

## PhD Project Proposal

### Funder details

**Studentship funded by:** Marie Skłodowska-Curie Actions (MSCA) Doctoral Networks

### Project details

**Project title:** Early-Onset Breast Cancer Risk: Age-Specific Genetic Susceptibility and Prediction in Diverse Populations

### Supervisory team

**Primary Supervisor:** **Montserrat Garcia-Closas**

**Associate Supervisor(s):** Nilanjan Chatterjee, John Hopkins University, Baltimore, USA  
Laura Blackburn PHG, Cambridge, UK

**Secondary Supervisor:** Clare Turnbull, Division of Genetics & Epidemiology, ICR

### Divisional affiliation

**Primary Division:** **Genetics & Epidemiology**

**Primary Team:** **Integrative Cancer Epidemiology**

**Site:** **Sutton**

### Project background

About 20% of breast cancers in Western countries, and an even higher proportion in other places such as African countries, occur in women under 50, with 5–10% attributed to inherited pathogenic variants in genes such as *BRCA1*, *BRCA2*, *ATM*, and *BARD1*. Younger women are often diagnosed with advanced disease, face higher mortality, and endure long-term treatment effects. Most national screening programs, however, target women aged 50–75 or younger women with a strong family history. Yet, previous studies have shown that incorporating additional risk factors beyond age and family history can identify many women under 50 who reach risk thresholds comparable to those of older women or those with family history. Risk-based strategies could therefore improve early detection, survival, and quality of life in younger women.

Improved risk assessment for younger women requires a deeper understanding of genetic susceptibility, including both rare and common variants and their interaction with other risk factors. Previous work has shown that polygenic risk scores (PRS) can modify the risks conferred by high- and moderate-risk mutations, many of which are linked to early-onset disease. It is also important to consider tumour subtypes and to identify genetic variants associated with aggressive forms, such as triple-negative breast cancer, which occur more frequently in younger women.

Using diverse data from the Confluence Project, this project will optimise PRS development in younger women, assess their joint effects with high- and moderate-risk mutations, and account for tumour

subtypes. This project will also evaluate how genetic associations and heritability vary across age groups to inform risk tools that are applicable across the lifespan.

This project is part of the [Marie Skłodowska-Curie Actions \(MSCA\) Doctoral Networks](#) initiative [HER-CARE](#) (Hereditary & Early Onset Breast Cancer: Comprehensive Personalized Assessment, Early Risk Evaluation, and Clinical Management). HER-CARE aims to advance early and hereditary breast cancer research by identifying unique risk factors and developing innovative tools for early detection and risk stratification. The doctoral candidate (DC) working on this project will join a network of 15 HER-CARE DCs and collaborate with other projects addressing breast cancer risk, stratification, diagnosis, and follow-up. Additional requirements for the MSCA doctoral network programme include:

- starting date will be before July 1, 2026
- eight-month secondment, including five months at John Hopkins University in the USA and three months at PHG Foundation in Cambridge, UK.

## Project aims

- Develop and validate multi-ancestry polygenic risk scores (PRS) for breast cancer across age groups, with a focus on early onset disease, both overall and by major subtypes.
- Evaluate joint associations of PRS and pathogenic variants in high and moderate risk genes across age groups, with a focus on early onset disease.
- Estimate age-specific heritability, enrichment and effect-size distribution of breast cancer polygenic risk overall and by major subtypes.

## Research proposal

### Study design:

International genome-wide association study.

### Data Sources:

The project will leverage large-scale individual-level and summary-level data from the [Confluence Project](#), an international collaboration integrating genetic and phenotypic data from more than 300,000 breast cancer cases and 300,000 controls across diverse ancestries (including of African, Asian, Hispanic/Latina, European and mixed ancestries). Confluence includes genome-wide genotype data (imputed to diverse reference panels, and including selected high- and moderate-risk mutations), harmonised epidemiological risk factors, and tumour characteristics.

### Analyses:

Analyses will include development and validation of multi-ancestry PRS for overall and subtype-specific breast cancer, with a particular focus on early-onset disease. Although we expect that most genetic variants contributing to breast cancer risk are shared across ages, their effect sizes may differ by age at onset. To account for this, analyses will first construct PRS using SNPs identified from the overall multi-ancestry GWAS and then re-calibrate their effect sizes for younger-onset disease. Recalibration will include (a) re-estimating SNP weights using younger-onset cases, and (b) applying empirical-Bayes and hierarchical modelling approaches that borrow information across age groups while allowing age-specific deviations.

Both ancestry-specific and trans-ancestry PRS using approaches such as LDpred2 will be used, along with advanced Bayesian and machine-learning frameworks that integrate information across ancestries and age strata. Prospective cohort data from Confluence will serve as independent validation sets to assess predictive performance. These analyses will yield calibrated, ancestry-informed PRS across age groups and form the basis for evaluating joint effects of PRS and pathogenic variants in subsequent aims.

