TITLE

Very rare pathogenic genetic variants detected by SNP-chips are usually false positives: implications for direct-to-consumer genetic testing

AUTHORS

Weedon MN[1]*, Jackson L[1]*, Harrison JW[1], Ruth KS[1], Tyrrell J[1], Hattersley AT[1], Wright CF[1$]

[1] Institute of Biomedical and Clinical Science, University of Exeter College of Medicine and Health, RILD Building, Royal Devon & Exeter Hospital, Barrack Road, Exeter UK, EX2 5DW

* These authors contributed equally to this work
[$] Correspondence: caroline.wright@exeter.ac.uk

ABSTRACT

Objective: To determine the diagnostic accuracy of SNP-chips frequently used by direct-to-consumer genetic testing companies for genotyping rare genetic variants.

Methods: We assessed the diagnostic accuracy of genotypes from SNP-chips (index test) with next generation sequencing data (reference test) in 49,908 individuals recruited to UK Biobank. We compared the genotyping accuracy of SNP-chip variants covered by the next generation sequencing data by allele frequency. We further used the ClinVar database to select rare pathogenic variants in the *BRCA1* and *BRCA2* genes as an exemplar for detailed analysis. Cancer registry data was gathered for BRCA-related cancers (breast, ovarian, prostate and pancreatic) across all participants.

Results: SNP-chip genotype accuracy is high overall, but the likelihood of a true positive result reduces substantially with decreasing allele frequency. The sensitivity, specificity, positive predictive (PPV) and negative predictive value (NPV) for heterozygous genotypes are all >99% for 108,574 single nucleotide variants directly genotyped by the UK Biobank SNP-chips. However, for variants with a frequency <0.001% in UK Biobank the PPV is very low, and only 16% of 4,711 heterozygote genotypes from the SNP-chip confirm with sequencing data. For pathogenic variants in the *BRCA1* and *BRCA2* genes, the overall performance metrics of the SNP-chips in UK Biobank are: sensitivity 34.6%, specificity 98.3%, PPV 4.2% and NPV 99.9%. Rates of BRCA-related cancers in individuals with a positive SNP-chip result are similar to age-matched controls (OR 1.28, P=0.07, 95% CI: 0.98-1.67), while sequence-positive individuals have a significantly increased risk (OR 3.73, $P=3.5\times10^{-12}$, 95% CI: 2.57-5.40).

Discussion: SNP-chips are extremely unreliable for genotyping very rare pathogenic variants and should not be used to guide health decisions without validation.

SUMMARY BOX

Section 1: What is already known on this topic

SNP-chips are an accurate and affordable method for genotyping common genetic variants across the genome. They are often used by direct-to-consumer genetic testing companies and research studies, but there several reports suggesting they perform poorly for genotyping rare genetic variants.

Section 2: What this study adds

Our study confirms that SNP-chips are highly inaccurate for genotyping rare and clinically-actionable variants. Using large-scale SNP-chip and sequencing data from UK Biobank, we show that the SNP-

chip false discovery rate is >84% for very rare variants (<0.001% frequency in UKB). Rare variants assayed using SNP-chips should not be reported back to patients without validation.

INTRODUCTION

Single nucleotide polymorphism genotyping arrays ("SNP-chips") test genetic variation at many hundreds-of-thousands of DNA locations [1]. They have proven to be excellent for studying the role of common genetic variation in many traits and diseases [2,3]. SNP-chips are widely used by both biobanks and direct-to-consumer (DTC) testing companies [4] as a cheap and reliable way to detect common genetic variation. However, information on common genetic predisposition is rarely used in clinical practice as most clinically relevant mutations are rare. In recent years, many SNP-chip designs have been augmented to include rare genetic variants that cause heritable genetic diseases and have direct clinical implications. As a consequence, consumers of DTC tests are increasingly receiving medically-actionable results, which they often take to healthcare professionals for advice [5]. These results can have implications for both prognosis and treatment: for example, a confirmed pathogenic variant in one of the breast-cancer genes, *BRCA1* or *BRCA2,* would need urgent follow-up with additional screening and potentially prophylactic surgical mastectomy and oophorectomy [6]. It is crucial to know how accurate these results are likely to be in order to interpret rare variants detected using SNP-chips.

