Title: Can a combination of tumour phenotyping and genotyping identify hereditary breast cancer predisposition?

Abstract of research Evidence is emerging that the molecular genetic events that underlie breast cancer may be determined to some extent by the genetic background on which these cancers develop. The characteristic features of tumours arising in carriers of the BRCA1 gene for example are typically somewhat different on average compared with sporadic or non BRCA1/BRCA2 tumours. Furthermore these differences in molecular evolution may lead to differential responses to chemotherapeutic agents that may be reflected not only in gene carriers but also in patients with breast cancer who have acquired BRCA gene dysfunction following bi-allelic somatic mutation in the gene. Current selection for mutation testing relies on a very strong family history of breast and (often) ovarian cancer which is a relatively non specific tool and will not detect women with somatic loss of BRCA1 function who might be expected to respond to adjuvant treatments in the same way as those with inherited mutation. The aim of this project is to determine whether profiling of tumour biology using two alternative strategies, array Comparative Genomic Hybridisation and immunoprofiling, may help to select patients with inherited BRCA1 mutations and further to determine whether this is also possible for BRCA2 gene carriers. The study design is designed to give results that will be directly translatable into clinical practice and suitable tumour based test has the potential to be incorporated directly into pre-chemotherapy tests in order to assist selection of patients for targeted chemotherapy trials.

Lay statement The appearance of a breast cancer when examined down the microscope is necessary for both diagnosis and prognosis and helps determine what additional treatments are likely to be beneficial to an individual eg. whether anti-oestrogen treatment or herceptin are likely to be effective. With more sophisticated approaches to genetic profiling of tumours, even more subtle and individualised treatments will become possible. A strategy starting with testing the tumour to find a profile of features characteristic of a cancer arising because of a specific underlying inherited gene fault will improve identification of women who would benefit from genetic counselling and improve targeted germline mutation testing for inherited susceptibility. This would allow more fully informed discussion about longer term risk reduction options. This study is aimed primarily at developing a robust method for clear identification of BRCA1 and BRCA2 carriers. In addition we anticipate being able to identify other subsets of hereditary risk from the tumour characteristics.

Statement of context: The aim of the POSH study overall is to look at clinical outcomes of breast cancer treatment in the context of genotype, this work will help to define genotype more effectively than currently available approaches and will give wider insights into the pathobiology of different subtypes of hereditary breast cancer.

Purpose of proposed investigation: There is evidence that carriers of mutations in the high risk BRCA1, BRCA2 genes develop breast tumours that progress down a limited molecular pathway dictated by the underlying genotype and identified as a specific molecular phenotype[1]. Immunoprofiling and array CGH are both techniques that can be applied to paraffin fixed material. Development of a simple and inexpensive test based on tumour molecular pheno/genotype and applicable to large numbers of tumours would allow rapid identification of individuals who may have hereditary cancers and molecular testing could then be targeted towards the most likely gene in those individuals. Since somatic silencing of BRCA1 probably occurs in a significant proportion of tumours, a successful means of identifying BRCA1-null tumours in gene carriers is also likely to be applicable to these tumours and would be a potentially useful tool for selecting patients for targeted treatment trials (e.g. see refs [2&3]) and future studies to investigate the events involved in somatic loss of BRCA1 function.

Background of the project:
In routine clinical practice a young breast cancer patient presenting with a family history of breast cancer may have an underlying BRCA1, BRCA2 or other (unidentified) high risk gene mutation, or a more moderate level of genetic susceptibility. The genetic status may have implications for treatment of that individual but current methods preclude rapid inexpensive testing of blood DNA for mutations because the genes are very large and mutations are scattered throughout. In reality genetic counselling and testing is restricted to those with a striking family history of breast and ovarian cancer and for the index case test results may not be available in time to influence treatment decisions. A tumour signature that has a high degree of specificity for an underlying high risk genotype will allow targeted and thus cost effective genetic testing by allowing the majority of samples not harbouring a germline mutation to be excluded from further exploration. Recent publications have suggested that a specific BRCA1 gene expression signature exists in tumour samples from gene carriers[4;5]. Array Comparative Genomic Hybridisation (aCGH) has been shown to separate BRCA1 tumours from cancers not known to be genetic and from familial non BRCA1/2 breast cancers although the few published studies involve small numbers [6]. Gene expression analysis and subsequent immunohistohemical profiling has demonstrated that breast cancers can be classified into distinct groups with clinical relevance[7]. One group, the basal phenotype, is more frequent in BRCA1-mutation-associated tumours, though the profile is not specific for this genotype[8]. However, recent more extensive immunoprofiling has shown that the basal phenotype can be further subdivided into groups with distinct clinical outcomes [I.O.Ellis, personal communication]. Thus, more detailed profiling, informed by CGH and existing immunohistochemical classifications, has the potential to identify more discriminating molecular signatures. A robust tumour phenotype, using routinely processed pathology blocks, which could specifically point to an individual's likely genetic background would be invaluable in initial treatment decision making and in directing referral for genetic counselling and testing. This study will identify a specific tumour signature for BRCA1 and BRCA2 gene carriers which will utilise tests possible to apply in any histopathology laboratory. Testing of tumour and blood will be truly prospective and will be rigorously validated.

