

Functional Breast Cancer Genomics

www.cambridgecancer.org.uk/carlosaldas

Group Leader **Carlos Caldas**

Clinician Scientist Fellow

John LeQuesne

Clinical Fellows

Nikita Makretsov

Stefanie Sassen

Clinical Training Fellows

Jean Abraham

Mahesh Iddawela

Sarah Jane Dawson

Postdoctoral Scientists

Alejandra Bruna

Suet-Feung Chin

Anna Git

Andrew Teschendorff

Bioinformaticians

Bin Liu

Lorna Morris

Irene Papatheodorou

Graduate Students

Daniel Holland

Gemma Sharp

Scientific Officers

Donna Grimmer

Inma Spiteri

Summer Placement Student

Lin Shao

Research Assistant

Helen Bardwell

Breast cancer remains a public health problem with over 40,000 new cases diagnosed every year in the UK. The molecular classification derived using modern genomics technology has highlighted the great heterogeneity of breast cancer, and this has profound clinical and biological implications.

This molecular heterogeneity needs to be taken into account to develop prognostic and predictive markers with clinical utility, to rationally stratify breast cancer patients into trials of novel targeted agents, and also to help characterisation of cancer stem cells from human breast tumours. A key question is whether all breast cancers start from the same cell type (presumably the breast stem cell), with distinct initiating mutational events leading to diverse subtypes, or whether the molecular heterogeneity results from different cells of origin.

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

We have contributed to the current understanding of the complex taxonomy of breast cancer by characterizing a common amplicon at 8p12 which appears to predominate in one subtype; we have also described novel breast cancer

subtypes, and showed that microRNA signatures distinguish subtypes. The 8p12 locus we characterised is one of the most common (and narrowest) amplifications in breast cancer and the future identification of the driver oncogene will have biological and clinical implications. PACK (profile analysis using clustering and kurtosis), an algorithm we developed, has unravelled a novel sub-classification of ER-breast cancer. Specifically, clustering over PACK-selected genes identified five different subtypes (CC+, CC+/IR+, IR+, ECM+, and SR+) characterised by the over-expression patterns of four distinct gene clusters, each enriched for inflammatory response (IR), extra-cellular matrix (ECM), cell cycle (CC), and steroid response (SR) genes, respectively. This analysis also showed that the previously called basal breast cancers are a heterogeneous group with at least four distinct subtypes (CC+, CC+/IR+, ECM+, and IR+). We also reported that microRNAs classify breast cancers into different molecular subtypes suggesting microRNA profiling might supersede mRNA profiling.

We have tested novel ways of identifying prognostic classifiers given that there are several known problems that hinder the identification of gene-expression signatures that validate across multiple data sets. One difficulty is that a large number of measurements and therefore potential correlations are derived. Another concern relates to the way cohorts are categorized into good and bad outcomes, which is usually performed by some rather arbitrary threshold, thus introducing significant study-specific bias. A natural way to circumvent this problem is to treat outcome as a continuous variable when ranking genes, and to use unsupervised clustering over these genes

to identify, without bias, subgroups of patients with different outcome. Using this approach we have identified prognostic classifiers that perform well in independent external validation. This work has also showed that while in ER+ breast cancer prognostic markers are associated mainly with cell cycle pathways; in ER- disease prognostic markers are associated with immune response pathways.

Our future plans are to complete a comprehensive multi-modality genomics profiling study – using array-CGH, mRNA and miRNA expression profiling and mutation analysis – to discover how many different types of breast cancer exist and to then test the clinical utility of the different subtypes for prognostication and prediction in large sample collections we have accrued. Molecular subtypes will also be used for patient stratification into prospective trials of targeted therapy. The following questions will be addressed: Are microRNAs robust subtype markers with the potential for use in paraffin-embedded tissue? Is the 8p12 amplicon a robust marker of luminal B tumours? What is the correlation between somatic mutations and breast cancer subtypes? Can histology-based markers that mirror the genomics classifiers be used to classify breast cancers?