There is no published systematic evaluation of the accuracy of SNP-chips for assaying rare genetic variants against the gold standard of DNA sequencing. False positive results of rare clinically-actionable mutations from DTC SNP-chips have been described in several case reports [7,8] and a small case series [9]. We previously showed by visual inspection of data from ~500,000 individuals in the UK Biobank (UKB) [10] that SNP-chips do not perform well for detecting rare genetic variants [11] due to inherent difficulties with clustering small numbers of data points (Figure 1). Next generation sequencing (NGS) is based on a completely different technology, and is used as the diagnostic test for many heritable genetic conditions [12,13] because it detects very rare variation extremely accurately. The recent release of NGS data on 49,960 individuals from UKB [14] – most of whom were previously genotyped with SNP-chips enriched for rare genetic variants – provides an ideal opportunity to empirically test the accuracy of SNP-chip genotyping against NGS.

Here, we use NGS data from 49,908 UKB participants as a gold standard to evaluate how well SNP-chips detect rare genetic variants. We use rare pathogenic variants in the *BRCA1* and *BRCA2* genes (collectively termed henceforth as BRCA) that cause dominant hereditary cancers as an exemplar to evaluate diagnostic test performance. We show that SNP-chips are extremely unreliable for genotyping rare variants.

METHODS

Study design: We performed a retrospective diagnostic accuracy study of SNP-chips (index test) using NGS (reference test) from UKB participants in whom both datasets were available. The SNP-chip data were generated centrally by UKB and the NGS data were generated externally by Regeneron and are returned to the UKB resource as part of an external access application request.

Participants: The UKB is a population-based research cohort of ~500,000 participants recruited in the UK between 2006 and 2010. Approximately 9.2 million individuals aged 40-69 years who lived within 40 km of one of 22 assessment centres were invited and 5.5% participated [10].

Test methods: We studied 49,960 individuals (55% female) from UKB with NGS data (reference test). Exome capture was performed using an approach developed at the Regeneron Genetics Centre and

samples were sequenced using 75-basepair paired-end reads on the Illumina NovaSeq 6000 platform [14]. Alignment, variant calling and quality control of the exome sequencing data are described in [14].

Of these individuals, 49,908 also had QC-passed SNP-chip data (index test). A subset of 4,037 individuals were previously genotyped using the Applied Biosystems UK BiLEVE Axiom Array by Affymetrix (807,411 genetic markers), and 45,871 individuals were previously genotyped using the Applied Biosystems UK Biobank Axiom® Array (825,927 genetic markers) that shares 95% of its marker content with the BiLEVE [10]. Individuals were genotyped in 106 batches of ~5000 samples. We included samples that passed central UKB quality control (QC) on either of the UKB SNP-chips and only included directly genotyped variants that had a genotype missingness rate <5% and Hardy Weinberg $P<1x10^{-6}$. We used the UCSC genome browser liftover tool to convert SNP-chip variant positions that were reported in hg37 to hg38 coordinates for direct comparison with NGS data.

Analyses: For genome-wide comparison with SNP-chip genotypes, we only included directly genotyped single nucleotide variants (SNVs) with genomic positions present in the gVCF files and covered by >15 reads in the NGS data. We used the minor allele frequency (MAF) from all 488,377 SNP-chip genotyped individuals in UKB [10]. For common and rare SNV subsets, we tested the genotyping quality of heterozygous SNVs on the SNP-chips versus NGS calls and calculated average performance metrics per variant.

For detailed gene-specific comparison with SNP-chip genotypes, we included directly genotyped SNVs, insertions and deletions in the *BRCA1* or *BRCA2* genes. Variants were defined as pathogenic if they were either predicted to result in a truncated protein or had previously been classified as likely or definitely pathogenic in the ClinVar database [15]; variants with conflicting reports of pathogenicity that included pathogenic assertions made within the last 5 years were included. NGS data was visually examined using the Integrative Genomics Viewer (IGV) [16] to determine whether the variant was present or not. Cancer registry data for breast, ovarian, prostate and pancreatic cancer was extracted for all participants. Logistic regressions were carried out to assess the relationship between test-positive participants and any BRCA-related cancer.

Results are presented in accordance with STARD guidelines for reporting diagnostic accuracy studies [17] using sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) to evaluate assay performance.

