**Supplementary Materials and Methods**

**Patient characteristics and generation of PDX**

PDX models were generated as previously described. Informed consent for PDX generation and whole exome sequencing was obtained from all subjects and the study received approval from the corresponding institutional Ethics Committee in accordance with the Declaration of Helsinki. All research involving animals was performed at the IDIBELL animal facility in compliance with protocols approved by the Institutional Committees on Animal Care and adhering to national and European Union regulations. Findings from the IDB models were evaluated in an interim analysis of two ongoing PDX-based preclinical chemotherapy trials being conducted at Baylor College of Medicine to be described in full elsewhere upon completion. A summary of all the TNBC PDX models and BCCLs used in this study is shown in Supplementary Table S1.1.

**Drug treatments**

IDB PDX and resistant variants were generated as previously described (1). Briefly, for the generation of the docetaxel resistant-derived tumors, treatment started in mice bearing tumors of 6x6 mm size (L\*l/100). When tumor volume decreased below 3x3 mm, treatment was interrupted, and reinitiated when tumors re-grew over of 6x6 mm. Mice were ethically sacrificed when the tumor size surpassed a diameter of 1,5 cm or mouse weight decreased by 20%. Tumors were excised and re-implanted into new hosts (passage 2) and docetaxel treatment was reinitiated.

Treatments performed at IDIBELL: Short-term intraperitoneally treatments with docetaxel (Hospira/Actavis, 20 mg/kg) or carboplatin (Accord, 10 and 50 mg/kg) every 5-7 days were performed in NOD/SCID or Foxn1nu tumor-bearing mice, respectively. In both cases, Fortecortin (Dexamethasone, 0.132 mg/kg, Merck) was administered 24 hr later. Treatment was initiated when tumors were bigger than 6x6 and stopped 10 to 50 days later, depending on tumors response. Tumors were collected 3-5 days after the final dose of docetaxel and frozen at -80ºC prior to DNA or RNA extraction.

Treatments performed at Baylor: among other treatments, PDX-bearing mice were treated either with vehicle (saline) or with four weekly intraperitoneal injections of either docetaxel (McKesson, 20mg/kg) or carboplatin (McKesson, 50mg/kg) as single agents. Tumor size was measured twice weekly and overall response was evaluated by RECIST 1.1 criteria (2) 28 days after the initiation of treatment.

*BRCA1* mutant patient: The TNBC patient was diagnosed in 2007. Surgical treatment consisted of a modified radical mastectomy. The patient received six courses of adjuvant docetaxel. Metastatic relapse was confirmed in 2012 (sternal mass and multiple bone, lymph node and liver metastases) and the patient was referred to Vall d’Hebron University Hospital for first line chemotherapy treatment. Written informed consent was obtained from the patient to perform exome sequencing of blood, primary tumor, metastatic relapse and liver metastatic progression within an Institutional Review Board- approved research project.

**Whole-exome sequencing of IDB PDX models and human samples of origin**

Exome capture and library construction were performed using Agilent SureSelect Human All Exon kit (Agilent, Santa Clara, CA, US) according to the manufacturer’s instructions. Paired-end sequencing was performed on a HiSeq2000 instrument (Illumina) using 100-base reads. Reads were aligned to the reference genome (GRCh37) using BWA (3) and a BAM file was generated using SAMtools (4). PCR duplicates were removed using SAMtools and custom scripts, and somatic substitutions and INDELs were called using Sidrón as described (5). For xenograft-derived samples, reads were first aligned to mouse genome (mm9), and those read-pairs that did not align to mouse were then aligned to the human genome with the GEM mapper (6). This procedure removed murine-derived reads, which might interfere in the analysis by artificially increasing the number of variants. Then, PCR duplicates were removed with Picard (Picard tools website) and variants were called using SAM tools. SnpEff and SnpSift were used to annotate and predict the impact of the called variants (7,8). Called variants in which all samples had a coverage ≥10 reads and a genotype quality ≥ 20 were used to calculate the genotype similarity between sample pairs. Copy number alterations (CNAs) were also assessed with Control-FREEC (9) using the same experimental design and using exome2CNV (10) by comparing each tumor sample or PDX to its normal counterpart, or with a normal sample in the case of the IBD-01 model. This study is registered in the European Genome-Phenome Archive (EGA) under the EGA ID EGAS00001003196.

