## 1. General information

### *Title*
Validating two blood tests for early breast cancer and gathering epigenetic information in the FH01 cohort

### *Sponsor*
University College, London

### *Sponsor's signatory*
TBA.

### *Sponsor's medical expert*
TBA.

### *Investigators*
Professor Stephen W. Duffy,
Cancer Research UK Centre for Epidemiology, Mathematics and Statistics,
Wolfson Institute of Preventive Medicine,
Charterhouse Square,
London EC1M 6BQ
Tel 44 (0)20 78823535

Professor Louise Jones,
Centre for Tumour Biology,
John Vane Science Centre,
Barts and the London,
Charterhouse Square,
London EC1M

Dr James Mackay,
Department of Biology,
University College London

Professor Attila Lorincz,
Cancer Research UK Centre for Epidemiology, Mathematics and Statistics,
Wolfson Institute of Preventive Medicine,
Charterhouse Square,
London EC1M 6BQ

### *Medical responsibility*
Prof Louise Jones (above)

### *Laboratories*

Molecular Epidemiology Laboratory
Cancer Research UK Centre for Epidemiology, Mathematics and Statistics,
Wolfson Institute of Preventive Medicine,
Charterhouse Square,
London EC1M 6BQ
Tel 44 (0)20 7014 0252

DiaGenic ASA
Grenseveien 92
NO-0663 Oslo
NORWAY

OncImmune Limited
Clinical Sciences Building
Nottingham City Hospital
Hucknall Rd.
Nottingham
NG5 1PB

## 2. Background information

### *2.1 Investigational product*

This is a study of potential markers for early breast cancer. The first is a test based on a gene expression panel (37 genes) in RNA from peripheral blood.[1] The second is of a combination of onco-antibody assays.[2] These two tests have already been fully developed and the purpose of this proposed study is to validate them. The third component of this study is more exploratory. It aims to describe the epigenetic profile of subjects at increased familial risk of breast cancer, in terms of a number of markers, notably methylation markers which have been shown to be influential in related malignancies such as prostate cancer.[3,4]

### *2.2 Summary of findings to date*

In internal k–fold cross-validation, the gene expression test was found to have 87% sensitivity and 86% specificity.[1] In particular, the test was sensitive to stage I carcinoma. The onco-antibody test was found to have 64% sensitivity and 85% specificity. The epigenetic studies are at an exploratory stage, so prior results are not available.

### *2.3 Risks and benefits*

The only risks to subjects from this study are those associated with having a blood sample drawn. No benefits are anticipated for the participants in this study during the period of the study. However, in the event of the study's results validating the test, those women with a positive blood test but no diagnosis of breast cancer as yet can be offered further diagnostic investigation and surveillance. Also, in this event, the potential benefits to future patients in terms of early diagnosis and treatment are substantial.

### *2.4 Administration of the intervention*

It is planned to take two blood samples every six months from the participants over two years, i.e. taking 4 double samples in total from each participant. Follow-up for occurrence of breast cancer will continue for a further two years thereafter. The subjects have already consented to flagging and follow-up for breast cancer. No further commitment is required from the participants for purposes of this study. Samples will be analysed for the gene expression and onco-antibody test for those who are subsequently diagnosed with breast cancer and for ten control subjects for each breast cancer case. The first sample will be taken using a kit designed for stable extraction of RNA. The second will use a kit which suitable for subsequent

separation, using serum for the onco-antibody assay with ELISA, and plasma for the epigenetic testing. Results will not be communicated to the subjects as at this stage, they have no implications for the treatment or surveillance of the subjects. For the epigenetic markers, we propose to test the first 1,000 samples to ascertain the profile of this risk group. As cases are diagnosed, their stored plasma samples will be tested, and compared with the population profile, to generate hypotheses on predisposition, for confirmation in future validation studies.

### *2.5 Compliance*

The study will comply with good clinical practice guidelines and all regulatory requirements.

