**Protocols**

**Molecular iodine: chemoresistance prevention and beneficial effects in a randomized breast cancer trial.**

**Abstract**

The present work analyzes the antineoplastic effects of molecular iodine (I2) alone and in combination with the neoadjuvant therapy FEC/TE (5-fluorouracil/epirubicin/cyclophosphamide or taxotere/epirubicin) in women with breast cancer. In this protocol the immunohistochemistry methodology for the quantification of the estrogen receptor and the proteins associated with the installation of the mesenchymal epithelium transition (e-cadherin and vimentin), immune response (CD8) and cell death (TUNEL) is described, and the methodology for transcriptomic analyses includes: RNA-Seq and transcriptomic analysis; pathway enrichment analysis; validation of the immune response and epithelium/stroma rate and, prediction of the transcription factor regulatory network and peroxisome proliferator-activated receptors type gamma (PPARγ) interaction

**Immunohistochemistry.**

Immunohistochemistry was performed on 3 or 5 µm-thick sections of formalin-fixed paraffin-embedded specimens that were float-mounted on silanized glass slides. The samples were dewaxed in xylene and rehydrated with a series of decreasing alcohol concentrations to water. All samples were subjected to antigen retrieval (10 mM sodium citrate pH 6 or 0.05% Citraconic Anhydride Buffer pH 7.4 (Sigma-Aldrich, St Louis, MO, USA) at 80°C for 10–25 minutes using a conventional pressure cooker. Non-specific binding in the sections was blocked with 3% (wt/vol) BSA for 30 minutes at 37ºC. Incubation with primary antibody was done overnight at 4°C. The primary antibodies included anti-ERα (polyclonal, diluted 1:200, HC-20 Santa Cruz Biotechnology, Dallas, Texas, USA), anti-CD8 (monoclonal, diluted 1:200, SC-7970 Santa Cruz Biotechnology), anti-E-cad (polyclonal, diluted 1:50, H-108 Santa Cruz Biotechnology), and Vim (monoclonal, ready to use, VIM 3B4 Dako Carpinteria, CA, USA). Incubation without primary antibody served as negative control.

For ERα and CD8, detection was carried out using an EnVision Detection System kit (Agilent Dako, Carpinteria, CA) based on the use of secondary antibodies conjugated to horseradish peroxidase-labelled polymers. Sections were counterstained with Mayer’s hematoxylin, rinsed, dehydrated in a graded alcohol series, mounted with Entellan (Merck, Darmstadt, Germany), and cover slipped. ERα-positive cells were identified by the presence of a brown stain over the nucleus and/or cytoplasm. CD8+ lymphocytes were identified by the presence of brown stain in the cytoplasm. Five regions were analyzed, and the labeling indices were expressed as the percentage of labeled cells among the total number of cells per region. Tumor samples were considered as ERα-positive when 20 or more cells were positively labeled in the nucleus, cytoplasm or both and were found in at least two fields.

For E-cad and Vim, preparation of the slides was carried out for visualization by confocal microscopy (Zeiss LSM 780). After antigen retrieval, slides were incubated for 20 minutes in 70% ethanol with 0.1% Sudan Black B (Sigma-Aldrich) to reduce autofluorescence based on the results by Baschong et al. (1). Detection of the primary antibodies was carried out using Alexa 546-conjugated goat anti-rabbit and Alexa 647-conjugated goat anti-mouse secondary antibodies. The nuclei were observed using DAPI (diluted 1:100, Life Technologies). The slides were mounted using 4% antifade reagent (4% n-propyl gallate, 1% DABCO) and observed at 630 magnifications (objective Plan-Apochromat 63x/1.40 Oil DIC M27). Samples were excited at 750 nm (0.9% laser), 633 nm (1.0% laser), 561 nm (1.0% laser), and 488 nm (1.0% laser), and a Z-stack compilation was taken for each slide. A total of four different fields were analyzed for each tumor sample. Quantification was performed by counting all of the epithelial cells present in one field (Z-stack projection), and the expression of the different markers (E-cad and Vim) is reported as the percentage of positively stained epithelial cells in the field. Positive E-cad expression was defined as a positive staining in the membrane, cytoplasm, or nucleus. Positive Vim expression was defined as positive staining in the cytoplasm of epithelial or stromal cells. Co-expression of E-cad + Vim were defined as positive staining in the cytoplasm of only epithelial cells. The final percentage of expression was reported as the mean of the four different fields randomly taken for each tumor.

Cell death was detected using the in-situ fluorescein/POD cell detection kit (Roche Molecular Biochemicals, Mannheim, Germany), which is based on the terminal deoxynucleotidyl transferase-mediated dUTP nick-end labeling (TUNEL) method. TUNEL-positive cells were identified by a brown stain over the nucleus. The final percentage of expression was reported as the mean of the four different fields randomly taken for each tumor.

