The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets

Title: The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets

Authors: E. Medico, M. Russo, G. Picco, C. Cancelliere, E. Valtorta, G. Corti, M. Buscarino, C. Isella, S. Lamba, B. Martinoglio, S. Veronese, S. Siena, A. Sartore-Bianchi, M. Beccuti, M. Mottolese, M. Linnebacher, F. Cordero, F. Di Nicolantonio and A. Bardelli

Year: 2015

Journal: Nature communications

Volume: 6

DOI: 10.1038/ncomms8002

The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets

Enzo Medico[1,2,*,§], Mariangela Russo[1,2,3,*], Gabriele Picco[1,2,*], Carlotta Cancelliere[2,*], Emanuele Valtorta[4], Giorgio Corti[2], Michela Buscarino[2], Claudio Isella[1], Simona Lamba[1,2], Barbara Martinoglio[2], Silvio Veronese[4], Salvatore Siena[4], Andrea Sarote-Bianchi[4], Marco Beccuti[5], Marcella Mottolese[6], Michael Linnebacher[7], Francesca Cordero[5], Federica Di Nicolantonio[1,2, §], Alberto Bardelli[1,2, §]

[1] University of Torino, Department of Oncology, SP 142, Km 3.95, 10060 Candiolo, Torino, Italy; [2] Candiolo Cancer Institute – FPO, IRCCS, Candiolo, Torino, Italy; [3] FIRC Institute of Molecular Oncology (IFOM), Milano, Italy; [4] Department of Hematology and Oncology, Niguarda Cancer Center, Ospedale Niguarda Ca' Granda, Milan, Italy; [5] University of Torino, Department of Computer Science, Torino, Italy; [6] S.C. Anatomia Patologica, Istituto Nazionale Tumori Regina Elena –Rome, Italy; [7] Division of Molecular Oncology and Immunotherapy, University of Rostock, D-18057 Rostock, Germany

***** co-first authors

§ co-senior authors

Correspondence to: Federica Di Nicolantonio (federica.dinicolantonio@unito.it) and Alberto Bardelli (alberto.bardelli@unito.it)

Short title: A colorectal cancer cell encyclopedia

Abstract

The development of molecularly targeted anticancer agents relies on large panels of tumor-specific preclinical models closely recapitulating the molecular heterogeneity observed in patients. Here we describe the mutational and gene expression analyses of 151 colorectal cancer (CRC) cell lines. We find that the whole spectrum of CRC molecular and transcriptional subtypes, previously defined in patients, is represented in this cell line compendium. Transcriptional outlier analysis identifies RAS/BRAF wild type cells, resistant to EGFR blockade, functionally and pharmacologically addicted to kinase genes including ALK, FGFR2, NTRK1/2 and RET. The same genes are present as expression outliers in CRC patient samples. Genomic rearrangements (translocations) involving the ALK and NTRK1 genes are associated with overexpression of the corresponding proteins in CRC specimens. The approach described here can be used to pinpoint CRCs with exquisite dependencies to individual kinases for which clinically approved drugs are already available.

Introduction

Knowledge of tumor biology and the development of new anticancer agents depend on robust preclinical model systems that reflect the genomic heterogeneity of human cancers and for which detailed genetic and pharmacologic annotations are available. Cell lines represent a mainstay to functionalize molecular data as they allow experimental manipulation, global and detailed mechanistic studies, and high-throughput applications[1, 2, 3, 4]. Virtually all commonly used cancer cells were continuously grown *in vitro* for years, and decades have often passed since they were originally derived from patients. The *in*

vitro culture conditions used to propagate cancer cells are far from the histological landscape in which they originated. These considerations are often cited to question the relevance of cell lines as cancer models. We further observe that, for a given cancer type, only few cell lines (<10) are employed in most preclinical studies. The evidence that cancer patients have heterogeneous genetic features implies that a large number of lineage-specific cell lines is needed to capture the diversity observed in the clinic. Furthermore, cells commonly used to model a tumor type may not represent the patients that they intend to recapitulate. For instance, colorectal cancer (CRC) cell lines ordinarily used in preclinical studies often display microsatellite instability (MSI). MSI colorectal tumors are found in 10-15% patients [5], but they show an indolent clinical behavior and are less prevalent in more advanced stages of the disease, accounting for less than 5% in the metastatic setting [6, 7]. Consequently, MSI cell lines do not properly recapitulate the clinical setting in which targeted agents are most commonly administered to CRC patients.

