Patient-derived xenografts and matched cell lines identify pharmacogenomic vulnerabilities in colorectal cancer

Patient-derived xenografts and matched cell lines identify pharmacogenomic vulnerabilities in colorectal cancer

Luca Lazzari[1,2], Giorgio Corti[1], Claudio Isella[1], Monica Montone[1], Pamela Arcella[1], Eugenia R. Zanella[1], Luca Novara[1], Fabiane Barbosa[4], Andrea Cassingena[4], Carlotta Cancelliere[1], Enzo Medico[1,2], Andrea Sartore-Bianchi[4,5], Salvatore Siena[4,5], Andrea Bertotti[1,2], Livio Trusolino[1,2], Federica Di Nicolantonio[1,2], Michael Linnebacher[6], Alberto Bardelli[1,2] and Sabrina Arena[1,2,*]

[1]Candiolo Cancer Institute, FPO - IRCCS, 10060 Candiolo (TO), Italy; [2]Department of Oncology, University of Torino, SP 142 km 3.95, 10060 Candiolo (TO), Italy; [3]Department of Interventional Radiology, Ospedale Niguarda Ca' Granda, [4]Niguarda Cancer Center, Grande Ospedale Metropolitano Niguarda, 20162 Milan, Italy; [5]Department of Oncology and Hemato-Oncology, Università degli Studi di Milano, 20133 Milan, Italy; [6]Department of General Surgery, Molecular Oncology and Immunotherapy, University of Rostock, D-18057 Rostock, Germany.

Running Title: PDX-derived cells support pharmacogenomic analysis in CRC

Keywords: Colorectal cancer, preclinical models, EGFR, HER2, resistance

Additional Information:

Financial support: This work was supported by European Community's Seventh Framework Programme under grant agreement no. 602901 MErCuRIC (A.B.); H2020 grant agreement no. 635342-2 MoTriColor (A.B. and S.S.); IMI contract n. 115749 CANCER-ID (A.B.); AIRC 2010 Special Program Molecular Clinical Oncology 5 per mille, Project n. 9970 Extension program (A.B. and S.S.); AIRC IG n. 16788 (A.B.); AIRC IG n. 17707 (F.D.N.); AIRC IG n. 20685 (S.S.); AIRC Special Program 5 per mille Metastases Project n 21091 (A.B., F.D.N., E.M., S.S., And.B. and L.T.); Terapia Molecolare Tumori by Fondazione Oncologia Niguarda Onlus (A.S.B. and S.S.); Genomic-Based Triage for Target Therapy in Colorectal Cancer Ministero della Salute, Project n. NET 02352137 (A.S.B., A.B. and S.S.). TRANSCAN-2 JTC 2014 contract no. TRS-2015-00000060 INTRACOLOR (S.A.); TRANSCAN-2 JTC 2014 TACTIC (L.T.); AIRC MFAG no. 20236 (S.A.); European Research Council Consolidator Grant 724748 – BEAT (And.B.); AIRC IG n. 18532 (L.T.); AIRC IG n. 20697 (And.B.).

***correspondence to:** Sabrina Arena, Dept of Oncology- University of Torino and Candiolo Cancer Institute , FPO - IRCCS, Strada Provinciale 142, 10060, Candiolo (Torino, Italy). E-mail: sabrina.arena@unito.it

Conflicts of Interest: A.B. is an advisory board member for Roche; S.S. is an advisory board member for Amgen, Bayer, BMS, Celgene, Incyte, Merck, Novartis, Roche, and Seattle Genetics; the other authors declare no potential conflicts of interest.

Word Count: 6702

Translational Relevance:

Progress in the development of effective cancer treatments is limited by the availability of tumor models. For decades, established cancer cell lines represented the mainstay preclinical tumor model, but they show limitations, such as limited capacity to recapitulate inter- and intra-tumor heterogeneity, adaptation to grow in two-dimensional cultures and the lack of interaction with the microenvironment. To overcome these restraints, patient-derived tumor xenografts (PDX) have been developed in recent years. Although they mirror histological and molecular features of the patient's tumor, PDXs also have limitations including maintenance costs and unsuitability for large-scale screenings. Here we describe a novel platform of PDX-derived cell lines, that retain the genomic and pharmacologic profile of the sample of origin and offer advantages such as reduced working costs and ease of handling. In conclusion, we provide a valuable preclinical model for large-scale functional gene validation and assessment of novel therapeutic strategies in colorectal cancer.

ABSTRACT

PURPOSE: Patient-derived xenograft (PDX) models accurately recapitulate the tumor of origin in terms of histopathology, genomic landscape, and therapeutic response, but they also retain some limitations due to costs associated with their maintenance and restricted amenability for large-scale screenings. To overcome these issues, we established a platform of 2D cell lines (xeno-cell lines, XLs), derived from PDXs of colorectal cancer (CRC) whose patient germline gDNA was available.

EXPERIMENTAL DESIGN: Exome and expression analysis were performed. Biomarkers of response and resistance to anti-HER therapy were annotated.

RESULTS: Molecular features were remarkably concordant between PDXs and matched cell lines. XLs recapitulated the entire spectrum of CRC transcriptional subtypes. Exome and RNA-seq analyses delineated several molecular biomarkers of response and resistance to EGFR and HER2 blockade. Genotype-driven responses observed *in vitro* in XLs were confirmed *in vivo* in the matched PDXs.

CONCLUSION: The XL platform represents a preclinical tool for functional gene validation and proof of concept studies to identify druggable vulnerabilities in CRC.

1. Introduction

The awareness that cancer is a genetic disease and the availability of genomic profiles of tumor samples paved the way to 'precision medicine' in oncology (1).

In the last ten years, major advances in complex genomic technologies have strongly contributed to the mapping of cancer molecular landscape (2). The need to understand the impact of the major drivers of cancer development and progression on therapeutic intervention has thus become of paramount importance. Functional characterization and target validation has required generation of reliable cellular and animal models to understand the role of genomic alterations on tumor onset and evolution (3). Functional studies have proven that the tumor is often dependent on oncogenic alterations for its growth and maintenance, providing a strong rationale for the development of therapies targeting cancer driver genes (4).

On these premises, modeling disease-associated alterations in models that reflect the clinical features of patients is vital to implement precision medicine. So far, different cellular and animal models have been generated, each showing both advantages and pitfalls.

Cancer cell lines have been extensively and routinely used for biomarker discovery and drug development. Indeed, the Cancer Cell Line Encyclopedia, the Genomics of Drug Sensitivity in Cancer (GDSC) Sanger Institute project, the National Cancer Institute-60 (NCI-60) cancer cell line screen, and the Cancer Therapeutic Response Portal (CTRP) constitute paradigmatic examples of how coupling cell lines to systematic compound screening can provide an informative clinical platform for pharmacogenomic analysis (5-9). Although these studies aim at providing tools to predict anti-cancer drug response, their complete success is limited by several intrinsic characteristics of cell lines, such as limited capacity to recapitulate inter- and intra-tumor heterogeneity and adaptation to growing in artificial conditions. Genomic instability, loss of molecular heterogeneity and lack of the tumor

microenvironment represent the three main issues thought to make a 2D cell line model divergent from the original tumor; for these reasons, the clinical relevance of cell line models has been repeatedly questioned (10). These drawbacks may play against the recapitulation of the evolutionary nature of cancer itself, thus limiting the translation of the *in vitro* results to the clinical response of the tumor in the patient. In addition, the frequent unavailability of matched germline tissue has often hampered the confident identification of somatic changes.

To obviate these issues and better bridge the bench-to-bedside gap, patient-derived tumor xenografts (PDXs, xenopatients) have been developed. Although these models had already been described in the seventies, they have become popular for assessing genotype-drug correlation response studies in the last decade (11). Comprehensive genetic and transcriptomic studies have demonstrated that PDXs maintain the majority of genetic alterations and global pathway activity of primary tumors (12,13); importantly, the histological structure and intratumoral clonal heterogeneity are usually also preserved (14,15), although recent literature reports the selection of specific copy number alterations during PDX propagation (16). In colorectal cancer (CRC), seminal studies have already set the stage for pharmacogenomic analyses and association of PDX genotype with targeted therapy response (12,17). PDXs have significantly contributed to shedding light on the mechanisms of resistance and on identification of predictive biomarkers of response, but large-scale screening is actually limited by the costs, size and efforts for animal maintenance and manipulation.

New models that can couple both the ease of handling of cell cultures with the preservation of intra-tumor heterogeneity of PDXs have been proposed. Bruna and colleagues (5) have generated a large biobank of breast cancer PDXs combined with short-term cultures of PDX-derived cells demonstrating that the corresponding cell models could be reliably used to assess drug response and to identify biomarkers of resistance *in vitro*, as paralleled *in vivo*. Similar studies have been recently performed also in gastrointestinal and

pancreatic cancer (18,19), thus confirming the value and applicability of this approach.

In this work, we describe the establishment of a panel of 29 cell lines obtained from a cohort of 29 CRC PDXs and provide a comprehensive genomic analysis of these models. Importantly, we show that the molecular features of xenopatient-derived cell lines closely parallel their matched PDX models. We also show that PDX-derived cell lines constitute a valuable model to interrogate CRC patient molecular vulnerabilities and provide an effective opportunity to reveal patient-specific drug responses and implement precision oncology in CRC.

2. Materials and Methods

Xenopatient-derived cell line establishment, culture and authentication

For xenopatient generation, fresh surgically resected primary CRC and metastases tissues or biopsies were obtained and collected from consenting patients. All procedures were approved by the Italian Ministry of the Health and the Ethics Committee of the Medical faculty of the University of Rostock, in accordance with generally accepted guidelines for the use of human material.

Tumor specimens were cut into small pieces and either frozen (biobanked) or prepared for implantation, as previously described (12,20). For tumor implantation, six-week-old female NMRI nu/nu or NOD/SCID mice were used as recipients. All experimental procedures were carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Committee on the Ethics of Animal Experiments of the University of Rostock and the Internal Ethical Committee from Animal Experimentation of the Candiolo Cancer Center.

Established tumors from xenopatients (1500 mm^3) were removed and processed for *in vitro* culture as described below.

PDX tissues were dissociated into single-cell suspension by mechanical dissociation using the Gentle MACS Dissociator (Miltenyi Biotec) and enzymatic degradation of the extracellular matrix using the Human Tumor Dissociation Kit (Miltenyi Biotec) according to the manufacturer's protocol. Cell suspension was centrifuged three times and finally the pellet was re-suspended with DMEM/F12 medium containing 10% FBS and 2 mM L-glutamine. The final cell suspension was then filtered through a 70-um cell strainer (Miltenyi Biotec) and cells that were not filtered out were re-

suspended in DMEM/F12 medium containing 10% FBS, 2mM L-glutamine, antibiotics (100 U/ml penicillin and 100 mg/ml streptomycin) and 10 μM ROCK inhibitor Y-27632 (Selleck Chemicals Inc) and cultured on collagen-coated plates (Corning) at 37°C in a humidified atmosphere of 5% CO_2. Medium was changed regularly every three days. Growing cell lines were further passaged and were subjected to three sequential cycles of freezing and thawing to ensure the stability of the cell lines. Growing cell lines were stocked at low passages by cryopreservation.
Only in one case (CRC0080), the cell line was established as xenosphere (21) and then adapted to grow in 2D as monolayer. HROC cell lines collection was derived by Michael Linnebacher's laboratory, as previously reported (20).
BT474 and SKBR3 cell lines were purchased by American Type Culture Collection (ATCC) and maintained in DMEM/F12 and DMEM media, respectively, containing 10% FBS, 2mM L-glutamine, antibiotics (100 U/ml penicillin and 100 mg/ml streptomycin) and cultured at 37°C in a humidified atmosphere of 5% CO_2.
XLs were routinely checked for mycoplasma contamination using the Venor GeM Classic Kit (Minerva Biolabs), according to the manufacturer's protocol.
XLs, matched PDXs and patient authentication was performed using the Gene Print 10 System (Promega), through Short Tandem Repeats (STR) at 10 different loci (D5S818, D13S317, D7S820, D16S539, D21S11, vWA, TH01, TPOX, CSF1PO and amelogenin). Amplicons from multiplex PCRs were separated by capillary electrophoresis (3730 DNA Analyzer, Applied Biosystems) and analysed using GeneMapperID v.3.7 software (Life Technologies).

