance to 80 years of 79.5% (95%CI 75.5–83.5%) for breast cancer and 65% (95%CI 75–84%) for ovarian cancer are reported (36), whereas population cohorts indicate lower risks. For example, reported penetrance of pathogenic/loss-of-function variants in *BRCA1* in population cohorts is heterogeneous (mean 38%, range 0%-100%) (34) and influenced by family history (37). Despite the large size of the kindred we report here, power is limited to precisely estimate the penetrance of V1736A. The available data on number of cases and age of onset suggest more modest breast cancer penetrance than is typically seen in genetic clinic families with *BRCA1*, which fits with a missense variant with some residual function. However, the penetrance data we present is similar to that of many other *BRCA1* pathogenic variants (2, 36).

Breast cancer risk in women from breast-ovarian cancer families born before 1940 is considerably less than in those born after (36). This is likely not only due to reduced longevity, but also to lower body mass, higher parity, prolonged breast feeding and dietary factors in earlier generations. This observation fits with our data (available on request) that indicate higher ovarian than breast cancer risk. Counselling in the family has highlighted this familial context. Most women with positive predictive tests in the family have chosen breast MRI screening and risk-reducing bilateral salpingo-oopherectomy by age 40. Uptake of risk reducing mastectomy has been limited, but in line with wider local experience. To date, none of those undergoing predictive testing for the variant have developed breast or ovarian cancer. VIKING II, which is

recruiting those of Northern Isles ancestry regardless of domicile (38), has highlighted scientifically for the first time the extent of Orcadian emigration, across Canada, New Zealand, and Australia but also in England and mainland Scotland.

High penetrance *BRCA1* and *BRCA2* founder pathogenic variants are described in several populations such as Iceland, the Ashkenazim, Poland, Norway and others, and testing for these is well established (5, 6, 8, 9). For example, in the French-Canadian founder population, twenty variants in *BRCA1*, *BRCA2*, and *PALB2* that predispose families to breast and ovarian cancer have been identified at increased frequencies. A recent paper demonstrated that genetic screening in that population could identify up to 10% of those who currently present with early-onset breast and ovarian cancer, prior to a diagnosis (39). However, the challenges of likely reduced penetrance in those without a known family history of cancer, and cost, have limited adoption of asymptomatic BRCA screening outwith selected founder populations in resource-limited healthcare systems. The carrier frequency of 1% that we observe for the c.5207T>C; p.Val1736Ala *BRCA1* variant in the Orkney population is higher than some of the founder variants reported in these populations, and cost effectiveness of population-based screening for *BRCA1* founder pathogenic variant at 1% frequency has been reported in Sephardi Jewish women (10). Recently, NHS England announced its first programme of targeted founder BRCA pathogenic variant screening for people with at least one Jewish grandparent (NHS to launch expanded BRCA genetic testing for Jewish community - The Jewish Chronicle (thejc.com)). In support of this approach, an economic evaluation of population-based *BRCA1*/*BRCA2* pathogenic variant testing across multiple countries and health systems has recently been published (40).

Although to date, the consent framework of most research cohorts does not allow the return of results about carrier status of actionable variants (19), participants in clinical and biobanking studies often wish to receive their results, particularly about “actionable” findings. This has recently stimulated publication an international policy for returning genomic research results (41). Others also recommend the return of results following detection of hereditary breast and ovarian cancer risk to adult population-based biobank participants (41-43). Our ongoing recruitment to a new Northern Isles cohort study, VIKING II, offers new

and existing cohort members the option of consent to return of selected clinically actionable results, and the return of this variant will be prioritised in that process. Relevant participants resident in Scotland will be offered clinically accredited verification on a new sample by the NHS clinical genetics service.

We propose that all women of Orcadian ancestry (worldwide) with a diagnosis of breast cancer should be offered a targeted test for this variant, if a *BRCA1/BRCA2* gene screen is not offered as part of their clinical care. This targeted test for Orcadians with a family history of breast or ovarian cancer is now routine practice in the NHS Grampian clinic, but we know this approach will miss many of those at risk.