To assess interactions between PRS, carrier status in established predisposition genes, age, and family history, models that incorporate these factors simultaneously will be used. Final models will be selected using 10-fold cross-validation, maximising the area under the curve as a function of the L1 penalty. The same modelling strategy will be applied to develop ER-subtype-specific PRS.

Model estimates will be combined with country-specific age-specific incidence rates to calculate five-year and lifetime absolute risks (up to age 80) for carriers and non-carriers of pathogenic variants. Pointwise confidence intervals for absolute risks will be derived using parametric bootstrap samples.

To characterise the genetic architecture of early-onset (18–49 years) versus later-onset (50+ years) breast cancer, genetic correlations using LD score regression will be estimated. Analyses will be conducted for overall breast cancer and separately by ER, PR, and HER2 status to assess the extent to which subtype composition explains genetic differences by age at onset. Local genetic correlation analyses will be performed to identify genomic regions driving shared or distinct age-specific effects. Stratified LD score regression will be used to test whether shared versus age-specific loci are enriched for functional or regulatory annotations. Finally, differences in polygenicity and effect-size distributions between early- and later-onset disease will be evaluated using methods designed to estimate genome-wide effect-size architectures.

### **Impact:**

By developing robust multi-ancestry PRS and quantifying how common variants modify risks associated with high- and moderate-risk genes, this research will enable more precise identification of younger women who could benefit from earlier screening or risk-reducing interventions. The work will also enhance understanding of the biological drivers of early-onset breast cancer by clarifying age-specific heritability and the genetic architecture of aggressive subtypes such as triple-negative disease. Ultimately, the project aims to enable risk prediction tools that are equitable, clinically relevant, and suitable for implementation in diverse populations. Outcomes will be directly translatable to risk-stratified screening strategies and personalised prevention efforts.

### **Research setting:**

The PhD student will form part of the within the HER-CARE MSCA Doctoral Network, joining an international cohort of 15 PhD researchers focused on hereditary and early-onset breast cancer. The primary host institution (ICR) will provide training in genetic epidemiology. In addition, as part of the MSCA programme, the student will undertake:

- a five-month secondment at John Hopkins University in the USA (hosted by Prof Nilanjan Chatterjee) that provide complementary experience in statistical genetics and data science; and
- a three-month secondment at PHG Foundation in Cambridge, UK (hosted by Dr Laura Blackburn) that will provide complementary experience in clinical translation and policy evaluation.

In addition, the student will interact closely with investigators at the National Cancer Institute – the coordinating centre for Confluence – and other international investigators and collaborators. The research environment emphasises interdisciplinary training, open science, and international collaboration, ensuring that the student develops strong expertise in both methodological innovation and applied cancer research. The starting date will be before July 1, 2026.

## **Literature references**

Ahearn TU, Zhang H, ..., Schmidt MK, García-Closas M, Chatterjee N. Common variants in breast cancer risk loci predispose to distinct tumor subtypes. *Breast Cancer Res.* 2022 Jan 4;24(1):2. doi: 10.1186/s13058-021-01484-x. PMID: 34983606; PMCID: PMC8725568.

Gao G, Zhao F, Ahearn TU, Lunetta KL, Troester MA, Du Z, Ogundiran TO, Ojengbede O, Blot W, Nathanson KL, Domchek SM, Nemesure B, Hennis A, Ambs S, McClellan J, Nie M, Bertrand K, Zirpoli G, Yao S, Olshan AF, Bensen JT, Bandera EV, Nyante S, Conti DV, Press MF, Ingles SA, John EM,

Bernstein L, Hu JJ, Deming-Halverson SL, Chanock SJ, Ziegler RG, Rodriguez-Gil JL, Sucheston-Campbell LE, Sandler DP, Taylor JA, Kitahara CM, O'Brien KM, Bolla MK, Dennis J, Dunning AM, Easton DF, Michailidou K, Pharoah PDP, Wang Q, Figueroa J, Biritwum R, Adjei E, Wiafe S; GBHS Study Team; Ambrosone CB, Zheng W, Olopade OI, García-Closas M, Palmer JR, Haiman CA, Huo D. Polygenic risk scores for prediction of breast cancer risk in women of African ancestry: a cross-ancestry approach. *Hum Mol Genet.* 2022 Sep 10;31(18):3133-3143. doi: 10.1093/hmg/ddac102. PMID: 35554533; PMCID: PMC9476624.

