Supplementary methods, tables and figures

Supplementary material and methods

The following tumor samples were collected at patient screening: one formalin-fixed, paraffin-embedded (FFPE) core and one frozen core embedded in optimal cutting temperature (OCT), the frozen core was not used in this report) from the local recurrence or the metastatic lesion, except for bone-only patients, and one archived FFPE core from the primary tumor. The FFPE samples could be a block or 25 unstained sections of 5 µm each on SuperFrost Plus slides (Thermo Fisher Scientific). Tumor cellularity had to be >10% of the whole tumor sample for the sample to be accepted. Biopsy samples were processed within 30 minutes from the procedure. 1x9mL whole blood was also collected in an EDTA tube for germline DNA sequencing, as well as 2x9mL blood in EDTA tube for plasma cfDNA sequencing and 1x9mL blood in serum clot activator (not used in this study). For plasma separation, the two EDTA blood tubes were centrifuged at 820g for 10 minutes at 4°C (advised) or at room temperature within 30 minutes of the blood draw. The supernatant from these tubes was then centrifuged a second time at 20,000g for 10 minutes at 4°C (advised) or room temperature. The plasma was then carefully transferred to 1.8mL cryovials and immediately frozen at -80°C. The frozen metastatic sample, blood and plasma samples were shipped on dry ice while the FFPE samples were shipped at room temperature.

IHC and HER2 FISH

At the central laboratory, IHC for ER, PR, HER2 and Ki67 were performed using the ER/PR pharmDx kit (Dako), the HercepTest kit (Dako) and the clone MIB-1 (Dako) respectively. The interpretation of the ER and PR staining was performed in accordance with the ASCO/CAP 2010 guidelines (1). Interpretation of the HER2 staining was performed in accordance with the ASCO/CAP 2013 guidelines (2). When required by the guidelines, FISH for *ERBB2* was performed using the HER2 FISH pharmDx kit (Dako).

Typing of each tumor was done in one of 3 subtypes: TNBC, HER2+ and HR+/HER2-. It was done using the local assessment, if available. When not available, central typing was used. HR+ tumors were defined as those being ER+ or PR+. Tumors that had equivocal HER2 status were considered HER2 negative. The subtype was assigned based on that of the primary tumor. If unavailable, the subtype of the metastasic sample was used.

Extraction of DNA, RNA and cfDNA

Wherever applicable, macrodissection to enrich for tumor cells was performed. DNA was extracted from FFPE samples using the QIAamp DNA FFPE tissue kit (QIAGEN) and for blood samples the QIAamp DNeasy Blood and Tissue kit (QIAGEN), following the manufacturer's instructions. DNA concentrations were measured using the Qubit fluorometer (Life Technologies). The cut-off values for tumor content and DNA quantity were 10% and 400 ng, respectively. RNA was extracted from FFPE samples using the RNeasy FFPE kit (QIAGEN) and eluded in RNAse-free water. RNA concentration was determined using Qubit Fluorometric Quantitation (Thermo Fisher Scientific) and its integrity (DV200) was assessed using the Agilent Bioanalyzer. Samples with an RNA DV200 <30% were discarded. cfDNA was extracted from plasma using the QIAsymphony DSP Circulating DNA Kit (QIAGEN) and quantified using the Thermo Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific), with readings done on a Berthold TriStar fluorometer.

Targeted gene sequencing using Ion Torrent NGS

Somatic mutations were assessed using the OncoDEEP clinical cancer panel (OncoDNA) which is a validated AmpliSeq design panel targeting the exonic regions of 409 cancer related genes to which probes specific for the BRCA1 and BRCA2 genes were added (supplementary table 3). The same protocol was applied to DNA extracted from FFPE tumor and whole blood normal matched samples. Briefly, the targeted sequencing libraries were generated using the Ion AmpliSeq library kit 2.0 according to the manufacturer's instructions (Life Technologies) using 80 ng of genomic DNA. The primers used for amplification were partially digested by *Pfu* restriction enzyme and the digestion

products were ligated to barcoded adaptors and purified using Ampure Beads. The purified products were amplified for five cycles and purified again using Ampure Beads. The quality of the libraries was assessed using a qPCR following which 10 pM of each library underwent emulsion PCR using an IonChef system. The chips were loaded on an Ion PGM and were sequenced at a target coverage of 500X.

Copy number aberration profiling using SNP arrays

Copy number aberration profiling using the Affymetrix OncoScan FFPE array were performed from 80 ng of DNA according to the manufacturer's instructions. In short, the molecular inversion probes (MIP) were incubated with the FFPE extracted DNA at 58°C overnight after an initial denaturation at 95°C for 5 min. Each sample was then split into two aliquots and a gap fill reaction was performed. Uncircularized MIP and genomic DNA were digested using a cocktail of exonucleases. The remaining circular MIP were then linearized using a cleavage enzyme and amplified by PCR. Following a second round of PCR amplification, the 120 bp amplicons were cleaved into two fragments with the *Hae*III enzyme. The samples were then mixed with the hybridization buffer and injected into the arrays where they were allowed to hybridize at 49°C for 16–18 h. At the end of the hybridization period, the arrays were stained and washed using the GeneChip Fluidics Station 450 and loaded into the GeneChip Scanner 3000 where array fluorescence intensity was scanned to generate binary CEL files using the Affymetrix GeneChip Command Console.

RNA-Seq

Sequencing libraries were prepared from 100ng of RNA from each FFPE sample using the transcriptome capture library Illumina TruSeq RNA Access library following the manufacturer's instruction and sequenced on Illumina HiSeq2500 in 2x100bp paired-end mode.

cfDNA sequencing

Libraries were prepared using the AmpliSeq Library Kit 2.0 (Thermo Fisher Scientific) with a combination of an AmpliSeq custom panel and the OncoTrace core panel (OncoDNA) (supplementary table 4). Libraries were quantified using the Thermo Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific), with readings done on a Berthold TriStar flurorometer. Sequencing was done using the Ion 540 Kit – Chef in combination with the Ion 540 Chip Kit.

Bioinformatics analyses

Mutation calling from Ion Torrent targeted gene sequencing

Sequence reads from the tumor and matched normal samples were aligned against the human genome reference version hg19/GRCh37 using the Ion Torrent TMAP aligner with default parameter settings. Mutations were called from the resulting BAM files using the Torrent Suite variant caller (Life Technologies) with the default settings of the 'Somatic High Stringency' pipeline and cross-checked using the NextGENe software (Softgenetics) using the 'Ion Torrent' predefined pipeline. Germline mutations were filtered by subtracting variants found in the matched blood sample (with variant allele fractions - VAF > 5%) from those called in the corresponding tumor sample. As in the AURORA pilot (3), the resulting somatic mutation calls were further cleaned using the data purveyor (oncoDNA) standard filters, i.e. 1) that were not sequenced in both sense with a minimum ratio of 10/90%, 2) with depth less than 100 read depth, 3) with VAF lower than 10% in the tumor sample or 4) present at 1% or more in ExAC (4) were excluded. Mutations that had less than 50x coverage in the normal were recovered if they appeared at least 10 times in COSMIC, and were not present at more than 1% in ExAC (5 mutations in total were recovered this way). For mutations found in a primary/metastasis, a lower VAF threshold of 1% was used to call the mutation in the corresponding metastasis/primary. For an alpha list of target genes (Supplementary Table 3), mutations occurring below 10% VAF were accepted. Several categories of variants are manually evaluated in IGV to rule out artefacts including all *BRCA1*, *BRCA2* and *PALB2* variants, and all indels.

Copy number alteration analysis using SNP arrays

Copy numbers were extracted from Affymetrix SNP arrays when available. ASCAT (5) was used with default parameters. When both primary and metastasis were available, they were co-segmented. The resulting fits were checked manually, and samples for which no copy number aberrations were found, because of lack of aberrations or lack of purity, were discarded. Samples for which the fits reported were not optimal were reoptimized. Cases were ASCAT would give different ploidies for the primary and the metastasis of the same patients were assessed. If there was no clear evidence for that change of ploidy, fit with similar ploidies were chosen instead.

Copy number aberration analysis using targeted NGS

Copy numbers were further extracted from TGS data using FACETS (6). Pileups were calculated at every SNPs and every 50 bp, with no maximum depth, and minimum depth of 20bp for the normal. For positions that were not SNPs, a rolling window smoothing of width 3 was used twice. FACETS preprocessing was done with a maximum depth of 10^5 and cval of 10, the processing with cval of 50 and min.nhet of 3. Fifty such fits were obtained with FACETS, and their qualities were recorded. In the case no Affymetrix data were available for the patient, the best fit among the primary and the meta sample was kept, and for the other sample the fit that matched best was kept. In the case Affymetrix data were available, fits that match Affymetrix data best were kept. From the FACET fits, copy numbers were derived for each gene of the targeted panel.

CNV analysis

To compare CNV between samples, we scaled those to get pseudo-diploid samples, hence we divided by half of their ploidy as estimated from the median CN across the genome. Three copy number aberration categories were considered: deletions (CN < 1.5), gain (CN > 2.5) and amplifications (CN > 4).

Power analysis for targeted sequencing

In our case, the sequencing depth was large, but it is likely that in general the library complexity was much lower, so that identical fragments were sequenced many times. Because of the library preparation method, it was not possible to flag duplicates based on their start and end positions. An alternate method to determine library complexity was developed, using heterozygous SNPs. The idea is that the distribution of the VAFs of those SNPs would have more variance that what would be expected based on the number of reads.

Our model is as follows. In the case of a diploid cancer sample, at a given SNP, there are N fragments, so that the distribution of the VAF at the fragment level (x) is $x \sim B(N, .5)$, where B is the binomial distribution. From that fragment level VAF, N_2 reads are sequenced ($N_2>N$ in general), leading the distribution of VAF at the read level (y): $y \sim B(N_2, x/N)$. The compression ratio is $c=N/N_2$, which must be determined from the data. In the case of a non-diploid tumor sample of purity p and copy number at the site of the SNP n, the distribution of x would be $x \sim B(N, (1+p(i-1))/(2+p(n-2)))$, where i is the number of fragments with the alternative allele. For each SNP, the most likely i from 0 to n is chosen. The full likelihood is optimized to find the sample compression ratio c. SNPs copy numbers and purity are taken from running FACET on the TGS data.

Once the compression ratio is obtained, the likelihood that a mutation seen in a sample but not in another was present but missed can be estimated. To do so, we hypothesize that the VAF is the same in both samples, corrected for their tumor purities. We then need to estimate the probability that the observed VAF is below 1% (our cut-off) knowing the expected VAF v. This is be done by summing the probabilities of observing such low VAFs for all possible fragment-level VAFs, taking into account the probability of having each fragment-level VAF, so the probability of having a false negative P is

$$P = \sum_{i=0}^{cN} \mathrm{B}(I, cN, v)\mathrm{PB}(0.01N, N, I/cN)$$

where B is the binomial pdf, PB is the binomial cdf, N is the number of reads, c is the compression factor and v the target VAF.

Among the 355 mutations that were found in the metastasis but not the primary, the average of the estimated false negative rate was 2.3%, for an expected total of 8.2 false negative. Among the 145 mutations that were found in the primary but not the metastasis the average of the estimated false negative rate was 5.9%, for an expected total of 8.2 false negative.

Driver genes identification

The dndscv R package (version 0.0.1.0)(7) was used to identify relevant genes in primary and metastatic samples. Statistically significant genes were identified using a restricted hypothesis testing (RHT) analysis. To do that, q-values by Benjamini-Hochberg (BH) procedure were recalculated using the list of BC relevant genes from IntOGen (August 2020, n=99) (8) included in the AURORA TGS (n=56) (figure S4).

Annotation of mutations

Single nucleotide and small indel variants were annotated as driver or passenger according to a set of rules. First, variants belonging to pathogenicity tiers 1, 2 or 3 in the COSMIC Cancer Mutation Census v92 (9) were annotated as drivers. Next, the category of the gene as a tumor suppressor gene or an oncogene was obtained from the COSMIC Cancer Gene Census as of January 2021 (10). For TSGs, frameshifting indels, nonsense mutations and canonical splice site mutations (positions +1, +2, -2 and -1 in the intron) were annotated as driver mutations. For oncogenes, reading frame-preserving indels were annotated as drivers. In the rare cases when a gene was identified as both TSG and oncogene in the Cancer Gene Census, the two rules applied. Finally, variants annotated as pathogenic by the pathology prediction programs PolyPhen (11), SIFT (12) or FATHMM (13) were annotated as drivers ("possibly_damaging" and "probably_damaging" in PolyPhen, or "deleterious" in SIFT or FATHMM).

Tumor Mutational Burden

Tumor Mutational Burden (TMB) was estimated as the sum of coding silent and non-silent mutations. To avoid biases in the estimation of TMB in matched versus non-matched samples, the

initial set of mutations with VAF≥10% (i.e., without the retrieval of mutations with VAF≥1% in the matched samples) was considered for this analysis. We classified samples into high or low TMB based on the 90th percentile (according to Fernandez et al (14) that estimated 11% of TMB-high breast cancer cases), corresponding to 8 and 11 mutations in primary and metastatic samples, respectively. TMB in TCGA dataset was estimated using the number of mutations considering only the genes included in the AURORA TGS panel.

Oncoplot

The oncoplot (Figure 2) was created using the ComplexHeatmap R package (15). The oncoplot includes data for patients with available TGS data for both primary and metastatic samples (n=242). It reports driver mutations (SNVs and InDels, see section Annotation of mutations) in driver genes (see section Driver genes identification), amplifications for a selection of known oncogenes (*CCND1, EGFR, ERBB2, FGFR1, FGFR2, FGFR3, FGFR4, FLT4, KAT6A, MDM4, MYC, PIK3CA*) and deletions for a selection of known tumor suppressor genes (*ARID1A, CDKN2A, NF1, PBRM1, PTEN, RB1, TP53*). For CNV events we considered CNAs from Affymetrix SNP array data and, if not available, the CNVs estimated from FACETS in TGS panel. Amplifications were defined as CN level (scaled by ploidy) > 4. For deletions, we considered scaled CN < 1 for Affymetrix SNP array data and CN < 0.7 for FACETS data. Concordance (measured by Cohen's κ) was acceptable (> 0.20) for 9 out of 15 genes. Discordances involve mainly genes most frequently involved in deletions, such as *RB1*, *TP53* and *PTEN* (supplementary figure S27). The analyses of CNVs using ASCAT and FACETS have therefore been combined in the oncoplot using different thresholds for deletions (0.7 for TGS, 1 for SNP Arrays).

Clonal analysis in matched primary and metastatic samples

Clonal variation between primary and metastatic paired samples (n=242) were evaluated using Cancer Cell Fraction (CCF) (16). For each variant, CCF was estimated using the gene-level copy number status and purity estimated by FACETS from TGS data. For each patient, sample-wise

(primary and metastatic) CCF was computed as the median of CCF of variants after the exclusion of variants with CCF=0 (i.e., variants private to primary or metastatic samples).

Homologous recombination deficiency determination

HRD status of the samples was determined with the SigMA (Signature Multivariate Analysis) computational tool (17). All mutations were used for the analysis, including matched mutations rescued at the 1% level. The three outputs of SigMA are reported: maximum likelihood (ML), and both versions of multivariate analysis (MVA), normal and strict. For the comparison between specific BRCA1/2 mutated samples and wild type samples, wild type was defined as samples with no BRCA1/2 mutations, germinal or somatic, at all. Only loss of function mutations (likely pathogenic and pathogenic) in BRCA 1 and 2 were considered.

Comparative analysis with TCGA-BRCA and METABRIC datasets

For TCGA-BRCA (18,19), mutations by Mutect2 (20) were downloaded using TCGABiolinks (21). For METABRIC (22), mutations were download from cBioPortal. Clinical data for TCGA-BRCA and METABRIC, including relapse (23,24), were downloaded from cBioPortal.

Transcriptome analyses

Sequencing reads were mapped to human genome version hg38|GRCh38 using STAR (Spliced Transcripts Alignment to a Reference) aligner (25) . Using gene models as reference, transcriptome alignments were generated and were used for transcript and gene quantification analysis.

Transcriptome alignments were then inspected for potential transcripts, annotated and quantified at transcript and gene level using RSEM software (26) . Transcripts were annotated based on gene model version GENCODE v27. Read counts were reported for all the genes found in GENCODE v27 database.

Gene counts were normalized by patient mean gene count and $\log_{10}$ normalized after adding 10^{-3}. PAM50 classifications were obtained from the genefu R package (27) , using the pam50.robust

method. Gene signatures were obtained as weighted means of the gene log expressions. UMAP analysis was performed on the 1000 more variable genes, if using more than 1000 genes. The immune score was obtained with the xCell R package (28). It is the sum of the concentrations of all immune cell types, as determined by xCell. Immune deconvolutions were also done with CIBERSORT (29).

Visualization of the biopsy site of the samples using UMAP showed that liver samples were clustered together (figure S28). We realized that the genes deriving this clustering were typically expressed in normal livers, hence we decided to remove those genes. Liver genes were determined by calculating the p-value of the association between gene expression in the metastasis and whether the biopsy site was liver, using the Mann-Whitney test. The 2855 genes with an FDR < 10^{-3} were deemed liver genes and reported supplementary file-liver genes. After this correction, visualization through UMAP did not highlight any clear clustering by site anymore. We avoided to correct for all sites as many had too few patients (e.g. brain), or were characterized by genes that are important for the biology of breast cancer (e.g. lymph nodes, characterized by higher expression of immune genes).