**Whole exome sequencing of Baylor models**

For DNA extraction from untreated PDX tissue, snap frozen early transplant generation (TG1 or TG2 when possible) xenograft-derived-tumors were harvested and saved at -80 °C prior to use. Exome sequencing libraries were prepared using the Agilent SureSelect system. Sequencing was performed on an Illumina NextSeq 500 (High Output, 300 cycle v2) Multiplexing 5 samples per run. Estimated reads per sample = 160M total PE. Raw FASTA data files were aligned to a human mouse concatenated genome reference sequence (hs37d5 and GRCm38) using the EdicoGenome DRAGEN system. The resulted bam files were sorted and duplicate-marked. Only the reads that were uniquely aligned to human genome were kept. The copy number aberration was analyzed by the R package CopywriteR (11). The segmentation was done using the circular binary segmentation (CBS) algorithm (12) which is also implemented in the CopywriteR package (parameters: alpha=0.0001, min.width=5). To summarize the copy number value for a given genome region, we used a weighted sum approach: all the segments located within the region were identified and their log2 ratios were weighted by their length before being summed to derive the final score.

**Exome sequencing of patient sample**

Patient provided written informed consent for somatic and germline DNA analysis. Samples were initially assessed for tumor content based on a hematoxylin and eosin staining. Samples were subjected to whole exome sequencing. Genomic DNA was extracted from frozen samples with the QIAamp DNA mini kit (Qiagen), from archival FFPE samples with the Maxwell FFPE Plus LEV DNA Purification kit (Promega) and from whole blood with the DNA QIAamp DNA blood midi kit (Qiagen), according to manufacturer’s instructions. Library preparation was performed following the standard Illumina protocol (Genomic Sample Prep). One µg of DNA was fragmented, ends repaired and an adenine was ligated to each of the 3’ends, where sample-specific adaptors were linked. Libraries were amplified using 8 cycles of polymerase chain reaction (PCR), and exome enrichment was performed using specific biotinylated probes (SureSelect XT Human All Exon 50Mb, Agilent). After enrichment, the exome libraries were PCR-amplified (8 cycles), quantified and loaded in a HiSeq2000 sequencer (Illumina). A quality check of the raw data was performed by the FastQC tool (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Paired-end 100 base sequences were aligned to the Sanger human reference genome (hg19) using the Burrows-Wheeler Alignment tool (BWA, v.0.7.12) (3). This yielded a depth of coverage in targeted regions of ~70%, with ~70% of the exome having at least 10 reads. The resulting BAM files were processed using SAMtools (v.0.1.19) (4,13) and the Genome Analysis ToolKit (GATK, v.3.2.0) (14). Mutations were called with VarScan2 software (v. 2.3.7) (15), with the following parameters: somatic, minimum 8X coverage, minimum 5 reads supporting the variant allele, minimum variant allele frequency (VAF) of 5%, strand-bias filtering and a *p* value below 0.05. Annotation of the vcf files was performed with ANNOVAR (16). Except for germline *BRCA1* and *BRCA2*, variants were filtered based on being annotated as exonic by RefSeq database (release 45) and non-synonymous. Final manual review was done to check for local indel misalignments and homopolymer false positive calls. Copy number alterations were identified using the tool SeqGene (17) and the R Bionconductor DNAcopy package (18).

**Sanger sequencing**

PCR reactions were set up using 25 µL final volume according with Immolase DNA Polymerase manufacturer’s recommendations (Bioline, BIO-21047), and 1 µl of template DNA (initial concentration 50 ng/µl) with specific primers for each point mutation or small INDEL and performed on an Applied Biosystems 2720 thermal cycler machine using the following conditions: 95**°**C for 10 min, 35 cycles of 94**°**C for 30 s, 55**°**C for 30 s, 72**°**C for 30 s, and a final extension of 5 min at 72**°**C. PCR products were separated by electrophoresis on a 2% agarose gel. Successful PCR products were cleaned with *in house*ExoSAP mix of Exonuclease I (New England Biolabs, M0293S) and Shrink Anctarctic Phosphatase (New England Biolabs, M0289S) at 37ºC for 15 min, and inactivated at 80ºC for 15 min. PCR products were subjected to Sanger sequencing using the following primers:

|  |  |
| --- | --- |
| *hKLHL42* Forward | AGCAGCAGATGGTGTCTGTG |
| *hKLHL42* Reverse | CCCTTGGAATGGGACACCAC |
| *hBRCA1* Forward | GCTTCTCTTTCTCTTATCCTGATG |
| *hBRCA1* Reverse | AATCCAAATTACACAGCCTCTC |

