Applying a PS4 likelihood ratio methodology (LR-Calc) for integrating enriched case data from national laboratory testing (NHS GLH Laboratories, Ambry Genetics) into large-scale case-control analyses

Sophie Allen[1], Charlie F. Rowlands[1], Alice Garrett[1,2], Fergus Couch[3], Marcy Richardson[4], Tina Pesaran[4], Joanna Pethick[5], Katrina Lavelle[5], Fiona McRonald[5], Sally Vernon[5], Miranda Durkie[6], George J. Burghel[7,8], Alison Callaway[9], Rachel Robinson[10], Joanne Field[11], Bethan Frugtniet[2], Shiela Palmer-Smith[12], Jonathan Grant[13], Judith Pagan[14], Trudi McDevitt[15], Katie Snape[2], Helen Hanson[16,17], Terri McVeigh[18], Steven Hardy[2], Clare Turnbull[1,18]

Introduction

- For v3.0 of the ACMG/AMP guidelines, **case-control data is applied under PS4 at a 'strong' level of evidence.**
- However, while quantitative methodologies have been developed for other evidence criteria to empower their use outside of the standard categorical evidence strengths, **no such quantitative methodology has been established for case-control data.**
- Large-scale unselected case-control datasets for breast cancer and **nationally-collected laboratory datasets** enriched for pathogenic variant carriers that have **never been analysed for variant interpretation** are now available
- We have combined datasets across varied ascertainment contexts in a quantitative manner using **novel likelihood ratio tools.**

Dataset	Ascertainment	Cases	Controls	Location	Cancer phenotype	Sex (cases)	Ancestry
UK Biobank	Unselected	18,487	419,717	UK	Breast	Female	White
CARRIERS	Unselected	32,247	32,544	USA	Breast	Female	Non-Hispanic White
BRIDGES	Unselected	42,062	44,035	Europe	Breast	Female	European
Ambry	Clinical	187,642	NA	USA	Breast	Female	White
NDRS/ UK Lab	Clinical	44,917	NA	UK	Breast	Female	White
gnomAD v4.1.0	Population	NA	175,054	Global	NA	NA	Non-Finnish European

Figure 1: Summary of all datasets included in this analysis.

Methods

Using the published PS4-LR-Calc methodology (a flexible methodology for quantifying case-control evidence as a likelihood ratio using prespecified odds ratio thresholds), rules were iterated for combination of unselected and enriched laboratory datasets:

- **Rule 1:** A variant must be seen in at least two datasets, or more than once in one dataset.
- **Rule 2:** The ratio of total cases:controls in a study must not exceed a maximum tolerated threshold.
- **Rule 3:** The observed odds ratio for a variant (and 95% confidence intervals for the observed odds ratio) must not be between the target odds of association and non-association.
- **Rule 4:** An Enrichment Factor must be calculated for Enriched datasets based on total allele frequency of truncating variants in Enriched versus Unselected datasets. This enrichment factor is used to proportionally increase the target odds of association for Enriched datasets to align with

Dataset	BRCA1	BRCA2	PALB2	ATM	CHEK2
NDRS/ UK Lab	4.99	4.07			
Ambry	2.66	2.51	1.93	2.03	1.83

Figure 2: Enrichment Factors calculated for each gene in the enriched laboratory datasets

- **Rule 5:** Enriched laboratory datasets must be paired with equivalent controls or population data.

Results

- Data was combined for **10,820** missense variants from **325,255 cases** female breast cancer patients and **671,350 controls** of Western European ancestry for five breast cancer susceptibility genes (***BRCA1, BRCA2, PALB2, ATM, CHEK2***).
- A combined log likelihood ratio (LLR) was produced for **5,360 missense variants**.
- Under an assumption of standard penetrance for each gene (OR≥4 for *BRCA1* and *BRCA2*, OR≥3 for *PALB2*, and OR≥2 for *ATM* and *CHEK2*):
  - **934** variants received evidence towards pathogenicity (LLR≥1)
  - **3,791** received evidence towards benignity (LLR≤1).

	Rare missense variants included in final dataset			Rare missense variants excluded from final dataset		Total rare missense variants
	Evidence towards pathogenicity (LLR≥1)	No evidence applied	Evidence towards benignity (LLR≤-1)	Single observation (cases or controls)	1≥OR≥target odds (suspected reduced penetrance)	
UKB	416	527	498	1406	16	**2863**
CARRIERS	186	180	508	1219	16	**2109**
BRIDGES	278	254	733	2372	15	**3652**
Ambry	661	377	2492	3985	15	**7530**
NDRS	128	32	1000	1030	12	**2202**
COMBINED	**934**	**635**	**3791**	**5444**	**16**	**10820**