Statistical analyses

Statistical analyses were performed with the R software environment for statistical computing and graphics (R core team, 2020 https://www.R-project.org/).

Survival analyses used the Cox proportional hazard method, with the likelihood test. Time to event was left truncated between the time of diagnosis of the metastatic disease and the time of inclusion in AURORA. Median duration of follow-up was defined as interval between enrollment and death or between enrollment and "database lock" for patients still alive on Feb 7th 2020. To control TMB for other covariates, ANOVA tests between the Cox regression models including relevant clinical covariates and these covariates plus TMB were used. Relevant covariates include i) stage (T1/T2 vs. others), node status (pos/neg), grade (1-4) for primary disease and ii) time to relapse (<24 vs. ≥24 months), number of metastatic sites (< 3 vs. ≥ 3) and metastatic disease (enrolled before first line vs.

enrolled after first line vs. de novo) for metastatic disease. Mann-Whitney (for 2 groups) or Kruskal-Wallis (for 3 groups or more) tests were used to assess the significance of the relationship between a continuous variable and a categorical variable. All correlations are Spearman correlations. Significances between two categorical variables were assessed using the Fisher exact test.

The significance of over-representation of apparition of CNAs or mutations in the metastasis compared to the primary was computed on paired samples by first obtaining a baseline probability that a CNA/mutation appear in the metastasis, as the mean number of such appearance across all genes. As noise can also be gene specific, a second estimate was taken as the number of CNA/mutations for the gene that were present in the primary but not in the metastasis. A binomial test was then used using the maximum of those two probabilities.

For analyses bringing multiple results, p-values were corrected using the Benjamini-Hochberg procedure. Corrected p-values were considered as significant if below 0.05. Boxplots shown are standard R boxplots, hence boxes span the 1st to 3rd quartile, horizontal lines are the median, and the whiskers extend to the most extreme data point which is no more than 1.5 times the length of the box away from the box.

Molecular advisory board

An AURORA-dedicated MAB convenes remotely and provides annotations on detected variants in a subset of the target sequencing panel genes deemed to have clinical interest (the star genes, supplementary table 3). The first step of the annotation process is to determine by consensus between the members of the MAB whether the variant is likely to have a pathogenic effect at the molecular level, with pathogenicity being scored on a five-point scale (benign, likely benign, variant of uncertain significance, likely pathogenic, pathogenic). Several online resources are used to determine pathogenicity, including but not restricted to COSMIC (9) , ClinVar (30) , ExAC (4) and relevant articles indexed in PubMed. Frameshift, nonsense and canonical splice site mutations in tumor suppressor genes are annotated as likely pathogenic in the absence of additional evidence.

The second step of the annotation process is to search clinical trial databases such as ClinicalTrials.gov to identify and inform local investigators about potential genotype-driven clinical trials recruiting patients in Europe. These trials include, but are not limited to, targets such as *PIK3CA*, *AKT1*, *ERBB2*, *ERBB3*, homologous recombination genes, and mismatch repair genes, including germline variants that warrant confirmation using an accredited assay in clinical practice. The consolidated genomic report is sent back to the investigators via the IT platform. The impact of the genomic results on the choice of therapy is left at the discretion of the treating physician. Furthermore, a disclaimer highlights the fact that the results are generated in a research environment.

Data availability

Instructions to access the manuscript processed data for reproducibility purposes are available at the webpage http://aurora.bigagainstbreastcancer.org and can be obtained upon signature of an appropriate data transfer agreement subject to applicable laws. Instructions to access processed or raw manuscript data to perform original research are also available on the webpage and investigators can contact aurora.researchproposals@bigagainstbc.org for enquiries. Access to data for research will be granted upon review of a project proposal & endorsement by the study Steering Committee, and after entering into an appropriate data access agreement subject to applicable laws.

Code availability

All the software used for the analyses is publicly available and can be accessed from the links provided in the cited references.

References

1. Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol. 2010;28:2784–95.

2. Wolff AC, Hammond MEH, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol. 2013;31:3997–4013.

3. Maetens M, Brown D, Irrthum A, Aftimos P, Viale G, Loibl S, et al. The AURORA pilot study for molecular screening of patients with advanced breast cancer-a study of the breast international group. NPJ Breast Cancer. 2017;3:23.

4. Exome Aggregation Consortium, Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.

5. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proceedings of the National Academy of Sciences. 2010;107:16910–5.

6. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res. 2016;44:e131–e131.

7. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017;171:1029-1041.e21.

8. Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A compendium of mutational cancer driver genes. Nat Rev Cancer. 2020;

9. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Research. 2019;47:D941–7.

10. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer. 2018;18:696–705.

11. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–9.

12. Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40:W452-457.

13. Shihab HA, Gough J, Cooper DN, Day INM, Gaunt TR. Predicting the functional consequences of cancer-associated amino acid substitutions. Bioinformatics. 2013;29:1504–10.

14. Fernandez EM, Eng K, Beg S, Beltran H, Faltas BM, Mosquera JM, et al. Cancer-Specific Thresholds Adjust for Whole Exome Sequencing-based Tumor Mutational Burden Distribution. JCO Precis Oncol. 2019;3.

15. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32:2847–9.

16. 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. 2016;7:11479.

17. Gulhan DC, Lee JJ-K, Melloni GEM, Cortés-Ciriano I, Park PJ. Detecting the mutational signature of homologous recombination deficiency in clinical samples. Nat Genet. 2019;51:912–9.

18. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61–70.

19. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell. 2018;173:291-304.e6.

20. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213–9.

21. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016;44:e71.

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

23. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2:401–4.

24. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6:pl1.

25. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

26. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.

27. Gendoo DMA, Ratanasirigulchai N, Schröder MS, Paré L, Parker JS, Prat A, et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. Bioinformatics. 2016;32:1097–9.

28. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017;18:220.

29. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol. 2018;1711:243–59.

30. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Research. 2018;46:D1062–7.

31. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer Cell. 2017;32:169-184.e7.

32. Robinson DR, Wu Y-M, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. Nature. 2017;548:297–303.

33. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med. 2017;23:703–13.

34. Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, Vasan N, et al. The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. Cancer Cell. 2018;34:427-438.e6.

35. Bertucci F, Ng CKY, Patsouris A, Droin N, Piscuoglio S, Carbuccia N, et al. Genomic characterization of metastatic breast cancers. Nature. 2019;569:560–4.

36. Angus L, Smid M, Wilting SM, van Riet J, Van Hoeck A, Nguyen L, et al. The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. Nat Genet. 2019;51:1450–8.

Supplementary tables, figures, and files

Supplementary table 1

Primary tumor subtype			
IHC		PAM50	
HR+/HER2-	246 (65%)	Luminal A	47 (12.5%)
TNBC	73 (19%)	Luminal B	79 (21%)
HER2+	60 (16%)	HER2-Enriched	24 (6.5%)
		Basal	52 (14%)
		Normal-like	9 (2%)
		N/A	168 (44%)
Time from early diagnosis to first relapse (days)			
Median (IQ range)	906 (1964)		
Metastatic treatment received			
Metastatic treatment naïve	277 (73%)		
1st line systemic treatment	102 (27%)		
Classes of treatments received prior to AURORA			
(Neo)-adjuvant chemotherapy	238 (63%)		
(Neo)-adjuvant endocrine therapy	201 (53%)		
(Neo)-adjuvant anti-HER2 therapy	31 (8%)		
Chemotherapy in the metastatic setting	69 (18%)		
Endocrine therapy in the metastatic setting	58 (15%)		
Anti-HER2 therapy in the metastatic setting	9 (2%)		
CDK4/6 inhibitor in the metastatic setting	5 (1%)		

Menopausal status	
Pre-Menopausal	89 (23%)
Peri-Menopausal	20 (5%)
Post-Menopausal	268 (71%)
Unknown	2 (<1%)
Age (years)	
Median (IQ range)	56 (19)
Histology (primary tumor)	
Ductal	313 (83%)
Lobular	36 (10%)
Ductal-Lobular	5 (1%)
Other	20 (5%)
Unknown	5 (1%)
De novo status	
De novo treatment naive	49 (13%)
De novo after one treatment line	28 (7%)
Metastatic relapse	302 (80%)
Bone-only status	
Patients with bone-only metastases	9 (2%)

Pathological and clinical variables of the tumor, and clinical features of the patients are reported.

Supplementary table 2

Breast cancer subtype	Genomic alteration	Matched therapy
TNBC	NOTCH1 mutation	Gamma-secretase inhbitor
ER+/HER2-	PIK3CA mutation	PIK3CA inhibitor
ER+/HER2-	PIK3CA mutation	PIK3CA inhibitor
ER+/HER2-	PIK3CA mutation	PIK3CA inhibitor
ER+/HER2-	Double PIK3CA mutations	PIK3CA inhibitor
ER+/HER2-	PIK3CA mutation	PIK3CA inhibitor
ER+/HER2-	Double PIK3CA mutations	PIK3CA inhibitor
TNBC	PIK3CA mutation	PIK3CA inhibitor
ER+/HER2-	ERBB2 mutation and ERBB3 mutation	Pan-HER tyrosine kinase inhibitor
TNBC	ERBB2 mutation	Pan-HER tyrosine kinase inhibitor
ER+/HER2-	Myc amlification	BET inhibitor
ER+/HER2-	ESR1 mutation	Oral selective estrogen receptor degrader (SERD)
HER2+	gBRCA2 mutation	PARP inhibitor
ER+/HER2-	gBRCA2 mutation	PARP inhibitor
ER+/HER2-	gBRCA2 mutation	PARP inhibitor
ER+/HER2-	gBRCA2 mutation	PARP inhibitor
ER+/HER2-	gPALB2 mutation	PARP inhibitor
ER+/HER2-	gBRCA2 mutation	PARP inhibitor
ER+/HER2-	sBRCA2 mutation and gBRCA2 mutation	PARP inhibitor
ER+/HER2-	gBRCA2 mutation	PARP inhibitor
ER+/HER2-	gBRCA2 mutation	PARP inhibitor
TNBC	gBRCA1 mutation	PARP inhibitor
ER+/HER2-	gBRCA1 mutation	PARP inhibitor
TNBC	sBRCA1 mutation	PARP inhibitor
TNBC	EGFR amplification	Anti-EGFR monoclonal antibody

Every line is a patient. We report the primary tumor subtype, the molecular alteration that informed therapy, and the targeted therapy that was prescribed.

Supplementary table 3

ABL1	ABL2	ACVR2A	ADAMTS20	AFF1	AFF3	AKAP9	*AKT1*	*AKT2*	*AKT3*	*ALK*
APC	*AR*	ARID1A	ARID2	ARNT	ASXL1	ATF1	*ATM*	ATR	ATRX	AURKA
AURKB	AURKC	AXL	BAI3	BAP1	BCL10	BCL11A	BCL11B	BCL2	BCL2L1	BCL2L2
BCL3	BCL6	BCL9	BCR	BIRC2	BIRC3	BIRC5	BLM	BLNK	BMPR1A	*BRAF*
BRCA1	***BRCA2***	BRD3	BRIP1	BTK	BUB1B	CARD11	CASC5	CBL	CCND1	CCND2
CCNE1	CD79A	CD79B	CDC73	CDH1	CDH11	CDH2	CDH20	CDH5	CDK12	*CDK4*
CDK6	CDK8	*CDKN2A*	*CDKN2B*	CDKN2C	CEBPA	*CHEK1*	*CHEK2*	CIC	CKS1B	CMPK1
COL1A1	CRBN	CREB1	CREBBP	CRKL	CRTC1	CSF1R	CSMD3	CTNNA1	CTNNB1	CYLD
CYP2C19	CYP2D6	DAXX	DCC	DDB2	DDIT3	*DDR2*	DEK	DICER1	DNMT3A	**DPYD**
DST	*EGFR*	EML4	EP300	EP400	EPHA3	EPHA7	EPHB1	EPHB4	EPHB6	*ERBB2*
ERBB3	*ERBB4*	ERCC1	ERCC2	ERCC3	ERCC4	ERCC5	ERG	*ESR1*	ETS1	ETV1
ETV4	EXT1	EXT2	EZH2	FAM123B	FANCA	FANCC	FANCD2	FANCF	FANCG	FAS
FBXW7	*FGFR1*	*FGFR2*	*FGFR3*	*FGFR4*	FH	FLCN	FLI1	FLT1	*FLT3*	FLT4
FN1	FOXL2	FOXO1	FOXO3	FOXP1	FOXP4	FZR1	G6PD	GATA1	GATA2	GATA3
GDNF	GNA11	GNAQ	GNAS	GPR124	GRM8	GUCY1A2	HCAR1	HIF1A	HLF	HNF1A
HOOK3	*HRAS*	HSP90AA1	HSP90AB1	ICK	*IDH1*	*IDH2*	*IGF1R*	IGF2	IGF2R	IKBKB

IKBKE	IKZF1	IL2	IL21R	IL6ST	IL7R	ING4	IRF4	IRS2	ITGA10	ITGA9
ITGB2	ITGB3	JAK1	JAK2	JAK3	JUN	KAT6A	KAT6B	KDM5C	KDM6A	KDR
KEAP1	KIT	KLF6	*KRAS*	LAMP1	LCK	LIFR	LPHN3	LPP	LRP1B	LTF
LTK	MAF	MAFB	MAGEA1	MAGI1	MALT1	MAML2	MAP2K1	MAP2K2	MAP2K4	MAP3K7
MAPK1	MAPK8	MARK1	MARK4	MDB1	MCL1	MDM2	MDM4	**MEN1**	MET	MITF
MLH1	MLL	MLL2	MLL3	MLLT10	MMP2	MN1	MPL	MRE11A	***MSH2***	***MSH6***
MTOR	MTR	MTRR	MUC1	**MUTYH**	MYB	MYC	MYCL1	MYCN	MYD88	**MYH11**
MYH9	NBN	NCOA1	NCOA2	NCOA4	NF1	***NF2***	NFE2L2	NFKB1	NFKB2	NIN
NKX2	NLRP1	*NOTCH1*	NOTCH2	NOTCH4	NPM1	NRAS	NSD1	NTRK1	NTRK3	NUMA1
NUP214	NUP98	PAK3	**PALB2**	PARP1	PAX3	PAX5	PAX7	PAX8	PBRM1	PBX1
PDE4DIP	PDGFB	*PDGFRA*	*PDGFRB*	PER1	PGAP3	PHOX2B	PIK3C2B	PIK3CA	PIK3CB	PIK3CD
PIK3CG	*PIK3R1*	PIK3R2	PIM1	PKHD1	PLAG1	PLCG1	PLEKHG5	PML	*PMS1*	*PMS2*
POT1	POU5F1	PPARG	PPP2R1A	PRDM1	PRKAR1A	PRKDC	PSIP1	PTCH1	***PTEN***	PTGS2
PTPN11	PTPRD	PTPRT	RAD50	*RAF1*	RALGDS	RARA	***RB1***	RECQL4	REL	***RET***
RHOH	RNASEL	RNF2	RNF213	ROS1	RPS6KA2	RRM1	RUNX1	RUNX1T1	SAMD9	SBDS
SDHA	**SDHB**	**SDHC**	**SDHD**	SEPT9	SETD2	SF3B1	SGK1	SH2D1A	SMAD2	SMAD4
SMARCA4	SMARCB1	SMO	SMUG1	SOCS1	SOX11	SOX2	*SRC*	SSX1	**STK11**	STK36
SUFU	SYK	SYNE1	TAF1	TAF1L	TAL1	TBX22	TCF12	TCF3	TCF7L1	TCF7L2
TCL1A	TET1	TET2	TFE3	**TGFBR2**	TGM7	THBS1	TIMP3	TLR4	TLX1	TNFAIP3

TNFRSF14	TNK2	TOP1	***TP53***	TPR	TRIM24	TRIM33	TRIP11	TRRAP	***TSC1***	***TSC2***
TSHR	UBR5	*UGT1A1*	USP9X	***VHL***	WAS	WHSC1	WRN	**WT1**	XPA	XPC
XPO1	XRCC2	ZNF384	ZNF521							

The list of genes composing the TGS gene panel used for tissue samples is reported. Genes for which germline mutations are reported are in bold. The alpha list of target genes annotated by the molecular advisory board are in italics.