RESULTS

Performance of SNP-chips appears excellent when assessed for all variants

We assessed NGS data from 49,908 individuals in UKB with NGS data, of whom 45,871 were genotyped using the Axiom® chip and 4,037 using the BiLEVE chip. We evaluated the accuracy of the SNP-chips to correctly classify heterozygote genotype calls for 108,574 unique SNVs directly genotyped on the UK Biobank SNP-chips in regions well covered in the NGS data. Overall performance across both chips for all SNVs is very good (Table 1), with $3.1 x10^{8}$ true positives, $4.6x10^{9}$ true negatives, $3.2x10^{6}$ false positives and $2.7x10^{6}$ false negatives (sensitivity 99.2%, specificity 99.9%, PPV 99.0% and NPV 99.9%). Performance of both chips for genotyping common variants with a MAF of >1% is also very good: Axiom® chip (average sensitivity 99.8% and specificity 99.7%) and BiLEVE chip (average sensitivity 99.7% and specificity 99.7%).

False positive results from SNP-chips greatly increases with decreasing allele frequency

The genotyping performance of the SNP-chips is strongly related to the frequency of the variant allele in the population (Table 1). There are 10,891 (Axiom®) and 7,408 (BiLEVE) SNVs on the two

UKB SNP-chips with a frequency <0.001% in the UK Biobank that are well covered in the NGS data. At these ultrarare variants the sensitivity of the SNP-chips to detect heterozygous genotypes is low (29.5% for Axiom® and 4.4% for BiLEVE), although because the variant alleles are very rare the specificity remains high (99.9% for both chips). Most striking is the substantially lower PPV (Table 1) and higher false discovery rate for rare SNVs compared with common variants (Figure 2). For the ultrarare SNVs, only 16% of Axiom SNP-chip heterozygous genotypes (708 true positives in 4,239 individuals at 3,422 SNVs) were confirmed by the NGS data and only 9% (46 true positives in 518 individuals at 488 SNVs) for the BiLEVE chip.

Pathogenic BRCA variants are very poorly genotyped by SNP-chips

We went on to evaluate the performance of the SNP-chips in detail for 1,139 pathogenic and likely pathogenic variants in BRCA that were included on the chips; 916 (80%) of these are rare with an allele frequency of <0.01% in UKB. The performance of both chips is very poor for genotyping pathogenic BRCA variants in UKB (overall sensitivity 34.6%, specificity 98.3%, PPV 4.2% and NPV 99.9%), with a very high false discovery rate (Table 1 and Figure 3). Across both SNP-chips there are 425 pathogenic BRCA variants in 889 individuals in UKB. Of these, just 17 variants in 37 individuals are present in the NGS data; the most common variant is present in 10 individuals and has conflicting and mostly uncertain interpretations in ClinVar. A further 43 pathogenic BRCA variants are present in the NGS data of 70 individuals but are not detected by either SNP-chip despite being assayed. The performance of both chips for genotyping pathogenic BRCA variants is very poor (Table 1 and Supplementary Figures 1 and 2): Axiom® chip (sensitivity 33.0% and specificity 99.7%) and BiLEVE chip (sensitivity 50.0% and specificity 82.7%).

Risk of BRCA-related hereditary cancer in individuals with a pathogenic BRCA variant genotyped by SNP-chips is similar to population risk

The risk of BRCA-related cancers in UKB individuals with a positive SNP-chip result for any pathogenic BRCA variant is similar to the age-matched risk in UKB (OR 1.28, P=0.07, 95% CI: 0.98-1.67). In contrast, those with a positive NGS result – including the 107 individuals with BRCA variants assayed by either chip, plus another 137 individuals with pathogenic BRCA variants not assayed by either chip – have a markedly increased risk (OR 3.73, $P=3.5\times10^{-12}$, 95% CI: 2.57-5.40).

DISCUSSION

The analytical validity of SNP-chips is extremely poor for correctly genotyping rare variants in the general population compared with gold standard sequencing data. Genetic variants that are rare in the population being assayed will tend to be badly genotyped, with a very high false discovery rate, while variants that are common in that population will tend to be well genotyped. This low analytical validity for rare genetic variants is an inherent issue with the clustering method upon which SNP-chip genotyping relies (Figure 1). Clustering of data from individuals with similar genotypes becomes harder as the number of individuals in a batch decreases, and extremely difficult to distinguish from experimental noise where there is only a single heterozygous carrier present. This means that genotyping by SNP-chip is dependent on a sufficient number of alternate alleles for genotyping clustering. Genotyping batch size will therefore also affect accuracy, with fewer individuals leading to more genotyping errors for rare variants. As a result, although the performance of chips from different manufacturers may differ somewhat, our findings are likely to be generalisable to any SNP-chip and indeed the results are similar across both SNP-chips used in UKB. NGS data are not affected by the same technical issues as SNP-chips and thus provide a much more accurate method for genotyping rare variants.