DNA extraction and MSI analysis

Genomic DNA samples were extracted from each XL and matched PDX using ReliaPrep gDNA Tissue Miniprep System (Promega). Patient's germline genomic DNA obtained from normal tissue (liver) or PBMCs of each patient was used as reference genome.
The MSI status was evaluated by mean of MSI Analysis System kit (Promega), according to manufacturer's protocol. The analysis requires a multiplex amplification of seven markers including five mononucleotide repeat markers (BAT-25, BAT-26, NR-21, NR-24 and MONO-27) and two pentanucleotide repeat markers (Penta C and Penta D). The products were analyzed by capillary electrophoresis in a single injection using ABI 3730 DNA Analyzer capillary electrophoresis system (Applied Biosystems). The results were analyzed using GeneMapper V5.0 software. Sample with instability in one or more markers are defined as MS-instable (MSI-H). Sample with no detectable alterations are MS-stable (MSS).
MSI events were evaluated by MSIsensor (22). Signatures were calculated using a custom Python script.

Exome sequencing and bioinformatic analysis

Exome sequencing was performed in outsource, while data analysis was performed at the Candiolo Cancer Institute. Each triplet of DNA samples (PDX, xeno-cell line, germline) was sent to GATC (Kostanz, DE) that performed library preparation, exome capture, sequencing on Illumina platform and data demultiplexing. Contaminating mouse reads were removed using Xenome (23). Human sequences were mapped to the hg38 human reference by BWA-mem (24) and PCR duplicates were removed using Picard Mark Duplicates function (broadinstitute.github.io/picard/). Somatic mutational and gene copy number profiles for each XL and matched PDX were obtained subtracting germline profiles as previously described (25).

Gene copy number analysis

Real-time PCR was performed with 10 ng of DNA per single reaction using GoTaq QPCR Master Mix (Promega) with an ABI PRISM 7900HT apparatus (Applied Biosytems; primers' sequences are available on request). Gene copy numbers were normalized to a control diploid cell line, HCEC.

RNA extraction, analysis and identification of cancer-cell intrinsic subtypes

RNA was extracted using miRNeasy Mini Kit (Qiagen), according to the manufacturer's protocol. The quantification and quality analysis of RNA was performed on a Bioanalyzer 2100 (Agilent), using RNA 6000 nano Kit (Agilent). Total RNA extracted from XLs was processed for RNA-seq analysis with the TruSeq RNA Library Prep Kit v2 (Illumina) following manufacturer's instruction and sequenced on a NextSeq 500 system (Illumina). Each fastq file was aligned using MapSplice (26). Version hg19 of the genome was used and Gencode v19 as reference transcriptome database and gene quantification was performed with RSEM (27). Assignments of each xeno-cell line to cancer-cell intrinsic subtypes and consensus molecular subtypes were performed as previously described (28,29). Hierarchical clustering of Xenopatient and matching XenoLine samples was performed with maximum cosine distance using the GEDAS software (30) based on genes whose variance in log2 expression signal was greater than 0.5 in both the Xenopatient and the XenoLine datasets and on CRIS geneset.

Drug proliferation assays

CRC XLs were seeded at different densities (5–7 x10^3 cells per well) in 100 µl complete growth medium in 96-well plastic culture plates at day 0. The following day, serial dilutions of the indicated drugs were added to the cells in serum-free medium, while medium-only was included as controls. Plates were incubated at 37 °C in 5% CO_2 for 5 or 6 days, after which cell viability was assessed by measuring ATP content through Cell TiterGlo Luminescent Cell Viability assay (Promega). Luminescence was measured by SPARK M10

(Tecan) plate reader. Treated wells were normalized to untreated/DMSO treated wells.
For long-term proliferation assays, cells were seeded in 24-wells plates ($1x10^4$ cells per well) and cultured in the absence and presence of drugs as indicated. Wells were fixed with 3% paraformaldehyde (Santa Cruz) and stained with 1% crystal violet-methanol solution (Sigma-Aldrich) after 2 weeks. All assays were performed independently at least three times.

Cetuximab and trastuzumab were obtained from the Pharmacy at Niguarda Cancer Center in Milan, Italy. Lapatinib was purchased from Selleck Chemicals.

Lentiviral Transduction of HER2 Mutant Colorectal XL-cells
The PIK3CA p.H1047R mutation was inserted into pLenti-PIK3CA-myc-DDK vector (Origene) using site directed mutagenesis with the Quick Change II Site-Directed Mutagenesis kit (Agilent), according to manufacturer's instructions. Primers sequences for PIK3CA site directed mutagenesis were the following: GAAACAAATGAATGATGCACGTCATGGTGGCTGGACAAC (PIK3CA_H1047R_F) and GTTGTCCAGCCACCATGACGTGCATCATTCATTTGTTTC (PIK3CA_H1047R_R). Lentiviral control pLenti-myc-DDK vector was purchased from Origene.
Lentiviral control vectors (pRLL empty and FG12 empty), pRLL-KRAS WT, pRLL-KRAS G13D and FG12-BRAF V600E vectors were previously exploited (31,32).
Lentiviral vector stocks were produced by transient transfection of the transfer plasmids, the packaging plasmids pMDLg/pRRE and pRSV.REV, and the vesicular stomatitis virus (VSV) envelope plasmid pMD2.VSV-G (12, 5, 2.5, and 3 µg, respectively, for 10 cm dishes) in HEK-293T cells. Determination of the viral p24 antigen concentration was done by ALLIANCE HIV-I P24 ELISA 2 PLATE KIT (PerkinElmer Life Science Inc.). Cells were transduced in six-well plates (3×10^5 per well in 2 ml of medium) in the presence of polybrene (8 mg/ml) (Sigma).

Western blotting analysis
Total cellular proteins were extracted by solubilizing the cells in boiling SDS buffer (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, and 1% SDS). Extracts were clarified by centrifugation and protein concentration was determined using BCA Protein Assay Reagent kit (Thermo). Western blot detection was performed with enhanced chemiluminescence system (GE Healthcare) and peroxidase conjugated secondary antibodies (Amersham). The following primary antibodies were used for western blotting (all from Cell Signaling Technology, except where indicated): anti-phospho-HER2 YY1221/1222);

anti-HER2 (Santa Cruz); anti-phospho-EGFR (Y1068); anti-EGFR (clone 13G8, Enzo Life Science); anti-vinculin (Millipore); anti-phospho-p44/42 ERK (Thr202/Tyr204); anti-p44/42 ERK; anti-phospho AKT (Ser473); anti-AKT; anti-PTEN; anti-HSP90 (Santa Cruz); anti-GAPDH (Abcam).

In vivo treatment

Established tumors (average volume 400 mm^3) were treated with the following regimens, either single-agent or in combination: lapatinib (Carbosynth) 100 mg/kg, daily (vehicle, 0.5% methylcellulose, 0.2% Tween-80); trastuzumab (Roche), 30 mg/kg once weekly (vehicle: physiological saline). Tumor size was evaluated once weekly by caliper measurements, and the approximate volume of the mass was calculated using the formula 4/3π (d/2)2 D/2, where d is the minor tumor axis and D is the major tumor axis. Animal procedures were approved by the Ethical Commission of the Candiolo Cancer Institute and by the Italian Ministry of Health.

Statistical Analysis

Results were expressed as means ± standard error of the mean (SEM) or standard deviation (SD) as indicated in the legend. Statistical significance was evaluated by t test or two-way ANOVA, using GraphPad Prism software. $P<0.05$ was considered statistically significant.

3. Results

Establishment of PDX-derived cell lines

We derived primary cell lines from a set of CRC PDX models with the goal of setting up a platform of *in vitro* CRC preclinical models closely resembling patient's features. The clinico-pathological characteristics of patients from whom the PDXs were initially obtained are listed in Supplementary Table S1; the establishment procedure, described in detail in the 'materials and methods' section, is schematically represented in Supplementary Fig. S1. Briefly, the PDX tumor is surgically excised and mechanically and enzymatically dissociated, to obtain cells that are subsequently cultured in 2D petri dishes. In this work, a total of 29 primary cell lines have been derived from a collection of 29 PDXs established at the Candiolo Cancer Institute (n= 17) and at Rostock University (n=12) (12,17,20). Among the 29 PDXs, 10 were originated from primary tumors and 19 from liver or peritoneum metastases of 28 CRC patients (two different metastatic lesions - IRCC-5A

and IRCC-5B - were obtained from the same patient) (Fig. 1A and Supplementary Fig. S1). To underline the origin of these 2D primary cell lines from xenopatients, we will refer to them as xeno-cell lines (XLs).

XLs, although presenting cell-specific morphological features, displayed a predominant and typical epithelial pattern characterized by islets of compact or round shaped cells growing predominantly as a monolayer or, in one case (HROC147), as cell-clumps in suspension (Supplementary Fig. S2). The epithelial phenotype was confirmed by analysis of expression of specific epithelial markers (e.g. E-cadherin, cytokeratin 8, 18 and 19 and EpCAM) with respect to mesenchymal markers (e.g. vimentin, SNAIL and TWIST1) (Supplementary Fig. S3).
All 29 XLs displayed the ability of long-term propagation (multiple passages) *in vitro* and their stability was ensured by at least three sequential freezing/thawing cycles prior to bio-banking. Cells were maintained in culture for at least 8 passages prior to nucleic acid extraction.

Xeno-cell lines recapitulate the genomic landscape of matched patient-derived xenografts

To assess whether the CRC molecular subtypes that are commonly ascertained in the clinic were preserved in both the PDXs and the matched 2D XLs, we performed genomic and transcriptomic analyses systematically comparing the two preclinical platforms.

Initially we analyzed microsatellite markers and found that 5 out of 29 (17%) PDXs displayed microsatellite instability (MSI-H), while the remaining were microsatellite stable (MSS). The MSI status was maintained in the matched XLs (Fig. 1B)(33).
RNA expression-based correlation analysis between XLs and their originating PDXs, which had been subject to gene expression analysis in a previous independent study (28), revealed higher correlation between matched XL-PDX pairs than between unmatched pairs (Kolmogorov-Smirnov, $P < 1x10^{-16}$, Supplementary Fig. S4A); furthermore, hierarchical clustering always

preferentially aggregated each XL with the PDX sample of origin (Supplementary Fig. S4B).

Next, parallel exome sequencing analyses of XLs and matched PDXs were performed. Importantly, to identify tumor-specific somatic mutations, germline genomic DNA (gDNA) extracted from normal tissues (liver) or blood (PMBCs) of the patient was used as reference genome.
Samples were sequenced with high depth and coverage to allow detection of low-frequency somatic mutations in xenopatient tumors and matched XLs despite contamination of human cancer cells with murine infiltrating cells in some samples. The median depth achieved was 143x and the median coverage was 97.8% on the coding regions of the reference genome (assembly hg38) with 90.3% of the bases covered at least 40 times.

A high level of correlation was found between cell lines and matched PDXs regarding the total number of somatic nonsynonymous single-nucleotide variants (SNVs) and insertions or deletions (INDELs) (Pearson r=.9870; P<.0001) (Supplementary Fig. S5A). Overall, somatic mutations detected in PDXs are highly preserved in cell lines (Supplementary Fig. S5B).
MSI-H models showed a higher number of MSI events compared to MSS cases (median value MSI-H = 7723.5 vs MSS = 263.5) (Fig. 1C). Moreover, the number of MSI events is maintained in each PDX and matched xeno-cell line (Pearson r=.998), in both MSS and MSI-H cases.
MSI-H cell lines and matched PDXs displayed a hypermutator phenotype with a higher mutational load (number of mutations/Mb) as compared to MSS models (median value MSI-H = 60.5 vs MSS = 5.5) (Fig. 1D), reflecting a higher number of non-synonymous single nucleotide variations (SNVs) (mutation burden) (median value MSI-H = 1691.5 vs MSS = 116) (Fig. 1E). This is consistent with the reported mutation frequency in CRC tissues and cell lines (2,7,34). No MSS model was found to be hypermutated, likely reflecting the lack of mutations in POLE or POLD1, which are known to generate mutations in the absence of MSI-H (2). Moreover, the INDEL burden was higher in MSI-H rather than MSS models (median value MSI-H = 562.5

vs MSS = 7), as previously reported in TCGA cancers and CRC cell lines (2,34) (Fig. 1E and Supplementary Fig. S5C).