Slightly over 11,000 females live in Orkney, of whom ~9,300 are adult, and ~70% of residents have two or more Orcadian ancestors. To date, we have identified less than half of the resident Orcadian V1736A carriers. We are therefore preparing a business case for population-based screening for the variant through primary care community hubs in Orkney, using the inexpensive Sanger sequencing assay that is established in the NHS Grampian genomics laboratory. We propose to pilot this program by offering a test to Westray residents of known Westray ancestry.

High penetrance genes contribute only a proportion of genetic cancer risk, and V1736A is only one of many contributors to breast and ovarian cancer risk in Orcadians. Common low penetrance variants identified through genome-wide association studies explain a further component. Polygenic risk scores (PRS) are being considered for enhancement of risk stratification, both in the general population and in BRCA1/2 carrier populations (44). Genetic drift of common low penetrance variants may limit the portability of scores developed elsewhere. Work is ongoing to assess the utility of PRS in Orcadians, and to determine if low penetrance breast cancer-associated SNPs are enhanced or reduced in this population. We are also examining in detail the clinical utility of testing for other Northern Isles drifted pathogenic variants identified through clinical practice and the Viking Genes studies.

In conclusion, we propose that women with two or more Orcadian grandparents should be offered testing for the V1736A variant, regardless of family history of breast or ovarian cancer. The analyses presented here of the *BRCA1* variant are relevant beyond the modern population of Orkney, both as an exemplar and due to emigration to elsewhere in the British Isles and the New World. Future research will explore

further genetically drifted loci observed as part of clinical care in Orkney and Shetland in the Viking Genes research cohorts.

Data Availability Statement

Some information (e.g. age and nature of a diagnosis) could potentially make individuals identifiable, so has not been shown, or is presented in aggregate form. These data can be made available to legitimate researchers affiliated to an academic organisation through application to the corresponding author. There is neither Research Ethics Committee approval, nor consent from ORCADES participants, to permit open release of the individual level research data underlying this study. The datasets generated and analysed during the current study are therefore not publicly available. Instead, the research data and/or DNA samples are available from accessQTL@ed.ac.uk on reasonable request, following approval by the Data Access Committee and in line with the consent given by participants.

Acknowledgements

The study team wish to thank staff from the NHS Grampian genetics team and the ORCADES Study for their contribution to these datasets, in particular, Barbara Gibbons for genetic counselling of family members, the NHS Grampian genomics laboratory team for finding and testing for the variant in the clinically ascertained cases, and Laura Taylor of NHS Grampian and the Public Health Scotland genealogy team for assembling the clinical pedigree. ORCADES DNA extractions were performed at the Edinburgh Clinical Research Facility, University of Edinburgh. ORCADES Sanger sequencing was performed by Camilla Drake and the technical services team at the MRC HGU. Emily Weiss and Reka Nagy assembled the ORCADES pedigree using records at the General Register Office and study information, building on earlier pedigree work by Ruth McQuillan and Jim Wilson (45). Regeneron Genetics Center performed the exome sequencing. We thank Thibaud Boutin for phasing the GSA chip data and Kiera Johnston for help with analysis of other cancer susceptibility genes. The data in the EHR was provided by patients and collected by the NHS as part of their care and support. The authors acknowledge the support of the eDRIS Team (Public Health Scotland) for their involvement in obtaining approvals, provisioning and linking this data.

We would also like to acknowledge the invaluable contributions of the research nurses in Orkney and the administrative team in Edinburgh. Finally and most importantly, we thank the people of Orkney for their involvement in and ongoing support for our research.