The poor genotyping quality of SNP-chips can be remedied through custom variant calling definitions, using multiple probes for individual variants or using positive laboratory controls to

improve variant clustering. Rare variants that are genotyped using these additional quality control methods will therefore have a substantially better performance. However, most DTC companies (including those focused on ancestry or other non-medical traits) also allow customers to download and analyse their raw data, which will often include many thousands of additional rare variants assayed on the SNP-chip that have not undergone stringent quality control and are therefore much more likely to be false positives [8].

Positive results from SNP-chips for rare pathogenic variants in dominant disease-causes genes are substantially more likely to be wrong than right. For pathogenic BRCA variants in UKB, the SNP-chips have a false discovery rate of 83-99% when compared with NGS. Were these results to be fed-back to individuals, the clinical implications would be profound. Women with a positive BRCA result face a lifetime of additional screening and potentially prophylactic surgery that is unwarranted in the case of a false positive result. Additionally, erroneous results in research biobanks can lead to false associations [18] and wasted resource in the development of new treatments against the wrong targets [19]. Moreover, although the false negative rate of SNP-chips is generally low, many very rare pathogenic variants are not included in the design and will therefore be missed.

Although there is an acknowledged healthy-volunteer bias in UKB [20], which will lead to a likely bias away from ultra-rare highly-penetrant variants in the current NGS dataset, our findings are nonetheless informative for management of incidentally detected pathogenic variants. We urge clinicians to validate any SNP-chip results from DTC companies or research biobanks using a gold standard diagnostic test prior to recommending any clinical action. In addition, individuals with symptoms or a family history of breast and/or ovarian cancer who have received a negative SNP-chip result should not be reassured that their risk is low and standard referral guidelines should be followed for diagnostic testing. We suggest that variants assayed using SNP-chips, where the heterozygous allele is very rare in the population being tested, should not be routinely reported back to individuals or used in research without validation. Clinicians and researchers should be aware of the poor performance of SNP-chips for genotyping rare genetic variants to avoid counselling patients inappropriately or investing limited resources into investigating false associations with badly genotyped variants.

ACKNOWLEDGEMENTS

This research has been conducted using the UK Biobank Resource under Application Number 49847 and 871. ATH is a Wellcome Trust Senior Investigator (098395) and an NIHR senior investigator.

TABLES

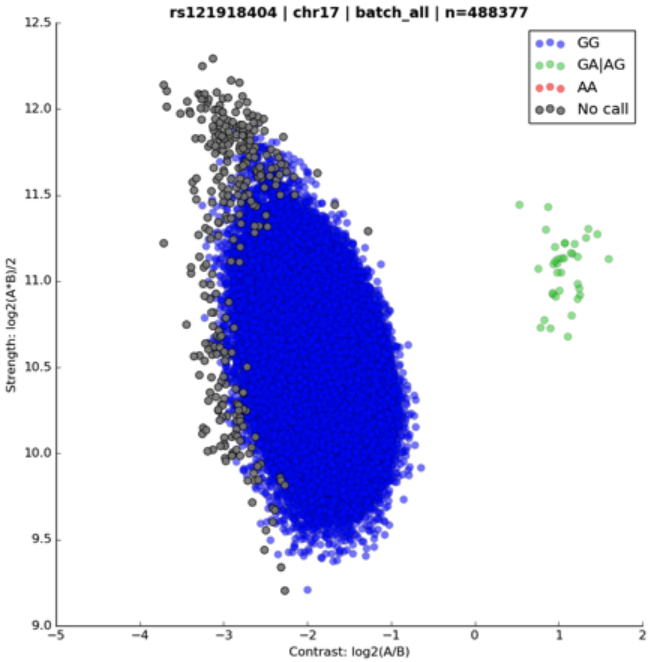
Table 1. Performance of SNP-chips in UKB versus gold standard NGS.
Results are split by SNP-chip (Axiom® or BiLEVE) and variant group: all SNVs, common SNVs (MAF <1%), rare SNVs (MAF <0.001%), and pathogenic BRCA variants. For BRCA variants, performance characteristics are based on participants rather than variants (see STARD flowcharts, Supplementary Figures 1 and 2).