Somatic variants within the coding regions in XLs were highly concordant with the corresponding xenopatient specimens for both MSS and MSI-H models (median value = 0.78 frequency of concordance, range 0.48-0.93) (Supplementary Fig. S5D). The allelic frequencies of alterations shared between XLs and corresponding PDXs were significantly correlated (Pearson correlation coefficient r = 0.8233)(Supplementary Fig. S5E).

Despite the global genetic correlation between the two models, we also detected some divergent mutations identified only in XLs or in PDXs (private mutations). The putative biological significance in cancer of these private mutations identified only in XLs (n=4174) or PDXs (n=1997) was assessed on Cancer Gene Census and data reported from CRC analysis (35,36). Only 5.6% (234/4174) of divergent mutations found in XLs affected cancer-related genes. Similarly, cancer-related genes that were discordant in the PDXs represented 5.2% (104/1997) (Supplementary Table S2). Of note, 74.04% (77/104) and 79.91% (187/234) of private mutations affecting cancer-related genes were identified in MSI-H PDXs and XLs, respectively. Moreover, private mutations had a median VAF of 31.69% and 32.88% for XLs and PDX, respectively (Supplementary Table S2).

While the SNVs spectrum was maintained between XLs and matched PDXs, it differed significantly between the MSI-H and MSS cases, with an increased proportion of C>G/G>C transversion (about 5-fold) and a decrease in T>C/A>G transition in the MSS cases (about 1.3-fold, Fig. 1F). Interestingly, MSI-H models present a SNV spectrum typically associated with defective DNA mismatch repair and found in microsatellite unstable CRCs (37) (Supplementary Fig. S6).

Finally, exome analysis revealed, when considering copy number variation (CNV) profiles, that the copy number profile of each PDX is also maintained in each derived cell line (Pearson correlation $0.82 \leq r \leq 0.98$). As expected, MSS xenopatients and matched XLs displayed a higher level of segments of loss of heterozygosity (LOH) or copy-number variations than MSI models, which showed a more stable CNV profile (Fig. 2A). In MSS cases, the most common

deleted chromosomal arms were 8p, 17p (including TP53), and 18q (including SMAD4), and the most common gained regions were chromosome 7, 8q (including MYC), 13, and 20q (Fig. 2B and 2C).

CRC pathway alterations are conserved in xeno-cell lines

Integration of genomic results allowed the identification of deregulated signaling pathways; the genes most commonly altered in CRC (2) were all represented in our panel of PDXs and matched cell lines (Fig. 3A). In particular, recurrent alterations in the DNA mismatch repair (MMR), WNT, MAPK, PI3K/PTEN, TGF-β and p53 pathways were identified (Fig. 3A).

Mutations in MMR genes were found in MSI-H tumors. These alterations involve missense mutations in *MSH2, MSH3, MSH6* and *POLE/POLD1* genes, as well as *MLH1* expression downregulation (Fig. 3A). Interestingly, 3 MSS cases displayed alterations in *MSH3, MSH6* and *POLE* MMR-associated genes in both xenopatients and matched XLs. These mutations were not reported previously, nor did they fall in regions important for protein activity, suggesting that these genes are not functionally affected by these mutations.

The WNT pathway is frequently altered (90%) in CRC. The most frequent alterations include inactivating mutations in *APC* as well as activating mutations in *CTNNB1* (2). Other variants affecting this pathway are inactivating mutations targeting regulators of the Wnt pathway, such as *TCF7L2, FBXW7, AXIN2* and *RNF43* genes.

APC mutants were identified in 14 out 29 cases (48.28%), while other alterations were found in the same pathway in wild-type APC cases, such as a *FBXW7* alteration (p.R465H) found in one case, and an *AXIN2* downregulation or mutation (p.G495A) in other three cases. Similarly, *RNF43* alterations were identified in one MSS APC wild-type cell line and in four out of five microsatellite unstable cases, highlighting the association of these alterations with the MSI-H phenotype, as previously reported (38). Of interest,

multiple alterations affecting this gene co-occurred in two *APC* wild-type cases.

In addition, a specific translocation of the Wnt-agonistic *RSPO3* with *PTPRK* was found in one *BRAF*-mutated *APC* wild-type case, as assessed by RNA-seq analysis on XLs and validated by quantitative RT-PCR (Fig.3B and Supplementary Fig. S7A and S7B). No *RSPO2* gene fusions were identified in our collection of XLs.

We also evaluated the genetic status of *TP53* that resulted altered in 79.3% of cases (23/29). In addition, alterations in *ATM* were found in 10.35% (3/29), one of these co-occurred with a *TP53* mutation, while the remaining two cases were *TP53* wild-type.

The TGF-β signaling pathway is also frequently deregulated in CRC (2). In our dataset, we identified 2 TGFBR1 mutated cases and four cases carrying TGFBR2 alterations. Consistent with previous findings (39), all of these models were MSI-H. Moreover, alterations in SMAD4 were identified in 7 cases, while other 4 cases (CRC0781-XL, CRC0691-XL, HROC59 and HROC239) did not express SMAD4, as assessed by RNA-seq analysis (Fig. 3A and Supplementary Fig. S7C); lack of SMAD4 expression was ascribed to LOH in three cases (Fig. 2C). No alterations in SMAD2 and SMAD3 were found; three MSI-H cases (IRCC-1-XL, HROC131 and HROC285) displayed deletions in the ACVR2A gene, a TGF-β family member, known as a potential player in CRC (2).

Genetic alterations in RAS-MAPK and PI3K-PTEN pathways are very frequent in CRC. Activating mutations in *KRAS* - including the less common mutations p.Q61K (exon 3), p.K117N and p.A146V (exon 4) - and in *NRAS* were identified in 13 out 29 XLs and matched PDXs, thus reflecting the prevalence of *RAS* mutations found in CRC patients. We identified five models harboring *BRAF* p.V600E mutations, two derived from MSS metastases and 3 from primary MSI-H CRC. While alterations in *KRAS*, *NRAS* and *BRAF* displayed a significant pattern of mutual exclusivity, 2 out of ten *KRAS* mutated models displayed also co-occurring activating mutations in the *PIK3CA* gene (codon 1047 and 1049). Only one PIK3CA variant (p.H1047R) was identified in a KRAS/NRAS/BRAF wild-type case.

Missense (p.A86E and p.D92Y) as well as nonsense (p.R233*) mutations affecting the *PTEN* gene were identified in three cases. Although both missense mutations fall in the proximity or within the WPD loop of the catalytic site, only alterations in codon 92 were reported to fully abrogate PTEN function (40). In addition, western blot analysis highlighted the loss of PTEN expression in two xeno cell lines, CRC0394-XL carrying the p.R233* nonsense mutation, and HROC277, in which PTEN expression was consistently decreased at the transcript level due to genetic loss of the PTEN locus (Fig. 2C, Supplementary Fig. S7C and S7D).

Alternative ways of activation for PIK3CA pathway include transcriptional upregulation of IGF2 (Supplementary Fig. S7E). In our panel of XLs, we identified seven cases (CRC0081-XL, CRC0438-XL, HROC46, HROC278Met, HROC59, HROC239 and HROC285) overexpressing the IGF2 transcript, as assessed by RNA-seq analysis (log2 ratio > 1.5) (Supplementary Fig. S7E). Of note, overexpression of IGF2 was not associated to genomic amplification of the IGF2 gene, as previously reported (Fig. 2C) (2,41).

Other identified alterations in the MAPK pathway included an activating mutation (p.K57N) in MAP2K1 (encoding the protein MEK1) observed in one XL and its matched PDX (CRC0740).

Evaluation of genetic alterations in receptors of the ERBB family receptors unveiled the presence of an *EGFR* extracellular mutation (p.G465E) in a *PIK3CA*-mutated case (CRC0104-XL). Of interest, this *EGFR* mutation was previously found to be associated with acquired resistance to anti-EGFR therapies (42,43). Indeed, its corresponding xenopatient was established from a liver metastasis of a patient who had received cetuximab treatment within six months prior to resection (17).

Another clinically actionable ERBB family member frequently altered in cancer is ERBB2 (also known as HER2); ERBB2 gene amplification was identified in two MSS wild-type RAS/BRAF cases (CRC0080-XL and CRC0186-XL) (Fig. 2C and Supplementary Fig. S8A). Interestingly, CRC0186-XL showed a

remarkable increase in ERBB2 gene copy number (n=33.2), while CRC080-XL displayed a less pronounced increase (n=6.1), comparable to that present in the breast cancer cell lines BT474 and SKBR3 (n=6.5 and n=5.6, respectively) (Supplementary Fig. S8B). In these cell lines, ERBB2 gene copy number gain led to protein overexpression (Supplementary Fig. S8C). Moreover, CRC0080-XL carried also an ERBB2 activating mutation (p.L866M) affecting the intracellular kinase domain of the receptor (44) (Fig. 3A).
Mutations in the extracellular region of the ERBB2 receptor (p.S310F and p.L465V) were identified in two additional xenopatients and matched cell lines (HROC87 and HROC112Met, respectively). Of note, ERBB2 p.S310F was co-present in a BRAF p.V600E mutated case. Similarly, in another BRAF mutated case (HROC131), we identified co-occurrence of the ERBB3 p.V104M variant (Fig. 3A).

Finally, expression analysis revealed that molecular subtypes found in XLs correlated with transcriptional classes identified by the recently described CRC intrinsic subtype (CRIS) analysis (28) (Fig. 3A and C). Indeed, hierarchical clustering based on CRIS-classifier genes, as already seen for the whole transcriptome, preferentially clustered XLs with the PDXs of origin (Supplementary Fig. S4). In total, 26 out of 29 XLs were significantly assigned to a specific CRIS class (Supplementary Table S3). In particular, MSI-H and KRAS/BRAF mutant cells were enriched in CRIS-A, while CRIS-C included KRAS/BRAF-WT lines. In addition, 2 cell lines that displayed TGF-β pathway alteration clustered in CRIS-B and IGF-2 overexpressing cells were included in CRIS-D group (Fig. 3 and Supplementary Table S3). A similar distribution in genetic alterations was also found for the consensus molecular subtypes (CMS): MSI and BRAF mutational status were enriched in CMS1, and KRAS mutant model were partially depleted in CMS2.
Of note, associations between CRIS and CMS transcriptional classes are also evident. Indeed CRIS-A samples are assigned to CMS1 and CMS3, while CRIS-C is predominantly composed of CMS2. Interestingly, in the context of stromal depleted samples, CRIS-D samples are also classified as CMS4, likely due to the similarities in stem cells rewiring program in intrinsic cancer

cell intrinsic traits – i.e. LGR5 pathway – captured by the two classifiers. (Fig. 3C).

Sensitivity to EGFR blockade is maintained in xeno-derived cell lines

Anti-EGFR monoclonal antibodies are approved to treat metastatic CRC and achieve a 10% objective response rate in unselected patients treated with antibody monotherapy (45). Clinical benefit to EGFR blockade increases up to 25% of cases when tumors are stratified for wild-type *KRAS*, *NRAS* and *BRAF* (45). To assess whether XLs could recapitulate this genotype-drug response, and therefore could be exploited as a reliable platform for drug testing, we checked the entire cell platform for cetuximab sensitivity.

Based on the level of response to cetuximab, the cell collection was classified in 3 different groups: sensitive (n=2), intermediate (n=5) and resistant (n=22) (Fig. 4). As observed in the clinic, *KRAS, BRAF* or *NRAS* mutations, as well as other genetic alterations (i.e. MAP2K1 p.K57N and EGFR p.G465E) associated with resistance to EGFR blockade in CRC cells (Fig. 4) (17,25,42,46). Similarly, ERBB2-amplified CRC0186-XL was refractory to cetuximab treatment, confirming previous findings obtained by pre-clinical and clinical data from the same individual patient (12,17). Conversely, the other ERBB2-amplified and mutated CRC0080-XL cell line displayed higher sensitivity to cetuximab treatment.

