Integrated Analysis of Copy Number Aberrations for Primary Breast Cancer Tumours in the HER2 region, miRNA and mRNA expression

Laura Winchester[1], Antoine De Weck[2], Jiannis Ragoussis[2], Adrian L Harris[1] and Francesca Buffa[1]

[1]*Molecular Oncology Laboratories, Department of Medical Oncology, University of Oxford, Weatherall Institute of Molecular Medicine, Oxford, OX3 9DS, UK*
[2]*Genomics Research, Wellcome Trust Centre For Human Genetics, University of Oxford, Oxford, OX3 7BN*

In this study we illustrate a system approach to genomic analysis where data from DNA, mRNA and miRNA are integrated and used together to enable more powerful inferences about genomic changes, and better understanding of their functional consequences and clinical implications. We chose a specific example relevant for breast cancer management.

One of the causes of Breast Cancer (BC) development is the accumulation of somatic and inherited genetic alterations. Somatic alterations or Copy Number Aberrations (CNA) are amplifications or deletions of genomic DNA. Analysis of CNA in BC is confounded by tumour heterogeneity where content is a mixture of genetically different cells with different phenotypic effects. Specifically, it is difficult to distinguish between a change in cancer cell copy number and a stromal cell contamination effect. Here, we apply a recently developed method, OncoSNP, which accounts for cell mixture in a tumour sample and successfully predicts CNA in the mutated sample. We apply results from other experiments to inform subtypes and biological heterogeneity.

The HER2 region contains a cluster of genes including *ERBB2,* which are reported to be over expressed in a subset of BC tumours and have a negative effect on patient prognosis. *ERBB2* encodes for an epidermal growth factor receptor responsible for cell development through enhancement of cellular function signaling pathways. HER2 status is used as a diagnostic test to assess targeted therapy options.

DNA and total RNA was extracted from 207 frozen breast cancer samples collected from 2000-2005. Illumina SNP arrays were used for copy number identification incorporating the OncoSNP detection algorithm[1]. Samples were assessed for contaminating cell content and highly heterogeneous samples were eliminated. Other outliers with increased overall CN were flagged as potential polyploidy samples. Illumina microarrays were used to generate mRNA and miRNA expression data. An integration of results from the miRNA and mRNA arrays has been published separately[2]. Distant relapse-free survival (DRFS) was calculated for each patient as defined by the STEEP criteria. Immunohistochemistry (IHC) classification was used to confirm amplification status.

The HER2 region contained amplifications in 26.8% of the sample set and deletions in 5.5%. From the subgroup of HER2 deletions, we show reduced HER2 expression has an adverse effect on patient survival. There were significant correlations between the copy number change predictions, mRNA HER2 expression (Spearman correlation rho = 0.473, p-value = 5.083×10^{-12}) and HER2 3+ IHC (rho = 0.471, p-value = 1.37×10^{-07}). In addition to comparisons of mRNA and miRNA expression, each miRNA was tested for correlation to HER2 copy number. Copy number predicted HER2+ samples were anti-correlated with

lower expression of hsa-miRNA-10b and let-7c (rho = -0.277, p-value = $8.06x10^{-05}$ and rho = -0.295, p-value = $2.62x10^{-05}$). The highly correlated miR-210 and miR-130b were part of the prognostic group studied in the previously published miRNA and mRNA integration (rho = 0.317, p-value = $6.10x10^{-06}$ and rho = 0.302, p-value = $1.70x10^{-05}$).

Using the mRNA data we compared the expression of all the genes in the HER2 region. In addition to *ERBB2,* other genes including *STARD3*, C17orf37 and *GRB7* were highly correlated to copy number for the region. Although *NEUROD2* is disrupted by similar CNA, it has a lower correlation suggesting the gene is not expressed as part of the HER2 region.

We have demonstrated the use of parallel datasets to validate and confirm hypothesis based on the CNA predictions from the SNP array dataset. For example, the correlation of mRNA HER2 expression and copy number state, confirming the increase in CNA is linked to the increase in gene expression. In addition, the SNP array results have been confirmed by IHC giving increased confidence in the validity of CNA calls, which gives increased weighting to genome–wide predictions where IHC data is not available.

Using a unified approach we have integrated data from DNA, mRNA and miRNA array datasets to analyse the variability of the effects of HER2 amplification on clinical outcome. We have detected miRNAs that correlate with HER2 amplification which are also in common with the prognostic miRNAs identified in miRNA – mRNA integration study (miR-210 and miR-130b)[2]. By looking at the three array datasets we can make inferences about changes in genomic material and potential impact on patient outcome.

Reference
(1) Yau *et al.* Genome Biol. 2010 11(9);
(2) Buffa *et al.* Cancer Res September 1, 2011 71; 5635