Chip	Individuals	Dataset	Sensitivity	Specificity	PPV	NPV
Axiom®	45,871	All SNVs	99.2%	99.9%	99.0%	99.9%
		UKB MAF >1%	99.8%	99.7%	99.0%	99.9%
		UKB MAF <0.001%	29.5%	99.9%	16.1%	99.8%
		BRCA pathogenic	33.0%	99.7%	16.9%	99.9%
BiLEVE	4,037	All SNVs	99.2%	99.9%	98.9%	99.9%
		UKB MAF >1%	99.7%	99.7%	98.7%	99.9%
		UKB MAF <0.001%	4.4%	99.9%	9.4%	99.7%
		BRCA pathogenic	50.0%	82.7%	0.7%	99.8%

FIGURES

Figure 1. Examples of two rare variants called using the UKB SNP-chip. **(a)** Well genotyped heterozygous variant, rs121918404 (UKB MAF = 0.004%) and **(b)** poorly genotyped heterozygous variant, rs80357065 (UKB MAF = 0.0007%). Automated clustering of dye signal intensities versus strength across multiple samples is used to determine the genotype in each individual. The rarer the variant, the fewer the datapoints, so the harder it is to cluster the data and distinguish real differences from experimental noise. Blue circle = homozygous reference; green circle = heterozygous variant; grey circle = uncertain or missing results. Figures created using Evoker Lite.

(a)



(b)

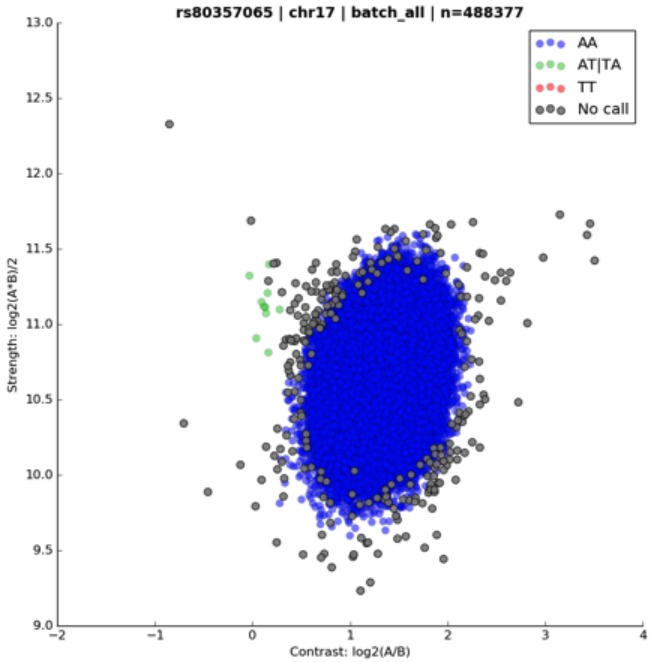
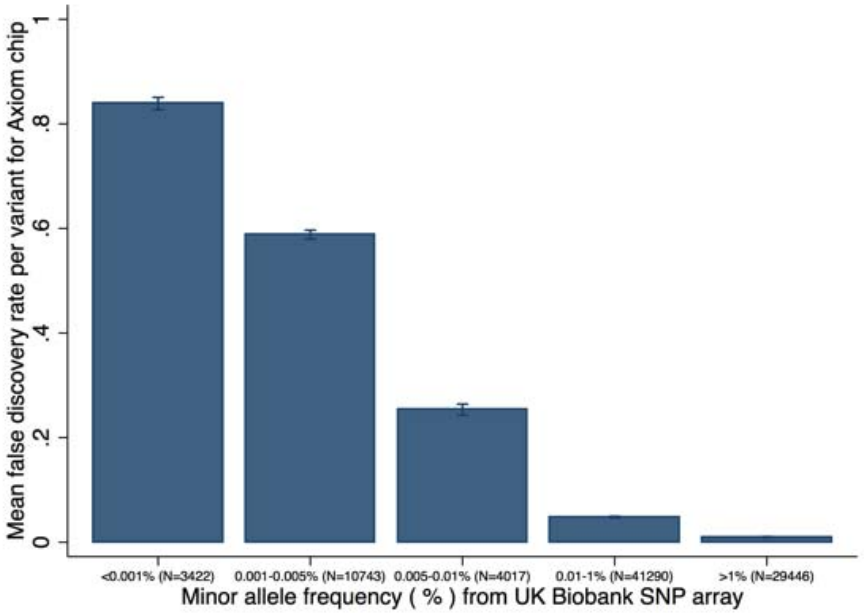