Integration of multi-layer data revealed also a positive correlation with the transcriptional classes obtained from the cancer cell intrinsic classifier (Supplementary Table S3). As previously reported, CRIS-A resulted enriched in both MSI-H and KRAS/BRAF mutated cases (IRCC-1-XL, HROC50, HROC277, HROC147, CRC0438-XL and CRC0781-XL) with marked resistance to cetuximab treatment (Fig. 3C), and CRIS-C depleted for KRAS mutant cells (IRCC-5A-XL, IRCC-5B-XL, CRC0078-XL, CRC0080-XL and CRC394-XL), showing sensitivity to anti-EGFR treatment. Confirming

previous evidences(28,47), both CRIS-C and CMS2 transcriptional classes are enriched for cetuximab responsive models (Supplementary Table S4).

Expression analysis identifies therapeutic targets in xeno-derived cell lines

RNA-seq data analysis unveiled potentially actionable targets in the XL platform. Since receptor tyrosine kinases are often implicated in cancer and are ideally suited for pharmacological inhibition, the analysis was focused on this class of proteins. Overall, 30 distinct RTKs passed the filtering process during RNA-seq analysis due to their expression level in our XL collection. Outlier expression analysis identified ten (log2 ratio ≥ 3.5) overexpressed receptor tyrosine kinase genes including ERBB2, AXL, FGFR1, NTRK1, NTRK2, ROR2, RET, EPHA4, EPHB1 and EPHB6 in individual xeno-derived cell lines (Fig.5).

Among these, only two identified RTKs were overexpressed in RAS/RAF wild-type cells ("WT-specific" RTKs), namely NTRK1 (TRKA) and ERBB2 (HER2). Integrated analysis showed that their overexpression is associated to different molecular alterations, namely, gene translocation (NTRK1) or gene amplification (ERBB2) respectively (Fig. 5 and Supplementary Fig. S8A, S8B and S9).

In the TRKA overexpressed case, we identified a genetic rearrangement involving exon 10 or 11 of LMNA and exon 11 of NTRK1 genes (48,49). The two distinct splice variants encoding exons 1-10 or 1-11 of LMNA gene fused to exons 11-16 of NTRK1 gene reported in the donor patient were identified by RNA-seq analysis in the matched xeno cell line (Supplementary Fig. S9). Remarkably, the pharmacological response of these xeno-cells resembled the one observed in the corresponding PDX upon entrectinib treatment (48).

In order to validate ERBB2 amplification as an actionable target, we first evaluated biochemical pathway activation. In contrast to CRC0186-XL, CRC0080-XL showed lower level of HER2 protein, but still an active downstream pathway due to the presence of the L866M mutation (Supplementary Fig. S8C).

This analysis confirmed that overexpression and constitutive signaling as well as downstream pathway activation of HER2 paralleled the extension of ERBB2 gene amplification.

To validate overexpressed ERBB2 as putative target for pharmacological inhibition (12), we challenged our *in vitro* models with rational combinations of drugs. We initially tested the effect of drugs (lapatinib and trastuzumab) that have previously shown activity on HER2 positive metastatic CRC patients (12,50). While monotherapy showed no or mild effects, both lines (CRC0080-XL and CRC0186-XL) displayed significant sensitivity to the combinatorial treatment (Fig. 5C) in short-term proliferation assays. We performed western blot analysis showing that the biological effect of the drugs on ERBB2-amplified CRC cell lines is matched by the quenching of HER2 biochemical downstream signals (Fig. 5D). Of note, similar results were also obtained *in vivo*, when the corresponding ERBB2-amplified xenopatients were challenged with the same treatments for 28-35 days (51) (Fig. 5E), thus confirming the XL as a valuable preclinical model to assess drug response and to dissect biochemical pathway activation.

Combinatorial treatment with lapatinib and trastuzumab of HER2+ patients has been exploited in the recently completed proof-of-concept clinical trial named HERACLES (50). Our previous work based on liquid biopsy analysis has shown that 32 out of 35 patients treated with anti-HER2 inhibitors relapsed after initial response revealing the presence of molecular mechanisms potentially responsible for resistance (primary or acquired) to HER2 blockade (52). In particular, either KRAS or BRAF mutated clones were preferentially found in patients who were initially refractory to anti-HER2 treatment (primary resistance), while PIK3CA mutations were detected at progression in patients who had responded to lapatinib+trastuzumab combinatorial treatment and then relapsed (acquired resistance). In order to validate these genetic players as putative drivers of resistance, we transduced HER2+ CRC XL-cells by means of lentiviral vectors and performed pharmacological and biochemical analysis of treated cells. We found that overexpression of mutant alleles of KRAS, BRAF and PIK3CA conferred

resistance to combinatorial treatment (Fig. 6A and Supplementary Fig. S10A), due to sustained ERK and/or AKT activation (Fig. 6B and Supplementary Fig. S10B).

4. Discussion

Human cancer cell lines represent a tool extensively exploited in the last half century to study the biology of different tumor types as well as for extensive drug screening efforts. Exceptional efforts such as the generation of the Cancer Cell Line Encyclopedia (7) have been pivotal for unveiling novel genomic correlates of drug sensitivities in 36 tumor types, thus coupling cell line genomic annotation to individual pharmacological profiles.

We have previously described a bowel cancer cell-line collection including 151 established cell lines (53). This platform can recapitulate the majority of CRC molecular and transcriptional subtypes previously defined in patients. Genomic analysis has allowed the identification of novel outliers in response to selected targeted agents. Yet, infrequent molecular subtypes or tumors carrying rare genetic lesions may be under-represented or completely absent in the above-mentioned CRC cell lines collection. For instance, while *ERRB2* amplified cell line models have been widely employed for breast cancer studies, none of the models in the 151-CRC cell collection were found to carry this alteration. In this work, we provide an accurate characterization of two novel cancer cell lines of colorectal origin bearing *ERRB2* amplification – which could be used in the future to further dissect the role of ERBB2 oncogenic signaling in the context of gut pathophysiology.

Cell lines may have intrinsic limitations that hamper representation of the whole tumor disease in the patient. Common immortalized cell lines have been adapted to growing on plastic for decades and this may have caused genetic drifts and the acquisition of phenotypic features different from original cancers in patients. It is now evident that CRC is indeed a heterogeneous disease whose development and progression rely on continuous evolution of

multiple clones under environmental and pharmacological pressure (54).For this reason, precision medicine, which is based on the design of personalized treatments, requires a repertoire of preclinical models that mirror the *in vivo* tumor more closely than conventional cancer cell lines. Derivation of long-term culture directly from tumor samples has proven to be difficult and inefficient, at least in CRC, not only because of contaminations but also due to failure to adhere to the culture dish or loss of proliferative capacity after a few passages (55). Establishment of short-term tumor cell culture has proven to be successful (56), but long-term experiments as well as gene editing by lentiviral transduction for experimental needs might not be feasible.

Generation of PDX models has contributed to the identification and validation of biomarkers of response and resistance to current anticancer therapies (17). However, they present some important shortcomings such as long time for establishment and expansion, cumbersome manipulation and cost-ineffective experiments, especially if set on a large scale.

Here we present a complementary approach based on PDX-derived cell lines. We have established 29 cell lines from PDXs originally derived from samples of primary or metastatic CRC tissue. Initially, to assess whether passage from 3D/*in vivo* to 2D/*in vitro* growing condition did not select for genomic alterations that may compromise the biology of the tumor, we performed genetic and transcriptomic analyses to compare samples. Whole exome sequencing confirmed preservation of the PDX main genetic features in 2D-derived cell lines, which remained stable during extended passaging. Importantly, in contrast to conventional cell lines, we could take advantage of the matched normal (germline) DNA to identify somatic tumor alterations versus germline variants. High concordance in genetic variation was evident between the two platforms when considering cancer-related genes, although we identified a variable range of private mutation variation in each PDX-XL pair. This variation could either reflect the enrichment/depletion of a sub-clonal population, or the acquisition of additional mutations during derivation or propagation of the XL-cells with respect to the matched xenopatient.

In line with these observations, gene expression profiles of matched XLs and PDX models were highly correlated, confirming the stability of the tumor models upon propagation *in vitro*. Likewise, the molecular classification of cell lines based on transcriptional traits did reproduce the occurrence observed in the original tumor, and conserved its distinct associations with genetic, molecular and pharmacological associations. Indeed, despite the small size of our dataset, we could confirm the association between CRIS-C assignment and higher probability of response to anti-EGFR treatment, as well as the association between CRIS-A and MSI status.
Of note, comparing the classification of CMS and CRIS in the context of cell models we unveiled correspondence in classification of CMS4 and CRIS-D, likely emerging due to absence of stromal cells in the cancer cells profiles and similarities in intrinsic traits, such as the LGR5 stem-like signature detected in both subtypes (29,57).

We next assessed whether these new preclinical models could represent a reliable tool for biomarker validation and pharmacological screening. The identification of genomic players that may confer sensitivities or resistance to selected therapies is of paramount importance for the design of effective therapeutic strategies. As a proof of concept, we have shown two independent examples in which the results observed with the XLs highly mirrored what was previously observed in PDXs. We focused our attention on anti-HER therapies, which are approved or investigational agents for patients with RAS wild-type and HER2-overexpressing CRC tumors. When we challenged XL-cells with the anti-EGFR monoclonal antibody cetuximab, we observed a distribution of responses that closely correlated with genetic and non-genetic biomarkers of resistance identified in the corresponding PDXs.

As a paradigmatic example, we have identified in our cohort two XLs harboring HER2 receptor overexpression due to ERBB2 gene amplification. When these cells where challenged with a rational combination of HER inhibitors, we observed that their proliferation was severely impaired, as previously seen with the matched PDXs treated with the same drugs. This further confirms the XL platform as a valuable tool for genotype-drug

correlation studies. Moreover, we have validated molecular determinants of resistance to dual HER2 blockade in CRC, as emerged from the HERACLES trial (52), suggesting that combinatorial treatment with other targeting agents might be beneficial for HER2+ patients (51).

In summary, we have established a compendium of 29 XLs that substantially recapitulate the genomic features of matched PDXs, with the great advantage of providing ease of handling and amenability for high-throughput compound screening. This may significantly accelerate the translation of drug-response information from *in vitro* testing to clinical practice.

Another notable advantage provided by this approach is the possibility of generating cell lines with genotypes that are either missing o rare in common commercial repositories (i.e. HER2 amplified or NRAS Q61 mutated cell lines), as well as increasing the number of MSS CRC cell lines, which are typically underrepresented in existing cell line collections (34,53). It derives that the more XLs we manage to establish, the more we will have the chance to understand and model the molecular underpinnings of CRC heterogeneity.
In conclusion, we present a novel preclinical platform that – owing to its reliability, large-scale use and proximity to original patient tumors – could offer easier and more rapid access to the dissection of genetic and patient-selective therapeutic vulnerabilities, thus leading to the design of novel effective treatment strategies.

Acknowledgments: The authors thank all the members of Molecular Oncology Laboratory at Candiolo Cancer Institute for critically reading the manuscript and, in particular, Mariangela Russo, Giulia Siravegna, Benedetta Mussolin, Francesca Lodi and Riccardo Volta for their help with experiments and constructive discussion. The authors also thank Dr. Paolo Luraghi for providing CRC0080-XL and for his helpful discussion on cell line derivation.