Material for study: The Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH) has recruited a cohort of newly diagnosed breast cancer patients all aged ≤40 years at diagnosis from across the UK over the last four years. There were over 1,300 recruits at the end of June 2005 and we anticipate completing recruitment of 2000 women in May 2006. Blood samples and pathology blocks are available from all recruits and detailed clinical information about diagnosis and treatment and outcomes are logged. Family history is recorded for each recruit and analysis to date shows that 22% of cases have a greater than 50% chance of being due to BRCA1, BRCA2 or other unknown hereditary breast cancer genes. We are in addition however recruiting known gene carriers who have developed breast cancer within the time frame of the project and aged up to and including 50 years at diagnosis so will have sufficient known gene carriers to select these specifically for the development phase of the study. The large cohort of controls (over 1000 with no family history of breast cancer) available in this study means we have the opportunity to match cases with controls for major histopathological features particularly ER status and grade. Matching is desirable since we wish to find critical new features that signify gene carrier status and ER status and grade are variables already known to have some correlation with BRCA1 status for example[9].
The POSH study material allows a much more efficient study design than would be possible using tumour bank material and we have the opportunity to include a much larger number of cases analysed with array CGH than has hitherto been reported in the literature and to compare profiles of tumours from gene carriers with matched controls leading to a much better chance of detecting a genotype specific profile with high predictive value.

Plan of investigation and methodolgy
Sample selection: Two sets of samples will be used for the investigation.

1. The initial "training" set will comprise 40 cases from known BRCA1 carriers, 40 cases from known BRCA2 carriers and for each set 40 controls (80 controls in total). Controls will be selected from breast cancer cases where BRCAPRO (software estimating

likelihood of being a BRCA1 or BRCA2 carrier based on family history data) predicts an extremely low likelihood of carrying a mutated BRCA1 or BRCA2 We will use the clinical pathology reports logged for each recruit to match controls to cases for grade, ER and LN status. These 160 cases and controls will form the first series of Tissue Micro Arrays (TMAs) to be assembled (in progress). This will be the training set from which genotype specific molecular signatures routinely applicable to breast tumour tissue will be identified.

2. For the validation phase of the study, immunohistochemistry (IHC) and chromogenic *in situ* hybridisation (CISH) will be used to detect the protein and chromosomal losses and gains that best reflect the genotype specific tumour signature. These tests will be applied to TMAs representing the entire cohort of 2000 tumours. This will be used to assess the sensitivity and specificity of the predictor to determine the correct underlying genotype.

Retrieval of pathology blocks: Pathology blocks from all cases are being collected as part of the POSH study and sections will be taken for H&E staining for a systematic pathology review. Cores of tissue will be taken for assembly into tissue microarrays (2-3 representative cores of tumour tissue per case will be selected from H&E stained sections).

Systematic histopathology review of all 2000 cases: Sample identification will be blinded (unique POSH study number used) for standard pathology review (H&E staining) and coding of TMAs. This process of systematic review will involve 10 breast pathologists internationally (panel to be confirmed but informal approaches to date have received favourable responses from several potential panel members). A rolling process of review will continue throughout the study but the initial cases reviewed will be the training set. Each case will be reviewed by two pathologists. A final meeting of the panel will allow any divergence of views to be resolved.

Phase 1 (0-12 months) – determining a genotype specific signature

Tissue microarrays (TMAs): TMAs will be constructed from 0.6mm cores using the MTA1 instrument from Beecher Instruments Inc. Cores from each case will be selected by Professor Louise Jones using the H&E sections from the corresponding blocks. There will be up to 3 cores per case (depending on tumour size and heterogeneity). TMAs will be assembled by an experienced histopathology technician based in Professor Jones' lab. Initially TMAs will be made from the training set of 160 samples. Each tumour group (BRCA1, BRCA2, non BRCA-associated) will be randomly represented on the array and identified by POSH study number. TMAs will then be constructed from the remaining 2000 cases to be available for phase 2 of the study.