Author Contribution Statement

SK managed the project and drafted the manuscript. EC analysed the clinical data. LK analysed the exome datasets and did the haplotype comparisons. CB and DO'S recognised and interpreted the variant and provided clinical expertise. DB managed and analysed the EHR data. JJG contributed data. CVvH, GT and ARS conceived and managed the ORCADES exome sequencing. JFW is the Chief Investigator of ORCADES, was awarded funding to implement the work, did genealogical analyses and interpreted the data. ZM recognised the family, initiated the work, led the clinical team, interpreted the data and proposed policy. All authors provided input and feedback on drafts of the manuscript.

Ethical Approval

It is clear that information robustly linking genetic variants (e.g. *BRCA1* V1736A) with specific conditions (e.g. breast and ovarian cancer) is fundamental biological knowledge, not personal information, and therefore should not require specific consent for clinicians to share (46). In contrast, neither identifiable medical details about the patients, nor their personal identifiers, were shared by the clinical team with the research team. Eligible participants were recruited to ORCADES, Research Ethics Committee references 26-11-2003 and 12/SS/0151. Research participants gave written informed consent for research procedures including electronic health record linkage. The data linkage and access to NHS Scotland-originated data for the ORCADES cohort was approved by the Public Benefit and Privacy Panel for Health and Social Care (Ref 1718-0380). This research has also been conducted using data from UK Biobank, as part of project ID number 19655.

Funding

This work was funded by the MRC University Unit award to the MRC Human Genetics Unit, University of Edinburgh, MC_UU_00007/10. LK was supported by an RCUK Innovation Fellowship from the National Productivity Investment Fund (MR/R026408/1). ORCADES was supported by the Chief Scientist Office of the Scottish Government (CZB/4/276 and CZB/4/710), a Royal Society URF to JFW and Arthritis Research UK.

Competing Interests

AS, CVVH and GT are former employees and / or stockholders of Regeneron Genetics Center or Regeneron Pharmaceuticals.

References

1. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet. 2003;72(5):1117-30.

2. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA. 2017;317(23):2402-16.

3. Cline MS, Liao RG, Parsons MT, Paten B, Alquaddoomi F, Antoniou A, et al. BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. PLoS Genet. 2018;14(12):e1007752.

4. Robson M, Im SA, Senkus E, Xu B, Domchek SM, Masuda N, et al. Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. N Engl J Med. 2017;377(6):523-33.

5. Rubinstein WS. Hereditary breast cancer in Jews. Fam Cancer. 2004;3(3-4):249-57.

6. Thorlacius S, Sigurdsson S, Bjarnadottir H, Olafsdottir G, Jonasson JG, Tryggvadottir L, et al. Study of a single BRCA2 mutation with high carrier frequency in a small population. Am J Hum Genet. 1997;60(5):1079-84.

7. Simard J, Dumont M, Moisan A-M, Gaborieau V, Vézina H, Durocher F, et al. Evaluation of BRCA1 and BRCA2 mutation prevalence, risk prediction models and a multistep testing approach in French-Canadian families with high risk of breast and ovarian cancer. Journal of Medical Genetics. 2007;44(2):107-21.

8. Crowgey EL, Washburn MC, Kolb EA, Puffenberger EG. Development of a Novel Next-Generation Sequencing Assay for Carrier Screening in Old Order Amish and Mennonite Populations of Pennsylvania. J Mol Diagn. 2019.

9. Shi L, Webb BD, Birch AH, Elkhoury L, McCarthy J, Cai X, et al. Comprehensive population screening in the Ashkenazi Jewish population for recurrent disease-causing variants. Clin Genet. 2017;91(4):599-604.

10. Patel S, Legood R, Evans DG, Turnbull C, Antoniou AC, Menon U, et al. Cost effectiveness of population based BRCA1 founder mutation testing in Sephardi Jewish women. American journal of obstetrics and gynecology. 2018;218(4):431.e1-.e12.

11. Gilbert E, O'Reilly S, Merrigan M, McGettigan D, Vitart V, Joshi PK, et al. The genetic landscape of Scotland and the Isles. Proceedings of the National Academy of Sciences. 2019:201904761.