Gao C, Polley EC, Hart SN, Huang H, Hu C, Gnanaolivu R, Lilyquist J, Boddicker NJ, Na J, Ambrosone CB, Auer PL, Bernstein L, Burnside ES, Eliassen AH, Gaudet MM, Haiman C, Hunter DJ, Jacobs EJ, John EM, Lindström S, Ma H, Neuhausen SL, Newcomb PA, O'Brien KM, Olson JE, Ong IM, Patel AV, Palmer JR, Sandler DP, Tamimi R, Taylor JA, Teras LR, Trentham-Dietz A, Vachon CM, Weinberg CR, Yao S, Weitzel JN, Goldgar DE, Domchek SM, Nathanson KL, Couch FJ, Kraft P. Risk of Breast Cancer Among Carriers of Pathogenic Variants in Breast Cancer Predisposition Genes Varies by Polygenic Risk Score. *J Clin Oncol.* 2021 Aug 10;39(23):2564-2573. doi: 10.1200/JCO.20.01992. Epub 2021 Jun 8. PMID: 34101481; PMCID: PMC8330969.

Jia G, Ping J, Guo X, Yang Y, Tao R, Li B, Ambs S, Barnard ME, Chen Y, Garcia-Closas M, Gu J, Hu JJ, Huo D, John EM, Li CI, Li JL, Nathanson KL, Nemesure B, Olopade OI, Pal T, Press MF, Sanderson M, Sandler DP, Shu XO, Troester MA, Yao S, Adejumo PO, Ahearn T, Brewster AM, Hennis AJM, Makumbi T, Ndom P, O'Brien KM, Olshan AF, Oluwasanu MM, Reid S, Butler EN, Huang M, Ntekim A, Qian H, Zhang H, Ambrosone CB, Cai Q, Long J, Palmer JR, Haiman CA, Zheng W. Genome-wide association analyses of breast cancer in women of African ancestry identify new susceptibility loci and improve risk prediction. *Nat Genet.* 2024 May;56(5):819-826. doi: 10.1038/s41588-024-01736-4. Epub 2024 May 13. PMID: 38741014; PMCID: PMC11284829.

Zhang H, Zhan J, Jin J, Zhang J, Lu W, Zhao R, Ahearn TU, Yu Z, O'Connell J, Jiang Y, Chen T, Okuhara D; 23andMe Research Team; Garcia-Closas M, Lin X, Koelsch BL, Chatterjee N. A new method for multi-ancestry polygenic prediction improves performance across diverse populations. *Nat Genet.* 2023 Oct;55(10):1757-1768. doi: 10.1038/s41588-023-01501-z. Epub 2023 Sep 25. PMID: 37749244; PMCID: PMC10923245.

Zhang, H. et al. Novel methods for multi-ancestry polygenic prediction and their evaluations in 5.1 million individuals of diverse ancestry. *Nat. Genet.* (in press) (2023). doi:10.1101/2022.03.24.485519.

Zhang H, Zhao N, Ahearn TU, Wheeler W, García-Closas M, Chatterjee N. A mixed-model approach for powerful testing of genetic associations with cancer risk incorporating tumor characteristics. *Biostatistics.* 2021 Oct 13;22(4):772-788. doi: 10.1093/biostatistics/kxz065. PMID: 32112086; PMCID: PMC8511944.

Zhang H, Ahearn TU, Lecarpentier J, ..., Easton DF, Chatterjee N, García-Closas M. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet.* 2020 Jun;52(6):572-581. doi: 10.1038/s41588-020-0609-2. Epub 2020 May 18. PMID: 32424353; PMCID: PMC7808397.

Zhang YD, Hurson AN, Zhang H, ... Chanock SJ, Chatterjee N, Garcia-Closas M. Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. *Nat Commun.* 2020 Jul 3;11(1):3353. doi: 10.1038/s41467-020-16483-3. PMID: 32620889; PMCID: PMC7335068.

## Candidate profile

**Note:** the ICR's standard minimum entry requirement is a relevant undergraduate Honours degree (First or 2:1).

### Pre-requisite qualifications of applicants:

Master in Epidemiology, Genetics, Biostatistics, Data Science or related field; or equivalent experience in these areas.

Additional requirements from MSCA doctoral networks programme:

- **No** doctoral degree at recruitment date.
- No residence or main activity (work, studies, etc.) in the UK for more than **12 months in the 36 months** before their recruitment date.

### Intended learning outcomes:

- Understand the genetic epidemiology of breast cancer.
- Critically read and analyse scientific literature.
- Develop hypotheses that build upon existing knowledge.
- Apply advanced methods for data analyses and interpretation.
- Understand clinical translation and policy implications of research
- Learn to work in a collaborative research environment, leveraging the support of internal teams and external collaborators to enhance research outcomes.
- Communicate research goals, methods, results and implications in both writing and orally.
- Understand and adhere to the ethical considerations and guidelines pivotal in research involving human samples and data.

## Advertising details

### Project suitable for a student with a background in:

- ☐ Biological Sciences
- ☐ Physics or Engineering
- ☐ Chemistry
- ☒ Maths, Statistics or Epidemiology
- ☒ Computer Science