**Transcriptomic analysis**

*RNA-Seq and transcriptomic analysis.* Total RNA was extracted with Qiazol and RNeasy (both from Qiagen). Two different pools of four individual tumor samples were used. As a normal control, we used a pool of two normal mammary gland samples from aesthetic surgeries (volume reduction). Poly-A enriched mRNA was used to construct stranded mRNA-Seq libraries following the manufacturer's instructions (KAPA Biosystems). Sequencing was carried out at Duke University Genome Sequencing Shared Resource Center (Durham, NC). The libraries were sequenced on an Illumina HiSeq 2500 platform in which 101 bases were determined in pair end mode. Data were assessed for quality and filtered with FastQC and Trimmomatic, respectively. Filtered reads were mapped to the human genome (GRCh38), and expression levels were determined by htseq-count. Differential expression analysis was performed using Fisher's exact and Benjamini-Hochberg (FDR) tests. Genes that were altered at least 2-fold or less than 0.5-fold with an FDR value equal or lower than 0.05 were considered biological and statistically significant. The full annotated sequences from the RNQ-sequencing are available at the European Nucleotides Archives web site (https://www.ebi.ac.uk/ena/submit/accession-number-formats) as erp110028.

*Unsupervised clustering and heat map visualization.* Hierarchical clustering based on Pearson distance and the war.D2 method was carried out according to the expression levels of genes with an FDR ≤ 0.05 for each treatment. Clustering and heat map construction were performed using the Bioconductor ComplexHeatmap R package.

*Pathway enrichment analysis.* Pathway and Gene Ontology (GO) functional enrichment analyses were carried out on the sets of DEGs (genes with 1 ≤ Log2FC ≤ -1; FDR ≤ 0.05) using the ToppGene suite (ToppFun) with the KEGG 2016 database (2). Pathways with a minimum of 5 genes and an FDR ≤ 0.001 were considered significant**.**

*EMT signature.* A set of EMT genes defined in *homo sapiens* was obtained from the GSEA MSigDB (3). Using our transcriptomic data, we focused on differentially expressed EMT genes that had an FDR ≤ 0.05 and used unsupervised clustering to analyze our datasets.

*Immune cell abundance estimation.* Using our transcriptomic data, immune and stromal cell abundance in the different treatment groups was calculated using the MCP-counter method available as an R package (4).

*Th1 and Th2 differentiation genes.* Genes known to be involved in CD4+ T cell differentiation towards Th1 or Th2 cells were obtained from public datasets (KEGG hsa04658, R&D systems Pathways) and were analyzed in our DEG sets.

*Validation of the immune response and epithelium/stroma rate.* The effects of I2, FEC/TE, or both (FEC/TE + I2) on the immune response and epithelial/stromal tumor ratio were evaluated. The number of tumor lymphocytes and CD8+ lymphocytes was estimated by two independent observers (XZ and CA) in anonymized and blinded samples using hematoxylin and eosin (H&E) and immunohistochemistry staining, respectively. The epithelial and stromal ratios were quantified in the final tumor tissue with Masson trichrome staining (Poly-sciences, Inc., Warrington, PA). Quantification was performed using ImageJ 1.47 (Wayne Rosband. NIH, USA). Three random regions (40X) were used per analysis.

*Prediction of the transcription factor regulatory network and PPARγ interaction.* The transcription factors involved in regulating the common DEGs in I2-treated tumors (I2 and FEC/TE+I2) were predicted using the iRegulon Cytoscape plugin (5). In this analysis, the transcription factors are predicted based on ChIPseq data as well as motif enrichment from different data collections, including ENCODE, TRANSFAC, and JASPAR. A functional interaction network between PPARγ and the top transcription factors from this analysis was then established using the Reactome FI Cytoscape plugin. Other possible relationships were established by Chilibot, a data-mining tool from Chen and Sharp (6).

*PPARγ-correlated genes in breast cancer.* Genes having a positive correlation with PPARγ (Pearson ≥ 0.3) in 2,506 breast cancer patients were obtained from cBioPortal, which uses Breast Cancer METABRIC data (7-9). In our analysis, we focused on DEGs from this subset that had an FDR < 0.05 and used unsupervised clustering to analyze our datasets.

**Bibliography**

1. Baschong W, Suetterlin R, Laeng RH. Control of autofluorescence of archival formaldehyde-fixed, paraffin-embedded tissue in confocal laser scanning microscopy (CLSM). *J Histochem Cytochem*. 2001;49(12):1565–1572.
2. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* 2009;37:W305–W311.
3. Ksiązkiewicz M, Markiewicz A, Zaczek AJ. Epithelial-mesenchymal transition: a hallmark in metastasis formation linking circulating tumor cells and cancer stem cells. *Pathobiology.* 2012;79(4):195–208.
4. Becht E, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression*. Genome Biol*. 2016;17(1):218–238.
5. Janky R, et al. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol*. 2014;10(7):e1003731.
6. Chen H, Sharp BM. Content-rich biological network constructed by mining PubMed abstracts. *BMC Bioinformatics.* 2004;5:147.
7. Pereira B, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun.* 2016;7:11479.
8. Gao J, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013;6(269):pl1.
9. Cerami E, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401–404.