References

1. Kelloff GJ, Sigman CC. Cancer biomarkers: selecting the right drug for the right patient. Nat Rev Drug Discov **2012**;11(3):201-14 doi 10.1038/nrd3651.
2. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature **2012**;487(7407):330-7 doi 10.1038/nature11252.
3. Friedman AA, Letai A, Fisher DE, Flaherty KT. Precision medicine for cancer with next-generation functional diagnostics. Nat Rev Cancer **2015**;15(12):747-56 doi 10.1038/nrc4015.
4. Weinstein IB, Joe A. Oncogene addiction. Cancer Res **2008**;68(9):3077-80; discussion 80 doi 10.1158/0008-5472.CAN-07-3293.
5. Bruna A, Rueda OM, Greenwood W, Batra AS, Callari M, Batra RN*, et al.* A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. Cell **2016**;167(1):260-74.e22 doi 10.1016/j.cell.2016.08.041.
6. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S*, et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res **2013**;41(Database issue):D955-61 doi 10.1093/nar/gks1111.
7. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S*, et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature **2012**;483(7391):603-7 doi 10.1038/nature11003.
8. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW*, et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature **2012**;483(7391):570-5 doi 10.1038/nature11005.
9. Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, Coletti ME*, et al.* Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. Cancer Discov **2015**;5(11):1210-23 doi 10.1158/2159-8290.CD-15-0235.
10. Gillet JP, Varma S, Gottesman MM. The clinical relevance of cancer cell lines. J Natl Cancer Inst **2013**;105(7):452-8 doi 10.1093/jnci/djt007.
11. Byrne AT, Alférez DG, Amant F, Annibali D, Arribas J, Biankin AV*, et al.* Interrogating open issues in cancer precision medicine with patient-derived xenografts. Nat Rev Cancer **2017**;17(4):254-68 doi 10.1038/nrc.2016.140.
12. Bertotti A, Migliardi G, Galimi F, Sassi F, Torti D, Isella C*, et al.* A molecularly annotated platform of patient-derived xenografts ("xenopatients") identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. Cancer Discov **2011**;1(6):508-23 doi 10.1158/2159-8290.CD-11-0109.
13. Julien S, Merino-Trigo A, Lacroix L, Pocard M, Goéré D, Mariani P*, et al.* Characterization of a large panel of patient-derived tumor xenografts representing the clinical heterogeneity of human colorectal cancer. Clin Cancer Res **2012**;18(19):5314-28 doi 10.1158/1078-0432.CCR-12-0372.
14. Galimi F, Torti D, Sassi F, Isella C, Corà D, Gastaldi S*, et al.* Genetic and expression analysis of MET, MACC1, and HGF in metastatic colorectal cancer: response to met inhibition in patient xenografts and pathologic

correlations. Clin Cancer Res **2011**;17(10):3146-56 doi 10.1158/1078-0432.CCR-10-3377.
15. Guenot D, Guérin E, Aguillon-Romain S, Pencreach E, Schneider A, Neuville A*, et al.* Primary tumour genetic alterations and intra-tumoral heterogeneity are maintained in xenografts of human colon cancers showing chromosome instability. J Pathol **2006**;208(5):643-52 doi 10.1002/path.1936.
16. Ben-David U, Ha G, Tseng YY, Greenwald NF, Oh C, Shih J*, et al.* Patient-derived xenografts undergo mouse-specific tumor evolution. Nat Genet **2017**;49(11):1567-75 doi 10.1038/ng.3967.
17. Bertotti A, Papp E, Jones S, Adleff V, Anagnostou V, Lupo B*, et al.* The genomic landscape of response to EGFR blockade in colorectal cancer. Nature **2015**;526(7572):263-7 doi 10.1038/nature14969.
18. Damhofer H, Ebbing EA, Steins A, Welling L, Tol JA, Krishnadath KK*, et al.* Establishment of patient-derived xenograft models and cell lines for malignancies of the upper gastrointestinal tract. J Transl Med **2015**;13:115 doi 10.1186/s12967-015-0469-1.
19. Knudsen ES, Balaji U, Mannakee B, Vail P, Eslinger C, Moxom C*, et al.* Pancreatic cancer cell lines as patient-derived avatars: genetic characterisation and functional utility. Gut **2017** doi 10.1136/gutjnl-2016-313133.
20. Maletzki C, Stier S, Gruenert U, Gock M, Ostwald C, Prall F*, et al.* Establishment, characterization and chemosensitivity of three mismatch repair deficient cell lines from sporadic and inherited colorectal carcinomas. PLoS One **2012**;7(12):e52485 doi 10.1371/journal.pone.0052485.
21. Luraghi P, Bigatto V, Cipriano E, Reato G, Orzan F, Sassi F*, et al.* A molecularly annotated model of patient-derived colon cancer stem-like cells to assess genetic and non-genetic mechanisms of resistance to anti-EGFR therapy. Clin Cancer Res **2017** doi 10.1158/1078-0432.CCR-17-2151.
22. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD*, et al.* MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics **2014**;30(7):1015-6 doi 10.1093/bioinformatics/btt755.
23. Conway T, Wazny J, Bromage A, Tymms M, Sooraj D, Williams ED*, et al.* Xenome--a tool for classifying reads from xenograft samples. Bioinformatics **2012**;28(12):i172-8 doi 10.1093/bioinformatics/bts236.
24. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.: arXiv:1303.3997v1; 2013.
25. Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G*, et al.* Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. Nat Med **2015**;21(7):795-801 doi 10.1038/nm.3870.
26. Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL*, et al.* MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res **2010**;38(18):e178 doi 10.1093/nar/gkq622.

27. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **2011**;12:323 doi 10.1186/1471-2105-12-323.
28. Isella C, Brundu F, Bellomo SE, Galimi F, Zanella E, Porporato R*, et al.* Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. Nat Commun **2017**;8:15107 doi 10.1038/ncomms15107.
29. Sveen A, Bruun J, Eide PW, Eilertsen IA, Ramirez L, Murumägi A*, et al.* Colorectal Cancer Consensus Molecular Subtypes Translated to Preclinical Models Uncover Potentially Targetable Cancer Cell Dependencies. Clin Cancer Res **2018**;24(4):794-806 doi 10.1158/1078-0432.CCR-17-1234.
30. Fu L, Medico E. FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. BMC Bioinformatics **2007**;8:3 doi 10.1186/1471-2105-8-3.
31. De Roock W, Claes B, Bernasconi D, De Schutter J, Biesmans B, Fountzilas G*, et al.* Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. Lancet Oncol **2010**;11(8):753-62 doi 10.1016/S1470-2045(10)70130-3.
32. Di Nicolantonio F, Arena S, Gallicchio M, Zecchin D, Martini M, Flonta SE*, et al.* Replacement of normal with mutant alleles in the genome of normal human cells unveils mutation-specific drug responses. Proc Natl Acad Sci U S A **2008**;105(52):20864-9 doi 10.1073/pnas.0808757105.
33. Gatalica Z, Vranic S, Xiu J, Swensen J, Reddy S. High microsatellite instability (MSI-H) colorectal carcinoma: a brief review of predictive biomarkers in the era of personalized medicine. Fam Cancer **2016**;15(3):405-12 doi 10.1007/s10689-016-9884-6.
34. Mouradov D, Sloggett C, Jorissen RN, Love CG, Li S, Burgess AW*, et al.* Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. Cancer Res **2014**;74(12):3238-47 doi 10.1158/0008-5472.CAN-14-0013.
35. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R*, et al.* A census of human cancer genes. Nat Rev Cancer **2004**;4(3):177-83 doi 10.1038/nrc1299.
36. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR*, et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. Nature **2014**;505(7484):495-501 doi 10.1038/nature12912.
37. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV*, et al.* Signatures of mutational processes in human cancer. Nature **2013**;500(7463):415-21 doi 10.1038/nature12477.
38. Giannakis M, Hodis E, Jasmine Mu X, Yamauchi M, Rosenbluh J, Cibulskis K*, et al.* RNF43 is frequently mutated in colorectal and endometrial cancers. Nat Genet **2014**;46(12):1264-6 doi 10.1038/ng.3127.
39. Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J*, et al.* Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science **1995**;268(5215):1336-8.
40. Rodríguez-Escudero I, Oliver MD, Andrés-Pons A, Molina M, Cid VJ, Pulido R. A comprehensive functional analysis of PTEN mutations: implications

in tumor- and autism-related syndromes. Hum Mol Genet **2011**;20(21):4132-42 doi 10.1093/hmg/ddr337.
41. Zanella ER, Galimi F, Sassi F, Migliardi G, Cottino F, Leto SM*, et al.* IGF2 is an actionable target that identifies a distinct subpopulation of colorectal cancer patients with marginal response to anti-EGFR therapies. Sci Transl Med **2015**;7(272):272ra12 doi 10.1126/scitranslmed.3010445.
42. Arena S, Siravegna G, Mussolin B, Kearns JD, Wolf BB, Misale S*, et al.* MM-151 overcomes acquired resistance to cetuximab and panitumumab in colorectal cancers harboring EGFR extracellular domain mutations. Sci Transl Med **2016**;8(324):324ra14 doi 10.1126/scitranslmed.aad5640.
43. Arena S, Bellosillo B, Siravegna G, Martínez A, Cañadas I, Lazzari L*, et al.* Emergence of Multiple EGFR Extracellular Mutations during Cetuximab Treatment in Colorectal Cancer. Clin Cancer Res **2015**;21(9):2157-66 doi 10.1158/1078-0432.CCR-14-2821.
44. Kavuri SM, Jain N, Galimi F, Cottino F, Leto SM, Migliardi G*, et al.* HER2 activating mutations are targets for colorectal cancer treatment. Cancer Discov **2015**;5(8):832-41 doi 10.1158/2159-8290.CD-14-1211.
45. Misale S, Di Nicolantonio F, Sartore-Bianchi A, Siena S, Bardelli A. Resistance to anti-EGFR therapy in colorectal cancer: from heterogeneity to convergent evolution. Cancer Discov **2014**;4(11):1269-80 doi 10.1158/2159-8290.CD-14-0462.
46. Russo M, Siravegna G, Blaszkowsky LS, Corti G, Crisafulli G, Ahronian LG*, et al.* Tumor Heterogeneity and Lesion-Specific Response to Targeted Therapy in Colorectal Cancer. Cancer Discov **2016**;6(2):147-53 doi 10.1158/2159-8290.CD-15-1283.
47. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C*, et al.* The consensus molecular subtypes of colorectal cancer. Nat Med **2015**;21(11):1350-6 doi 10.1038/nm.3967.
48. Russo M, Misale S, Wei G, Siravegna G, Crisafulli G, Lazzari L*, et al.* Acquired Resistance to the TRK Inhibitor Entrectinib in Colorectal Cancer. Cancer Discov **2016**;6(1):36-44 doi 10.1158/2159-8290.CD-15-0940.
49. Sartore-Bianchi A, Ardini E, Bosotti R, Amatu A, Valtorta E, Somaschini A*, et al.* Sensitivity to Entrectinib Associated With a Novel LMNA-NTRK1 Gene Fusion in Metastatic Colorectal Cancer. J Natl Cancer Inst **2016**;108(1) doi 10.1093/jnci/djv306.
50. Sartore-Bianchi A, Trusolino L, Martino C, Bencardino K, Lonardi S, Bergamo F*, et al.* Dual-targeted therapy with trastuzumab and lapatinib in treatment-refractory, KRAS codon 12/13 wild-type, HER2-positive metastatic colorectal cancer (HERACLES): a proof-of-concept, multicentre, open-label, phase 2 trial. Lancet Oncol **2016**;17(6):738-46 doi 10.1016/S1470-2045(16)00150-9.
51. Leto SM, Sassi F, Catalano I, Torri V, Migliardi G, Zanella ER*, et al.* Sustained Inhibition of HER3 and EGFR Is Necessary to Induce Regression of HER2-Amplified Gastrointestinal Carcinomas. Clin Cancer Res **2015**;21(24):5519-31 doi 10.1158/1078-0432.CCR-14-3066.
52. Siravegna G, Lazzari L, Crisafulli G, Sartore-Bianchi A, Mussolin B, Cassingena A*, et al.* Radiologic and Genomic Evolution of Individual Metastases during HER2 Blockade in Colorectal Cancer. Cancer Cell **2018**;34(1):148-62.e7 doi 10.1016/j.ccell.2018.06.004.

53. Medico E, Russo M, Picco G, Cancelliere C, Valtorta E, Corti G, *et al.* The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets. Nat Commun **2015**;6:7002 doi 10.1038/ncomms8002.
54. Amirouchene-Angelozzi N, Swanton C, Bardelli A. Tumor Evolution as a Therapeutic Target. Cancer Discov **2017** doi 10.1158/2159-8290.CD-17-0343.
55. Dangles-Marie V, Pocard M, Richon S, Weiswald LB, Assayag F, Saulnier P, *et al.* Establishment of human colon cancer cell lines from fresh tumors versus xenografts: comparison of success rate and cell line features. Cancer Res **2007**;67(1):398-407 doi 10.1158/0008-5472.CAN-06-0594.
56. Lee JK, Liu Z, Sa JK, Shin S, Wang J, Bordyuh M, *et al.* Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. Nat Genet **2018**;50(10):1399-411 doi 10.1038/s41588-018-0209-6.
57. Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, *et al.* Stromal contribution to the colorectal cancer transcriptome. Nat Genet **2015**;47(4):312-9 doi 10.1038/ng.3224.