Supplementary table 4

AKT1	ALK	AR	BRAF	BTK	CTNNB1	DDR2	EGFR
ERBB2	ESR1	EZH2	FBXW7	FGFR1	FGFR2	FGFR3	FOXL2
GNA11	GNAQ	GNAS	HRAS	IDH1	IDH2	JAK2	JAK3
KIT	KRAS	MAP2K1	MAP2K2	MET	MPL	MTOR	NPM1
NRAS	PDGFRA	PIK3CA	PTEN	RAF1	RET	ROS1	TP53

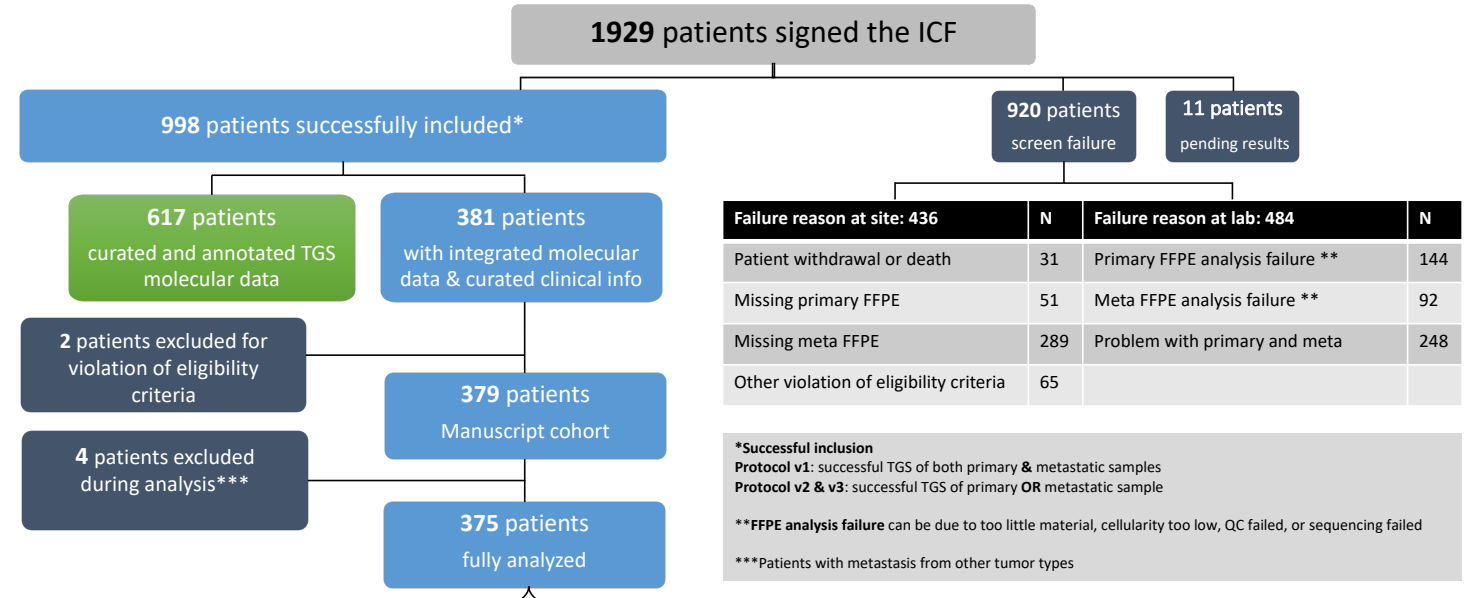
The list of genes composing the TGS gene panel used for liquid biopsies is reported.

Supplementary Data Files

Supplementary file – liver genes

List of liver genes excluded from the analysis of transcriptomic data.

Supplementary figure S1



Failure reason at site: 436	N	Failure reason at lab: 484	N
Patient withdrawal or death	31	Primary FFPE analysis failure **	144
Missing primary FFPE	51	Meta FFPE analysis failure **	92
Missing meta FFPE	289	Problem with primary and meta	248
Other violation of eligibility criteria	65		

***Successful inclusion**
Protocol v1: successful TGS of both primary **&** metastatic samples
Protocol v2 & v3: successful TGS of primary **OR** metastatic sample

****FFPE analysis failure** can be due to too little material, cellularity too low, QC failed, or sequencing failed

*****Patients with metastasis from other tumor types

	Techniques	All patients (both | prim | meta)	HR+/HER2- (both | prim | meta)	HER2+ (both | prim | meta)	TNBC (both | prim | meta)	De novo untreated (both | prim | meta)	De novo treated (both | prim | meta)
DNA-sequencing	TGS (411 genes) Ion Proton	242 | 19 | 114	150 | 14 | 82	43 | 0 | 15	49 | 5 | 17	44 | 3 | 6	23 | 2 | 7
RNA Sequencing	Illumina Hi-Seq 2500	152 | 53 | 109	97 | 34 | 72	24 | 8 | 17	31 | 11 | 20	31 | 8 | 9	18 | 3 | 7
CNV	Affymetrix array OncoScan	67 | 110 | 67	47 | 74 | 47	9 | 14 | 10	11 | 22 | 10	8 | 13 | 10	6 | 10 | 6

Data as of 13 June 2020

Figure S1: AURORA consort diagram

Consort diagram representing the status of the program as of 13 June 2020. For the cohort of patients analyzed in the manuscript, details are provided for the availability of TGS, RNASeq and CNV data overall and for subpopulations of interest, and according to their availability in primary only, metastasis only or in both samples. For the patients considered as screening failure, reasons for failure at site and at the central lab are provided.

Supplementary figure S2

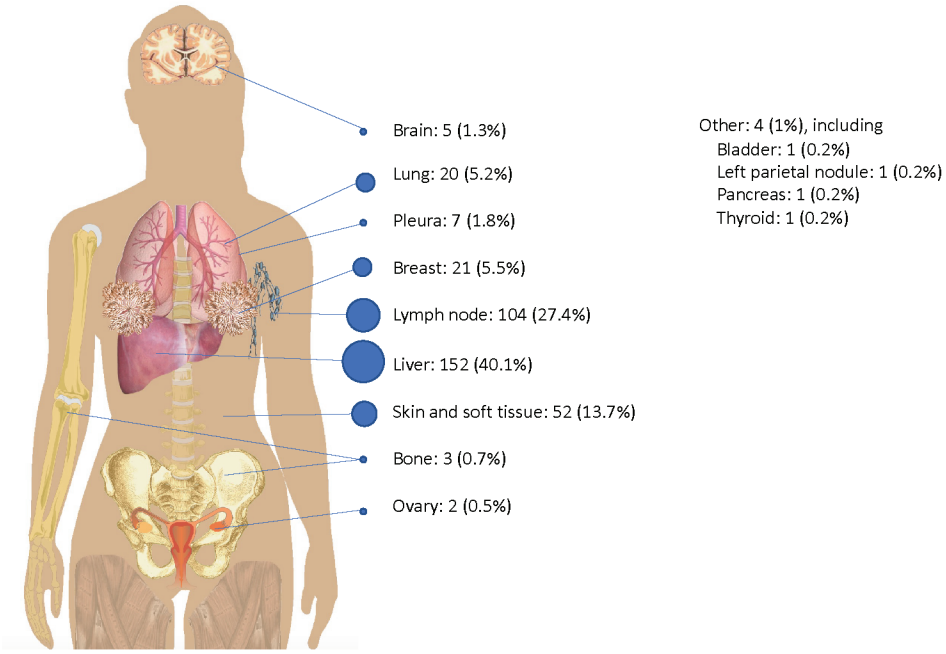


Figure S2: AURORA biopsy sites

AURORA biopsy sites for the 379 patients of the manuscript cohort. Nine patients were enrolled as having a bone only disease and no metastatic biopsies were available.

Supplementary Figure S3

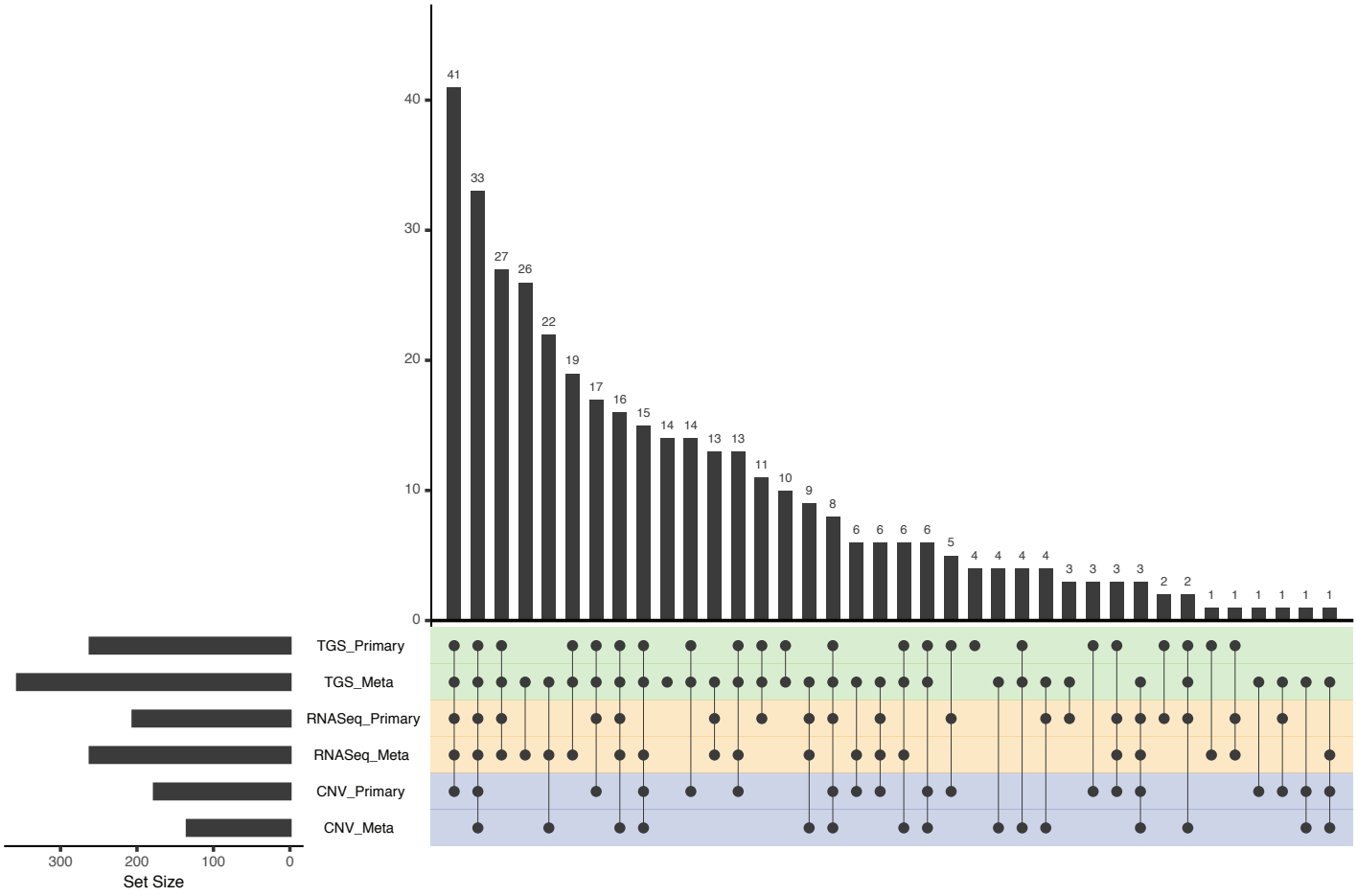
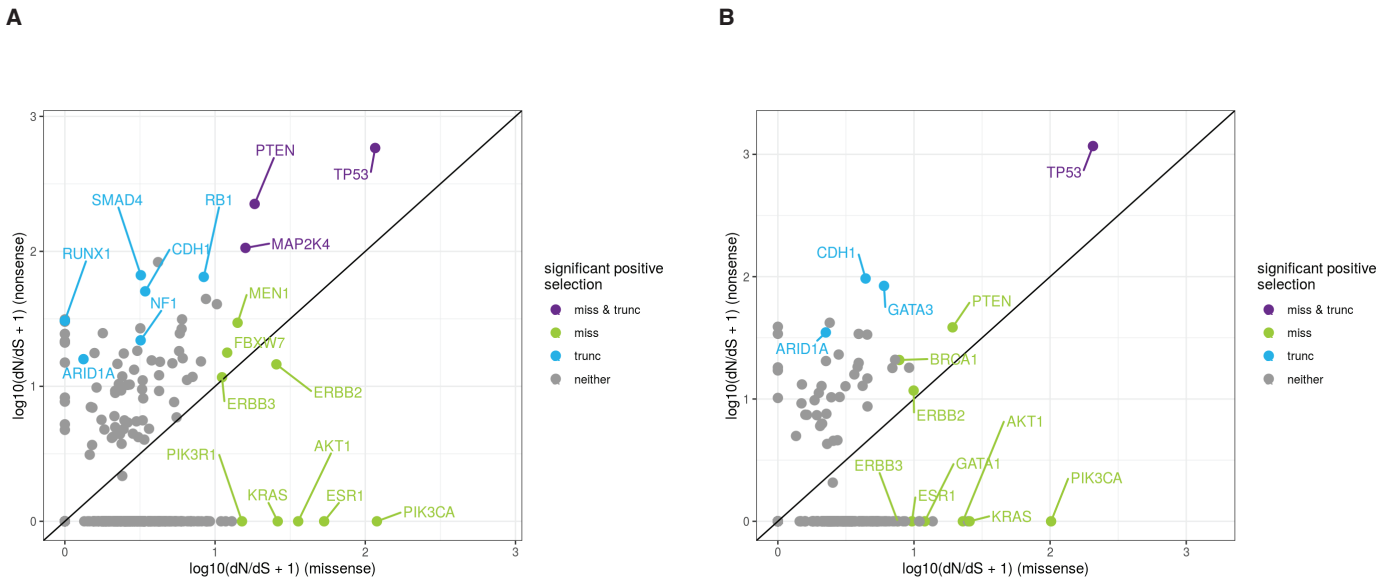
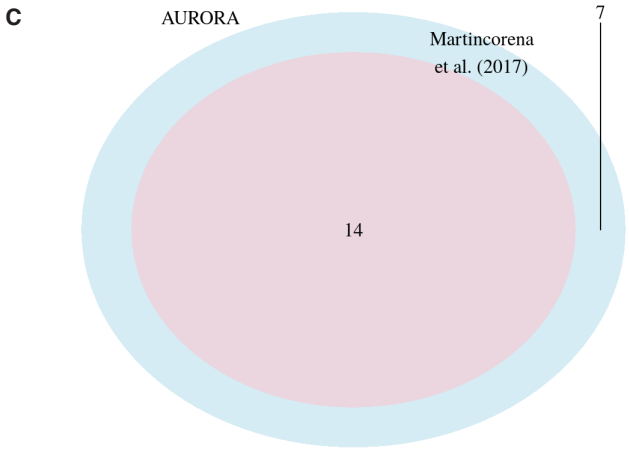


Figure S3: data availability

Upset diagram of the molecular data availability for the 375 AURORA patients fully analyzed in the manuscript.

Supplementary figure S4





AURORA only
ERBB3
ESR1
FBXW7
GATA1
KRAS
MEN1
NF1

Intersection
PIK3CA
TP53
GATA3
CDH1
PTEN
RUNX1
MAP2K4
ARID1A
PIK3R1
ERBB2
RB1
AKT1
BRCA1
SMAD4

Figure S4: cancer driver identification in the AURORA dataset

A: Scatter plot of log transformed dN/dS values for missense and truncating substitutions for primary samples. Labeled are genes with significant positive selection on missense mutations and/or truncating substitutions; B: scatter plot of log transformed dN/dS values for missense and truncating substitutions for metastatic samples. Labeled are genes with significant positive selection on missense mutations and/or truncating substitutions, C: Venn diagram showing the overlap between positively selected genes from this study (AURORA, light blue) and the breast cancer genes from Martincorena, et al. (2017) also included in the AURORA TGS panel (pink).