### *2.6 Population to be studied*

FH01 is a study of annual mammography in 6,669 women aged under 50 with an enhanced risk of breast cancer due to family history of the disease.[5] Women of this age tend to have radiologically dense tissue, reducing the sensitivity of mammography, and a biological marker of early breast cancer would be of great value in this population. We propose to investigate the blood tests in a subgroup of 4,000 of the FH01 population, taking 6-monthly blood samples over two years.

### *2.7 References*


1. Sharma P, Sahni NS, Tibshirani R, *et al*. Early detection of breast cancer based on gene expression patterns in peripheral blood cells. *Breast Cancer Res* 2005;**7**:R634–44.
2. Chapman C, Murray A, Chakrabarti J, *et al*. Autoantibodies in breast cancer: their use as an aid to early diagnosis. *Ann Oncol* 2007;**18**:868–73.
3. Nelson WG, Yegnasubramanian S, Agoston AT, Bastian PJ, Lee BH, Nakayama M, De Marzo AM. Abnormal DNA methylation, epigenetics, and prostate cancer. *Fron Biosci* 2007;**12**:4254–66
4. Cooper CS, Foster CS. Concepts of epigenetics in prostate cancer development. *Br J Cancer* 2009;**100**:240–5.
5. FH01 management committee, steering committee and collaborators. The challenge of evaluating annual mammography screening for young women with a family history of breast cancer. *J Med Screening* 2006;**13**:177–182.


## 3. Objectives and purpose

As noted above, cross-validation exercises suggest that the gene expression test can identify early breast cancers with approximately 87% sensitivity and 76% specificity.[1] Recent improvements suggest that 86% specificity can be achieved (Christensen, in preparation). The corresponding results for the onco-antibody test indicate 64% sensitivity and 85% specificity.

A number of questions regarding the first two tests can be clarified using the FH01 cohort.[5] In the first instance, are the promising first results replicated in FH01: can the tests prospectively detect early cancers in this particular familial risk group as already observed in other settings? A related important point is: are they equally efficacious in women with dense breast tissue as in women with fatty replaced breast tissue?

More interesting, do the tests show positive results in some time in advance of suspicious imaging followed by a diagnosis of breast cancer? Does this apply equally to those with dense and non-dense breasts?

The epigenetic work is at the exploratory, hypothesis-generating stage. It is anticipated that a small number of DNA methylation markers will be identified as possible risk markers. These will need to be confirmed in future validation studies.

Thus, the objectives can be summarised as:

1. Prospective estimation of sensitivity of the tests in women with a significant family history of breast cancer.
2. Prospective estimation of specificity of the tests in women with a significant family history of breast cancer.
3. If the test proves sensitive to breast cancer in this setting, estimation of how far in advance of diagnosis a positive test is observed.
4. Do the answers to (1–3) above differ between those women with dense breasts and those with fatty replaced breasts?
5. What is the typical profile of this moderate familial risk group with respect to the epigenetic markers?
6. Do the epigenetic markers correlate with risk factors for breast cancer?
7. Are any of the epigenetic markers correlated with risk of breast cancer in this population?

## 4. Study design

From the seven highest recruiting Cancer Research networks to FH01 (Wales, Greater Manchester, Scotland, Avon, Merseyside/Cheshire, South-West London and Mid Trent) we can recruit 4,200 subjects so far free of cancer. We assume 4,000 to be conservative. These subjects will be asked to provide a blood sample every six months for two years. In addition to mammographic surveillance within FH01, the subjects are flagged with ONS, so that all breast cancers, whether detected at surveillance or outwith the study, will be ascertained. A dedicated member of staff at each centre will be trained by Diagenic staff in the use of the sample kits, and will be responsible for taking all samples.

The Diagenic staff will also provide training to laboratory staff in extraction and storage of the RNA. RNA will be extracted within a month of taking the samples, however the extracted RNA will be analysed for the gene expression test only for those who are subsequently diagnosed with breast cancer and for ten control subjects for each breast cancer case. Results will then be analysed for questions (1–4) in section 3 above.