Figure Legends

Figure 1. Exome Sequencing analysis of XLs matches the profile of the corresponding PDX of origin.

A. Site of origin of PDX samples used to generate XLs.

B. Relative distribution of microsatellite status in XL-cells.

C. Number of MSI events in the 29 PDX-XL pairs, stratified by the length of the repeat units identified by exome analysis.

D. Mutational load (number of mutations per Mb) in the 29 PDX-XL pairs, as calculated by exome analysis.

E. The counts of non-synonymous mutations (blue bar) and INDELs (red bar) in the 29 PDX-XL pairs, as calculated by exome analysis.

F. Mutational spectra in the 29 PDX-XL pairs, as relative proportion of the six different possible base-pair substitutions (nucleotide transitions and trasversions).

Microsatellite stable (MSS) and unstable (MSI-H) cases are highlighted in light blue and red, respectively.

Figure 2. Comparison of copy-number variation unveils gain or loss of selected genome segments in PDX-XL matched models.

A. Heat map comparing copy-number aberration between matched PDX and XL models in log2 scale, as assessed by WES. Red colors indicate gains, while blue colors indicate losses. The band on the left represents chromosomes 1-22.

B. Plot of frequency of gains and losses for each position per group (PDX and XL) along the genome, as assessed by WES. Each horizontal dotted line represents chromosomal boundaries.

C. Heat map comparing copy-number between matched PDX and XL models in log2 scale for the indicated genes.

Figure 3. Genomic landscape of deregulated signaling pathways in PDX-XL pairs.

A. Landscape of the gene alterations found in the PDX-XL pairs; missense mutations (gray boxes), nonsense mutations (black boxes), insertion (yellow

boxes) and deletion (violet boxes) as well as co-occurrence of multiple alterations (green boxes) are displayed. Orange squared boxes indicate gene amplification, while blue squared boxes indicate gene downregulation. Genes are grouped according to signaling pathways; cases are grouped according to transcriptional CRIS classification of XLs (as assessed in (28)). Primary tumor site (right and left) for each cell line are also reported. MSI-H cases are also shown.

B. Percentage of gene alterations in PDX-XL pairs (gray bars) are compared to percentage of cases reported in TCGA CRC cases (red bars).

C.Caleydo view of correspondences between the CRIS vs CMS class assignments of XLs.

Figure 4. Cetuximab screening in XLs. Bar graph representing the effect of cetuximab treatment normalized on control untreated cells for the XLs. Viability was assessed by measuring ATP content after 6 days of treatment with cetuximab at the clinically relevant dose of 10 ug/ml. Data represent the average of value obtained from at least three independent experiments, error bars represent SD. Based on their response to cetuximab, cell lines were subdivided in three different groups: resistant (first tertile, range 66-100%), intermediate (second tertile, range 65-34%) and sensitive (third tertile, range 33-1%). MSI-H cells are highlighted in light yellow and for each cell line its CRIS classification is reported.

Figure 5. Identification of overexpressed receptor tyrosine kinases (RTKs) in CRC XLs and effects of HER blockade in ErBB2 amplified cell models.

A. Heatmap showing the relative expression (log2 ratio FPKM) of receptor tyrosine kinases (RTKs) as identified by RNA-seq analysis of 29 XLs. RTKs listed here (n=32) represent the only members of this class of protein passing the filtration process during RNA-seq analysis. KRAS, BRAF and NRAS mutated cases are also indicated.

B. Scatter plot displaying the correlation between absolute ERBB2 expression (FPKM, x-axis) and relative ERBB2 copy-number variation (CNV, log2ratio) in the 29 XLs, as assessed by RNA-seq and WES sequencing

analyses, respectively. ERBB2-overexpressing and amplified xeno-cell lines (CRC0080-XL and CRC01086) are indicated.

C. Short-term viability assay showing the effects of single dose of trastuzumab (TMAB, 25 μg/m), lapatinib (LAP, 120 nM), and combination of lapatinib and trastuzumab (LAP+TMAB) on HER2-amplified CRC0186-XL (left panel) and CRC080-XL (right panel). Viability was assessed by measuring ATP content after 6 days of treatment. Data represent the values obtained from at least three independent experiments in triplicate, normalized on control DMSO-treated cells; error bars indicate SD (lapatinib vs lapatinib+trastuzumab in CRC0186: ***, *P* = 0.001; lapatinib vs lapatinib+trastuzumab: **, P = 0.0036 in CRC080; unpaired Student's t test. IC50 values were calculated from dose inhibition curves using GraphPad Prism software.

D. Activation status/phosphorylation of HER2 receptor and downstream pathway in dose-response experiments in CRC0186-XL (left panel) and CRC080-XL (right panel). Cells were treated with the indicated concentrations of trastuzumab (TMAB), lapatinib (Lap), or the combination of both drugs for 4 hours. Cell extracts were immunoblotted with the indicated antibodies. HSP90 was used as loading control.

E. In vivo tumor growth curves of CRC0186 PDX (n=6) and CRC080 PDX (n=6) treated with the indicated modalities. Data represent the mean of tumor volume at each time point; error bars indicate S.E.M. (lapatinib vs lapatinib+trastuzumab: ***, *P* < 0.001 by two-way Anova Test. Response curves for both models were previously reported in (51) and are reshown here for comparative purposes.

Figure 6. Constitutive expression of mutant RAS/BRAF/PIK3CA mediates resistance to lapatinib+trastuzumab combinatorial treatment in ErBB2 amplified cells.

A. Long term viability assay showing the effect of different indicated doses of lapatinib in absence or in presence of doses of 25 μg/ml of trastuzumab on HER2-amplified CRC0186-XL transduced with lentiviral vector overexpressing BRAF V600E, KRAS G13D or PIK3CA H1047R mutants. Cells transduced

with empty lentiviral vector were used as control (CTRL). Cell viability was measured by crystal violet staining after 15 days of treatment. Representative wells were photographed and reported.

B. Activation status/phosphorylation of HER2 receptor and downstream pathway in dose-response experiments in CRC0186-XL transduced with lentiviral vector overexpressing the indicated alleles. Cells transduced with empty lentiviral vector were used as control. Cells were treated with the combination of 500 nM of lapatinib and 10 μg/ml of trastuzumab (L+T) for 4 hours. Cell extracts were immunoblotted with the indicated antibodies. Vinculin was used as loading control.

Figure 1

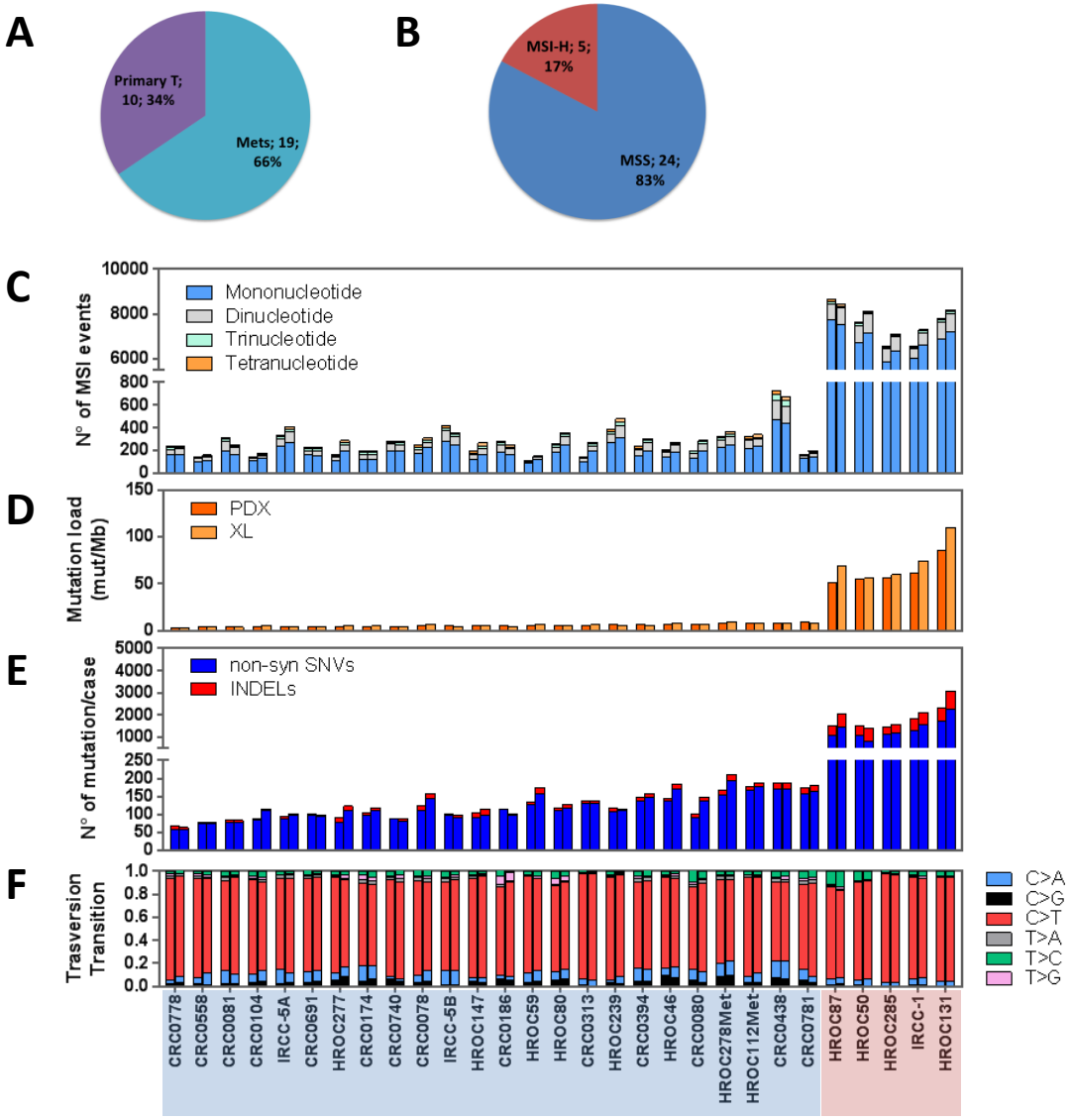

A
Primary T; 10; 34%
Mets; 19; 66%
B
MSI-H; 5; 17%
MSS; 24; 83%
C
N° of MSI events
Mononucleotide
Dinucleotide
Trinucleotide
Tetranucleotide
D
Mutation load (mut/Mb)
PDX
XL
E
N° of mutation/case
non-syn SNVs
INDELs
F
Trasversion
Transition
C>A
C>G
C>T
T>A
T>C
T>G
CRC0778
CRC0558
CRC0081
CRC0104
IRCC-5A
CRC0691
HROC277
CRC0174
CRC0740
CRC0078
IRCC-5B
HROC147
CRC0186
HROC59
HROC80
CRC0313
HROC239
CRC0394
HROC46
CRC0080
HROC278Met
HROC112Met
CRC0438
CRC0781
HROC87
HROC50
HROC285
IRCC-1
HROC131

