

The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups

Christina Curtis^{1,2†*}, Sohrab P. Shah^{3,4*}, Suet-Feung Chin^{1,2*}, Gulisa Turashvili^{3,4*}, Oscar M. Rueda^{1,2}, Mark J. Dunning², Doug Speed^{2,5†}, Andy G. Lynch^{1,2}, Shamith Samarajiwa^{1,2}, Yinyin Yuan^{1,2}, Stefan Gräf^{1,2}, Gavin Ha³, Gholamreza Haffari³, Ali Bashashati³, Roslin Russell², Steven McKinney^{3,4}, METABRIC Group†, Anita Langerød⁶, Andrew Green⁷, Elena Provenzano⁸, Gordon Wishart⁸, Sarah Pinder⁹, Peter Watson^{3,4,10}, Florian Markowetz^{1,2}, Leigh Murphy¹⁰, Ian Ellis⁷, Arnie Purushotham^{9,11}, Anne-Lise Børresen-Dale^{6,12}, James D. Brenton^{2,13}, Simon Tavaré^{1,2,5,14}, Carlos Caldas^{1,2,8,13} & Samuel Aparicio^{3,4}

The elucidation of breast cancer subgroups and their molecular drivers requires integrated views of the genome and transcriptome from representative numbers of patients. We present an integrated analysis of copy number and gene expression in a discovery and validation set of 997 and 995 primary breast tumours, respectively, with long-term clinical follow-up. Inherited variants (copy number variants and single nucleotide polymorphisms) and acquired somatic copy number aberrations (CNAs) were associated with expression in ~40% of genes, with the landscape dominated by *cis*- and *trans*-acting CNAs. By delineating expression outlier genes driven in *cis* by CNAs, we identified putative cancer genes, including deletions in *PPP2R2A*, *MTAP* and *MAP2K4*. Unsupervised analysis of paired DNA–RNA profiles revealed novel subgroups with distinct clinical outcomes, which reproduced in the validation cohort. These include a high-risk, oestrogen-receptor-positive 11q13/14 *cis*-acting subgroup and a favourable prognosis subgroup devoid of CNAs. *Trans*-acting aberration hotspots were found to modulate subgroup-specific gene networks, including a TCR deletion-mediated adaptive immune response in the ‘CNA-devoid’ subgroup and a basal-specific chromosome 5 deletion-associated mitotic network. Our results provide a novel molecular stratification of the breast cancer population, derived from the impact of somatic CNAs on the transcriptome.