**Gene expression microarray**

Genome-wide gene expression was analysed in sensitive IDB-02S and resistant IDB-02R tumors. 200-ng aliquots of total RNA were used to produce fluorescent complementary RNA following the Two-Color Microarray-Based Gene Expression Analysis v. 6.5 (Agilent) protocol following manufacturer’s instructions. All samples were hybridized to the SurePrint G3 Human Gene Expression 8 × 60 K microarray (Agilent Technologies). The signal values were extracted using the Feature Extraction software (Agilent Technologies). After scanning and normalization processes, statistical analyses were carried out in the R programming environment using the Bioconductor’ package for gene expression analysis: Limma, RankProd, Marray, affy, pcaMethods, EMA and RamiGO. The original gene expression datasets are available in the GEO repository and can be retrieved from GSE110153 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110153> token: gtajoqgurtkzjgx.

**GSEA analysis**

Gene Set Enrichment Analysis (GSEA) is freely available and is supported by the Broad Institute website (<http://software.broadinstitute.org/gsea/index.jsp>) and includes versions compatible with Java, R or Gene Pattern. All GSEA analyses presented here were performed using the R GSEA implementation.

**qPCR using TaqMan probes and data analysis**

Copy number assays were performed in a duplex real-time PCR reaction using a QuantStudio 5 Real-Time PCR System (Life Technologies) in a standard 384 well format in a total volume of 10 μl. *GABARAPL1* (Life Technologies, 4400291, Hs02787462\_cn), *ETV6* (Life Technologies, 4400291, Hs02036151\_cn) and *KLHL42* (Life Technologies, 4400291, Hs01571843\_cn) pre-designed copy number assays with a FAM™ dye-labeled MGB probe were used and a *TERT* pre-designed copy number reference assay with a VIC® dye-labeled TAMRA™ probe (TaqMan® Copy Number Reference Assay TERT, Life Technologies, 4403315) were used.

The PCR was carried out using 5 μl of 2x TaqMan® Universal PCR Master Mix (cat. no. 4304437, Life Technologies), 0.5 μl of 20x TaqMan® Copy Number Reference Assay TERT (Life Technologies), 0.5 μl of 20x TaqMan® Copy Number Assay mix, 2 μl of DNase free water and 2 μl of genomic DNA (initial concentration 5ng/µl). The final amount of genomic DNA was 10 ng/well in all reactions. All samples were amplified in three replicates and one non-template control per primer pair was included in each run together with a calibrator sample. The cycling conditions comprised 10 min polymerase activation at 95°C followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. The number of copies of the target sequence in each test sample is determined by relative quantitation (RQ) using the comparative Ct (ΔΔCt) method. With this approach, the predicted copy number of normal samples with two copies of each gene will be 2 for all assays. The predicted copy number of a sample with one gene deleted will be 1 in the respective assay. Predicted copy numbers are based on the calculated copy numbers (2-ΔΔCq).

**Public data analyses**

The METABRIC dataset (n = 2509) (19,20), consisting of copy number, gene expression and follow-up, data were downloaded from cBioPortal (<http://www.cbioportal.org/>). Similarly, TCGA data (21) were downloaded from the Genomics Data Commons portal [(https://portal.gdc.cancer.gov/](file:///C%3A%5CUsers%5Cgyoldi%5CAppData%5CLocal%5CMicrosoft%5Cjgomezm%5CDownloads%5C%28https%3A%5Cportal.gdc.cancer.gov%5C)). **A matrix with copy number results from TCGA/METABRIC breast cancer primary tumors was constructed covering chromosome 12 from 1 to 33,000,000 bp. This range was subsequently divided into 100 windows each of 33,000 bp and the corresponding copy number values defined from the segment mean values.** Mutational signatures were defined using the [“deconstructSigs”](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4762164/) R package (22). **Sample classification based on copy number profiles or mutational signatures were performed using hierarchical clustering with Euclidean distance and complete agglomeration.** The Kaplan-Meier plots and log-rank tests were computed using the survfit function in the same package.