Figure 2

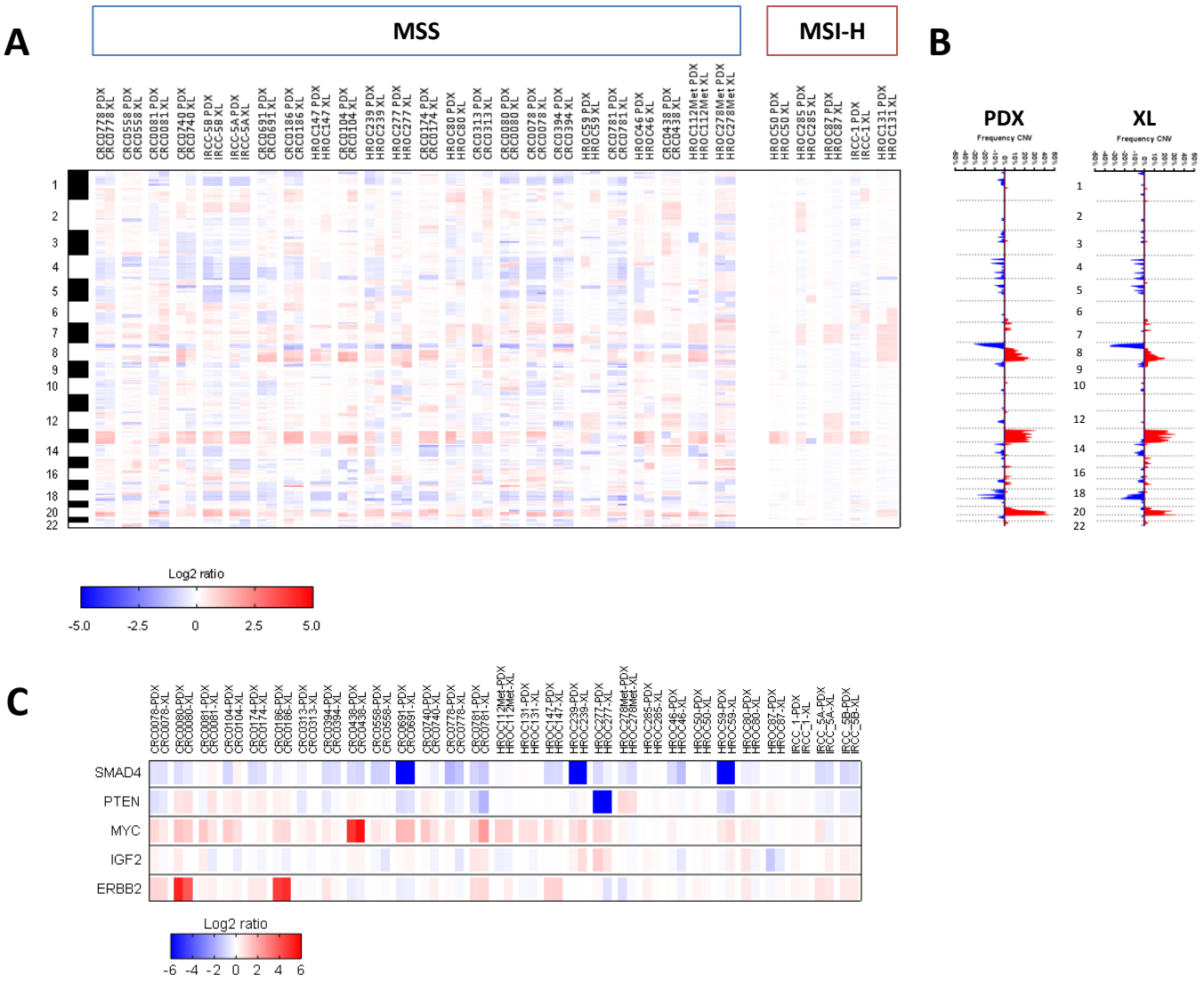

A
MSS
MSI-H
B
PDX
XL
Frequency CNV
Log2 ratio
-5.0
-2.5
0
2.5
5.0
C
SMAD4
PTEN
MYC
IGF2
ERBB2
Log2 ratio
-6
-4
-2
0
2
4
6

Figure 3

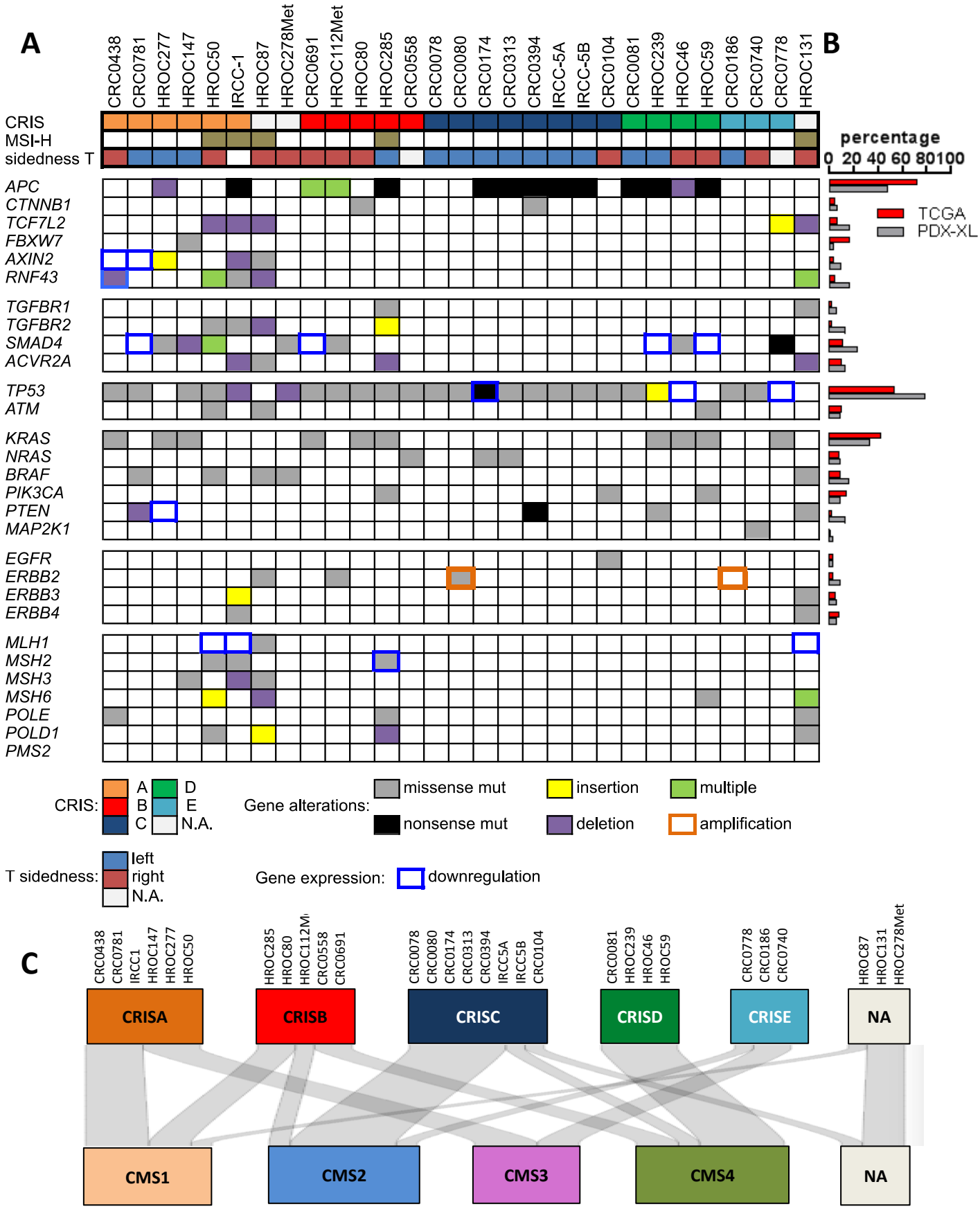

A
B
C
CRC0438
CRC0781
HROC277
HROC147
HROC50
IRCC-1
HROC87
HROC278Met
CRC0691
HROC112Met
HROC80
HROC285
CRC0558
CRC0078
CRC0080
CRC0174
CRC0313
CRC0394
IRCC-5A
IRCC-5B
CRC0104
CRC0081
HROC239
HROC46
HROC59
CRC0186
CRC0740
CRC0778
HROC131
CRIS
MSI-H
sidedness T
percentage
0 20 40 60 80100
TCGA
PDX-XL
APC
CTNNB1
TCF7L2
FBXW7
AXIN2
RNF43
TGFBR1
TGFBR2
SMAD4
ACVR2A
TP53
ATM
KRAS
NRAS
BRAF
PIK3CA
PTEN
MAP2K1
EGFR
ERBB2
ERBB3
ERBB4
MLH1
MSH2
MSH3
MSH6
POLE
POLD1
PMS2
CRIS:
A
B
C
D
E
N.A.
Gene alterations:
missense mut
nonsense mut
insertion
deletion
multiple
amplification
T sidedness:
left
right
N.A.
Gene expression:
downregulation
CRISA
CRISB
CRISC
CRISD
CRISE
NA
CMS1
CMS2
CMS3
CMS4
NA

