

Array-CGH profiling

We used human *Integrachip V2* to establish genomic profiles (IntegraGen SA, Evry, France, <http://www.integragen.com/>). IntegraChip V2 is composed of 3172 BAC clones including 2862 sequenced clones with a median gap of 1 clone/875 Kb. BAC clones are spotted in quadruplicate. IntegraGen SA provides preblocked slides. Genomic DNA (300 ng) was labeled by random-priming in a 45µl reaction containing: 0.12mM dATP, 0.12mM dGTP, 0.12mM dTTP; 0.04mM dCTP; 0.05mM Cy3-dCTP (tumor samples) or Cy5-dCTP (normal sample: pool of 20 normal reference female DNAs); 18 Units of Klenow Fragment exo- (5U/µl, Euromedex, Souffelweyersheim, France), 10 mM β-mercaptoethanol, 5mM MgCl₂, 50mM Tris-HCL pH 6.8 and 110µg/ml random octamers. The reaction was incubated at 37°C for 20 h and stopped by adding 2.5µl EDTA 0.5M pH8. We purified labeled products using microcon 30 filters (Millipore, Molsheim, France). Abundance of the labeled DNA is checked using a spectrophotometer and incorporation of dyes is calculated. A mix of 700 pmol Cy5 and 700 pmol Cy3 labeled probes was ethanol precipitated in the presence of 250µg of human Cot-1 DNA (Roche Diagnostics, Meylan, France) and 100µg herring sperm DNA (Promega, Charbonnières, France). The pellet was dried and resuspended in 110µl Hybridization buffer composed of 50% desionised formamide, 2XSSC, 10% Dextran Sulfate. The probes were denatured at 80°C for 10 min, and repetitive sequences were blocked by pre-annealing at 37°C for 30 min. Slide processing was done using a HS4800 hybridization station (Tecan, Lyon, France). Preannealed probes were injected in the chambers and hybridization took place at 37°C for 16h with mild agitation. Posthybridization washes were as follows: 3 washes at 52.5°C in solution 1 (2X SSC, 0.2%SDS), followed by 3 washes in solution 2 (0.5X SSC, 0.2%SDS) and 1 wash in solution 3 (0.5X SSC) at 52.5°C. To remove salt and detergent residues a brief wash with dd water was performed at 37°C and slides were dried within the chambers by an injection of N₂ at 30°C. Arrays were then scanned with an Axon 4000B scanner (Dipsi, Chatillon, France).

Array-CGH image processing

Images were analyzed using the Genepix 6.0 software. Spots were defined by use of the automatic grid feature of the software. Normalization, replicate filtration, segmentation and graphical representation were done using CGH-Array Analysis

Platform (CAPWeb, <http://bioinfo-out.curie.fr/CAPweb/>). Genomic clones were positioned according to the Build 36 of the human genome sequence (Hg18). Clones with missing values in over 50% of the tumors were discarded. Gains and losses were defined at 0.25 and -0.25 as \log_2 ratio thresholds.

Analysis of cDNA microarray data

Gene filtering before hierarchical clustering

Before clustering, we applied a gene filtering to the 8305 Image clones present on our microarrays to exclude genes with low variation in expression level and genes poorly measured across all 50 samples. Inclusion criteria were expression present in at least 80% of the samples and with standard deviation >0.1 . By applying these filtering, we conserved in the final analysis 7782 genes/ESTs.

Identification of molecular subtypes of tumors

Five biologically relevant tumor subtypes (luminal A, luminal B, basal, *ERBB2*+, and normal-like) have been identified using an intrinsic set of ~500 genes in DBC (Sorlie et al., 2001). We looked the distribution of these subtypes in our series of IDCs and ILCs by using the 169 common to our gene set and the Stanford/Norway intrinsic 500-gene set.

Fisrt, we verified that this 169-gene set was sufficient to identify biologically relevant subtypes. Based on a hierarchical clustering applied to these genes and the 122 Sorlie and coworkers' samples (Sorlie et al., 2003), we defined five sets of centroids representing the median expression of each subtype. By measuring the correlation of each of the 122 samples with each centroid, we assigned a molecular subtype to each sample, and observed 92% of concordance between these subtypes and those defined on the same samples by using the whole intrinsic gene set. This result suggested that the 169-gene set was sufficient to identify these five biologically relevant subtypes.

We next measured the correlation of each of our 50 tumor samples with each centroid. This allowed assigning each tumor to a molecular subtype (Figure 4B; Supplementary Table 1): 10 samples (5 IDCs and 5 ILCs) were closer to the luminal A centroid, 7 closer to the luminal B centroid (7 IDCs), 4 closer to the basal centroid (3 IDCs and 1 ILC), 3 closer to the *ERBB2* centroid (2 IDCs and 1 ILC) and 12 closer to the normal-like centroid (4 IDCs and 8 ILCs). Fourteen samples (8 IDCs and 6

ILCs) displayed a correlation inferior to 0.15 with any centroid and were not attributed any subtype with this 169-gene set.