Immunohistochemistry: Immunostaining of the TMAs using an extensive panel of markers shown to categorise breast cancers into distinct subgroups will be carried out. The choice of markers will be informed by mining the multiple data sets available in the literature, and from our own studies[7;8;10;11]**.** Among others, the markers will include CK5, 6, 15, 17, 18 and 19, ER, PgR, Her2, EGFR, p53, SMA, p63, P-Cadherin and □v□6 integrin.

DNA extraction from tumour blocks: Tumour tissue sections, 5x5 µm thick, will be taken from a homogeneous region in each case for harvesting tumour tissue and DNA extraction. Where a high tumour content (>90%) is not present needle dissection under a low power microscope will be used to obtain suitable material. Sufficient tissue will be taken to ensure the extraction of at least 12µg of genomic DNA; this will avoid any need for whole genome amplification since this may compromise the efficiency of aCGH.

Comparative Genomic Hybridisation: Array CGH will be performed on DNA extracted from tumour tissue from all 160 cases in the training set. The Breakthrough Centre team are currently successfully producing arrays in house comprising 3000 BAC clones located at 1Mb intervals throughout the human genome. BAC clones have been mapped to the latest human genome assembly (hg17) on the basis of in-house resequencing (n=700), publicly available sequence data (n=2,500) and both clone ID and STS marker accessions. Currently, they are at the final stages of production of whole genome arrays comprising 30,000 overlapping BAC clones covering the entire human genome and anticipate that these will be available by the initiation of this project.

DNA labelling and Array hybridisation

Labelled reference and tumour DNA will be denatured then hybridised onto microarray slides then scanned with a GenePix 4000A scanner (Axon Instruments, Inc., Union City, CA) and fluorescence data processed with GenePix 3.0 image analysis software (Axon Instruments, Inc.). For dye bias correction, in separate reactions, the same reference and sample DNAs will be labelled with the opposite dye ('dye-swap').

12-18 months: Development of the predictor profile

Data analysis

Development of the panel of tests that best predict genotype will be carried by the current study co-ordinator (Sue Gerty) in the second year of the study using the data from the arrays and IHC profiling linked to known genotype (see below). Due to the somatic occurrence of null BRCA tumours, it is anticipated that the inclusion of family history in the predictor will improve selection however the study design allows us to identify how frequently inherited mutations occur where there is no family history reported at all (currently thought to be fairly uncommon).

IHC - immunostaining for each marker will be scored semiquantitatively by two pathologists. Supervised hierarchical cluster analysis will be carried out to identify marker profiles associated with different genotypes (see ref [10]).

Array CGH: data analysis will be performed as previously described[12]. A categorical analysis will be applied to the BACs after classifying them as representing gain, loss, or no-change according to their smoothed log2 ratio values. The threshold values are chosen to correspond to 3 standard deviations of the normal ratios obtained from the filtered clones mapping to chromosomes 1-22, assessed in comparisons between DNA extracted from a pool of male and female blood donors. Data transformation and statistical analysis will be carried out in R 2.0.1 (http://www.r-project.org/) and BioConductor 1.5 (http://www.bioconductor.org/).

Chromogenic *in situ* hybridisation (CISH): CISH has a similar degree of accuracy to fluorescence *in situ* hybridization (FISH), but does not require fluorescence microscopy so is a more practical and economical method and is applicable to TMAs[13]. The predictor profile will combine IHC and CISH to give the most discriminating combination for each underlying genotype.

18-24 months

Phase 2 – Validation of genotype predictor profile

The ability of the predictor profile to discriminate between BRCA1, BRCA2 and non-BRCA-associated tumours will be tested by applying the CISH and IHC panels to TMAs representing the the entire POSH cohort of ~2000 cases. All cases will be categorised according to the predictor profile into (i) predicted BRCA1, (ii) predicted BRCA2, (iii) non BRCA-associated. It is expected that for those predicted a low likelihood of being gene carriers, some will and some will not have a family history of breast cancer. We will select 150 cases for genotyping for mutations in BRCA1 and BRCA2; 50 cases will be those predicted most likely to be BRCA1 carriers, 50 those most likely to be BRCA2 carriers and 50 unlikely to carry either gene, half with and half without a family history of breast cancer. BRCAPRO likelihood will also be assessed in each case.