In addition to 'genetic' driven subtypes, tumor lineages can be subdivided based on their transcriptional profiles. Transcriptional profiling has been recently used to identify distinct CRC subtypes [8, 9, 10, 11, 12]. The subtypes are associated with biological and clinical features such as cell of origin, microsatellite instability, prognosis and response to treatments. Whether the transcriptional subtypes are maintained in CRC cell models and how this knowledge can be used to discover novel pharmacogenetic relationships has not been explored. To address these challenges we assembled and annotated a comprehensive collection of 151 human CRC cells.

Please add a short paragraph (2-3 sentences) to summarize the main findings of your manuscript. Those must be in the present tense.

We find that the molecular heterogeneity (oncogenic mutations and transcriptional subtypes) previously defined in CRC patients, is maintained in colorectal cancer cells. Individual lines can be stratified as responders or non-responders to EGFR blockade on the basis of clinically validated biomarkers. Transcriptional outlier analysis identifies CRC cells, resistant to EGFR blockade, pharmacologically addicted to kinase genes including *ALK, FGFR2, NTRK1/2* and *RET*.

Results

Genetic and transcriptional profiling of CRC cells

A collection of 152 CRC cell lines was initially genotyped at ten different microsatellite loci (STR profile) to unequivocally define their genetic identities (Supplementary Data 1). This analysis revealed that a few cell lines, previously thought to be unrelated, were derived from the same individual (Supplementary Data 1). Global mRNA expression profiling was carried out on the entire cell bank and TiGER (tissue-specific gene expression database)[13] was exploited to perform a "tissue of origin" analysis (Supplementary Fig. 1a). The analysis revealed a probable non-intestinal origin for one line, COLO741, found to express skin-specific genes, including Tyrosinase (Supplementary Fig. 1b,c). This, together with the lack of expression of the neuroendocrine markers CHGA, SYP, ENO2, ACPP and NCAM1, suggested that these cells are not derived from a CRC nor from a neuroendocrine tumor, but rather from a melanoma. For this reason, COLO741 cells were withdrawn from further analyses. HuTu80 cells displayed a small intestine-like signature (Supplementary Fig. 1d) and were included in subsequent analyses due to their intestinal origin. Gene expression-based hierarchical clustering revealed that cell lines with identical genetic background consistently clustered together (Fig. 1), indicating that the genotype maintained strong control over the transcriptome. The only exception was the COGA5/COGA5L pair; these lines were derived respectively from the primary tumor and a lymph node metastasis of the same patient [14].

The 151 CRC cell line collection was then assessed for the occurrence of microsatellite instability (MSI). MSI was detected in 63/151 cases (42%) (Supplementary Data 1). It has been previously noted that a higher than expected fraction of CRC cell lines are MSI,

conceivably because MSI tumors can be more easily propagated *in vitro* [15]. We also observed that the rare hypermutator/MSI-negative subgroup, identified in about 2% of CRC cases, is represented by three lines, HT115, HCC2998 and HT55 [16].

The mutational status of *KRAS, NRAS, BRAF,* and *PIK3CA* was next assessed as these genes are implicated in the response of CRC to EGFR targeted therapies and are often ascertained in the clinic (Supplementary Data 1). *KRAS, NRAS* and *BRAF* mutations were mutually exclusive and were present in 47%, 0.7% and 18% of the lines, respectively. *PIK3CA* mutations and PTEN status were detected in 18% and 14% of the cells, respectively, and occurred together with *KRAS* and *BRAF* at rates comparable to those previously reported[17, 18, 19] (Supplementary Data 1). Overall, the complexity reached in this compendium effectively reflects the genetic heterogeneity recognized in sporadic CRCs.