**Culture and treatment of human breast cancer cells**

All cell lines were purchased from the American Type Culture Collection (Rockville, MD), except for UACC3199 which was obtained from the Arizona Cancer Center (Tucson, AZ). ATCC provides molecular authentication in support of their collection through their genomics, immunology, and proteomic cores, as described, by using DNA barcoding and species identification, quantitative gene expression, and transcriptomic analyses. UACC3199 cells harbor a methylated *BRCA1* promoter suppressing gene transcription (23). UACC3199 was authenticated by its ability to re-express BRCA1 after DNA demethylation treatment with 5-aza-2'-deoxycytidine. All lines were expanded and frozen within 2 weeks of purchase and used for a maximum of 4 months after resuscitation of frozen aliquots. All cells but HCC1143 and HCC1937, which was cultured in RPMI 1640, were maintained in DMEM high glucose, containing 10% FBS (Gibco), L-glutamate (Gibco), and penicillin/streptomycin (PAA Laboratories) at 37°C in 5% CO2. At 60%–70% confluence the indicated concentrations of docetaxel or carboplatin were added. Cells were collected at the indicated time points and counted with trypan blue to exclude dead ones. All cell lines were routinely tested for mycoplasma (Biotools, B & M Labs, #4542) every month and before each experiment and were shown to be free of contamination. For all *in vitro* studies, a minimum of three biological replicates were performed. Experiments were not performed in a blinded fashion.

**Plasmid generation and Lentiviral infection**

ETV6 and GARAPL1 overexpressing plasmids were generated using Gateway technology. Briefly, the ORFeome collaboration sequence of human *ETV6* and *GABARAPL1* inserted in a pENTR223.1 vector (clone ID: 100015850) was inserted into pSD69 (PGK promoter, generously donated by S Duss and M Bentires-Alj) expressing vector and pLENTI-CMV (CMV promoter, Addgene, #17452) expressing vector using Gateway LR Clonase Enzyme Mix (Invitrogen). Control vectors were generated inserting lacZ from a pLENTI6-v5-lacZ vector into p201 donor vector using Gateway BP Clonase Enzyme Mix (Invitrogen) and then, into pSD69 and pLENTI-CMV using Gateway LR Clonase Enzyme Mix (Invitrogen).

For knock-down variants, commercial pGIPZ-shGABARAPL1, pGIPZsh-ETV6 and control vectors (Dharmacon GE) were used.

Lentiviral infection was done following the manufacturer´s indications (Invitrogen). Briefly 293FT cells were used for the production of the virus. 293FT cells (5x106) were transfected with the lentiviral vectors and packaging (gag-pol, vsvg, rev) plasmids (Addgene) by calcium phosphate method. 25mM HEPES was added 16 h later. Virus supernatants were harvested 72h post transfection, centrifuged at 250G 5’ and filtered with 0.22 µm filters. TNBC cell lines were transduced in a ratio 1:3 with fresh growth medium and with 8µg/ml of polybrene. Plates were centrifuged 1 hour at 1.000 rpm at 37ºC to improve the infection. Selection started with puromycin antibiotic (Sigma-Aldrich) at 1,5µg/ml. The resulting stable cell lines infected were maintained with 0,5 µg/ml. Medium was refreshed every three days. The following shGABARAPL1 sequences were tested 117161: TACCTTACTTCATACTTGC; 225602: TATTTCCTTGGCCTGATGG; 313119: AGGTACTTCCTCTTGTCCA; 313120: TGACAAAGAAGAATAAGGC; 313121: TAGACACTCTCATCACTGT and based on the results 225602 and 313121 were selected. The following shETV6 sequences were tested: 321568: TTGTCGTGATAGGTGACCT; 321569: AATGCTACACTGAGCAGGA; 321570: AGTTTGTAGTAGTGGCGCA; 321571: TGAAATCCACGGAGTGCCG; 321572: TGTCGTGATAGGTGACCTG and based on the results 321568 and 321570 were selected.