Figure 4

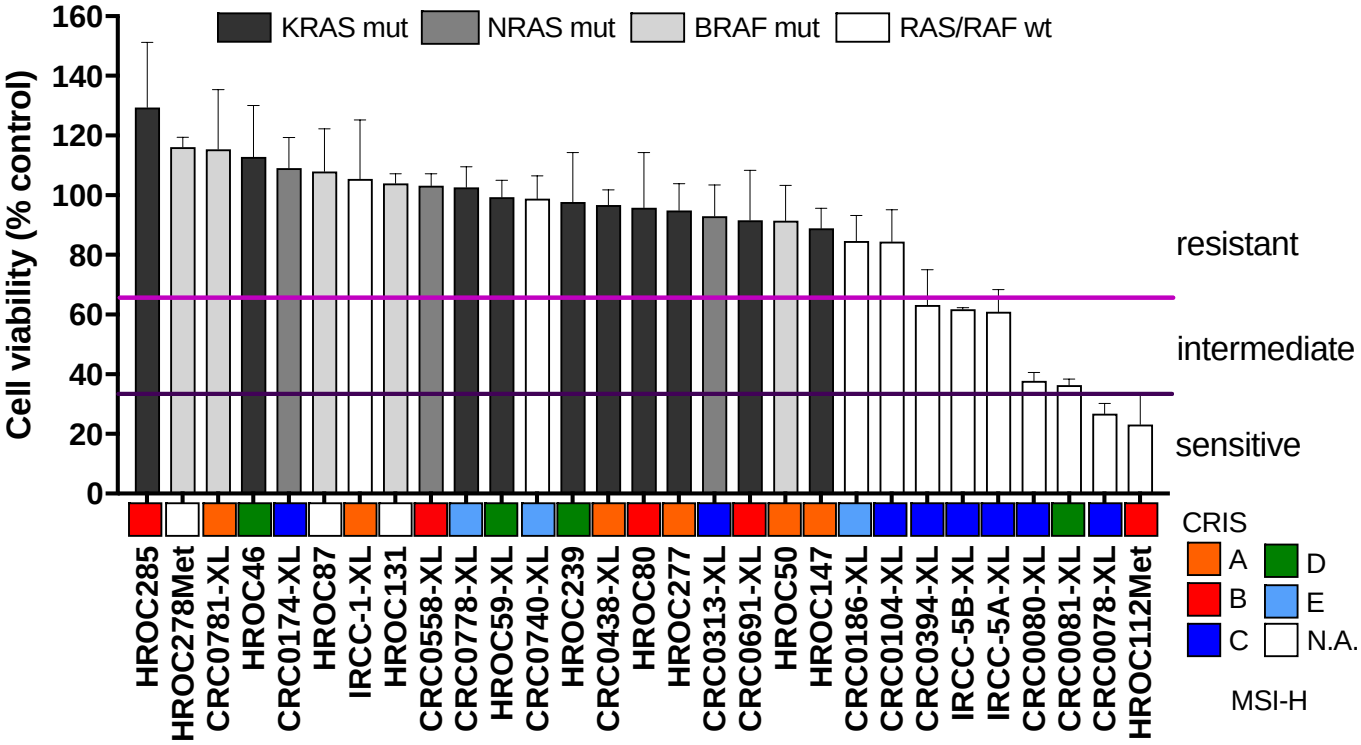
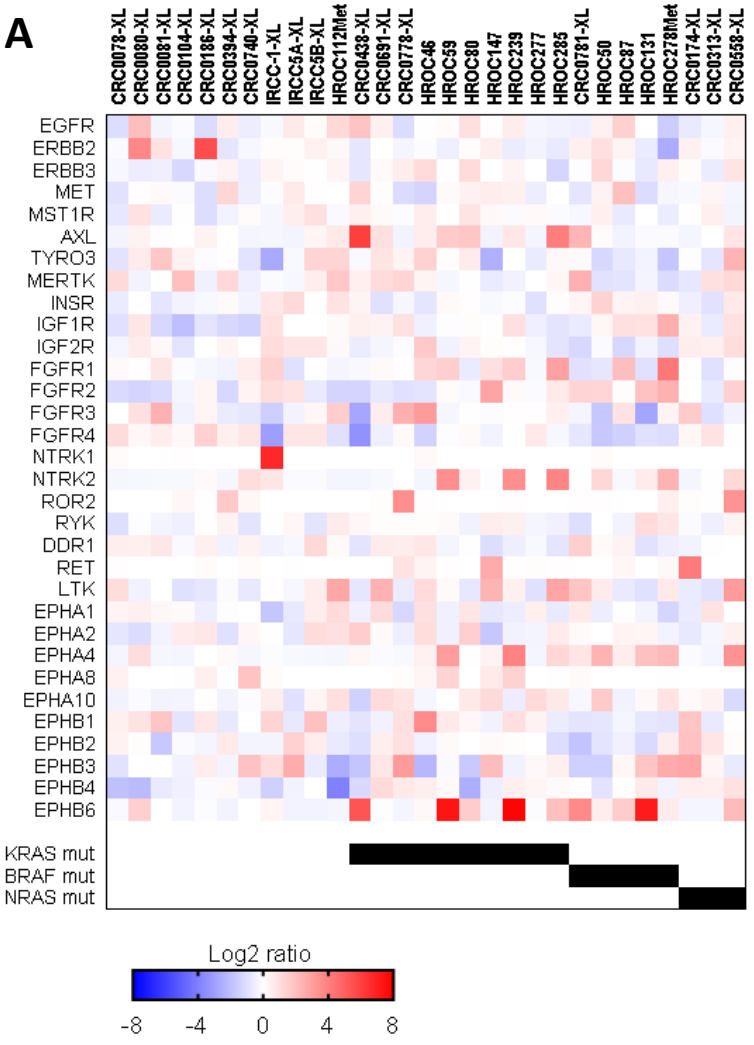
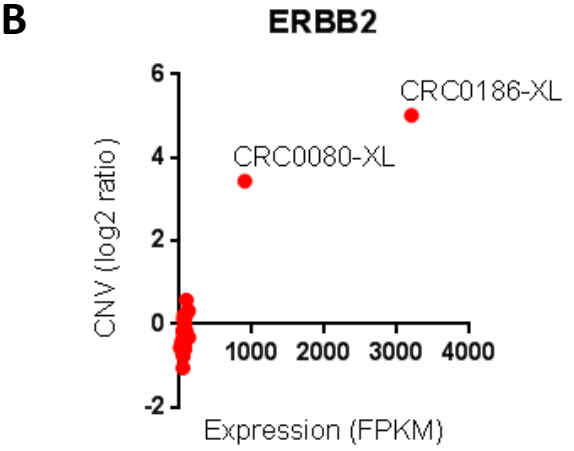
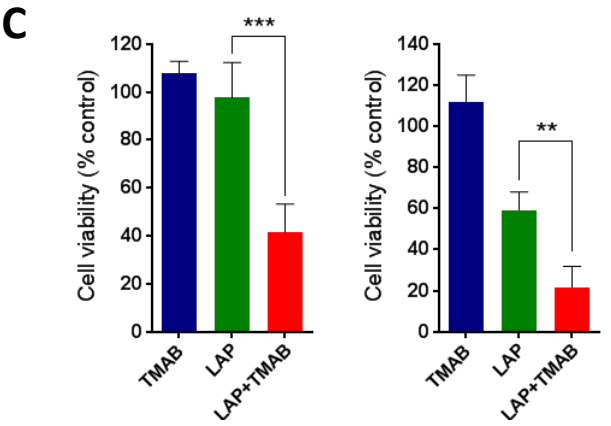

KRAS mut
NRAS mut
BRAF mut
RAS/RAF wt
Cell viability (% control)
160
140
120
100
80
60
40
20
0
resistant
intermediate
sensitive
HROC285
HROC278Met
CRC0781-XL
HROC46
CRC0174-XL
HROC87
IRCC-1-XL
HROC131
CRC0558-XL
CRC0778-XL
HROC59-XL
CRC0740-XL
HROC239
CRC0438-XL
HROC80
HROC277
CRC0313-XL
CRC0691-XL
HROC50
HROC147
CRC0186-XL
CRC0104-XL
CRC0394-XL
IRCC-5B-XL
IRCC-5A-XL
CRC0080-XL
CRC0081-XL
CRC0078-XL
HROC112Met
CRIS
A
B
C
D
E
N.A.
MSI-H

Figure 5

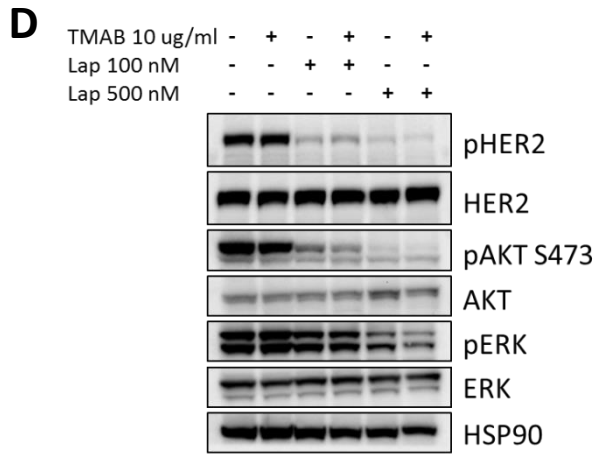






IC50	
drugs	μM
trastuzumab	> 3.43
lapatinib	0.35
lapatinib+trastuzumab	0.16

IC50	
drugs	μM
trastuzumab	> 3.43
lapatinib	0.13
lapatinib+trastuzumab	0.05



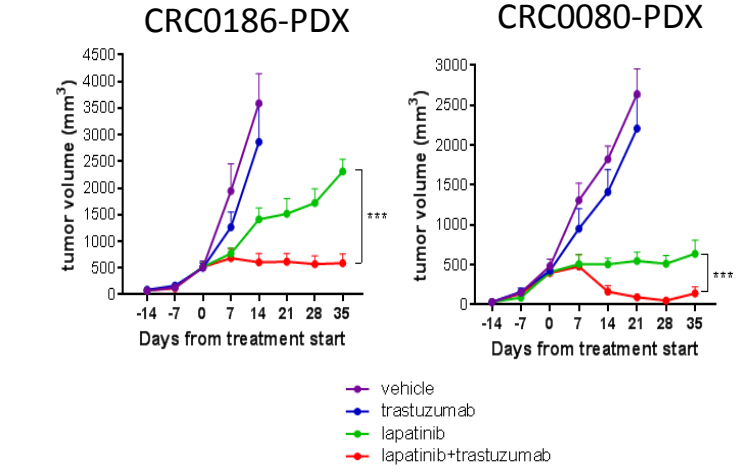
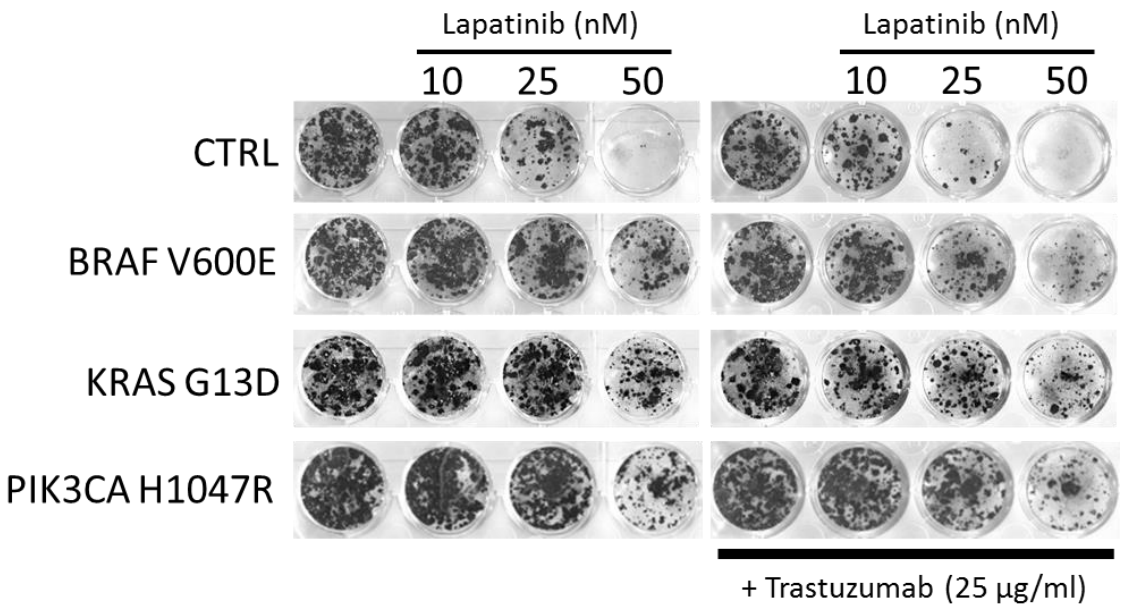
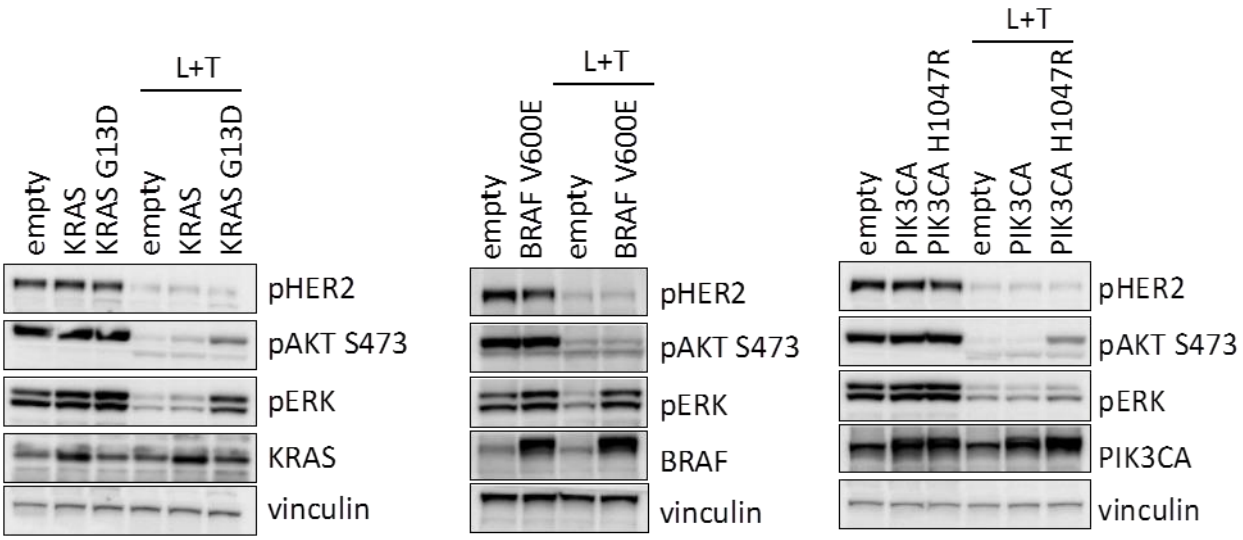


Figure 6

A


Lapatinib (nM)
10
25
50
Lapatinib (nM)
10
25
50
CTRL
BRAF V600E
KRAS G13D
PIK3CA H1047R
+ Trastuzumab (25 µg/ml)

B


L+T
empty
KRAS
KRAS G13D
empty
KRAS
KRAS G13D
pHER2
pAKT S473
pERK
KRAS
vinculin
L+T
empty
BRAF V600E
empty
BRAF V600E
pHER2
pAKT S473
pERK
BRAF
vinculin
L+T
empty
PIK3CA
PIK3CA H1047R
empty
PIK3CA
PIK3CA H1047R
pHER2
pAKT S473
pERK
PIK3CA
vinculin