Figure 2. Average false discovery rate of heterozygous genotypes for UKB SNP-chips. SNVs from **(a)** the Axiom® chip and **(b)** the BiLEVE chip (index tests) are compared against NGS (reference test). Variants are grouped by SNP-chip MAF from 488,377 UKB individuals. N is for number of SNVs where there is at least 1 heterozygous genotype from the SNP-chip in the 49,908 individuals with NGS data.

(a)



(b)

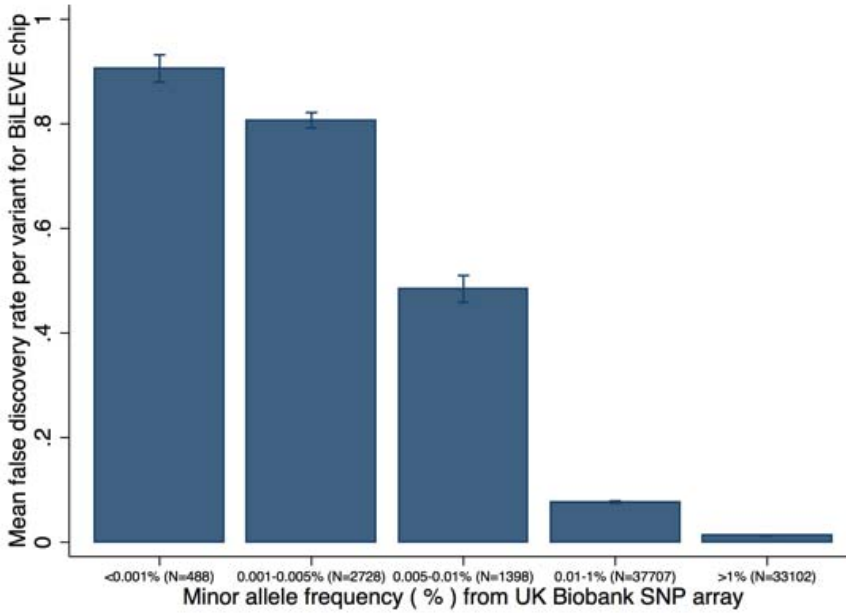
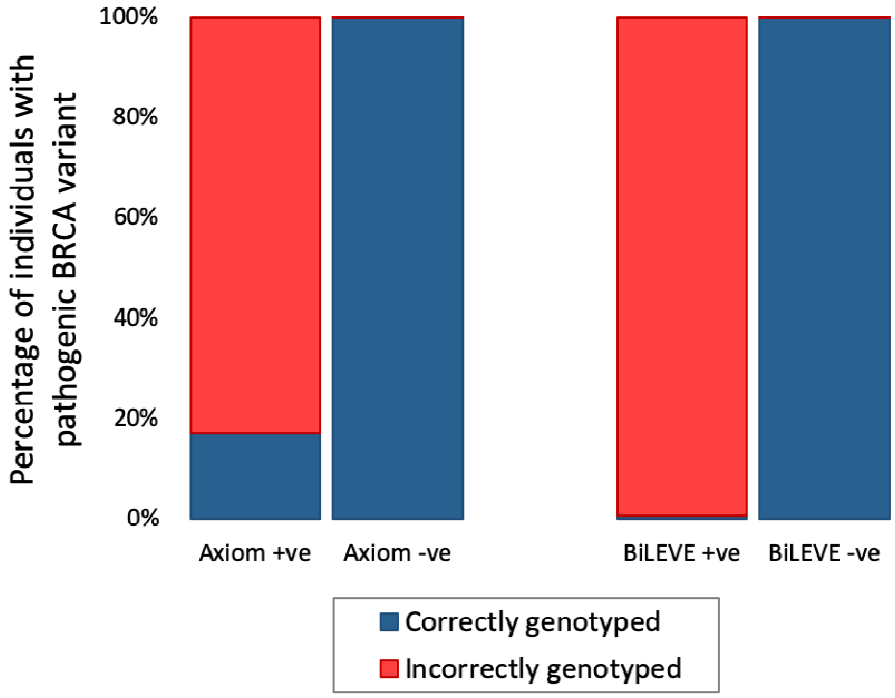
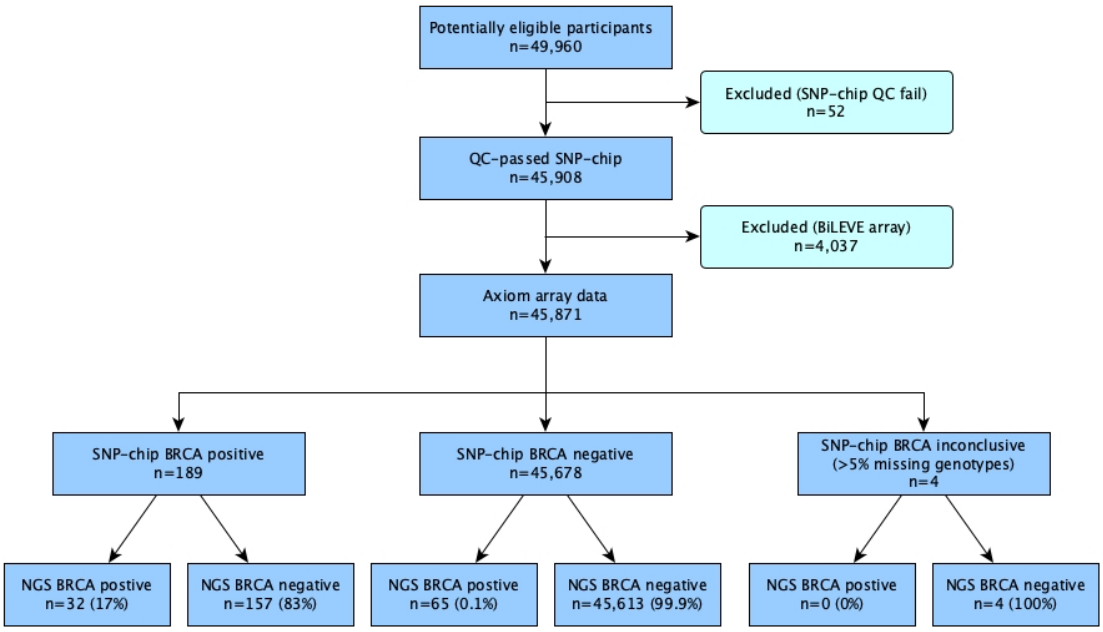