**DNA extraction and RNAse treatment**

Frozen tumor tissues were fractionated using the POLYTRON® system PT 1200 E (Kinematica) and incubated 12-16 hours with 0.25% SDS (Invitrogen, 24730020), 0.25 mg/ml Proteinase K (Sigma Aldrich, P4850) and RNase A (Sigma Aldrich, R5503) at 55ºC in a thermal block in the lyses buffer (100mM NaCl, 10mM Tris-Cl pH 8, 25mM EDTA) while TNBC cell lines were directly resuspended in the lyses buffer; the same procedures were applied to both sample types from this step.. For DNA purification, the homogenized sample was transferred to a phase lock gel heavy tube (VWR, 713-2538). Two steps of removal of proteins from nucleic acids using Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (Sigma Aldrich, P3803) and three steps of nucleic acid washing with chloroform (VWR, 1024311000) were performed, centrifuging each time at 1500G during 5 minutes. After last spinning, remaining volume containing DNA was recovered and transferred to a solution containing 2.5x absolute ethanol (Merck Millipore, 1009832500) and 30 mM of sodium acetate pH 5.2 (Sigma Aldrich, S2889), mix and centrifuge at maximum for 5 minutes. Wash two times with 70% ethanol centrifuging at maximum for 5 minutes. Finally, resuspension in TE buffer (10 mM Tris-HCl, 1 mM disodium EDTA, pH 8.0) or ultrapure water (MilliQ purification system) was performed. DNA concentrations were determined using a Nano Drop ND-1000 Spectrophotometer (Thermo Scientific, Inc. USA). The DNA was diluted to a final concentration of 50 ng/μL prior to conventional PCR and to 5 ng/μL prior to quantitative real-time PCR.

**RNA extraction, DNase treatment and qRT-PCR**

Frozen tumor tissues were fractionated using the POLYTRON® system PT 1200 E (Kinematica) and Tripure Isolation Reagent (Roche) and treated with DNA-free DNase I kit (Ambion, AM1906) while TNBC cell lines were directly resuspended in Tripure and treated with DNA-free DNAseI kit. cDNA was produced by reverse transcription using 1 μg of DNA-free RNA in a 35 μL reaction following TaqMan™ Reverse Transcription instructions (Applied Biosystems, N8080234). 20 ng/well of cDNA were used for the analysis performed in triplicates. Quantitative PCR was performed using the LightCycler® 480 SYBR green. Primer sequences are indicated below. Ct analysis was performed using LightCycler 480 software (Roche). All primers indicated below are in 5’ 🡪 3’ direction.

|  |  |
| --- | --- |
| *hA2ML1* Forward | GGCAGCAAGTGTATTTCCGC |
| *hA2ML1* Reverse | ATGCCTTGCTCAGGTACCAC |
| *hDERA* Forward | AACAGAAGCTTGGTGCTGACA |
| *hDERA* Reverse | GCCTTGCGAAACTGACGAATC |
| *hETV6* Forward | TGACAGCAACACGTTTGAAAT |
| *hETV6* Reverse | AGGAGTTCATAGAGCACATCACC |
| *hGABARAPL1* Forward | ACCATGGGCCAACTGTATGA |
| *hGABARAPL1* Reverse | TGGGCTTCCAACCACTCATTT |
| *hITPR2* Forward | CCTTGGGGTTAGTGGATGACAG |
| *hITPR2* Reverse | TGGCTTGCTTTGCTTTCCAAT |
| *hKLHL42* Forward | CGCCCTTACCCAATCCTCTG |
| *hKLHL42* Reverse | GTCCACATGTCGGAAGAGGG |
| *hM6PR* Forward | GCTACTCCAGTTTCCCACGA |
| *hM6PR* Reverse | GTAGCAGTCCAGTCCTCCAG |
| *hMGST1* Forward | AATTGTATTTCTGTCCCCGTGC |
| *hMGST1* Reverse | TCCATTACCTGGGTGAGGTCAA |
| *hSTRAP* Forward | ACAGCAGCTGCAGATTTCACA |
| *hSTRAP* Reverse | CCTGTCCCCCGGTTAACAA |
| *hPPiA* Forward | ATGCTGGACCCAACACAAAT |
| *hPPiA* Reverse | TCTTTCACTTTGCCAAACACC |

**Statistical Analyses**

The association between the copy number of chromosome 12p and the drug response was estimated by the Spearman’s rank correlation test. All data are expressed as mean ± SEM. Group differences were compared using Student’s unpaired samples t test and GraphPad Prism version 5.04. Values of p ≤ 0.05 were considered statistically significant. Levels of significance are expressed as: \*0.01 < p < 0.05; \*\*0.001 < p < 0.01; \*\*\*0.001 < p < 0.0001; \*\*\*\*p < 0.0001.

