Supplement: Immune and genetic signature of breast carcinomas triggering anti-Yo associated paraneoplastic cerebellar degeneration

1. eMethods

1.1. Immunohistochemical (IHC) investigations

IHC staining for classical diagnostic markers (p53, WT1) and oestrogen receptors, but also for progesterone receptors and HER2 was performed using a routine protocol, respectively on a BenchMark ULTRA Diagnostics (Roche, Meylan, France) and a PT Link Autostainer and Autostainer Link 48 (DAKO, Les Ulis, France) in the Biopathology department of the Centre Léon Bérard (Lyon, France).

1.2. Immunofluorescence investigations

Four µm FFPE tumour sections were deparaffinised, rehydrated and antigen retrieval treatment was performed using ER1 (for CD3, CD20, IgA and IgG) or ER2 (for DClamp and CK) buffer. The sections were sequentially stained with each primary antibody, followed by OPAL-HRP secondary antibody incubation then revealed with tyramide signal amplification and OPAL fluorophore, in the following order : DClamp (clone 1010E1.01; Dendritics; 1/100 dilution)/OPAL 520; CD20 (clone L26; DAKO; 1/600 dilution)/OPAL 690; IgG (PC; DAKO; 1/5000 dilution)/OPAL 620; IgA (PC; DAKO; 1/4000 dilution)/OPAL 570 ; CD3 (PC; DAKO; 1/60 dilution)/OPAL 480 and CK (clone AE1/AE3; DAKO; 1/100 dilution)/OPAL 780. The sections were then counterstained with spectral DAPI (Akoya Biosciences, Malborough, MA, USA) and mounted with a coverslip. Whole slides were imaged at a 20x magnification using the Vectra Polaris multispectral scanner (Akoya Biosciences) and digital images were visualised with the Phenochart viewer (Akoya Biosciences) and unmixed using the spectral library from the software.

1.3. Sampling and DNA extraction

Ten out of the 22 collected Yo-PCD breast tumour samples were eligible for a CGH-array analysis (4 were incubated in Bouin solution, 2 did not have a sufficient quantity of tumour material available, 2 had a too low DNA concentration). Four µm FFPE tumour sections were stained with haemotoxylin, phloxine and saffron (HPS), and the proportion (%) of tumour cells was determined by a trained pathologist (IT, Centre Léon Bérard). Tumour tissue was collected by scrapping 20 unstained 8-µm

FFPE sections. DNA extraction was performed using the QIAamp DNA Micro kit (Qiagen,, Manchester, UK) according to the manufacturer's instructions, and DNA was eluted in 25µl of water. DNA concentration was measured using Qubit® 2.0 fluorometer and the quality of DNA was checked using a Bioanalyzer 2200 TapeStation system (PerkinElmer, Villebon S/Yvette, France).

1.4. CGHa

Sample labelling

Using 1.5 µg of tumour genomic DNA and 1.5 µg of male reference DNA (Promega, Charbonniere, France), fragmentation and labelling were performed according to the manufacturer's recommendations for the CGH arrays (Agilent Technologies, Santa Clara, CA, USA) with some modifications. In brief, reference and tumour DNA were heat-denatured and fragmented for 10 min at 95°C. Then, tumour DNA was chemically labelled with a Cy5-dye using the Universal Linkage System (ULS™; Agilent Technologies), whereas reference DNA was labelled with Cy3-dye. The labelled samples were then purified using Agilent KREApure columns (Agilent Technologies). Labelling efficiency was calculated using Nanodrop ND2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), measuring A260 (DNA), A550 (Cy3) and A649 (Cy5).

Hybridisation, scanning, and data extraction

Co-hybridisation was performed on 4x180K Agilent SurePrint G3 Human whole-genome oligonucleotide arrays (Agilent Technologies) at 65°C for 24 hours, in a rotating oven at 20 rpm. Slides were then washed with Oligo aCGH Wash Buffer 1 at room temperature for 5 min followed by 1 min wash with Oligo aCGH Wash Buffer 2 at 37°C. Finally, slides were dried and scanned using Agilent Surescan scanner with 3 µm resolution. Scan images were processed using Agilent Feature Extraction Software V11.0 and the analysis was carried out using the Agilent Genomic Workbench software V7.0. The identification of aberrant copy number segments was based on ADM-2 segmentation algorithm with a threshold of 7. Basically, chromosomal loci were considered as amplified if log2 ratio was ≥ 1.6, and were considered as homozygously deleted if log2 ratio was ≤ -1. However, these thresholds were manually adapted according to the proportion (%) of tumour cells, and a copy number was estimated for *CDR2* and *CDR2L* genes. Using this approach, a gene gain corresponds to a copy number of 3, a high level gain to a copy number > 3.