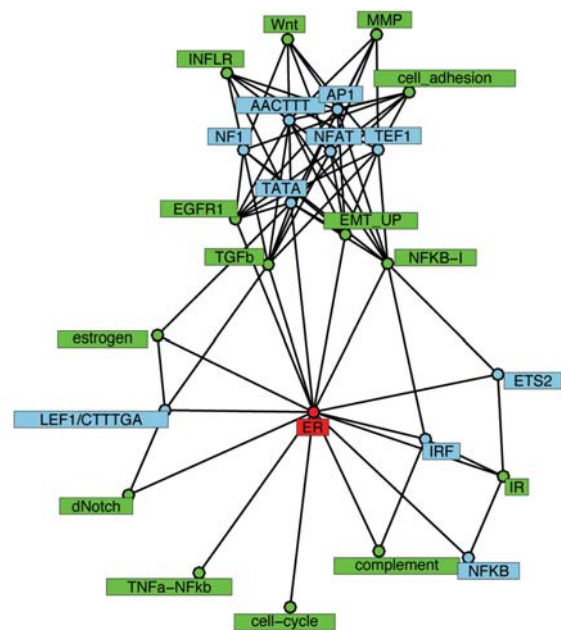
Collaborators: Paul Pharoah (Department of Oncology, University of Cambridge), Helena Earl (Department of Oncology and Addenbrooke's Hospital), Gordon Wishart (Addenbrooke's Hospital), Ian Ellis (University of Nottingham), Simon Tavaré, James Brenton and Bruce Ponder (CRI).

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

Breast cancer is not one disease, it is many different diseases, and appropriate models need to take this into account. Progress will require isolating and characterizing tumour initiating cells from different tumour types. The initiating mutation events will probably differ between subtypes and hence their functional characterization will require developing specific distinct models. In particular we want to isolate cancer stem cells from tumours/cell lines with 8p12 amplification to help characterize the driving oncogene and from luminal and basal tumours to address the role of microRNAs in cancers with luminal versus basal differentiation. An extremely interesting observation which might have profound biological implications is that most primary tumours with 8p12 amplification (high-level copy number gain) in our series are luminal B type. This suggests that amplification of this locus either drives tumours down the luminal B differentiation pathway or that only luminal B committed cells get a selective advantage with the amplicon. The roles of microRNAs in fine tuning gene expression and canalization of development hints they might also have a central role in regulating normal and cancer stem cell plasticity. Our unpublished data reveals highly regulated microRNA expression during adult murine

mammary development, differential expression of microRNAs in the normal breast epithelial hierarchy in both humans and mice, and differential expression in basal versus luminal tumours and cell lines.

We are using flow cytometry in combination with functional assays to detect and purify tumour stem cells. The assays to detect and enumerate the tumour stem cells are a novel *in vivo* xenotransplant assay and the *in vitro* mammosphere assay. The mammosphere assay we optimized using MCF7 cells is a relatively easy and reproducible method to study cancer stem cells and we now use it routinely with breast cancer cell lines and pleural effusions from patients. We are using established methodology to functionally identify the driver oncogene at 8p12 by targeting the three candidate oncogenes in both knockdown and transformation assays. For the microRNA experiments we use LNA/2'-O-methyl oligonucleotide mixmers, Peptide Nucleic Acids (PNA) and PNA-Peptide conjugates for downregulation and murine



Average association network for ER status derived from gene expression data. Only edges between phenotype, pathways, and transcription factors are shown. The diagram is colour-coded as follows: phenotype (red), pathways (green), and transcription factors/binding motifs (blue). Figure modified from: Teschendorff et. al., *PLoS Comput Biol* 2007; 3:e161.

stem cell virus (MSCV) vectors to direct expression of single or cluster microRNAs.

Collaborators: Eric Miska (Wellcome Trust/Cancer Research UK Gurdon Institute), John Stingl (joins the CRI in January 2008), Mike Gait (MRC Laboratory of Molecular Biology, Cambridge) and Shankar Balasubramanian (Department of Chemistry, University of Cambridge).