**Supplementary References**

1. Gómez-Miragaya J, Palafox M, Paré L, Yoldi G, Ferrer I, Vila S, et al. Resistance to Taxanes in Triple-Negative Breast Cancer Associates with the Dynamics of a CD49f+ Tumor-Initiating Population. Stem Cell Reports. 2017;8:1392–407.

2. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). European Journal of Cancer. 2009;45:228–47.

3. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.

4. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

5. Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011;475:101–5.

6. Marco-Sola S, Sammeth M, Guigó R, Ribeca P. The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods. 2012;9:1185–8.

7. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 2012;6:80–92.

8. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet [Internet]. 2012 [cited 2018 Apr 5];3. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3304048/

9. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012;28:423–5.

10. Valdés-Mas R, Bea S, Puente DA, López-Otín C, Puente XS. Estimation of copy number alterations from exome sequencing data. PLoS ONE. 2012;7:e51422.

11. Kuilman T, Velds A, Kemper K, Ranzani M, Bombardelli L, Hoogstraat M, et al. CopywriteR: DNA copy number detection from off-target sequence data. Genome Biology. 2015;16:49.

12. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004;5:557–72.

13. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987–93.

14. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.

15. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568–76.

16. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.

17. Deng X. SeqGene: a comprehensive software solution for mining exome- and transcriptome- sequencing data. BMC Bioinformatics. 2011;12:267.

18. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics. 2007;23:657–63.

19. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486:346–52.

20. Pereira B, Chin S-F, Rueda OM, Vollan H-KM, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun [Internet]. 2016 [cited 2017 Sep 29];7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4866047/

21. Network TCGA. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61–70.

22. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol [Internet]. 2016 [cited 2018 Mar 29];17. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4762164/

23. Rice JC, Massey-Brown KS, Futscher BW. Aberrant methylation of the BRCA1 CpG island promoter is associated with decreased BRCA1 mRNA in sporadic breast cancer cells. Oncogene. 1998;17:1807.

**Supplementary Figure S1. Point mutations or small INDELs are not associated to docetaxel chemoresistance acquisition in TNBC PDX.**

**A)** Diagrams showing the transplant history from IDB-01 and IDB-02 TNBC PDX models. Black spheres indicate the passage of a tumor to another set of mice. Clear lines correspond to sensitive tumors while dark lines correspond to docetaxel treated tumors during chemoresistance acquisition. Tumors used for whole exome sequencing are indicated with a yellow circle and their name. Passages of tumors used for Sanger sequencing validation are indicated with a box under the genealogical tree. **B)** Scheme of patient´s treatment and follow-up and PDX generation**. C)** Unsupervised hierarchical clustering using genotype similarity extracted from whole-exome sequencing data between metastatic sample of origin, sensitive and resistant TNBC PDX tumors from IDB-01 and IDB-02. Note sensitive and resistant samples do not cluster separately, while metastasis in IDB-02 does. **D, E)** Venn’s diagrams showing the comparisons between point mutations or small INDELs; (**E)** includes only those potentially affecting protein function present in the three resistant PDX tumors and absent in metastases or sensitive PDX tumors from IDB-01 and IDB-02. **F)** Sanger sequencing traces for *KLHL42* point mutation in sensitive (left) and resistant (right) IDB-02 tumors. Arrows indicate the base changed and the expected > alternative nucleotide is indicated at the top. \* indicates *BRCA1*-mutated models.

**Supplementary Figure S2. Copy number stability during passages and acquisition of resistance to docetaxel in the IDB-01 model and indirect validation of chr12p amplification using gene expression data.**

**A)** Copy number analysis performed in metastasis of origin (M) and sensitive (S) and resistant (R) IDB-01 PDX tumors compared with a normal control sample. **B)** Unsupervised hierarchical clustering using CNV similarity extracted from whole-exome sequencing data between metastatic sample of origin, sensitive and resistant TNBC PDX tumors from IDB-01 and IDB-02. CNV from each tumor was extracted (log2ratio>0.5). Regions with normal copy number were not considered. Note that in IDB-02 115 regions were altered whereas in IDB-01 only 12 regions were altered, making any disparity in copy number profile a high percentage of difference in similarity, which would not reflect reality. **C)** Comparisons between all genes located and overexpressed in chromosome 12 and specifically in chr12p amplified region in the IDB-02 PDX model. Chi-square and p-value statistics are shown. **D**) mRNA expression levels of indicated genes located at chr12p amplification relative to *PPiA* in IDB-02S and IDB-02R tumors. Each dot represents a tumor. Mean, SEM and t-test p-values are shown. \* indicates *BRCA1*-mutated models.