Figure 3. Performance of UKB SNP-chips (Axiom® and BiLEVE chips) versus NGS data for genotyping pathogenic BRCA variants (see Supplementary Figures 1 and 2 for details).

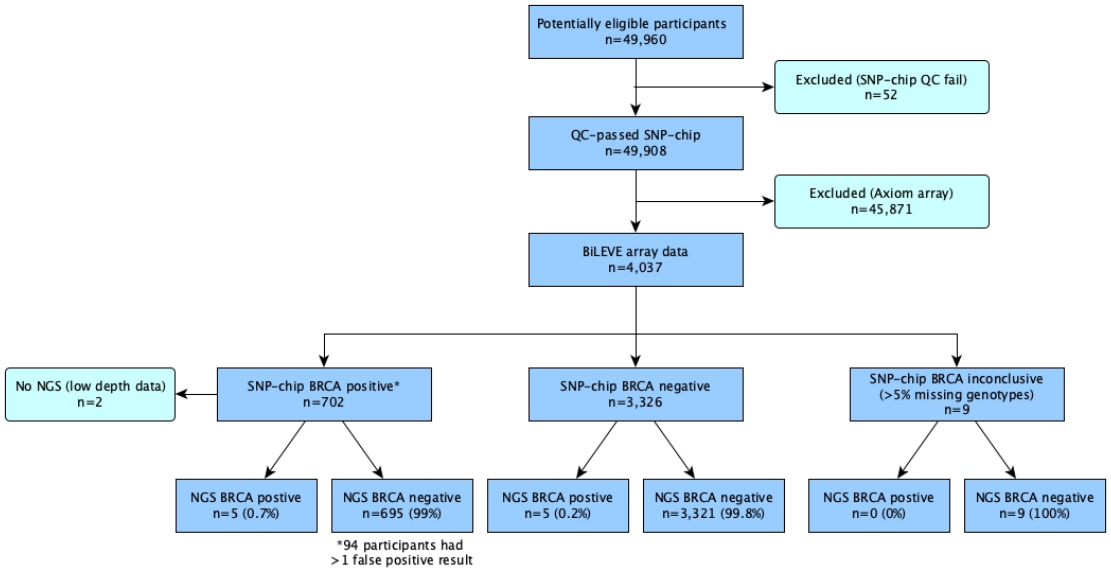


SUPPLEMENTARY FIGURES

Supplementary Figure 1. STARD diagram to report flow of participants with a pathogenic BRCA variant on the Axiom® chip compared with NGS.



