

Functional Breast Cancer Genomics

www.cambridgecancer.org.uk/carlosaldas

Group Leader **Carlos Caldas**

Associate Scientist

Suet-Feung Chin

Bioinformaticians

Zhihao Ding*

Bin Liu

Lorna Morris

Irene Papatheodorou

Clinician Scientist Fellow

John LeQuesne

Research Assistant

Helen Bardwell

Clinical Research Training Fellows

Jean Abraham

Mahesh Iddawela

Sarah Jane Dawson

Clinical Fellows

Stefanie Sassen†

Nikita Makretsov†

Postdoctoral Scientists

Anna Git

Alejandra Bruna

Andrew Teschendorff†

Graduate Students

Daniel Holland

Gemma Sharp

Angelika Modelska*

Scientific Officers

Wendy Greenwood*

Joanna Warren*

Donna Grimmer†

Sarah McGuire*

Inma Spiteri†

Visiting Worker

Stefanie Sassen*

Breast cancer survival continues to improve with the latest data from Cambridge showing that over 85% of patients are alive five years after diagnosis.

Despite this progress we are still a long way away from truly individualised treatment. This means that there are patients being over-treated, while for others all currently available modalities fail to control the disease. This great heterogeneity of breast cancer has challenging clinical and biological implications. Our laboratory is tackling this problem by developing better ways of stratifying breast cancers by identifying improved biomarkers with proven clinical utility and by characterizing the initiating events that transform cells within the breast epithelial hierarchy to give rise to cancer.

Translational breast cancer genomics: applications of molecular profiling in prognosis, prediction and novel therapeutics

The current understanding of the complex molecular taxonomy of breast cancer remains limited since to date most studies have not been adequately sized to robustly identify markers that have true clinical utility. We are currently conducting (in collaboration with colleagues in Vancouver, Nottingham and London) a comprehensive multi-modality genomics profiling (array-CGH, mRNA and miRNA expression profiling, mutation analysis) study of 2000 frozen breast cancers. Our aims with this study are:

- (a) to define the number of identifiable molecular subsets of breast cancer and to correlate these with the distinct clinical-pathological phenotypes
- (b) to identify and develop molecular signatures that will identify different prognosis groups
- (c) to develop surrogate markers for stage and grade

The molecular heterogeneity of breast cancer, which the above study will unravel, means that profiling large numbers of samples from patients treated in clinical trials will be needed. This is to validate molecular markers that are highly correlated with outcome (disease-free and overall survival) independently of treatment received (prognostic markers) or markers that can be used to determine which of the two treatment arms is better for a given group (predictive markers). We are generating a unique resource of clinically annotated tumours (in total more than 10,000 paraffin-embedded samples) from a population-based cohort and from four randomized clinical trials, to robustly address these questions. To handle these large numbers of cancer tissues we have constructed a tissue microarray workflow for immunohistochemistry and *in situ* hybridisation analysis. This analysis includes the appropriate relational databases and image analysis platforms. We are also optimising methods for the profiling of DNA and RNA extracted from these paraffin-embedded tissue cores.

Building on a previously identified seven-gene prognostic immune response module for ER- breast cancer, we developed a novel statistical tool based on mixture discriminant analysis in order to build a classifier that could