Genotyping: DNA will be extracted from peripheral blood leukocytes and analysed for BRCA1/2 mutations using MLPA and high throughput Conformation Sensitive Capillary Electrophoresis (CSCE) (in collaboration with the National Genetics Reference Laboratory in Salisbury).

Diagnostic ability of the predictor

Assessment of the sensitivity and specificity of the predictor to "call" gene carriers for each gene can be estimated from {sensitivity % = true positives/[true positives + false negatives] x 100} and {specificity % = true negatives/[true negatives + false positives] x 100). In this context true positives would be gene carriers correctly identified by the technique, true negatives would be sporadic cases not picked out for mutation testing and where no mutation is found on testing, false positives will be those identified as gene carriers who have a negative mutation screen (accepting that a small percentage of mutations may be missed by our approach to mutation screening and false negatives will be mutation carriers not picked as such by the technique). The performance of the molecular predictor will be compared to the performance of family history based methods (BRCPRO).

Out of 150 tumours selected for genotyping by the predictor, as described from the validation phase, 50 will be the tumours predicted to be highly likely to have arisen in BRCA1 carriers. If for example the predictor was able to select carriers of a particular genotype with 90% accuracy,

then we would expect that 45 of the 150 tested genotypes would be identified as carrying a BRCA1 mutation if our genotyping were 100% sensitive for all mutations. If our genotyping of blood DNA were only 90% sensitive for causative mutations then 40.5 (say 40) mutations would be found on genotyping in this group, 5 would be wrongly called (or be due to somatic BRCA1 mutations) and 5 would be occult mutations. The table of results would appear as follows.

	Predictor +ve	Predictor –ve	Total
BRCA1 +ve	40	10	50
BRCA1 –ve	10	90	100
Total	50	100	150

Specificity = 90/(90+10) x 100 = 90% with a 95% confidence interval of 83% to 95% for this sample size. Sensitivity = 40/(40 + 10) x 100 = 80% (CI = 71% to 87%).

Additional outcomes

An additional outcome from these experiments will be determination of regions of interest in the genome flagged by critical regions of loss or gain of material associated with tumours developing in young women with a family history of cancer but no detectable gene mutation and where tumour pathology predicts a low likelihood of an underlying BRCA1 or BRCA2 mutation. This information will inform current and future linkage studies in such families.

Ethical approval

The study was granted full MREC approval on 9th February 2001 (MREC 00/06/69). Consent covers all aspects of the work to be conducted in this sub study.

Timescale, milestones and potential problems:

The pathology block retrieval and TMA assembly for the training set is already in progress. The proposal aims at parallel processing in separate labs each with the relevant expertise allowing the work to progress rapidly. Professor Jones has extensive experience of using TMA's and immunohistochemistry so we do not envisage significant difficulties The extraction of DNA from paraffin blocks and its use for aCGH is routine and this work will be carried out by a technician working in the Breakthrough Centre microarray core facility. Finally the National Genetics Reference Laboratory in Salisbury is responsible for establishing a new high throughput mutation scanning facility for BRCA1 and BRCA2 by the end of 2005 which will be run by the NHS service laboratory (Wessex Regional Genetics Laboratory; WRGL). The MTO working on this project will work with the WRGL to analyse BRCA1 and BRCA2 mutations for this study. All data management and analysis, both to develop the predictor and the validation analysis, will be undertaken by the POSH study co-ordinator under the supervision of the POSH study statistician. The main risk is if no specific signature emerges for BRCA1 but this is unlikely in light of preliminary reports from the literature; the main area of higher risk is whether a signature for BRCA2 or other high risk genetic predisposition can be elucidated. However the work still has intrinsic value since it will lead to further relevant studies but these would form the basis of a subsequent application. If the predictor identifies a significant proportion of apparently sporadic cancers as having a BRCA1 or BRCA2 specific signature, the specificity of the predictor for underlying genotype may be reduced. The most likely reason for this occurring would be somatic loss of function in both alleles of one or other gene. It is anticipated that family history will help resolve this. However these BRCA null tumours would be an extremely interesting focus for a future study looking at the mechanism of somatic loss but would be outside the scope of the current project.