Sensitivity to cetuximab is an intrinsic trait of CRC cells

The anti-EGFR monoclonal antibodies cetuximab or panitumumab are approved to treat metastatic CRC (mCRC) but achieve less than 10% objective response rates in unselected CRC patients when given in monotherapy [20]. Clinical benefit to EGFR blockade is confined to 25% of cases carrying tumors wild-type for *KRAS*, *NRAS* and *BRAF* [21]. To assess whether CRC lines recapitulated the above pharmacogenetic relationships, we determined sensitivity to cetuximab in the entire cell collection over a wide range of drug concentrations (Fig. 2, Supplementary Fig. 2 and Supplementary Data 1). Ten *RAS/BRAF* wild-type cell lines (7% of the entire collection or 20% of *RAS/BRAF* wild-type cells) were highly susceptible to cetuximab inhibition (Fig. 2, Supplementary Fig. 2a). Pharmacological parameters such as the area under the curve (AUC) or the percentage of cell growth inhibition at a clinically relevant drug concentration of 10 μg/ml [22] closely paralleled the

response rates observed in patients [20, 21](Supplementary Fig. 3a). Cells classified as sensitive died upon anti-EGFR treatment (Supplementary Fig. 3 b,c,d), while partially sensitive cells were mainly growth impaired (Supplementary Fig. 3 e). As observed in the clinical setting, *KRAS, BRAF* or *NRAS* mutations conferred complete resistance to EGFR blockade in CRC cells (Fisher exact text p value<0.001) (Fig. 2).

CRC molecular subtypes are maintained in cell lines

Colorectal tumors can be classified in up to six unique subtypes with diverse clinical features according to the genes they express [8, 9, 10, 11, 12]. To what extent the transcriptional subtypes are cell-intrinsic or depend upon the microenvironment is unclear. Expression profiles were therefore used to assign each cell line to a molecular subtype [8, 9, 10, 11, 12] by the Nearest Template Prediction (NTP) algorithm, which also estimates the classification false discovery rate (FDR) [23]. Each of the five classifiers was able to assign (FDR < 0.2) the large majority of the cell lines to a subtype. Classifications of cell lines according to the five systems, together with the respective FDRs, are reported in Supplementary Data 2. Significant overlaps were detected among distinct classifiers. We found that the CRC-assigner (CRCA) classification system established by Sadanandam and colleagues[12] shared the highest overlap with all other classifiers (Supplementary Fig. 4a). It was therefore considered as a reference for reconciling the five classifiers into a unified consensus (Supplementary Fig. 4b).

Molecular and pharmacological associations, which previously emerged in CRC samples using the CRCA-based classification, were recapitulated in the cells collection (Fig. 3). MSI cells were significantly enriched in inflammatory and goblet subtypes (p-values: 0.0334 for inflammatory, 0.008 for goblet) and less prevalent in the Transit Amplifying (TA) and Stem

groups, which were instead mainly composed of microsatellite stable (MSS) cell lines (p-value<0.0005). *BRAF* mutated cell lines clustered in the inflammatory subtype (p-value = 0.0005), while *RAS* mutations were equally distributed among subgroups. Notably, 6 out of 9 *RAS/BRAF* wild type cetuximab sensitive cells belonged to the TA subtype (p=0.038). Such enrichment is in line with data on CRC specimens [12] (Fig. 3).

Outlier tyrosine kinases overexpressed in CRC cells To unveil actionable targets likely to be driver oncogenes in CRC, we carried out an outlier expression analysis focused on protein kinases [24], as the latter are frequently implicated in cancer and are ideally suited for therapeutic inhibition. Of 448 kinase genes expressed in the dataset [25], 32 (7.1%) displayed an outlier expression profile. A striking increase in the prevalence of outliers was observed in the tyrosine kinase (TK) subfamily (15 out of 74, 2.84x enrichment, hypergeometric p-value<0.0001). Of these 15 TKs, 9 were outliers only in *RAS/BRAF* wild-type cells, and were therefore named «WT-specific». The enrichment for WT-specific outliers in the TK subgroup was significant (9 out of 15, 3.21x enrichment, hypergeometric p-value<0.0005). Interestingly, 8 out of the 9 WT-specific outlier TKs clustered in cell lines resistant to cetuximab. The list of WT-specific outliers' genes comprises *ALK, FER, FGFR2, KIT, NTRK1, NTRK2, PDGFRA* and *RET* (Fig. 4a).