Supplementary Figure S5

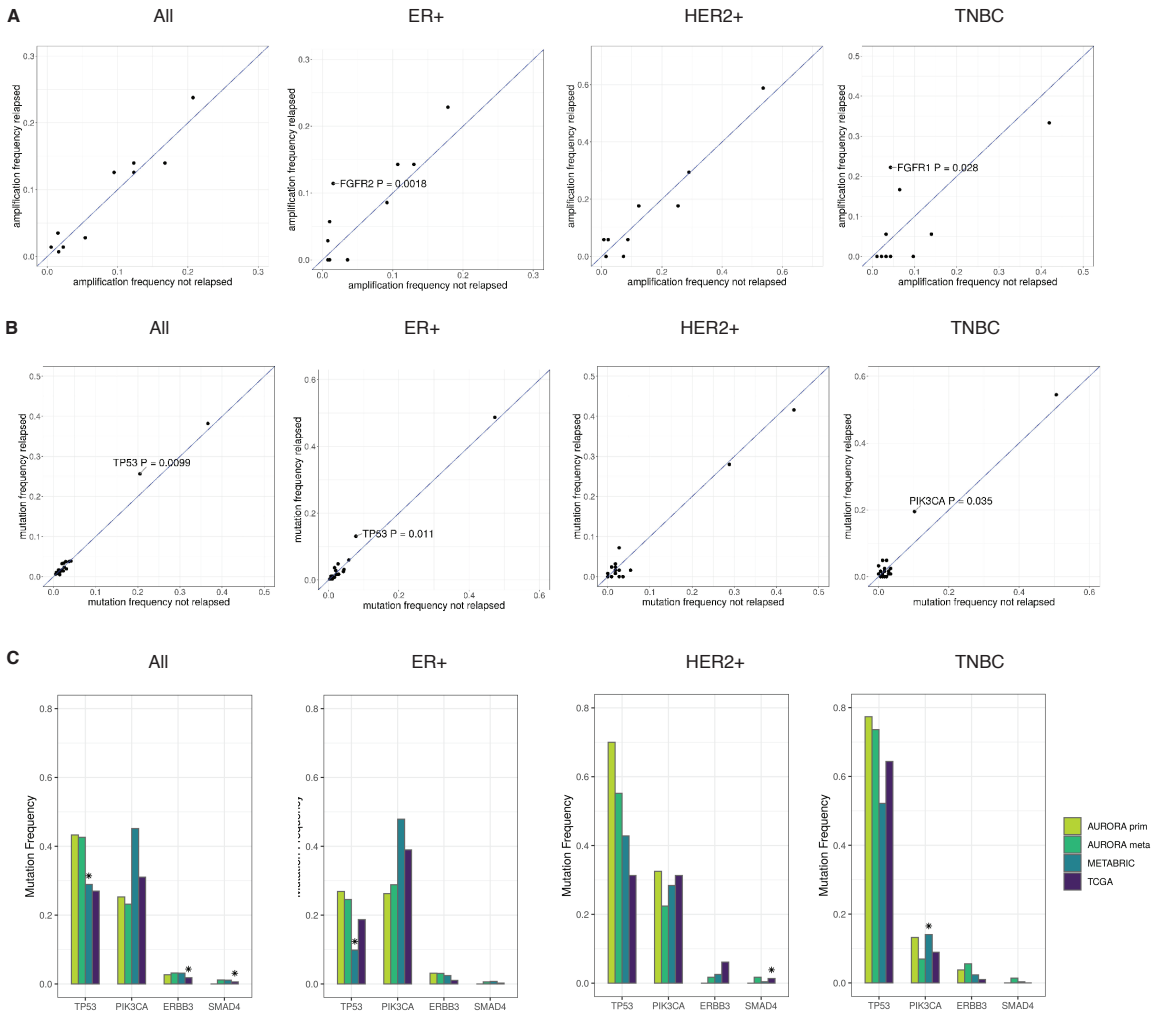


Figure S5: intra- and inter-dataset comparison of the prevalence of point mutations in the driver genes identified in AURORA

A: Scatter plots showing the frequency of alterations of driver genes identified in AURORA for TCGA-BRCA patients who relapsed (y-axis) versus patients that did not relapse (x-axis) for all types (left), HR+/HER2- cases (middle-left), HER2+ (middle right) and TNBC (right). Only genes showing statistically significant difference estimated by proportion test are shown.
B: Scatter plots showing the frequency of alterations of driver genes identified in AURORA for METABRIC patients who relapsed (y-axis) versus patients that did not relapse (x-axis) for all types (left), HR+/HER2- cases (middle-left), HER2+ (middle right) and TNBC (right). Only genes showing statistically significant difference estimated by proportion test are shown.
C: Bar plots showing the frequency of alteration of driver genes identified as statistically significant in patients who relapsed from TCGA-BRCA and METABRIC datasets in primary and metastatic samples from AURORA, METABRIC and TCGA for all types (left), HR+/HER2- cases (middle-left), HER2+ (middle right) and TNBC (right). The asterisk above the bar refers to the analysis from which the gene was identified as statistically significant in relapsed versus not-relapsed patients.

Supplementary Figure S6

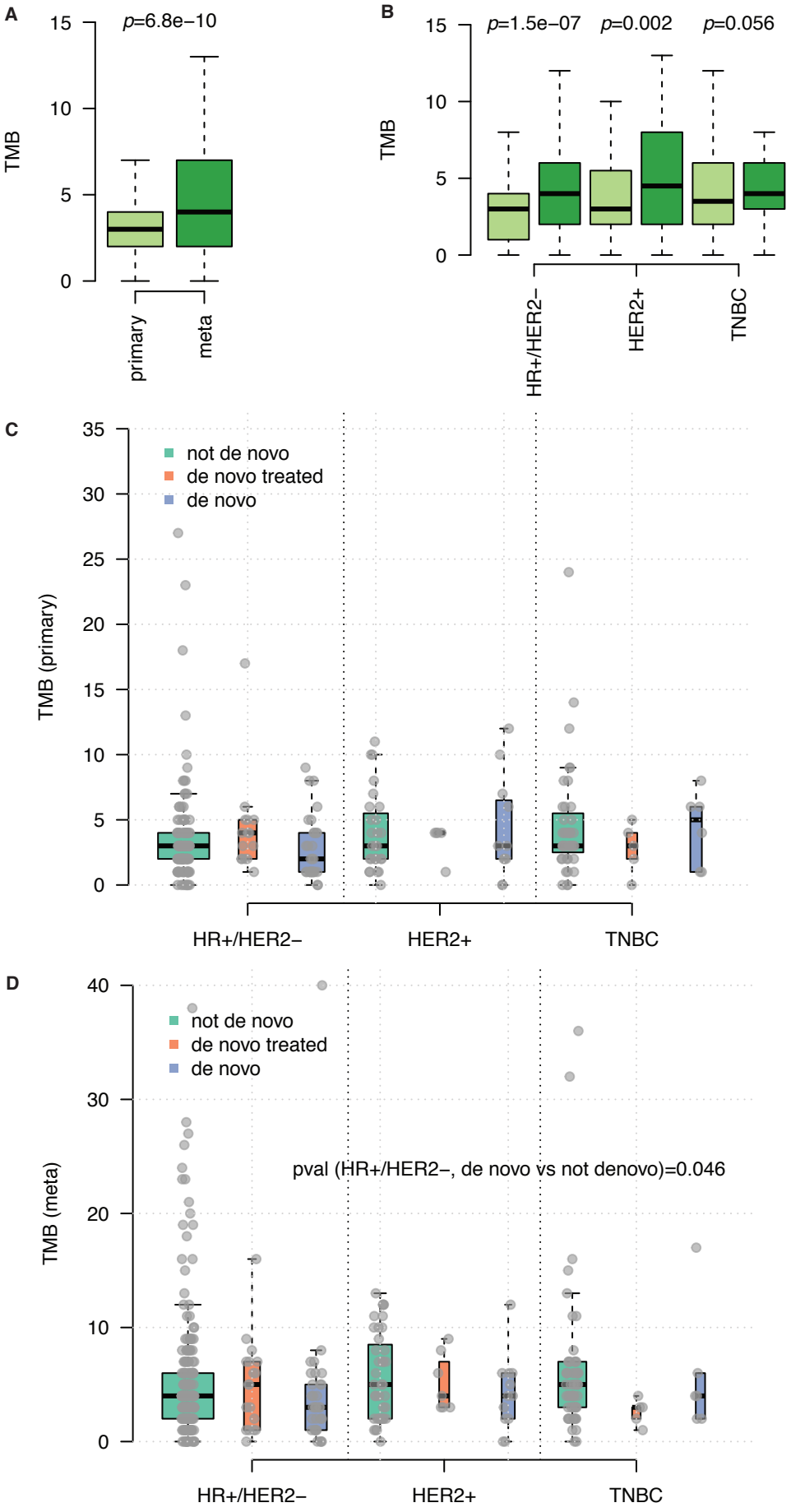


Figure S6: TMB in primary and metastatic samples

A: Box plots showing the distribution of TMB estimated in primary (light green) and metastatic (dark green) samples and B: across the different types (HR+/HER2-, HER2+, TNBC). C: Box plots showing the distribution of TMB estimated in primary samples in patients with not de novo, de novo treated and de novo disease across the different types. D: Box plots showing the distribution of TMB estimated in metastatic samples in patients with not de novo, de novo treated and de novo disease across the different types. P-values are estimated by Wilcoxon-Mann-Whitney test.

Supplementary figure S7

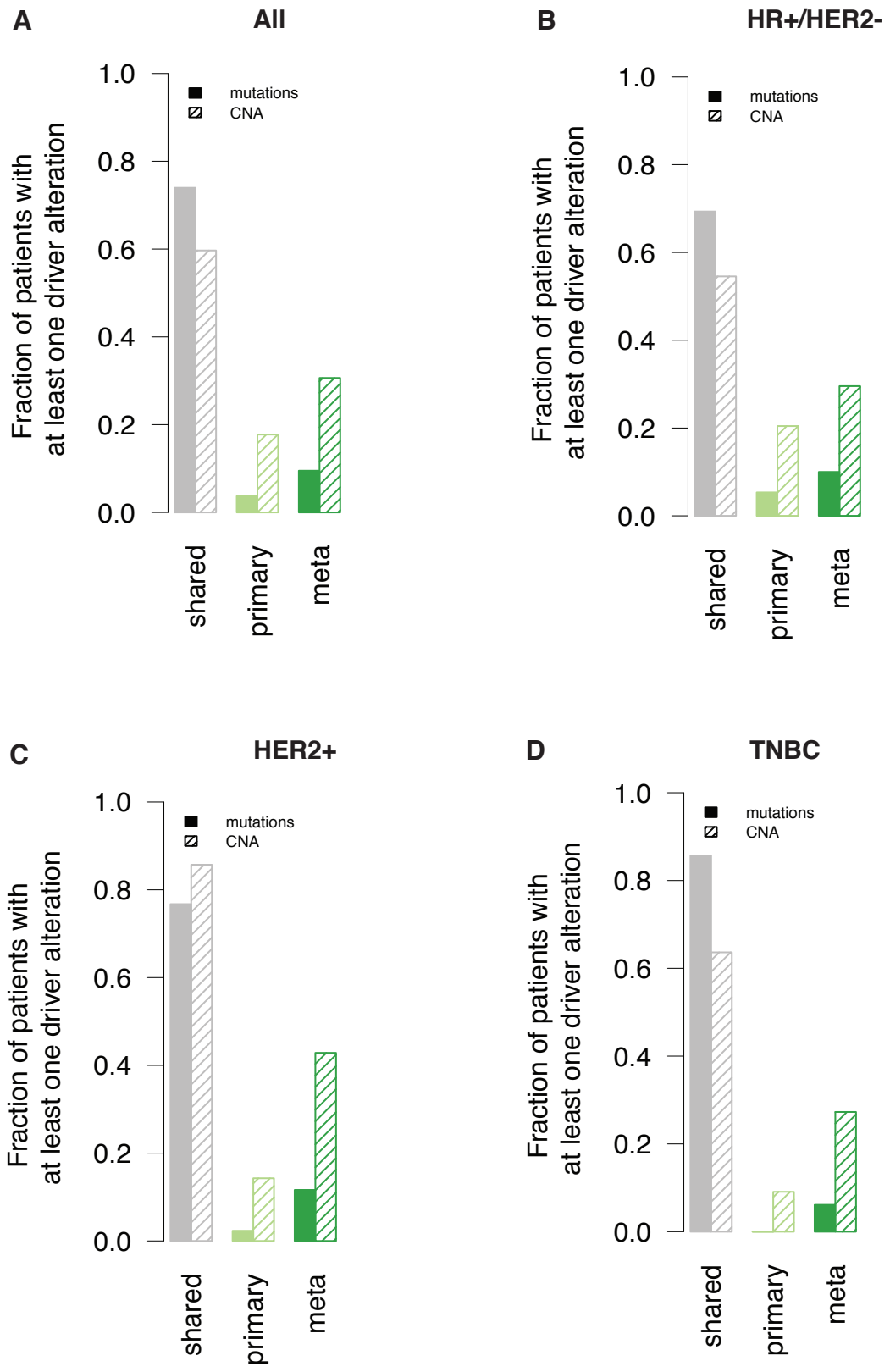
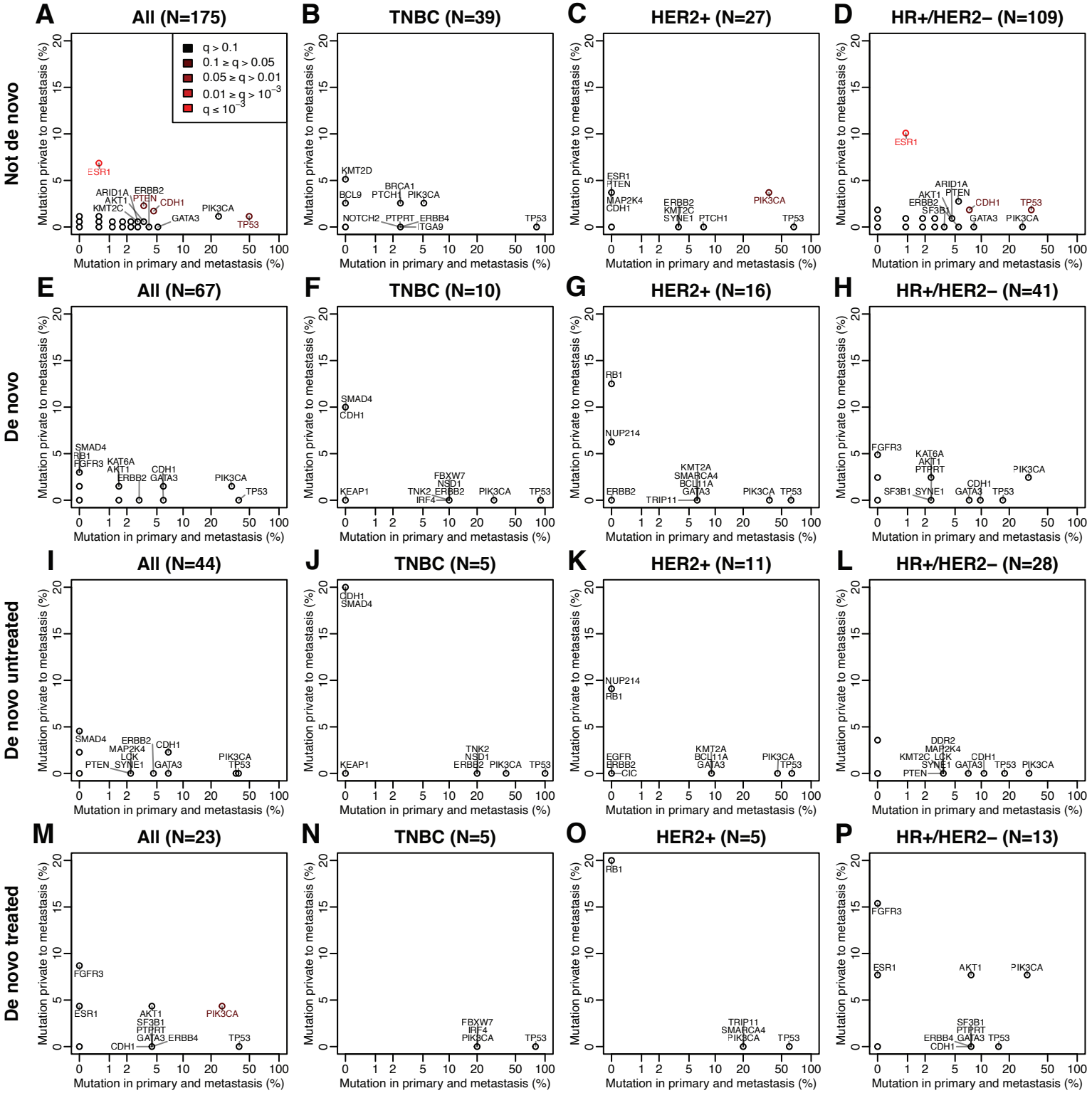


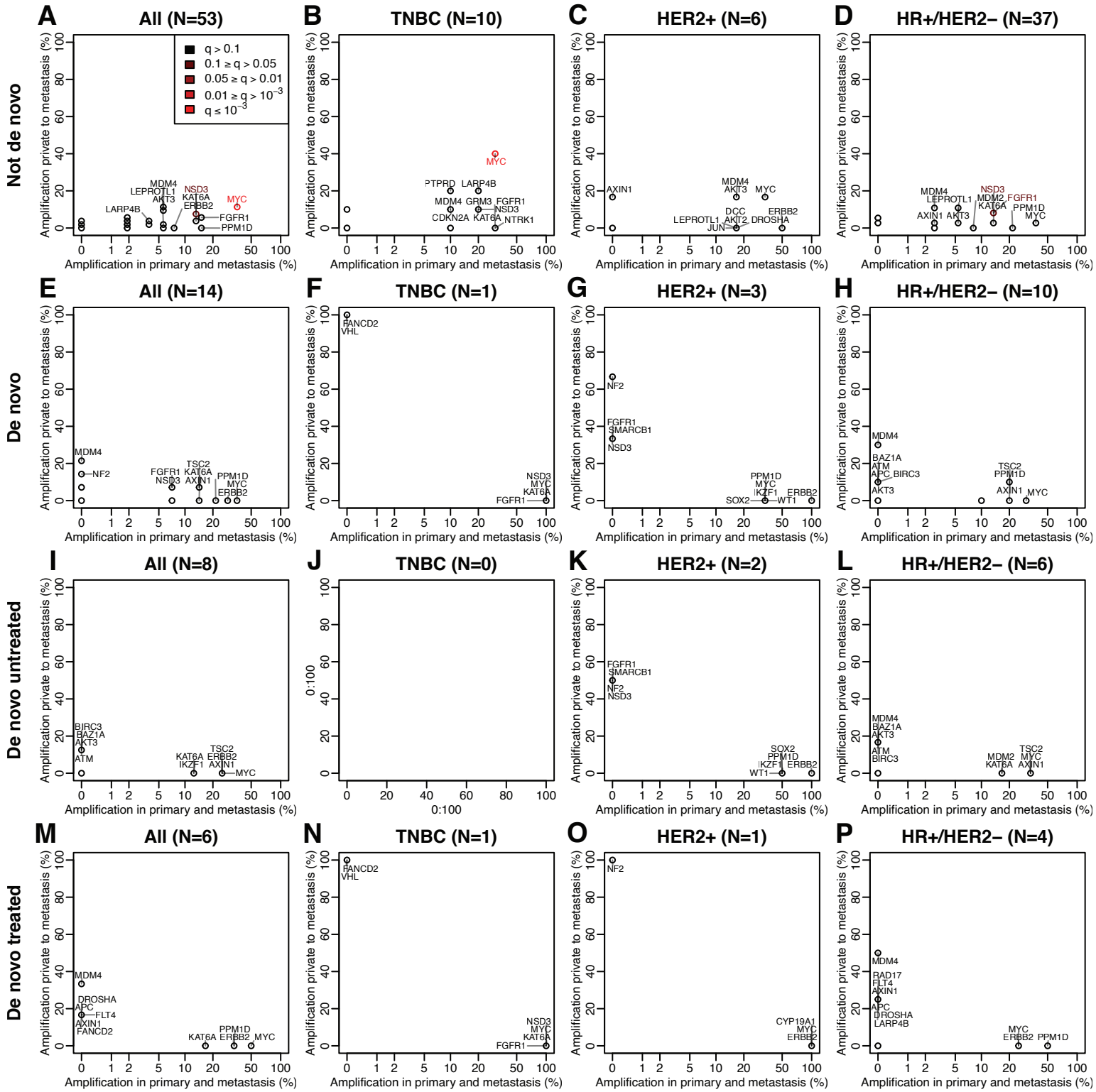
Figure S7: fractions of patients with at least one private alteration

Bar plot of the fraction of patients with available TGS for primary and metastatic samples (paired) (N=242) with at least one driver mutation in driver genes and amplification in oncogenes based on ASCAT (N=67) considering shared (grey), private to primary (light green) and private to meta (dark green) events for A: all subtypes, B: HR+/HER2-, C: HER2+, D: TNBC.