The second blood sample will be separated, with plasma stored for epigenetic testing and the serum for the onco-antibody assays. The latter will be carried out by Oncimmune, a small company in Nottingham, UK. Under the scientific direction of Professor John Robertson at the University of Nottingham (UoN), the company has been working for over 4 years, optimizing the blood based assay for the detection of breast cancer.

The epigenetic markers will be measured in the Molecular Epidemiology Laboratory, Wolfson Institute of Preventive Medicine, which has considerable experience in these markers.

A basic summary of the study milestones is given below:

1. Obtain consent for four 6-monthly blood samples in 4,000 unaffected recruits in the seven largest participating CRN's in FH01.
2. The 4,000 subjects are already flagged for cancer and death along with the rest of the FH01 cohort.
3. Continue to take 6-monthly samples from the 4,000 for two years.
4. RNA will be extracted from the first blood sample, and stored. The second sample will be separated, and the plasma and serum stored separately.
5. The first 1,000 plasma samples will be tested for the epigenetic markers and results correlated with breast cancer risk factors.
6. As breast cancers are diagnosed, retrieve the stored samples for the entire sequence of samples for each cancer and for 10 controls per cancer, and apply the expression and antibody tests to these. For the epigenetic markers, only the cases will be tested, since 1,000 test results will already be available for comparison.

7. Calculate prospective estimates of sensitivity, specificity and predictive values, from the cancers and the 930 controls, first for the overall performance of the expression and antibody tests, then for the performance by time before diagnosis of the cancers, and by breast density. This analysis is to be carried out at two years and four years. Assess associations of the epigenetic markers with risk using unconditional logistic regression.

### *4.1 Endpoints*

In view of the objectives above, the primary endpoints are the test results up to the time of diagnosis in breast cancer cases, and in disease free controls. From these, prospective sensitivity and specificity will be estimated, along with the lead time of the tests. Comparison of expression and antibody results between cases and controls will be made using conditional logistic regression, taking account of the matching. Unmatched comparison for the epigenetic markers will be carried out using unconditional logistic regression.

### *4.2 Type of study*

This is a single-arm prospective cohort study.

### *4.3 Measures taken to avoid bias*

The gene expression testing of the blood samples will be carried out blind to the diagnostic category (cancer/no cancer). All samples tested from cancer cases will be taken prediagnosis. PaxGene blood kits will be used which contain a solution to prevent RNA deterioration. Extraction of RNA will be carried out using state of the art Qiagen equipment. Separation and storage of serum and plasma are standard. Testing for the oncoantibodies and the epigenetic markers will be carried out in two laboratories with proven track experience in the respective areas (Oncimmune and the Molecular Epidemiology Laboratory, London).

### *4.4 Regimen*

Blood samples of 2.5 ml will be taken every six months using PaxGene blood kits. A second sample of 5 ml will be taken using standard equipment.

### *4.5 Duration*

Four sets of two blood samples will be taken over two years. Aside from providing these blood samples, no further commitment is asked of the subjects.

### *4.6 Stopping rules*

NA.

### *4.7 Accountability procedures*

The products under investigation are potential diagnostic tests which are under validation. They have no bearing on the treatment or clinical outcome of the study subjects. The only indemnity issues relate to taking the blood samples, and these will be covered by the sponsor. Results of the blood tests will not be communicated to the subjects, as at this validation stage they have no implications for the treatment or surveillance of the subjects.

### *4.8 Maintenance of treatment codes*

NA.

### *4.9 CRF data*

NA.

## 5. Selection and withdrawal of subjects

### *5.1 Inclusion criteria*

Subjects must be participants in FH01.

Only subjects with no prior diagnosis of breast cancer will be included.

Only subjects who give informed consent will be included.

### *5.2 Exclusion criteria*

Prior breast cancer.