12. Xue Y, Mezzavilla M, Haber M, McCarthy S, Chen Y, Narasimhan V, et al. Enrichment of low-frequency functional variants revealed by whole-genome sequencing of multiple isolated European populations. Nat Commun. 2017;8:15927.

13. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.

14. Carvalho M, Pino MA, Karchin R, Beddor J, Godinho-Netto M, Mesquita RD, et al. Analysis of a set of missense, frameshift, and in-frame deletion variants of BRCA1. Mutation research. 2009;660(1-2):1-11.

15. Lee MS, Green R, Marsillac SM, Coquelle N, Williams RS, Yeung T, et al. Comprehensive analysis of missense variations in the BRCT domain of BRCA1 by structural and functional assays. Cancer research. 2010;70(12):4880-90.

16. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, et al. Accurate classification of BRCA1 variants with saturation genome editing. Nature. 2018;562(7726):217-22.

17. Domchek SM, Tang J, Stopfer J, Lilli DR, Hamel N, Tischkowitz M, et al. Biallelic deleterious BRCA1 mutations in a woman with early-onset ovarian cancer. Cancer Discov. 2013;3(4):399-405.

18. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D7.

19. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK, et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. Nature. 2020;586(7831):749-56.
20. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. Nature. 2021;599(7886):628-34.
21. Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. Haplotype estimation using sequencing reads. Am J Hum Genet. 2013;93(4):687-96.
22. O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, et al. A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet. 2014;10(4):e1004234.
23. Kerr SM, Klaric L, Halachev M, Hayward C, Boutin TS, Meynert AM, et al. An actionable KCNH2 Long QT Syndrome variant detected by sequence and haplotype analysis in a population research cohort. Sci Rep. 2019;9(1):10964.
24. Grzymski JJ, Elhanan G, Morales Rosado JA, Smith E, Schlauch KA, Read R, et al. Population genetic screening efficiently identifies carriers of autosomal dominant diseases. Nat Med. 2020;26(8):1235-9.
25. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26(22):2867-73.
26. Kerr SM, Campbell A, Marten J, Vitart V, McIntosh AM, Porteous DJ, et al. Electronic health record and genome-wide genetic data in Generation Scotland participants. Wellcome Open Res. 2017;2:85.
27. Liede A, Cohen B, Black DM, Davidson RH, Renwick A, Hoodfar E, et al. Evidence of a founder BRCA1 mutation in Scotland. Br J Cancer. 2000;82(3):705-11.
28. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.
29. Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science. 2016;354(6319).
30. Schiabor Barrett KM, Masnick M, Hatchell KE, Savatt JM, Banet N, Buchanan A, et al. Clinical validation of genomic functional screen data: Analysis of observed BRCA1 variants in an unselected population cohort. HGG Adv. 2022;3(2):100086.

31. Jurgens SJ, Choi SH, Morrill VN, Chaffin M, Pirruccello JP, Halford JL, et al. Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. Nat Genet. 2022;54(3):240-50.

32. Chubb D, Broderick P, Dobbins SE, Houlston RS. CanVar: A resource for sharing germline variation in cancer patients. F1000Res. 2016;5:2813.

33. Breast Cancer Association C, Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Luccarini C, et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. N Engl J Med. 2021;384(5):428-39.

34. Forrest IS, Chaudhary K, Vy HMT, Petrazzini BO, Bafna S, Jordan DM, et al. Population-Based Penetrance of Deleterious Clinical Variants. JAMA. 2022;327(4):350-9.

35. Manchanda R, Lieberman S, Gaba F, Lahad A, Levy-Lahad E. Population Screening for Inherited Predisposition to Breast and Ovarian Cancer. Annu Rev Genomics Hum Genet. 2020;21:373-412.