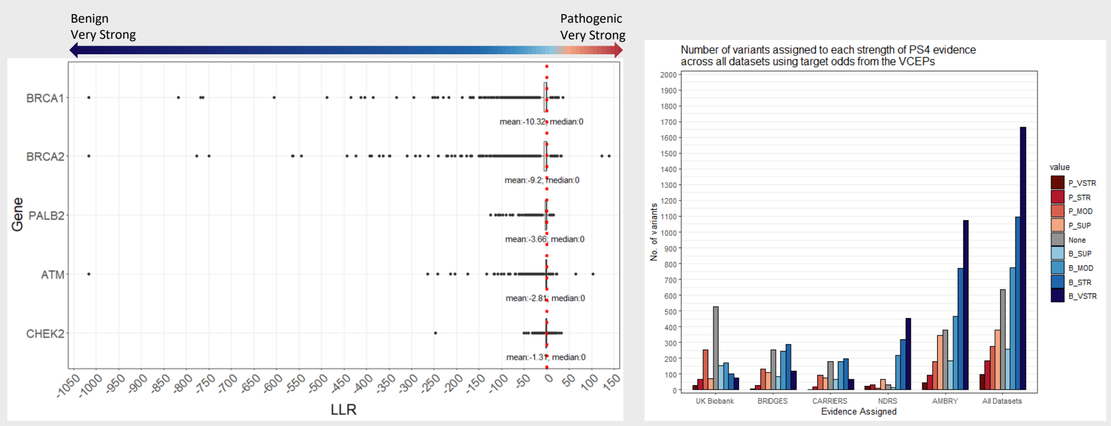

Figure 3: Overview of evidence applied for all 5,360 rare missense variants in the combined dataset. Target odds of association was set at OR≥4 for *BRCA1* and *BRCA2*, OR≥3 for *PALB2*, and OR≥2 for *ATM* and *CHEK2*.
3A: Number of variants attaining a combined LLR and variants excluded from analysis per the defined rules (See Methods).
3B: Per-gene breakdown of LLR (evidence points) in the combined dataset. Each point represents a variant. Mean and Median LLR applied is given for each gene.
3C: Per-dataset breakdown of evidence.

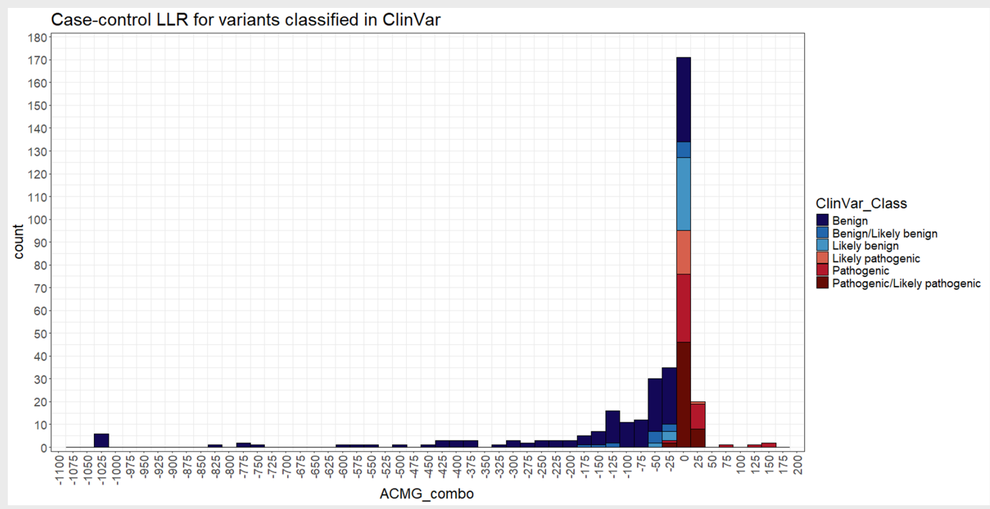

Figure 4: LLR for 350 missense variants classified as either pathogenic, likely pathogenic, likely benign, or benign with at least one star in ClinVar. Sensitivity = 71.3%, Specificity = 91.7%.

Conclusion

- This analysis forms the **largest case-control analysis of female breast cancer** to date, including **novel national datasets** never analysed for variant interpretation before.
- The next steps will be to analyse this large dataset to identify potential variants of **reduced penetrance in high penetrance genes** (*BRCA1, BRCA2, PALB2*), and **high penetrance variants in moderate penetrance genes** (*ATM, CHEK2*)

Contact

Sophie Allen, MSc | PhD Student
Division of Genetics and Epidemiology
The Institute of Cancer Research
15 Cotswold Road, Sutton, London, UK
Sophie.Allen@icr.ac.uk

Acknowledgments

This research has been conducted using data from the UK Biobank Resource, the BCAC BRIDGES Breast Cancer Risk Genes Analysis, the CARRIERS Consortium, Ambry Genetics, and the National Disease Registration Service (NDRS).

Author Affiliations

1 The Institute of Cancer Research, Sutton London, United Kingdom; 2 St George's University Hospitals NHS Foundation Trust, Tooting London, United Kingdom; 3 Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester Minnesota, USA; 4 Ambry Genetics, Aliso Viejo California, USA; 5 NHS England, National Disease Registration Service, United Kingdom; 6 Sheffield Children's NHS Foundation Trust, Sheffield , United Kingdom; 7 Manchester University Hospitals NHS Foundation Trust, Manchester , United Kingdom; 8 The University of Manchester, Manchester, United Kingdom; 9 University Hospital Southampton NHS Foundation Trust, Salisbury, United Kingdom; 10 Leeds Teaching Hospitals NHS Trust, Leeds , United Kingdom; 11 Nottingham University Hospitals NHS Trust, Nottingham , United Kingdom; 12 Cardiff and Vale University Health Board, Cardiff , United Kingdom; 13 NHS Greater Glasgow and Clyde, Glasgow , United Kingdom; 14 Western General Hospital, Edinburgh, United Kingdom; 15 CHI at Crumlin, Dublin, Ireland; 16 Royal Devon University Healthcare NHS Foundation Trust, Exeter, United Kingdom; 17 University of Exeter, Exeter, United Kingdom; 18 The Royal Marsden NHS Foundation Trust, London, United Kingdom