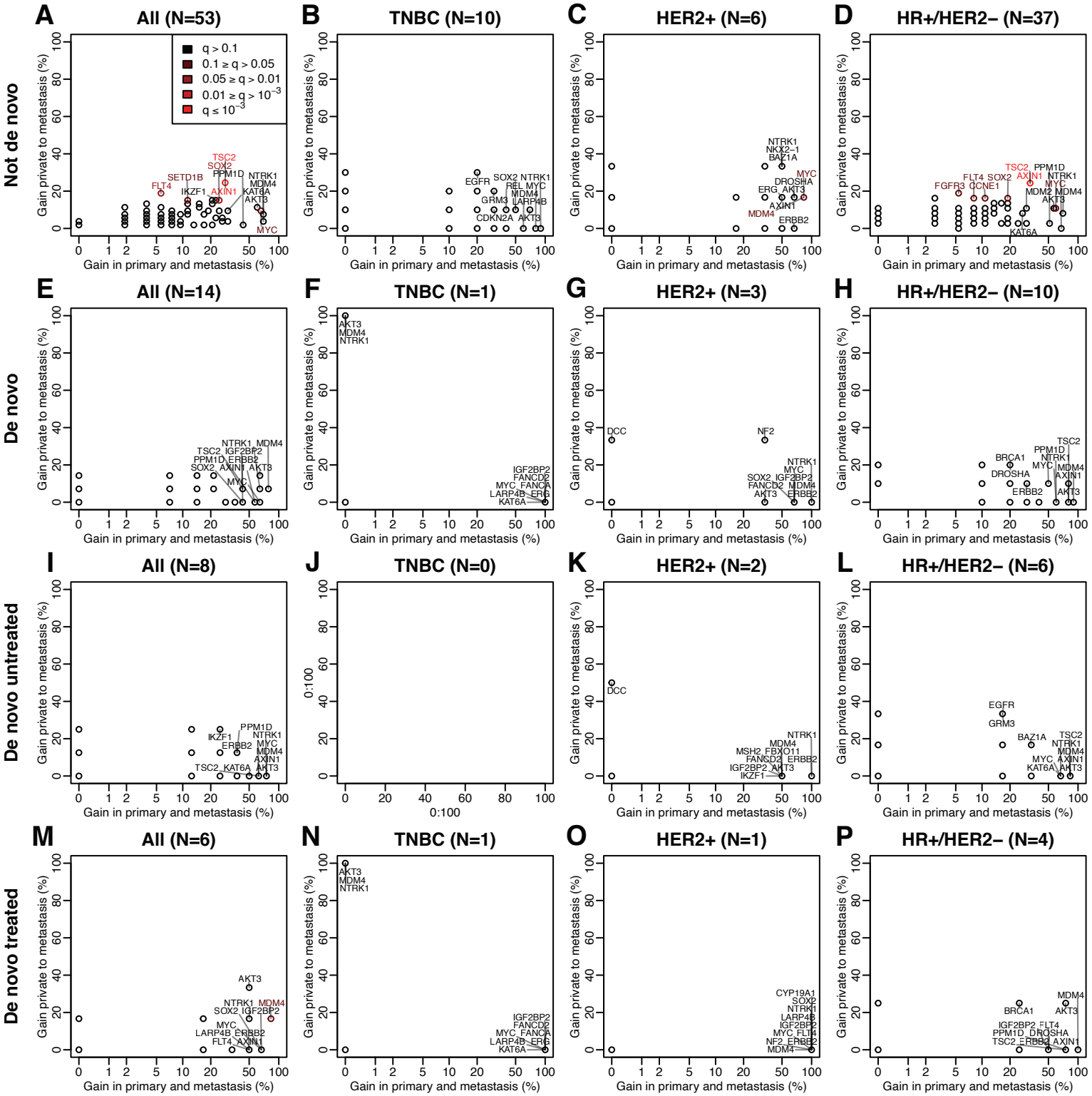
Supplementary figure S8



Supplementary figure S9



Supplementary figure S10



Supplementary figure S11

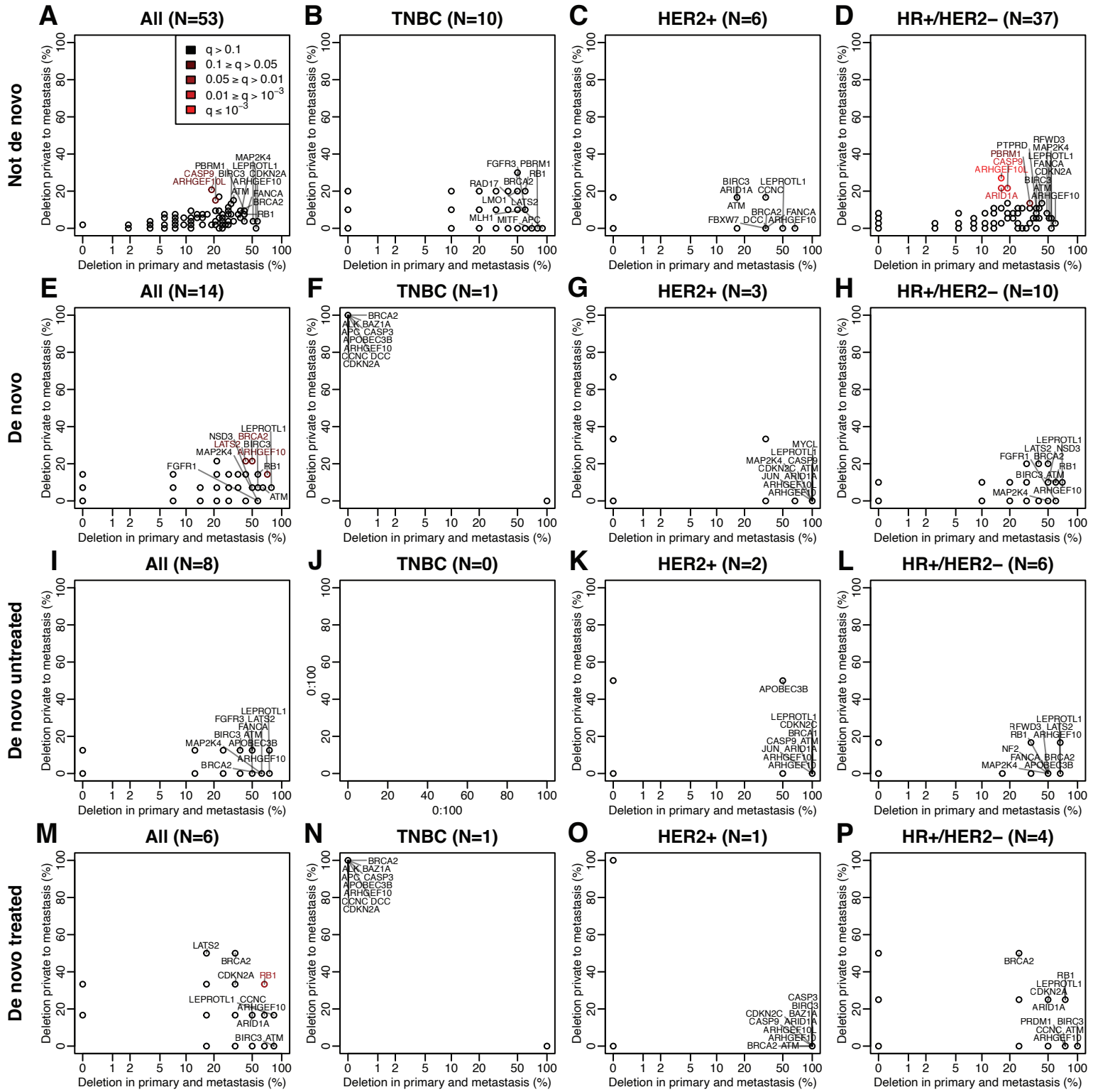


Figure S8 – S11

Comparison for each gene of the proportion of tumors with truncal alterations with the proportion of tumors with alterations private to the metastasis, showing the differences between de novo untreated, de novo treated and not de novo patients. Figure S8 is for point mutations (not de novo - A: all subtypes, B: TNBC, C: HER2+, D: HR+/HER2-; de novo - E: all subtypes, F: TNBC, G: HER2+, H: HR+/HER2-; de novo untreated - I: all subtypes, J: TNBC, K: HER2+, L: HR+/HER2-; de novo treated - M: all subtypes, N: TNBC, O: HER2+, P: HR+/HER2-), Figure S9 for amplifications (normalized CN>4; not de novo - A: all subtypes, B: TNBC, C: HER2+, D: HR+/HER2-; de novo - E: all subtypes, F: TNBC, G: HER2+, H: HR+/HER2-; de novo untreated - I: all subtypes, J: TNBC, K: HER2+, L: HR+/HER2-; de novo treated - M: all subtypes, N: TNBC, O: HER2+, P: HR+/HER2-), Figure S10 for gains (normalized CN>2.5; not de novo - A: all subtypes, B: TNBC, C: HER2+, D: HR+/HER2-; de novo - E: all subtypes, F: TNBC, G: HER2+, H: HR+/HER2-; de novo untreated - I: all subtypes, J: TNBC, K: HER2+, L: HR+/HER2-; de novo treated - M: all subtypes, N: TNBC, O: HER2+, P: HR+/HER2-) and Figure S11 for deletions (normalized CN<1.5; not de novo - A: all subtypes, B: TNBC, C: HER2+, D: HR+/HER2-; de novo - E: all subtypes, F: TNBC, G: HER2+, H: HR+/HER2-; de novo untreated - I: all subtypes, J: TNBC, K: HER2+, L: HR+/HER2-; de novo treated - M: all subtypes, N: TNBC, O: HER2+, P: HR+/HER2-). All plots are on paired samples, each point representing the percentage of tumors with an alteration common between the primary and the metastasis versus the percentage of tumors with an alteration found only in the metastasis. The points are colored in function of their q-values, which assess whether a given alteration is more often private to the metastasis than expected by the play of chance, corrected for multiple testing by panel.

Supplementary figure S12

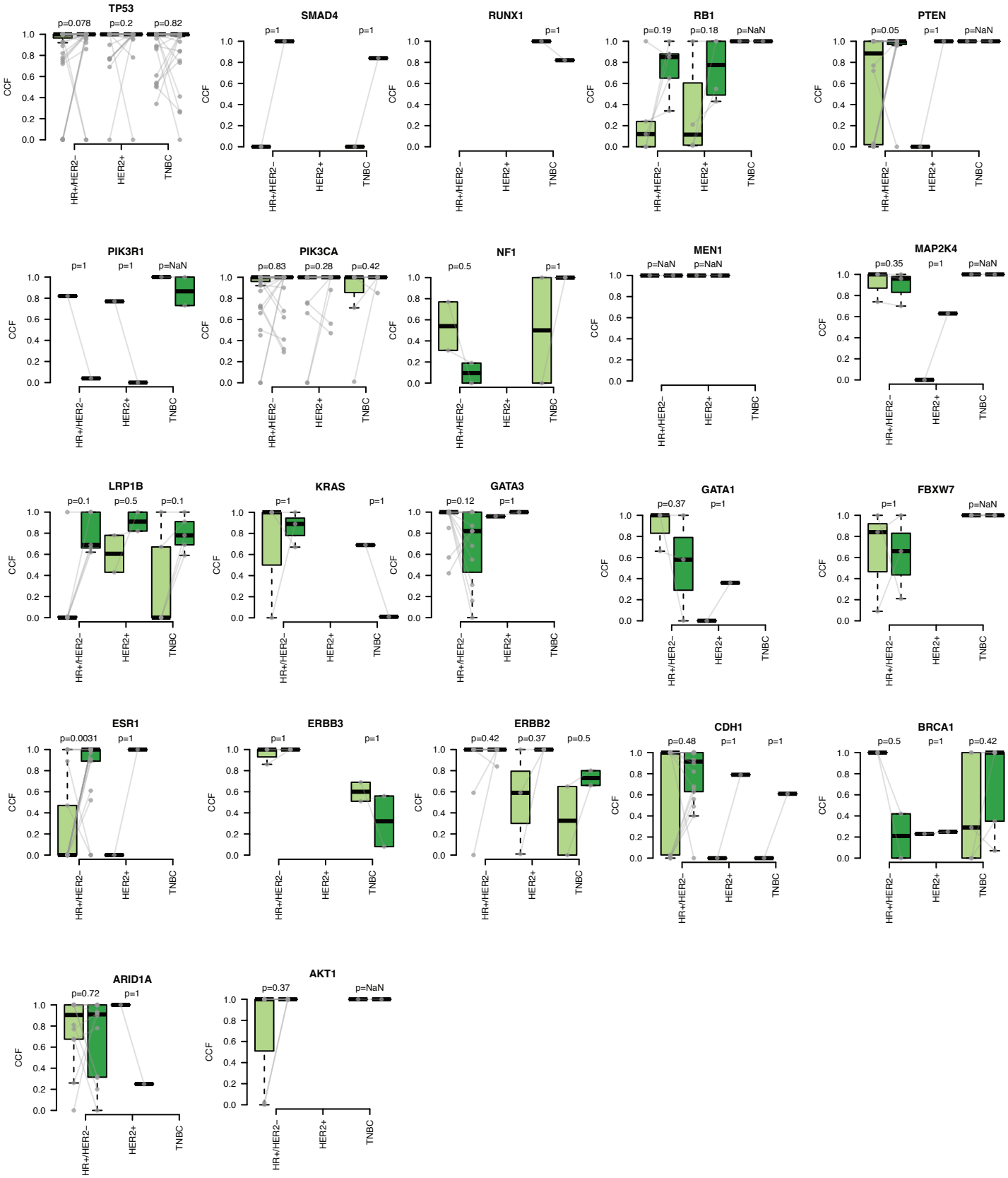


Figure S12: CCF for driver genes in paired primary and metastatic samples

Box plots of the distribution of Cancer Cell Fraction (CCF) in paired primary (light green) and metastatic samples (dark green) (n=242) for driver genes included in the oncoplot and other genes across the different types (HR+/HER2-, HER2+, TNBC). P-values are estimated by paired Wilcoxon-Mann-Whitney test.