Failure to give informed consent.

Presence of any condition contraindicating regular venepuncture.

### *5.3 Withdrawal criteria*

- Subjects who decide to withdraw will be removed from the study, and their wishes respected with regard to treatment of data and biological material collected so far.
- Subjects for whom complications or adverse events occur as a result of the blood sampling will be withdrawn.
- Subjects who develop a breast cancer will no longer be asked to provide blood samples, and their active period in the study will be truncated at their date of surgical treatment.
- Subjects who develop comorbidity contraindicating the taking of blood samples or rendering it personally inconvenient for them to continue will be removed from the study and their wishes respected with regard to treatment of data and biological material collected so far.
- Subjects who are removed from the study for any reason will not be replaced.

## 6. Treatment of subjects

NA.

## 7. Assessment of efficacy

In this context it must be emphasised that we are referring to screening and diagnostic efficacy, not treatment efficacy. The subjects in this study are healthy female population members, not patients.

### *7.1 Efficacy parameters*

Sensitivity, specificity, positive and negative predictive values of the test, in the population as a whole and separately for those with dense and fatty replaced breast tissue. Odds ratio estimates of relative risk will also be calculated.

### *7.2 Timing, assessing and recording efficacy parameters*

When a cancer is notified to the CPTU, the informatic staff will select ten controls matched for date of birth within one year, who have not as yet developed breast cancer. The identification numbers of the eleven subjects will be notified to the laboratory staff who will arrange for testing of the entire blood sample history of these subjects for the gene expression and onco-antibody tests. The laboratory staff will not be informed which subject is the cancer case. The laboratory staff will then communicate the test results to the informatic staff at CPTU. These will be stored securely for future statistical analysis when the study is complete.

## 8. Assessment of safety

### *8.1 Safety parameters*

The only such parameter is the outcome of each blood sampling episode.

### *8.2 Methods for timing, assessment, recording and analysis of safety parameters*

The study will have an independent data monitoring and ethics committee (DMEC). Any adverse events in relation to blood samples will be recorded and reported immediately to the investigators and the DMEC chair (see below). The DMEC will meet six-monthly.

### *8.3 Procedures for eliciting reports, etc.*

The staff taking the blood samples will also receive as part of their training instructions as to immediate reporting of any adverse events or any condition arising in a participant which might affect her suitability for continued blood sampling. A standard adverse event and intercurrent illness form will be provided for reporting. The only adverse events related to the research are untoward sequelae of venepuncture, including vasovagal attack (faint) and possible consequences such as indirect injury, bruising, tissue injury, cellulitis and accidental arterial puncture.

### *8.4 Follow-up of subjects with adverse events*

All blood sampling will cease in those subjects with serious adverse events consequent upon having a sample taken. They will be referred immediately for medical attention. Their follow-up in the study for breast cancer incidence will continue unless they request otherwise.

## 9. Statistics

### *9.1 Statistical methods and timing of analysis*

#### Gene expression and onco-antibody tests

We shall calculate sensitivities, specificities and predictive values, and will carry out ROC analysis in relation to test cut-off points, estimating the area under the ROC curve. We shall also estimate the likely benefit in terms of additional cases detected early as a result of augmenting mammography with the test. This will be based on the performance characteristics of testing one year or more before cancers are diagnosed. Analyses will be carried out twice, once of cases and controls accrued at two years, when the blood sampling ceases, and again after a further two years of follow-up, i.e. after four years. We shall also compare cases and controls with respect to test status using conditional logistic regression, including adjusting for risk factors and testing for effect modifications (interactions).

#### Epigenetic markers

We shall first estimate associations of the markers with risk factors for breast cancer using unconditional logistic regression. Cases will be compared with the original sample of 1,000 (excluding any who have subsequently been diagnosed with breast cancer) also using unconditional logistic regression.