**Supplementary Figure S3. Increased copy number of genes located in chr12p after short-term treatment with docetaxel and acquisition of resistance in TNBC PDX models and cell lines.**

**A)** Relative mRNA expression levels of indicated genes located at chr12p amplification in residual disease (RD) (n=4) compared with sensitive untreated tumors in IDB-02S (n=4). **B)** Representative kinetics of tumor growth during acquisition of resistance to docetaxel in the basal-like PDX VHIO98. Each line represents one tumor and each color represents an independent sensitive PDX of origin. Mice were sacrificed and tumors were transplanted into other mice to continue treatment as described [17]. Ps = passage treated with docetaxel. Arrows represent docetaxel treatment. Only mice bearing tumors bigger than 6x6 received docetaxel. Red circles indicate the tumors that were transplanted in another set of mice to continue the treatment. **C**) Copy number analysis by qPCR using TaqMan probes located at different genomic positions of chr12p in the indicated models. S (sensitive, untreated), RD (residual disease), R (acquired resistance). Each dot represents a tumor. Mean, SEM and t-test p-values are shown. **D**) Scheme of patient´s treatment and follow-up and generation of the PDX VHIO-127. **E)** Copy number analysis by qPCR using TaqMan probes located at different genomic positions of chr12p in the indicated TNBC cell lines. **F)** Copy number analysis by qPCR using TaqMan probes located at different genomic positions of chr12p in the indicated cell lines 72 h after treatment with the indicated doses of docetaxel. \* indicates *BRCA1*-mutated models.

**Supplementary Figure S4. Integrative genomic analysis using cBioPortal and METABRIC samples.**

**A)** Representation of most common cancer types harboring amplification of three selected genes located in the chr12p amplified region. The black line shows that breast cancer databases have around 4% of tumors with amplification. **B)** Representation of breast tumors from the METABRIC dataset harboring amplification and gain of three selected genes located at the chr12p amplified region. **C)** Association between copy number of three selected genes located at chr12p amplified region and 3-gene classifier subtype in METABRIC dataset. A Fisher exact test showing association between amplification and ER-/HER2- subtype is indicated below. **D**) Association between copy number and mRNA expression levels of three selected genes located at the chr12p amplified region in the METABRIC dataset. Each dot represents a tumor. Box and whiskers plots, SEM and t-test p-values are shown. **E)** GSEA analysis comparison between IDB-02R tumors and clinical TNBC harboring amplification of chr12p. Pathways overlapping between both groups are highlighted.

**Supplementary Figure S5. Chr12p amplification predicts better response to carboplatin.**

**A)** Differential copy number (vs diploid) of the surrogate genes *GABARAPL1*, *ETV6* and *KLHL42* relative to *TERT* in the Baylor PDX cohort. **B)** Relative tumor volume over time in sensitive untreated and sensitive carboplatin-treated IDB-02S tumors. Mice were sacrificed one week after the last dose. Each arrow indicates a carboplatin dose. Mean, SEM and t-test p-values are indicated. **C)** Copy number as determined by qPCR using three different Taqman probes located at different genomic positions of chr12p-amplified region during residual disease (RD) after carboplatin treatment and untreated IDB-02S tumors. Mean and SEM are shown. **D)** Percentage of surviving MDA-MB-436 TNBC cells treated with the indicated concentrations of carboplatin for 72 h. Three independent experiments were performed. **E)** Copy number analysis by qPCR using TaqMan probes located at different genomic positions of chr12p in the MDA-MB-436 TNBC cell line 72h after treatment with the indicated doses of carboplatin. \* indicates *BRCA1*-mutated models