*Joined during 2008 †Left during 2008

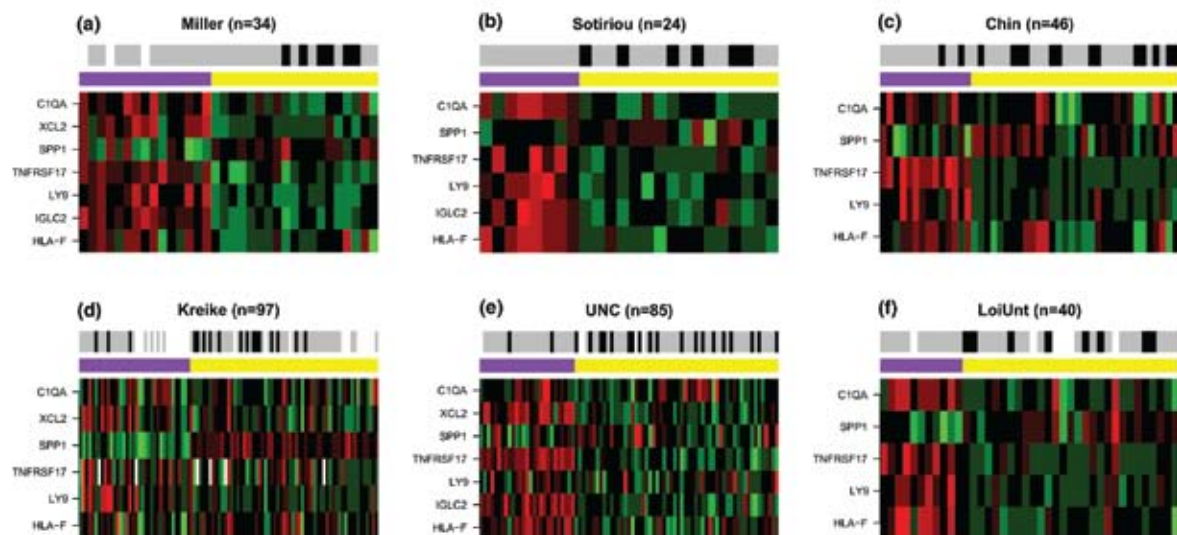


Figure 1. Heatmaps of gene expression of the seven-gene immune response-module for the training and six test cohorts (red = high relative expression, green = low). Samples are clustered into two groups according to the partitioning around medoids algorithm (purple = group overexpressing the immune response-module, yellow = group underexpressing the immune response-module). Clinical outcome as defined by a disease-specific death event (or distant metastasis if the former is not available) is also shown (black = poor, grey = good, white = missing data). Note that in some cases not all seven genes could be mapped to the external platform. C1QA = complement component 1, q subcomponent, A chain; HLA-F = major histocompatibility complex, class I, F; IGLC2 = immunoglobulin lambda constant 2; LY9 = lymphocyte antigen 9; TNFRSF17 = tumour necrosis factor receptor superfamily member 17; SPP1 = secreted phosphoprotein 1 (osteopontin); XCL2 = chemokine (C motif) ligand 2. Figure reproduced from: Teschendorff AE, Caldas C. A robust classifier of high predictive value to identify good prognosis patients in ER negative breast cancer. *Breast Cancer Research* 2008; 10:R73

accurately identify ER⁻ patients with a good prognosis (Figure 1). The classifier accurately predicts, across a training cohort of 183 ER⁻ tumours and six independent validation cohorts (a total of 469 ER⁻ tumours), ER⁻ patients of good prognosis (in validation sets, average predictive value = 94% [range 85 to 100%], average hazard ratio = 0.15 [range 0.07 to 0.36] $p \leq 0.001$) independently of lymph node status and treatment. This seven-gene classifier could be used in a polymerase chain reaction-based clinical assay to identify ER⁻ patients with a good prognosis, who may therefore benefit from less aggressive treatment regimens. This work highlights the new statistical approaches we are developing to train and validate novel classifiers. We are also optimising methodology using Illumina next generation sequencing for mutation analysis, for the identification of structural genomic rearrangements and for transcriptomic profiling.

Collaborators: Paul Pharoah (Strangeways Research Laboratory, Cambridge), Simon Tavaré (CRI), Helena Earl (Department of Oncology and Addenbrooke's Hospital), Gordon Wishart, Elena Provenzano (Addenbrooke's Hospital), Sam Aparicio (University of British Columbia), Paul Edwards (Hutchison/MRC Research Centre, Cambridge), and Ian Ellis (University of Nottingham)

Functional breast cancer genomics: characterising cancer stem cells in breast cancer subtypes

We have continued our work to identify and characterise which cells initiate different types of breast cancer and what is the nature of the pathways that are disrupted. In breast cancer cell lines, using the mammosphere assay as a read out for cancer stem cells, our data show that there is no uniform set of markers that can be universally applied. We are now testing different sub-populations for their 'stemness' using *in vivo* assays and will then characterise the transcriptome of these cells. We will soon initiate experiments with primary human tumour material. We are particularly interested in the role of miRNAs, TGF β and NOTCH signalling in the regulation of normal and malignant breast stem cells. We are testing whether disruption of these pathways alters the self renewal of cancer stem cells, which could lead to novel therapeutic approaches to eradicate tumours.

Collaborators: Eric Miska (Wellcome Trust/Cancer Research UK Gurdon Institute); John Stingl, Jason Carroll and Bruce Ponder (CRI); and Shankar Balasubramanian (Department of Chemistry, University of Cambridge)

Publications listed on page 56