Supplementary figure S13

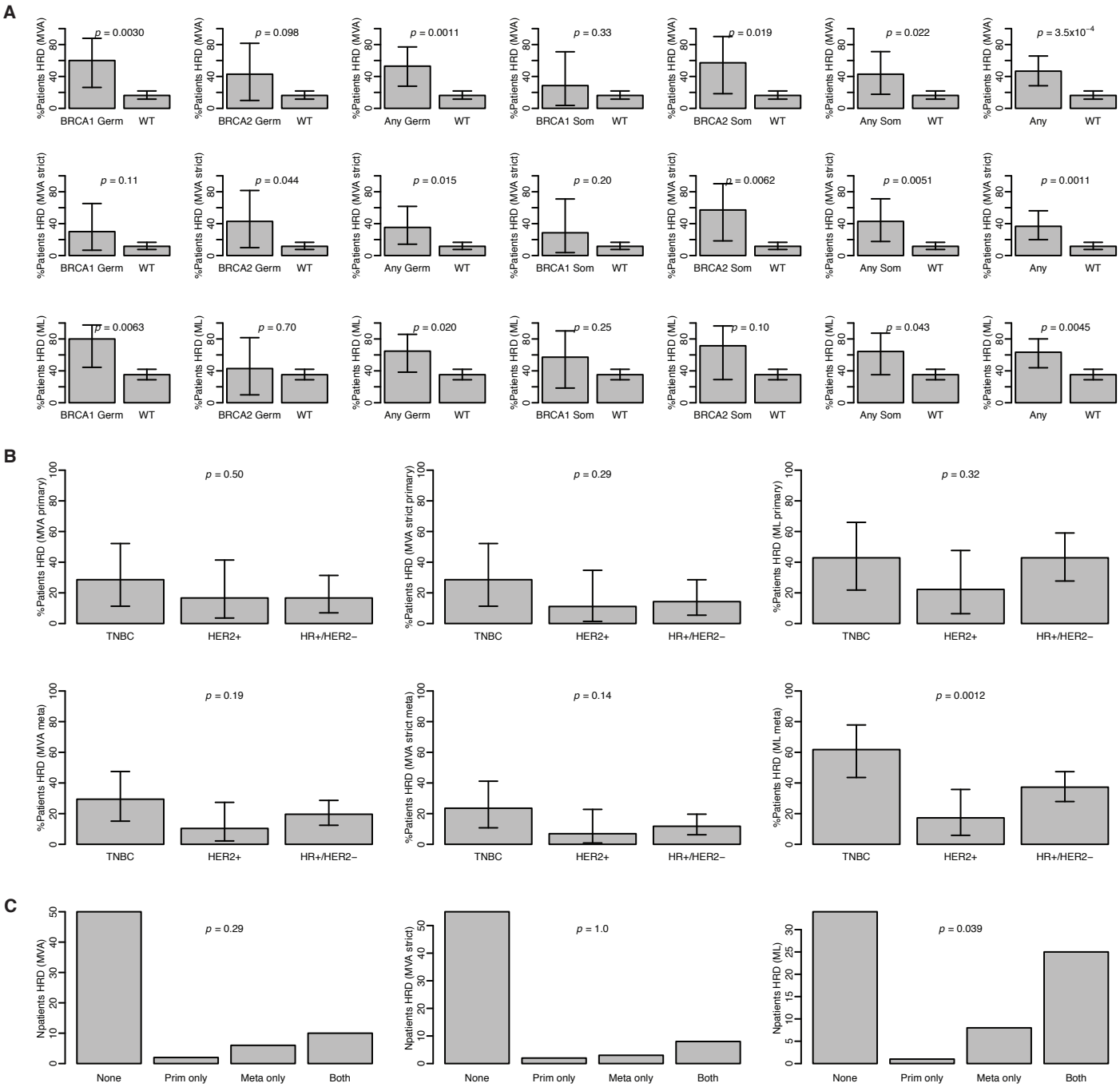


Figure S13: HRD signature by SigMA

A: HRD prediction by SigMA in patient samples with germline or somatic BRCA 1 or 2 mutations; B: HRD prediction by SigMA by IHC subtype; C: HRD prediction by SigMA primary samples versus metastatic samples.

Supplementary figure S14

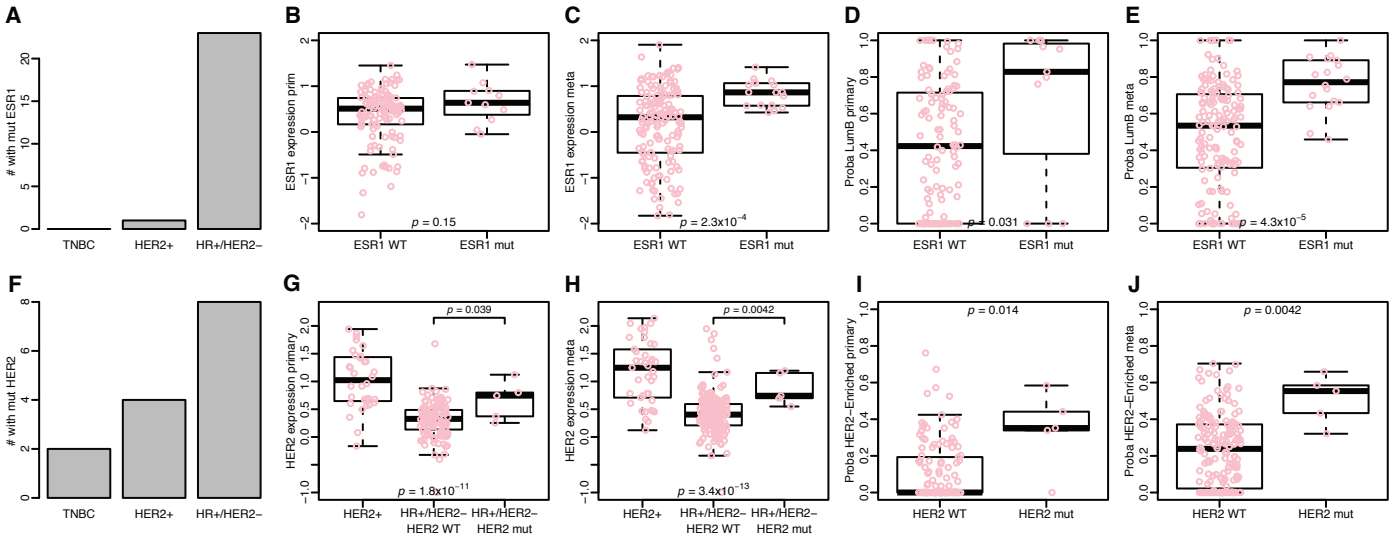


Figure S14: ESR1 and ERBB2 mutations versus gene expression

A: Number of patients harboring an ESR1 mutation per subtype; B, C: ESR1 expression in the primary/metastasis in function of ESR1 mutation status, in HR+/HER2- patients; D, E: probability to be Luminal B, as given by genefu, in function of ESR1 mutation status, in HR+/HER2- patients; F: number of patients harboring an ERBB2 mutation (in the primary or the metastasis) by subtype; G, H: comparison of ERBB2 expression in the primary/metastasis in function of ERBB2 mutation status and clinical subtype; I, J: comparison of the probability to be Her2-Enriched, as given by genefu, with ERBB2 mutation status in HR+/HER2- patients.

Supplementary figure S15

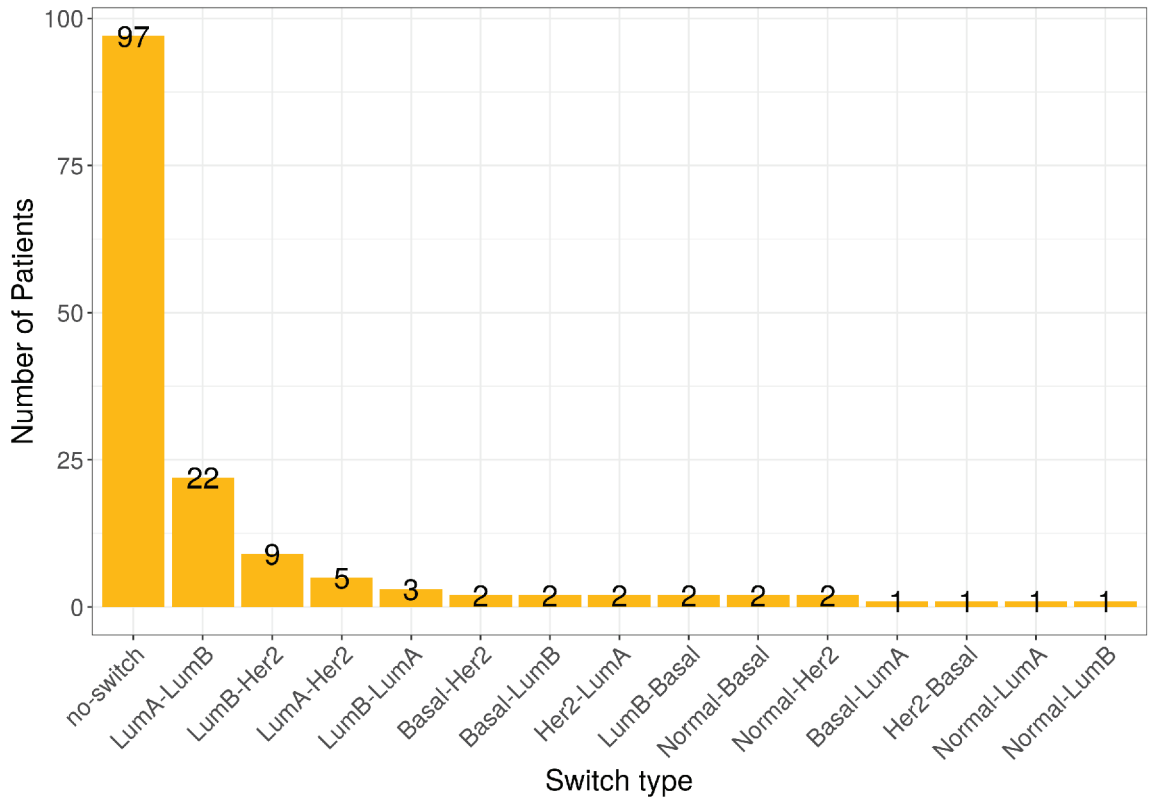
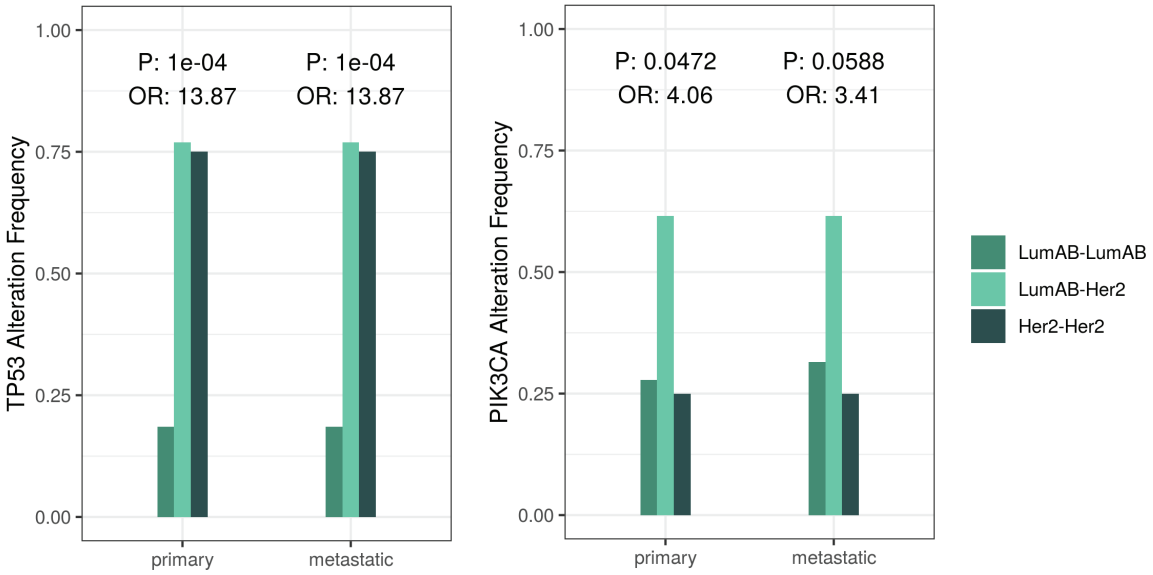


Figure S15: distribution of primary to metastasis intrinsic subtype switching

Bar plots of intrinsic subtype switch types in all patients with available paired RNA-seq data (n=152).

Supplementary figure S16



switch type	sample type	TP53 altered	total
LumAB-LumAB	prim	10	54
LumAB-LumAB	meta	10	54
LumAB-Her2	prim	10	13
LumAB-Her2	meta	10	13
Her2-Her2	prim	9	12
Her2-Her2	meta	9	12

switch type	sample type	PIK3CA altered	total
LumAB-LumAB	prim	15	54
LumAB-LumAB	meta	17	54
LumAB-Her2	prim	8	13
LumAB-Her2	meta	8	13
Her2-Her2	prim	3	12
Her2-Her2	meta	3	12

Figure S16: TP53 and PIK3CA mutations in Luminal A/B and HER2-E tumors

Bar plots reporting the alteration frequency of TP53 and PIK3CA mutations for patients with paired TGS and RNA-seq data (n=117) in primary and metastatic samples for patients with Luminal A/B subtype in both primary and metastatic samples, patients with Luminal A/B primary tumors switching to HER2-E metastases, and patients with HER2-E in both primary and metastatic samples. P-values are estimated by Fisher's Exact test between Luminal A/B to Luminal A/B and Luminal A/B to HER2-E.

Supplementary figure S17

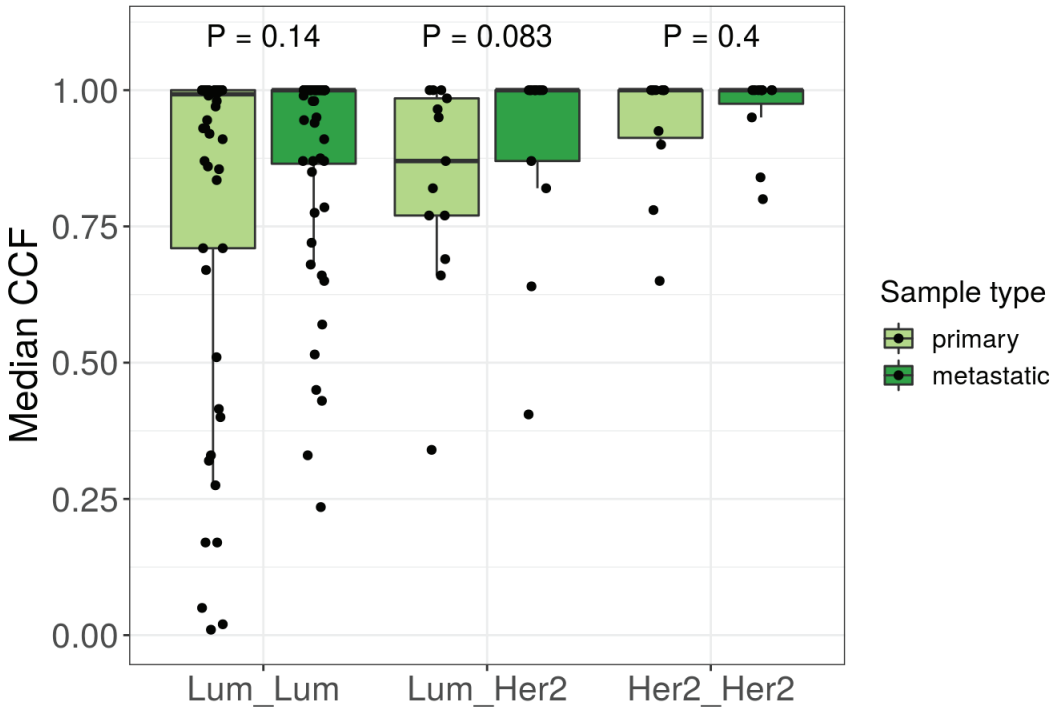


Figure S17: subtype switching and median CCF

Box plots showing the distribution of median CCF by patient in primary (light green) and metastatic (dark green) samples in patients with Luminal A/B subtypes in both primary and metastatic samples, patients with Luminal A/B primary tumors switching to HER2-E metastatic samples, and patients with HER2-E subtype in both primary and metastatic samples.

Suppelementary figure S18

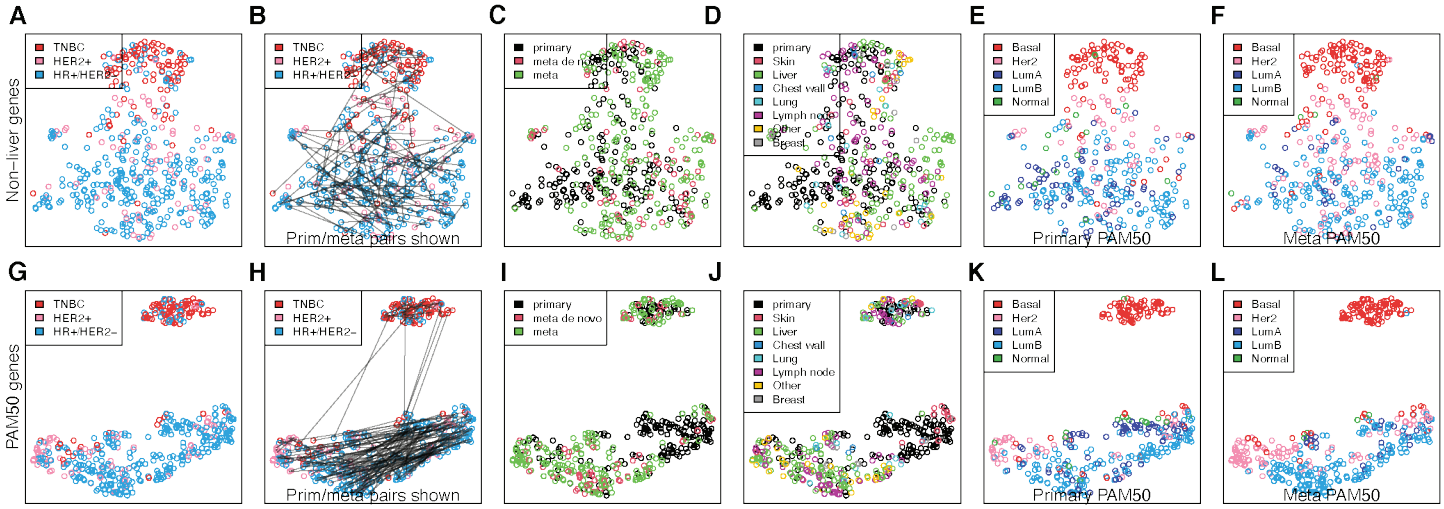


Figure S18: UMAP visualization of gene expression data

UMAP representation of gene expressions in all samples, using A-F: all non-liver genes or G-L: genes in the PAM50 classifier. Points are colored A, G: by IHC subtypes; B, H: by IHC subtypes with straight lines connecting primaries and their metastasis; C, I: by sample type; D, J: by metastasis site; E, K: by PAM50 primary subtype; F, L: by PAM50 metastasis subtype.