36. Evans DG, Shenton A, Woodward E, Lalloo F, Howell A, Maher ER. Penetrance estimates for BRCA1 and BRCA2 based on genetic testing in a Clinical Cancer Genetics service setting: risks of breast/ovarian cancer quoted should reflect the cancer burden in the family. BMC Cancer. 2008;8:155.

37. Jackson L, Weedon M, Harrison J, Wood A, Ruth K, Tyrrell J, et al. Influence of family history on penetrance of hereditary cancers in a population setting. medRxiv. 2022:2022.07.08.22277415.

38. Kerr SM, Edwards R, Buchanan D, Dean J, Miedzybrodzka Z, Wilson JF. VIKING II, a Worldwide Observational Cohort of Volunteers with Northern Isles Ancestry. medRxiv. 2021:2021.10.15.21265045.

39. Behl S, Hamel N, de Ladurantaye M, Lepage S, Lapointe R, Mes-Masson A-M, et al. Founder BRCA1/BRCA2/PALB2 pathogenic variants in French-Canadian breast cancer cases and controls. Scientific Reports. 2020;10(1):6491.

40. Manchanda R, Sun L, Patel S, Evans O, Wilschut J, De Freitas Lopes AC, et al. Economic Evaluation of Population-Based BRCA1/BRCA2 Mutation Testing across Multiple Countries and Health Systems. Cancers. 2020;12(7):1929.

41. Lewis ACF, Knoppers BM, Green RC. An international policy on returning genomic research results. Genome Med. 2021;13(1):115.

493 42. Leitsalu L, Palover M, Sikka TT, Reigo A, Kals M, Pärn K, et al. Genotype-first approach to the
494 detection of hereditary breast and ovarian cancer risk, and effects of risk disclosure to biobank
495 participants. European Journal of Human Genetics. 2021;29(3):471-81.
496 43. Manickam K, Buchanan AH, Schwartz MLB, Hallquist MLG, Williams JL, Rahm AK, et al. Exome
497 Sequencing–Based Screening for BRCA1/2 Expected Pathogenic Variants Among Adult Biobank
498 Participants. JAMA Network Open. 2018;1(5):e182140-e.
499 44. Mars N, Widen E, Kerminen S, Meretoja T, Pirinen M, Della Briotta Parolo P, et al. The role of
500 polygenic risk and susceptibility genes in breast cancer over the course of life. Nat Commun.
501 2020;11(1):6383.
502 45. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of
503 homozygosity in European populations. Am J Hum Genet. 2008;83(3):359-72.
504 46. Wright CF, Ware JS, Lucassen AM, Hall A, Middleton A, Rahman N, et al. Genomic variant sharing:
505 a position statement. Wellcome Open Res. 2019;4:22.
506

Figure Legends

Figure 1. Grandparental ancestry of carriers in ORCADES. The first eight columns are parishes or isles of Orkney. The remainder are locations elsewhere in Scotland, or unknown. Of all 80 grandparents of the carriers, 60% were from Westray, with the majority of the remainder coming from other parishes or isles of Orkney

Figure 2. Outline pedigree of two kindreds (A and B) from the ORCADES study. Filled circles are breast or ovarian cancers, red outlines are sequenced V1736A carriers, dotted red outlines are obligate carriers. The founders of kindred A, the largest, were born in Westray in the 1760s. All four of the other kindreds also eventually lead back to Westray common ancestors, in the 19th century (but with deeper ancestry there back to the same time depth). In kindred C, mostly resident in the East Mainland of Orkney, the Westray common ancestors were born in the early 1800s.

Figure 3. Haplotype sharing. (a) Genome-wide identity-by-descent sharing between two ORCADES carriers from different kindreds. In addition to the shared *BRCA1* haplotype on chromosome 17, background sharing due to Westray ancestry can be seen across the genome. (b) Haplotype sharing across chromosome 17 for all pairwise combinations of representatives of each of the four kindreds in ORCADES. Mb, megabase; IBD, identity-by-descent; * denotes the shortest shared haplotype