### *9.2 Number of subjects*

Approximately half of our cohort have mixed or dense breast patterns. The incidence of breast cancer in the last four years in FH01 was 4.2 per 1000 per year. In the coming four years, we can expect this to be 38% higher due to the ageing of the cohort. We would therefore expect 5.8 cancers per 1000 per year in the next four years. Thus in a subcohort of 4,000 followed up over four years we would expect 93 cancers.

We propose to select 10 controls from the population for each breast cancer case. Then the entire blood sample history of the cases and controls will be tested. From the anticipated 93 cancers and 930 controls, we shall estimate the sensitivities and specificities specified in objectives (1–4) above.

Based on the previous 87% estimate of sensitivity and 76% estimate of specificity, this would have a 95% confidence interval of no more than 7% in either direction on the sensitivity, and a 95% CI of no more than 3% in either direction on the specificity. For the antibody study, 93 cancers and 930 controls will give a 95% confidence interval on the sensitivity of 64% of no more than 10% in either direction, and on a specificity of 85% of no more than 3% in either direction.

This will give prospective confirmation of the tests' predictive power for breast cancer. If we assume that testing one year before mammographic diagnosis has a lesser sensitivity, say 70%, 93 cancers would give a 95% CI on the sensitivity of no more than 10% in either direction.

Because density is associated with increased risk of breast cancer one would expect 60% of the tumours to be in those with dense breast tissue, even though only 50% of the screened subjects in FH01 are likely to have dense tissue. So, to assess sensitivity to tumours in dense breasts, we might expect to have 60% of 93 cancers, i.e. 56 cancers, which would give a 95% CI on a sensitivity of 87% with no more than 9% in either direction.

### *9.3 Level of significance*
5% will be used throughout, although the emphasis in the statistical analysis will be point and interval estimation rather than significance testing.

### *9.4 Criteria for termination*
None. The study will run for four years and no early stopping is anticipated.

### *9.5 Missing, unused and spurious data*
The FH01 co-ordinator in Breast Test Wales will continue to actively solicit routine FH01 data, particularly on mammography episodes and cancer diagnoses, which is delayed in arriving at the data centre. FH01 data is subject to routine logic cross-checks. The data manager in CPTU will monitor data on blood sampling and will investigate if a particular centre appears to have a hiatus in recruitment or inflow of data.

### *9.6 Procedures for reporting any deviations from the statistical plan*
No such deviations are anticipated. Loss of information or failure to recruit the anticipated numbers will be reported in full in the final report.

### *9.7 Selection of subjects*
All subjects at each of the seven centres listed in section 4 above will be considered eligible unless otherwise indicated. For details of selection and withdrawal, see section 5 above.

## 10. Direct access to source data/documents

The investigators undertake to comply with authorised audits, regulatory inspections, ethics committee reviews, etc. All documents and data will be made available to the inspecting authorities on request.

## 11. Quality control and quality assurance

The RNA extraction procedures in the laboratory have inbuilt quality indicators which will be monitored in real time. Any centre which is out of the control range for these will be subject to immediate retraining.

## 12. Ethics

The major issue is that the participants are asked to provide eight blood samples over two years (two samples every six months) for no clinical benefit to themselves. This will be made clear in the information sheet. In addition, data will be stored on subjects and they are flagged and followed up for breast cancer. They have already given consent for this. The study will undergo ethical scrutiny by NRES.

## 13. Data handling and record keeping

Data for statistical analysis will be pseudonymised. Data will be held, password and firewall protected, at the Cancer Trials Prevention Unit (CPTU), Wolfson Institute of Preventive Medicine.

## 14. Financing and insurance

FH01 is financed by HTA. The blood study will be conducted with considerable contributions in kind (testing and training) from DiaGenic and oncimmune. We believe there are no serious insurance/indemnity issues but will take advice from the R and D department of the sponsoring institution.

## 15. Publication policy

Results, whether positive or negative, will be published in peer-reviewed medical journals.

## 16. Supplements

See attached protocol of FH01 main study.