Justification of costs

The cost of the type of work proposed (especially array CGH and large gene mutation analysis) is fairly high. To minimise cost and maximise the likelihood of a clear outcome that can be transferred to clinical usage to benefit patients, we have opted for a plan that involves the three distinct parts of the study being executed in three centres where the relevant expertise and experience mean that the work can be carried out quickly and efficiently and data then co-ordinated from the study centre in Southampton. The project should be completed in two years using this design but consequently costs are more in line with a three year project grant application. Funding for technician time to cover each of the three aspects of the work is requested, 12 months for the array CGH work, 12 months (2 years part time) for all the pathology,

TMA and IHC work, and 6 months for the molecular genetics analysis in blood DNA samples. The study co-ordinator will co-ordinate the pathology block recall and for the second year work full time on data analysis and co-ordinate the validation study – her salary is requested for the second year of the study. Consumables are calculated to cover each of the three main laboratory inputs to the project. The cost of processing each case with array CGH is at least £200 (£32,000 for 160 cases) however the Breakthrough Breast Cancer facility will cover the outstanding additional costs of these experiments above the £10,000 requested, making the consumables cost for the array CGH work very reasonable. The cost of reagents and antibodies for the full pathology review and immunohistochemistry on a total of 70-80 TMAs is conservatively estimated and the consumables estimated for the WRGL is also a standard amount for a project of this size. Travel is requested specifically to facilitate the systematic pathology review; most of the cases will be distributed by post but for all cases where there is a discrepancy between two reviewers, a final meeting of the full panel will be arranged to resolve discrepancies.

Reference List

1. Turner N, Tutt A, Ashworth A. Hallmarks of 'BRCAness' in sporadic cancers. *Nat.Rev.Cancer* 2004;**4**:814-9.

2. Couzin J. Choices--and uncertainties--for women with BRCA mutations. *Science* 2003;**302**:592.

3. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB *et al*. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 2005;**434**:917-21.

4. Hedenfalk I, Ringner M, Ben Dor A, Yakhini Z, Chen Y, Chebil G *et al*. Molecular classification of familial non-BRCA1/BRCA2 breast cancer. *Proc.Natl Acad.Sci.U.S.A* 2003;**100**:2532-7.

5. Jazaeri, A. A., Yee, C. J., Sotiriou, C., Brantley, K. R., Boyd, J., and Liu, E. T. Gene Expression Profiles of BRCA1-Linked, BRCA2-Linked, and Sporadic Ovarian Cancers. JNCI Cancer Spectrum 94, 990-1000. 2002.

6. Wessels LF, van Welsem T, Hart AA, van't Veer LJ, Reinders MJ, Nederlof PM. Molecular classification of breast carcinomas by comparative genomic hybridization: a specific somatic genetic profile for BRCA1 tumors. *Cancer Res.* 2002;**62**:7110-7.

7. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H *et al*. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc.Natl Acad.Sci.U.S.A* 2001;**98**:10869-74.

8. Foulkes WD, Brunet JS, Stefansson IM, Straume O, Chappuis PO, Begin LR *et al*. The prognostic implication of the basal-like (cyclin E high/p27 low/p53+/glomeruloid-microvascular-proliferation+) phenotype of BRCA1-related breast cancer. *Cancer Res.* 2004;**64**:830-5.

9. Lakhani SR, Van De Vijver MJ, Jacquemier J, Anderson TJ, Osin PP, McGuffog L *et al*. The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2. *J Clin Oncol* 2002;**20**:2310-8.

10. Makretsov NA, Huntsman DG, Nielsen TO, Yorida E, Peacock M, Cheang MCU *et al*. Hierarchical Clustering Analysis of Tissue Microarray Immunostaining Data Identifies Prognostically Significant Groups of Breast Carcinoma. *Clin Cancer Res* 2004;**10**:6143-51.

11. 't Veer LJ, Dai H, Van De Vijver MJ, He YD, Hart AA, Mao M *et al*. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;**415**:530-6.

12. Reis-Filho JS, Simpson PT, Jones C, Steele D, Mackay A, Iravani M *et al*. Pleomorphic lobular carcinoma of the breast: role of comprehensive molecular pathology in characterization of an entity. *J Pathol.* 2005.

13. Tanner M, Gancberg D, Di Leo A, Larsimont D, Rouas G, Piccart MJ *et al*. Chromogenic in Situ Hybridization : A Practical Alternative for Fluorescence in Situ Hybridization to Detect HER-2/neu Oncogene Amplification in Archival Breast Cancer Samples. *Am J Pathol* 2000;**157**:1467-72.