**Supplementary Figure S6. Overexpression and knockdown of *ETV6* and *GABARAPL1* in different TNBC BCCLs and docetaxel response.** **A)** Relative mRNA expression levels of indicated genes in a panel of TNBC BCCLs. qRT-PCR determinations were done in triplicate. Means and SEM are shown. **B)** *ETV6* and *GABARAPL1* mRNA expression levels in three different TNBC BCCLs stably infected with two independent *ETV6* or *GABARAPL1* overexpression and control vectors (pSD69-LacZ and pLenti6-LacZ). qRT-PCR determinations were done in triplicate. Means and SEM are shown. **C)** Percentage of surviving *ETV6* or *GABARAPL1* overexpressing cells and controls treated with indicated doses of docetaxel for 72 hr. Mean values of three independent experiments are indicated. **D)** *ETV6* and *GABARAPL1* mRNA expression levels in MDA-MB-436 cells stably infected with *ETV6*, *GABARAPL1* overexpressing or control vectors. qRT-PCR determinations were done in triplicate. Means and SEM are shown. **E)** Percentage of surviving *ETV6* or *GABARAPL1*- MDA-MB-436 overexpressing cells and controls treated with indicated doses of docetaxel or carboplatin for 72 hr. Mean and SEM values of three independent experiments are shown. **F)** *ETV6* and *GABARAPL1* mRNA expression levels in MDA-MB-436 cells stably infected with two independent shETV6 or shGABARAPL1 knockdown and control vectors (pGIPZ). qRT-PCR determinations were done in triplicate. Means and SEM are shown. **G)** Percentage of surviving shETV6 or shGABARAPL1-MDA-MB-436 cells and controls treated with indicated doses of docetaxel or carboplatin for 72 hr. Mean and SEM values of three independent experiments are shown.

**Supplementary Tables**

**Supplementary Table S1. TNBC models used, somatic point mutations and copy number alterations in the human lymphocytes, metastatic samples of origin, sensitive and resistant PDX.**

**1.1** TNBC PDX models and cell lines used in this study and main characteristics, including *BRCA1* and chr12p status and primary response to docetaxel and carboplatin.

**1.2.** Putative somatic point mutations estimated in the human metastatic sample of origin (M), and tumors from sensitive IDB-01S and resistant IDB-01R PDX.

**1. 3.** Somatic point mutations list in metastatic sample of origin and PDX from model IDB-02. L, lymphocytes; PT, peritumoral tissue; M, metastasis; IDB-02S, sensitive PDX; IDB-02R, resistant PDX.

**1.4.** Copy number alterations (CNA) list estimated in metastatic sample of origin and PDX model IDB-01. Green, common CNA between PDX tumors absent in metastasis.

**1.5.** Copy number alterations (CNA) list in metastatic sample of origin and PDX from model IDB-02. M, metastasis; IDB-01S, sensitive PDX; IDB-01R, resistant PDX. Grey, common CNA between metastasis and PDX tumors; green, common CNA between PDX tumors absent in metastasis; red, specific CNA present in resistant PDX tumors and absent in sensitive PDX tumors.

**Supplementary Table S2. Differentially expressed genes in chr12p between IDB02R and IDB-02S tumors, GSEA in TNBC patients harbouring chr12p amplification and co-occurrence/mutual exclusivity analyses of alterations in queried genes.**

**2.1** Gene expression matrix of IDB-02S and IDB-02R tumors of genes contained in chromosome 12. Statistically significant differentially expressed genes are indicated in bold.

**2.2** Gene expression matrix of IDB-02S and IDB-02R tumors of genes contained in chromosome 12p amplified region. Statistically significant differentially expressed genes are indicated in bold.

**2.3** Retrieval pathways from GSEA of TNBC patients from METABRIC dataset harbouring chr12p amplification.

**2.4.** Retrieval pathways from GSEA of differentially expressed genes from IDB-02R tumors.

**2.5.** Co-occurrence and mutual exclusivity analysis of alterations in queried genes *GABARAPL1*, *ETV6* and *KLHK42*.

**2.6.** Analysis of co-occurrence and mutual exclusivity enrichment of copy number alteration in breast cancers with alteration in the queried genes *GABARAPL1*, *ETV6* and *KLHK42*.

**2.7.** Co-occurrence and mutual exclusivity mutations of tumors in breast cancers with alteration in the queried genes *GABARAPL1*, *ETV6* and *KLHK42*.