Supplementary figure S19

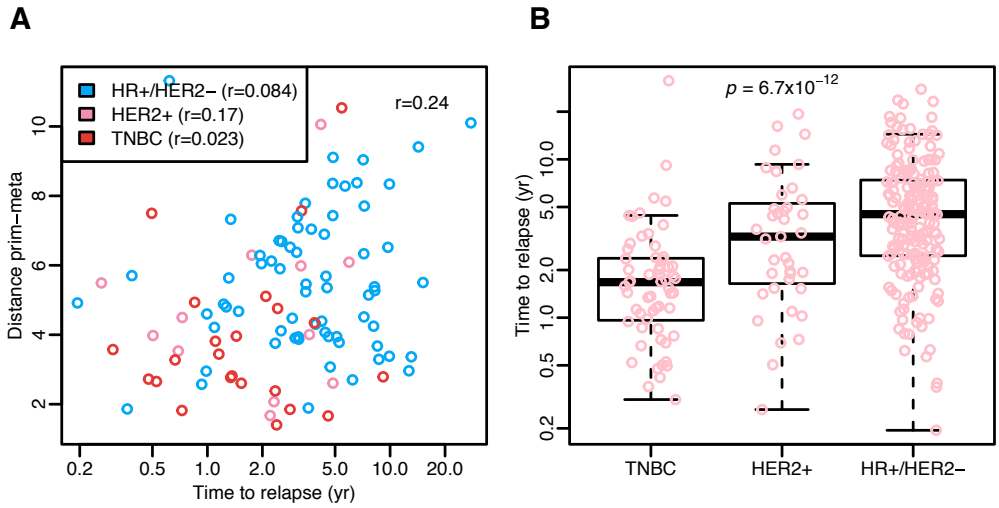


Figure S19: time to relapse and gene expression distance between primary and metastasis

A: Comparison of time to relapse with the distance between primary and metastasis, estimated from expression on the PAM50 genes, de novo patients being excluded. Spearman's correlations (r) are given, in general and by subtypes; B: Time to relapse by subtype.

Supplementary figure S20

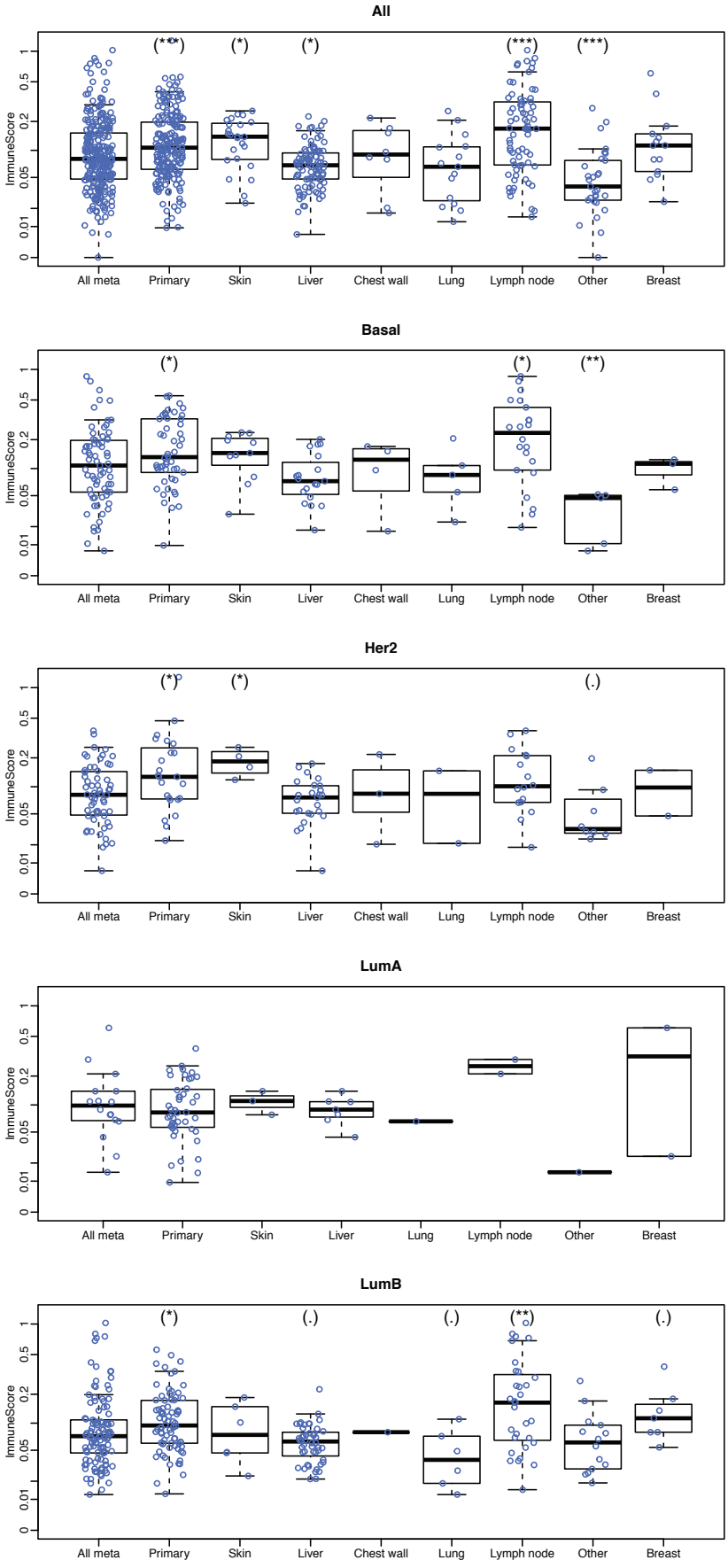
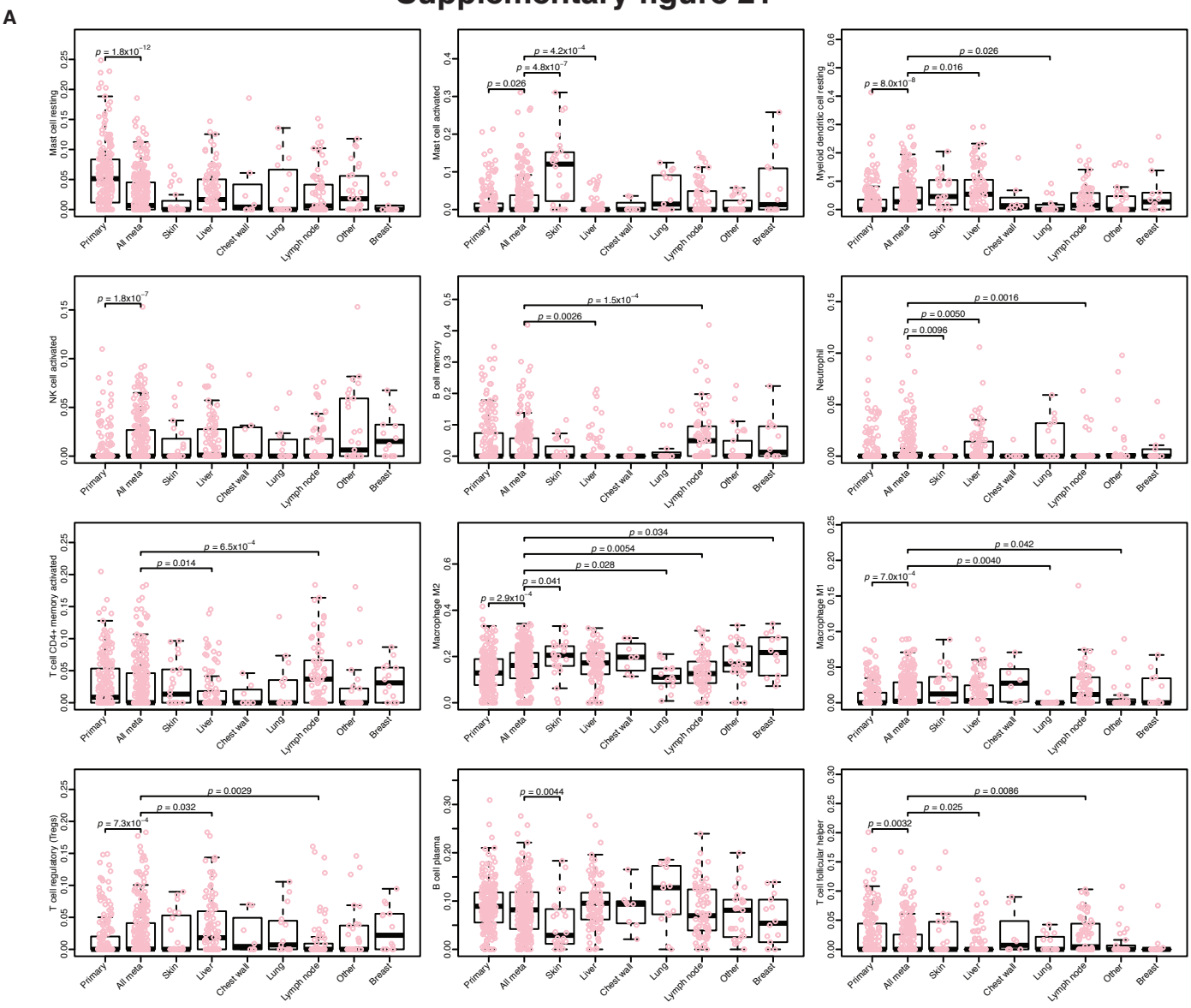


Figure S20: immune score by metastatic site

Distribution of the immuneScore, as obtained by xCell, for all metastasis biopsy sites represented at least 10 times, for all patients as well as by PAM50 subtype. The “star” notation indicates whether a site or the primary is different from all the metastases: *** is p≤0.001, ** is 0.001<p≤0.01, * is 0.01<p≤0.05, . is 0.05<p≤0.1.

Supplementary figure 21



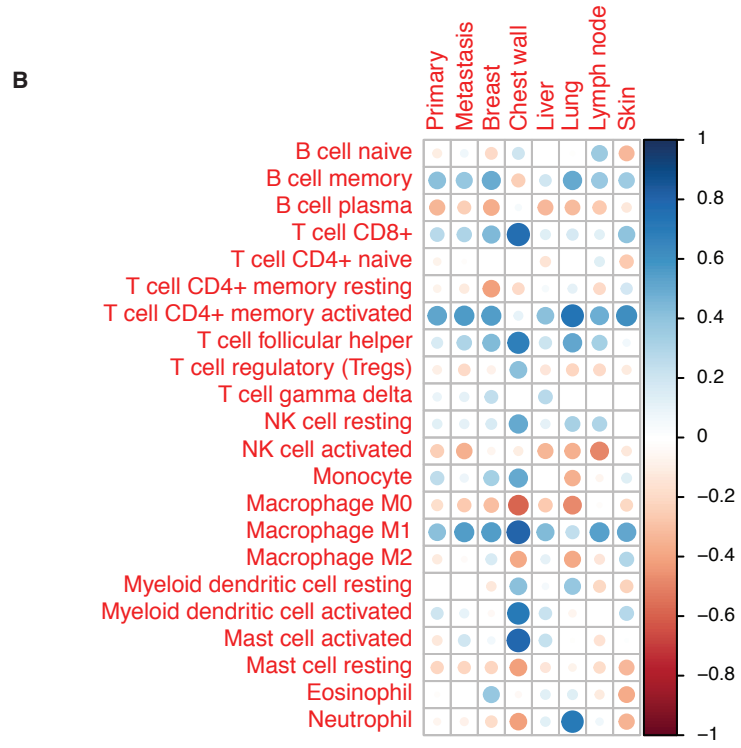


Figure S21: immune cell fractions in the primary tumor and metastatic sites

A: Comparison of the distribution of 12 immune cell fractions, as obtained by CIBERSORT, between the primary, all the metastases, as well as metastases divided by biopsy site. The p-values indicate the cases where a difference is observed between all the metastases and either the primary or a given metastatic site. No multiple testing correction was performed.
B: Correlations between the estimated fraction of the immune cells, as obtained by CIBERSORT, and the global immuneScore, as obtained by xCell, in the primaries, the metastases, or metastases from a given organ.

Supplementary figure S22

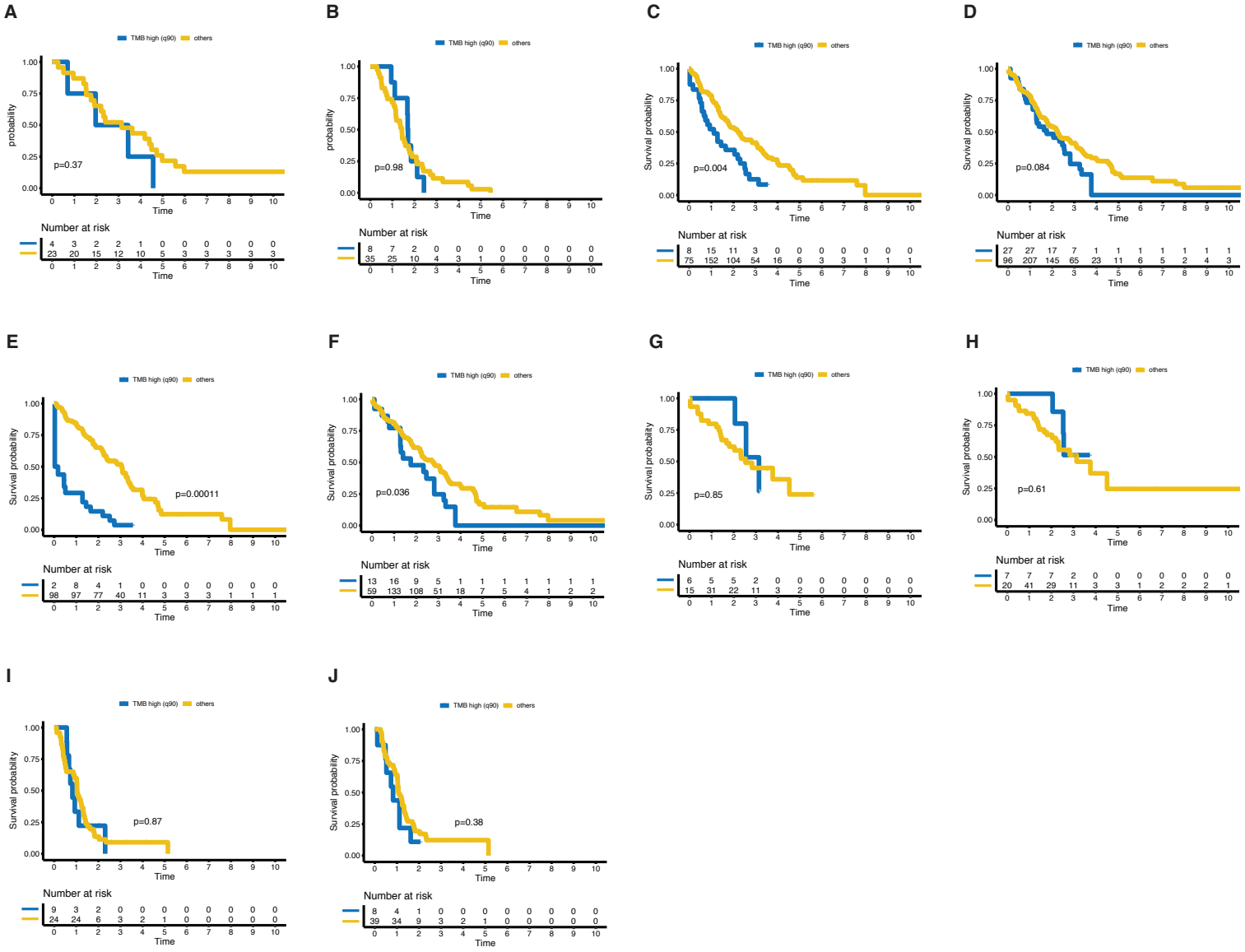


Figure S22: TMB and outcome

A: Time to relapse (TTR) by TMB in HER2+ primary samples; B: TTR by TMB in TNBC primary samples; C: Overall survival (OS) by TMB in all subtypes primary samples; D: OS by TMB in all subtypes metastatic samples; E: OS by TMB in HR+/HER2- primary samples; F: OS by TMB in HR+/HER2- metastatic samples; G: OS by TMB in HER2+ primary samples; H: OS by TMB in HER2+ metastatic samples; I: OS by TMB in TNBC primary samples; J: OS by TMB in TNBC metastatic samples.

