ICR The Institute of Cancer Research

PhD Project Proposal

Funder details

Studentship funded by:	Institute of Cancer Research

Project details

Project title:	Tissue-based biomarkers of risk and progression to breast cancer using pathomics

Supervisory team

Primary Supervisor:	Professor Montserrat Garcia-Closas, ICR
Associate Supervisor(s):	Dr. Mustapha Abubakar, National Cancer Institute, USA
Secondary Supervisor:	Professor Amy Berrington, ICR

Divisional affiliation

Primary Division:	Division of Genetics & Epidemiology
Primary Team:	Integrative Epidemiology Team
Site:	Sutton

Project background

Many women in the UK undergo breast biopsies each year to evaluate abnormalities or suspicious imaging findings. While some find cancerous lesions, the majority are normal or benign breast diseases (BBD) that, although non-cancerous, can progress or predict progression to breast cancer (Hartmann 2005). Histologic features in BBD H&E images of breast tissue could reveal footprints of past exposures to risk factors and early biological effects (Abubakar 2023a, that could be used as biomarkers for increased risk of breast cancer development (Abubakar 2021; Vellal 2021). Previous studies have shown that tumour-associated stromal cellular density in breast tumours reflects prior exposures to risk factors (Abubakar 2023b). This suggest that both epithelial hyperplasia (a known biomarker of risk) and stromal disruption are involved in the aetiopathogenesis of invasive breast carcinoma. Advances in radiomics are enabling AI analyses of BBD images with epithelial and stromal components to identify novel biomarkers of progression to subsequent invasive breast cancer in women with non-proliferative BBD lesions. Such tissue-based biomarkers can be used to identify high-risk women who could benefit from enhanced screening and risk-reduction strategies.

Project aims

- To evaluate reproductive, hormonal, lifestyle, genetic and mammographic breast density in relation to the epithelial and stromal changes in BBD using imaging AI analyses.
- To estimate the association of BBD epithelial and stromal changes with risk of developing different types of breast cancer.

- To identify relationships between epithelial and stromal changes in BBD and those in subsequent tumours.

Research proposal

Using novel machine-learning algorithms for image analyses and leveraging rich resources from the Generation study, the PhD candidate will explore epithelial and stromal features in BBD tissue slides to identify biomarkers of breast cancer risks and progression.

Study Population

Research will be based on data and tissue images from participants in the Generations Study, a prospective cohort of 110,000 women in the UK (Swerdlow et al. 2011). Participants have provided information about their BBD and the hospital where they underwent surgical or biopsy procedures through questionnaires. About 12,000 study participants in Generations Study have reported having a breast biopsy or surgery for a BBD lesion. The resource includes comprehensive and repeated information on risk factors, serial mammographic images, circulating hormone levels, genotyping, as well as paired samples of both BBD and tumour tissues. The richness of data makes this resource ideally suited to study predictors of risk and progression of BBD to breast cancer.

A nested case-control study within the sub-cohort of 12,000 women with BBD in the Generation study will be designed, and it is expected to include 675 women who developed breast cancer and 675 controls without breast cancer. Pathology reports will be used to confirm the BBD diagnosis, and H&E slides will be digitised for AI analysis. For women who develop cancer, H&E slides from tumour formalin-fixed, paraffin-embedded (FFPE) blocks will also be available for AI analyses.

Plan of investigation and methodology

The proposed project will involve designing a nested case-control study, data curation and performing AI image analyses of H&E slides from BBD and tumour tissues. Data analysis will use integrative statistical and data science methodologies will be conducted in collaboration with pathologists, statisticians and data scientists. Machine-learning analysis, including tissue classification and cell detection scripts, will be applied to histologic images to characterize epithelial and stromal features. This could include random forest or neural network algorithms optimized and validated to identify, segment, and quantify areas of interest. By analysing these data, the project will aim to investigate the relationship between epithelial and stromal changes in BBD and the subsequent development of tumours, explore the influence of genetic susceptibility loci and hormone levels on tissue characteristics in BBD, and their connection to cancer progression. This project will also evaluate how imaging features in BBD relate to those in subsequent breast tumours to evaluate possible pathways to progression. In addition, it will evaluate whether genetic (determined by polygenic risk scores and family history) and non-genetic risk factors (e.g. reproduction history, hormone levels, BMI, hormone replacement therapy use) are associated with imaging features in BBD and subsequent tumour, and whether the expected increase in risk is modified by risk factors.