Quality

The mean proportion (%) of tumour cells by sample was 68 %. The mean Derivative Log Ratio Spread of the samples was 0.41.

1.5. RNA Sequencing

Total RNA was extracted from macrodissected FFPE tumour sections using the FormaPure RNA kit (Beckman Coulter, Brea, CA, USA). RNase-free DNase set (Qiagen) was used to remove DNA. RNA quantification was assessed using NanoDrop 2000 (Thermo Fisher Scientific) measurement and RNA quality using the DV200 value (the proportion of the RNA fragments larger than 200nt) assessed by a TapeStation with Hs RNA Screen Tape (Agilent Technologies). Samples with sufficient RNA quantity (> 0.5 µg) and quality (DV200 > 30%) were further analysed by RNA sequencing. One hundred ng of total RNA was used to prepare libraries with TruSeq RNA Exome (Illumina, San Diego, USA). 12 libraries were pooled at a concentration of 4nM each together with 1% PhiX. Sequencing was performed (paired end, 2 × 75 cycles) using NextSeq 500/550 High Output V2 kit on a NextSeq 500 machine (Illumina). The mean number of reads per sample was around 80 million. Alignments were performed using STAR on the GRCh38 version of the human reference genome. The number of duplicate reads was assessed using PICARD tools. Samples with a number of unique reads below 10 million (5 million paired-reads) were discarded from the analysis. Expression values were extracted using Kallisto version 0.42.5 tool17 with GENECODE release 23-genome annotation based on GRCh38 genome reference.

1.6. FISH analysis

Copy number alterations of the *CDR2L* gene were assessed using a dual colour probe where the *CDR2L* gene is labelled in orange, and the centromere of chromosome 17 is labelled in green (Empire Genomics). FISH was performed on 4-µm sections of FFPE tissue, using the ZytoLight FISH-Tissue Implementation Kit (Zytovision, Bremerhaven, Germany) as per manufacturer's instructions. Regarding the enumeration of FISH signals, at least 40 non-overlapping intact nuclei were investigated. The *CDR2L* gene was considered amplified if the mean number of orange signals was ≥ 6.

1.7. Bioinformatical analysis

Gene fusion and variant analysis

Gene fusions were investigated with five different algorithms, STAR-Fusion, FusionMap, FusionCatcher, TopHat-Fusion and EricScript. Variant calling was performed using the GATK HaplotypeCaller[13], and filtering was achieved with the ANNOVAR Annotation tool.[14]

Differential genes expression analysis

Differential genes expression analysis was performed using *DESeq2* v1.32.0[1] R package by comparing genes expression over the two cohort, Yo-PCD vs Control. '*DESeq'* function was used with default values to fit the model and '*lfcShrink'* was used using the '*apeglm'* method to obtained the shrunken

log2 fold changes. P-values were adjusted using the Benjamini & Hochberg method. Genes with an adjusted p-value ≤0.001 and an absolute log Fold-Change ≥1 were defined as differentially expressed.

Hierarchical clustering

Kallisto TPM expression values were transformed in log10 (TPM+0.01), and transformed into a Z-score. Hierarchical clustering was performed with the *ComplexHeatmap* v2.8.0[2] R package using the Euclidean distance and the Ward clustering method.

Gene Ontology enrichment

Gene Ontology enrichment was performed using *clusterProfiler* v4.0.0[3] R package on Yo-PCD overexpressed and under-expressed genes separately and all genes used for differential genes expression were used as background. Gene Ontology Biological Process with a Benjamini & Hochberg adjusted p-value ≤0.05 and a q-value ≤0.05 were considered as significantly enriched. Raw results of significant annotations are available in Supplementary Table 1.

2. eTables

2.1. eTable 1

eTable 1 Results of DESeq2 on the differential gene expression analysis between Yo-PCD and control tumours for all genes and for differentially expressed genes (available as online Supplementary Information: https://zenodo.org/record/6477807, doi: 10.5281/zenodo.6477807)

2.2. eTable 2

eTable 2 Results of the Gene Ontology enrichment performed by clusterProfiler on genes over-expressed in Yo-PCD tumours (available as online Supplementary Information: https://zenodo.org/record/6477807, doi: 10.5281/zenodo.6477807)

2.3. eTable 3

eTable 3 Results of the Gene Ontology enrichment performed by clusterProfiler on genes under-expressed in Yo-PCD tumours (available as online Supplementary Information: https://zenodo.org/record/6477807, doi: 10.5281/zenodo.6477807)

REFERENCES

1. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1-21.

2. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847-2849.

3. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J Integr Biol*. 2012;16(5):284-287.