Supplementary figure S23

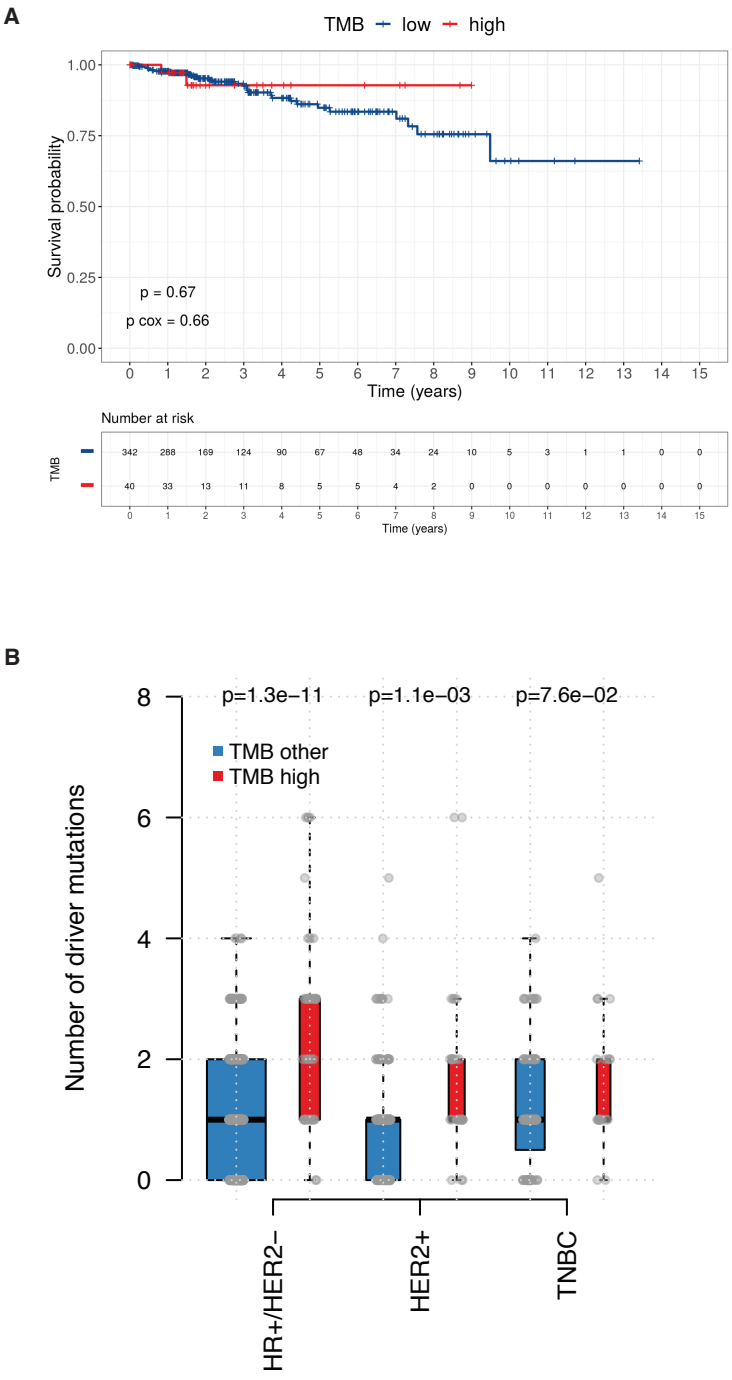


Figure S23: TMB, drivers and survival in TCGA

A: Kaplan-Meier curves showing the time to relapse by TMB classes estimated in primary samples for patients with HR+/HER2- disease in TCGA dataset; B: Box plots of the distribution of the number of driver mutations in driver genes by TMB classes estimated in primary samples across the different types (HR+/HER2-, HER2+, TNBC) in the TCGA dataset.

Supplementary figure S24

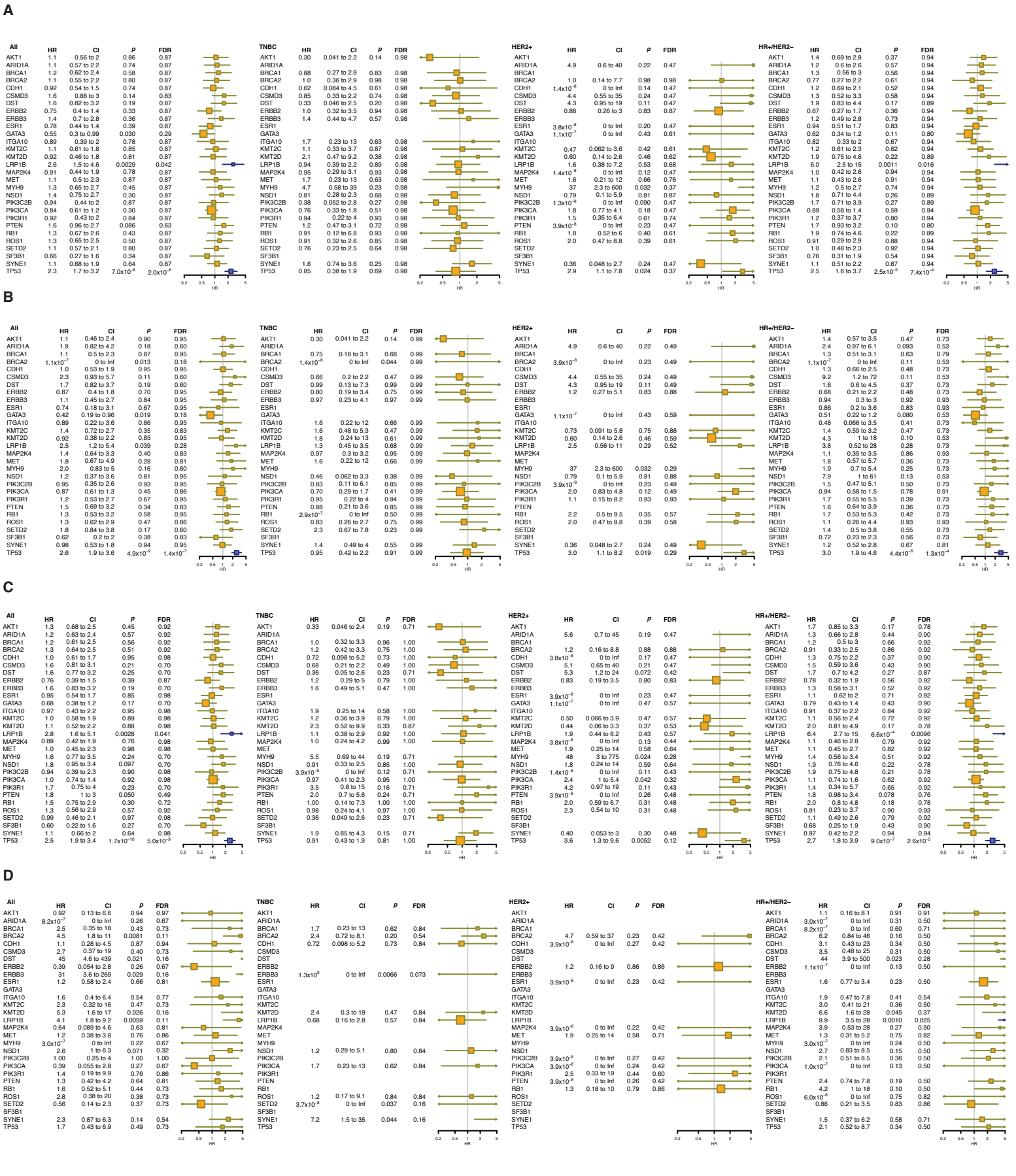


Figure S24: genomic landscape and overall survival

Forest plots indicating the relationship between the presence of a given mutation and overall survival, both across subtypes and by subtype. All genes mutated in at least 10 patients are shown. Analysis was done for mutations found in the primary A, the metastasis B, the primary or the metastasis C, and mutations private to the metastasis D. Multiple testing correction was performed by panel.

Supplementary figure S25

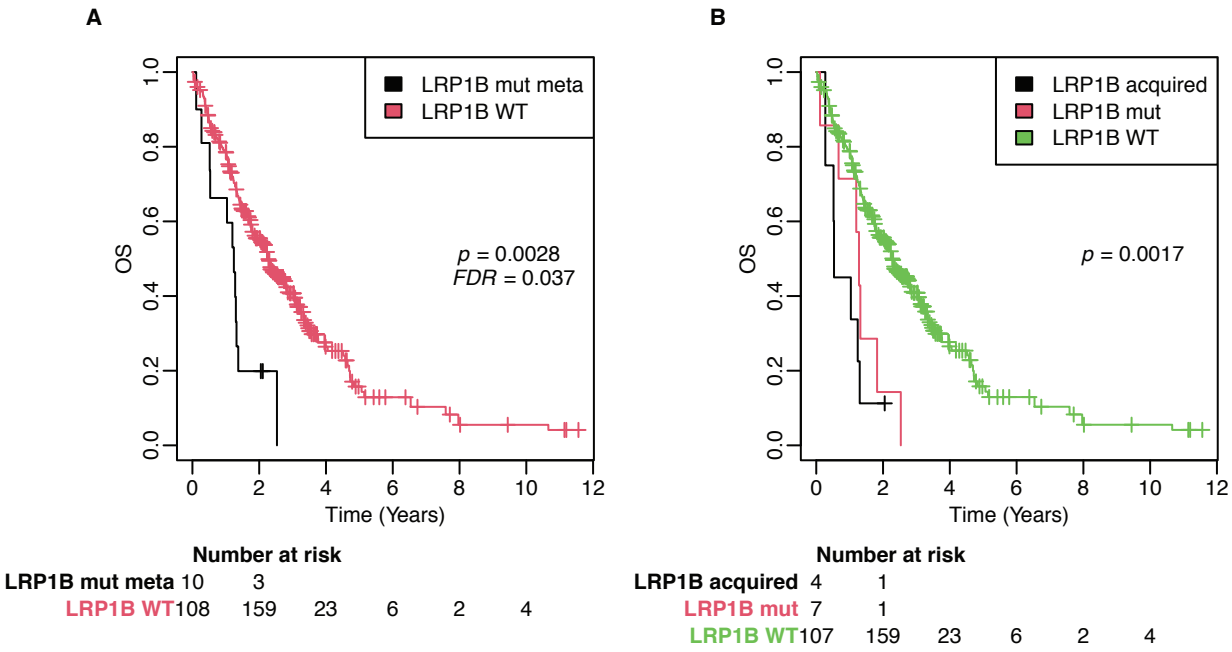


Figure S25: LRP1B mutations and survival

A: LRP1B mutations in the metastasis and OS; B: LRP1B mutations shared versus acquired and OS.

Supplementary figure S26

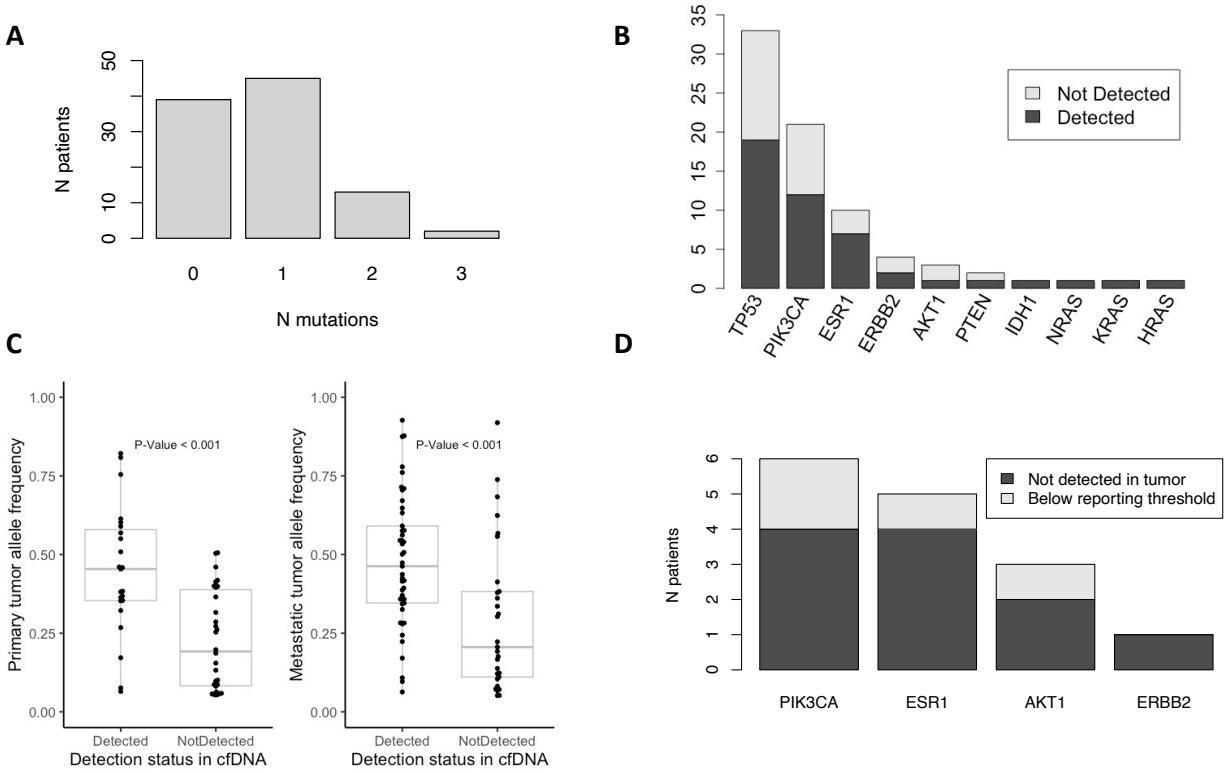


Figure S26: identification of variants in tissue and cfDNA (n=99)

A: Number of cfDNA variants detected per patient; B: detection in plasma of variants found in tumor tissue; C: for variants detected in solid tumors with an allele frequency above the reporting threshold of 0.05%, boxplot of the allele frequency in tumor vs the detection status in cfDNA. The P-value represented is from the comparison of allele frequencies between the cfDNA-detected and cfDNA-not detected groups (Mann-Whitney U test); D: variants detected only in cfDNA (ESCAT Tier I and II). Patients for whom the variant was not detected at all in the solid tumor (primary or metastatic) are represented in dark grey, whereas patients for whom the variant was present in the sequencing raw data but below the reporting threshold of 0.05% are represented in light grey.

Supplementary figure S27

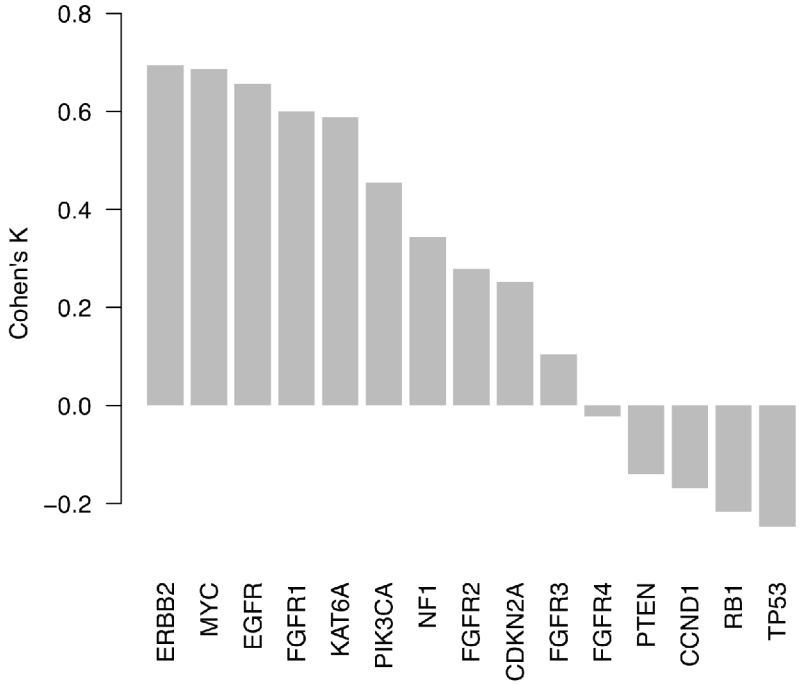


Figure S27: bin-wise concordance between FACETS and ASCAT

Concordance between CNVs measured by ASCAT and FACETS was measured by Cohen's κ.

Supplementary figure S28

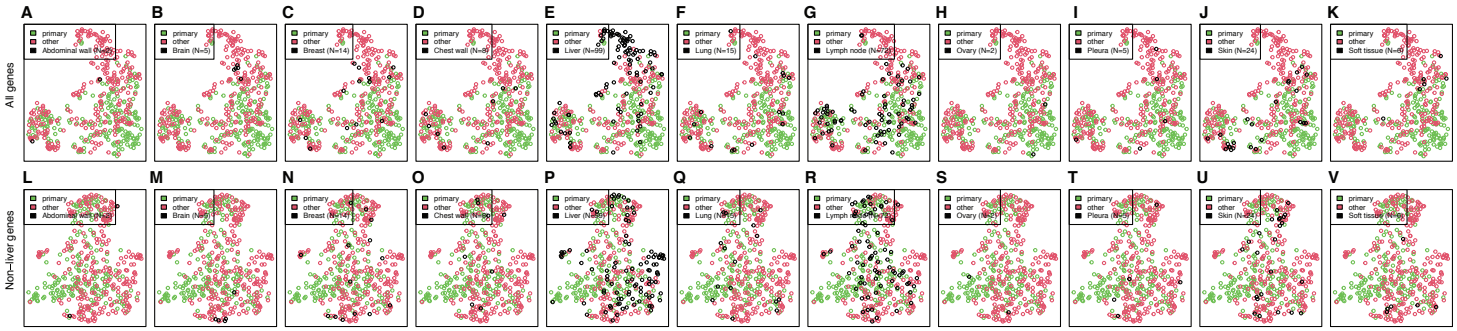


FIgure S28 UMAP visualization by biopsy site

UMAPs obtained on all genes (A-K) or only non-liver genes (L-V). For each panel, the metastatic samples coming from one biopsy site are highlighted in black.