Impact

Tissue-based biomarkers for women with BBD could be used to identify high-risk women who could benefit from enhanced screening and risk-reduction strategies. This has practical clinical significance considering that the NHS breast screening program conducts approximately 40,000 breast biopsies annually, with 28,000 of them resulting in non-proliferative diagnoses (NHS Breast Screening Program, England 2018-19). The outcome of this research has the potential to offer women with valuable insights into their individual risk profiles, reducing the uncertainty associated with receiving a diagnosis of BBD.

Literature references

[1] Abubakar M, Klein A, Fan S, Lawrence S, Mutreja K, Henry JE, Pfeiffer RM, Duggan MA, Gierach GL. Host, reproductive, and lifestyle factors in relation to quantitative histologic metrics of the normal breast. Breast Cancer Res. 2023a Aug 15;25(1):97. doi: 10.1186/s13058-023-01692-7. PMID: 37582731; PMCID: PMC10426057.

[2] Abubakar M, Ahearn TU, Duggan MA, Lawrence S, Adjei E, Clegg-Lamptey JN, Yarney J, Wiafe-Addai B, Awuah B, Wiafe S, Nyarko K, Aitpillah F, Ansong D, Hewitt SM, Brinton LA, Figueroa JD, Garcia-Closas M*, Edusei L*, Titiloye N*. Associations of breast cancer etiologic factors with stromal microenvironment of primary invasive breast cancers in the Ghana Breast Health Study. Res Sq [Preprint]. 2023b Apr 14:rs.3.rs-2791342. doi: 10.21203/rs.3.rs-2791342/v1; * co-senior authors

[3] Abubakar M, Fan S, Bowles EA, Widemann L, Duggan MA, Pfeiffer RM, Falk RT, Lawrence S, Richert-Boe K, Glass AG, Kimes TM, Figueroa JD, Rohan TE, Gierach GL. Relation of Quantitative Histologic and Radiologic Breast Tissue Composition Metrics With Invasive Breast Cancer Risk. JNCI Cancer Spectr. 2021 Feb 6;5(3):pkab015. doi: 10.1093/jncics/pkab015. PMID: 33981950; PMCID: PMC8103888.

[4] Hartmann LC, Sellers TA, Frost MH, Lingle WL, Degnim AC, Ghosh K, Vierkant RA, Maloney SD, Pankratz VS, Hillman DW, Suman VJ, Johnson J, Blake C, Tlsty T, Vachon CM, Melton LJ 3rd, Visscher DW. Benign breast disease and the risk of breast cancer. N Engl J Med. 2005 Jul 21;353(3):229-37. doi: 10.1056/NEJMoa044383. PMID: 16034008.
[5] Vellal AD, Sirinukunwattan K, Kensler KH, Baker GM, Stancu AL, Pyle ME, Collins LC, Schnitt SJ, Connolly JL, Veta M, Eliassen AH, Tamimi RM, Heng YJ. Deep Learning Image Analysis of Benign Breast Disease to Identify Subsequent Risk of Breast Cancer. JNCI Cancer Spectr. 2021 Jan 11;5(1):pkaa119. doi: 10.1093/jncics/pkaa119. Erratum in: JNCI Cancer Spectr. 2021 Aug 16;5(4):pkab055. PMID: 33644680; PMCID: PMC7898083.
[6] Swerdlow AJ, Jones ME, Schoemaker MJ, Hemming J, Thomas D, Williamson J, Ashworth A. The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer aetiology. Br J Cancer. 2011 Sep 27;105(7):911-7. doi: 10.1038/bjc.2011.337. Epub 2011 Sep 6. PMID: 21897394; PMCID: PMC3185950.

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

Intended learning outcomes:

- Understand the epidemiology and pathology of breast cancer.
- Critically read and analyse scientific literature, fostering a deep understanding and the ability to integrate current research with historical perspectives.
- Develop hypotheses that build upon existing knowledge.
- Apply rigorous epidemiological methods for study design, data generation, analyses and interpretation, accounting for potential biases.
- Develop machine-learning scripts for AI image analyses.
- 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