Supplementary Figure 2. STARD diagram to report flow of participants with a pathogenic BRCA variant on the BiLEVE chip compared with NGS.



REFERENCES

1 LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res* 2009;**37**:4181–93. doi:10.1093/nar/gkp552
2 Price AL, Spencer CCA, Donnelly P. Progress and promise in understanding the genetic basis of common diseases. *Proc Biol Sci* 2015;**282**:20151684. doi:10.1098/rspb.2015.1684
3 Visscher PM, Goddard ME. From R.A. fisher's 1918 paper to GWAS a century later. *Genetics* 2019;**211**:1125–30. doi:10.1534/genetics.118.301594
4 Gurwitz D, Bregman-Eschet Y. Personal genomics services: whose genomes? *Eur J Hum Genet* 2009;**17**:883–9. doi:10.1038/ejhg.2008.254
5 van der Wouden CH, Carere DA, Maitland-van der Zee AH, *et al.* Consumer Perceptions of Interactions With Primary Care Providers After Direct-to-Consumer Personal Genomic Testing. *Ann Intern Med* 2016;**164**:513–22. doi:10.7326/M15-0995
6 AlHilli MM, Al-Hilli Z. Perioperative Management of Women Undergoing Risk-reducing Surgery for Hereditary Breast and Ovarian Cancer. *J Minim Invasive Gynecol* 2019;**26**:253–65. doi:10.1016/j.jmig.2018.09.767
7 Schleit J, Naylor LV, Hisama FM. First, do no harm: direct-to-consumer genetic testing. *Genet Med* Published Online First: 14 June 2018. doi:10.1038/s41436-018-0071-z
8 Moscarello T, Murray B, Reuter CM, *et al.* Direct-to-consumer raw genetic data and third-party interpretation services: more burden than bargain? *Genet Med* Published Online First: 12 July 2018. doi:10.1038/s41436-018-0097-2
9 Tandy-Connor S, Guiltinan J, Krempely K, *et al.* False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care. *Genet Med* 2018;**20**:1515–21. doi:10.1038/gim.2018.38
10 Bycroft C, Freeman C, Petkova D, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;**562**:203–9. doi:10.1038/s41586-018-0579-z
11 Wright CF, West B, Tuke M, *et al.* Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *Am J Hum Genet* 2019;**104**:275–86. doi:10.1016/j.ajhg.2018.12.015
12 Clark MM, Stark Z, Farnaes L, *et al.* Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med* 2018;**3**:16. doi:10.1038/s41525-018-0053-8
13 Ellard S, Kivuva E, Turnpenny P, *et al.* An exome sequencing strategy to diagnose lethal autosomal recessive disorders. *Eur J Hum Genet* 2015;**23**:401–4. doi:10.1038/ejhg.2014.120
14 Van Hout CV, Tachmazidou I, Backman JD, *et al.* Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *BioRxiv* Published Online First: 9 March 2019. doi:10.1101/572347
15 Landrum MJ, Lee JM, Benson M, *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016;**44**:D862-8. doi:10.1093/nar/gkv1222
16 Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinformatics* 2013;**14**:178–92. doi:10.1093/bib/bbs017
17 Cohen JF, Korevaar DA, Altman DG, *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;**6**:e012799. doi:10.1136/bmjopen-2016-012799
18 van de Putte R, Wijers CHW, Reutter H, *et al.* Exome chip association study excluded the involvement of rare coding variants with large effect sizes in the etiology of anorectal malformations. *PLoS One* 2019;**14**:e0217477. doi:10.1371/journal.pone.0217477
19 Border R, Johnson EC, Evans LM, *et al.* No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *Am J Psychiatry* 2019;**176**:376–87. doi:10.1176/appi.ajp.2018.18070881

20 Fry A, Littlejohns TJ, Sudlow C, *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* 2017;**186**:1026–34. doi:10.1093/aje/kwx246