Outlier kinase genes are therapeutic targets in CRC cells

We next performed genetic and functional validation of the TK outliers. We decided to focus on TKs for which targeted agents are in clinical trial or are already approved for treatment, namely ALK, FGFR2, KIT, NTRK1, NTRK2, RET and PDGFRA. This analysis revealed that overexpression of NTRK1 and FGFR2 is associated to molecular alterations, such as gene translocation (*NTRK1*) or gene amplification (*FGFR2*) (Supplementary Figs

5,6), previously described in cellular models and in cancer patients [26, 27, 28], thus validating the experimental approach. The other tyrosine kinase outlier genes were not previously reported in CRC cells. We detected an in-frame gene fusion event between exon 20 of ALK and exon 13 of EML4 in C10 cells (Fig.4b). Break-apart FISH, cDNA PCR and western blot analysis confirmed the EML4-ALK translocation in C10 cells (Fig.4c-e). We reasoned that if the outlier kinase genes were causally implicated as oncogenic drivers, they should be functionally relevant in the corresponding cell models. We also postulated that the functional dependencies would be cell specific. To formally test these possibilities we used two complementary approaches: reverse genetics and pharmacological inhibition. Candidate-specific gene suppression was achieved with siRNA. In all cases reduced protein expression of the 'outlier' TK resulted in significant impairment of cell growth, which was often accompanied by downstream signaling inhibition and apoptosis (Fig. 4f; Supplementary Figs. 5-8).

Pharmacological validation of the TK outliers was performed using the following kinase-targeted drugs: CEP701 (TRK inhibitor), AZD4547 (FGFR2 inhibitor), crizotinib (ALK inhibitor), ponatinib (RET inhibitor) and imatinib and nilotinib (KIT and PDGFRA inhibitor). The drug inhibition profiles were cell specific and paralleled the expression profiles of individual TKs confirmed by RNAseq in the corresponding cell models (Fig. 4g-h, 5 and Supplementary Fig. 5-8). In two cases (KIT and PDGFRA) drug inhibition did not affect cell growth (Supplementary Fig. 8d,h). Nonetheless, in KIT overexpressing cells a reverse genetic experiment revealed functional dependency (Supplementary Fig. 8b,c).

Translocation of ALK and NTRK1 genes in colorectal cancer samples

To establish the clinical relevance of the cell line-based findings, we assessed whether the outlier TKs might be identified in CRC specimens using the same methodology. To this end we downloaded from cBbioPortal [29, 30] (http://www.cbioportal.org) the RNAseq expression Z-scores for 352 CRC tissue samples generated by the Cancer Genome Atlas (TCGA) network [5]. The entire set of candidates described above, with the exception of KIT, was independently validated as RNA expression outliers in the TCGA CRC dataset (Fig. 6a). As a further validation, we verified whether the 'outlier' strategy could lead to retrieval of TK candidates in archival CRC specimens for which RNA is typically not readily available. In this scenario, the most practical screening method should be capable of interrogating protein (rather than RNA) expression, on formalin-fixed paraffin embedded (FFPE) samples. As a proof of principle, we focused on *NTRK1*/Trk-A and ALK. An immunohistochemistry (IHC) based screen was applied to 742 CRC FFPE samples using a Trk-A (the protein encoded by *NTRK1* gene) specific antibody. One Trk-A outlier sample was unequivocally identified (Fig. 6b,c). The IHC profile was followed by break-apart FISH analysis, which detected *NTRK1* genetic rearrangement in the TrK-A positive sample (Fig. 6d). Screening with and ALK specific antibody of 742 CRC FFPE retrieved a sample with high levels of the ALK protein (Fig.6e,f). In the same specimen, break-apart FISH highlighted a genetic translocation involving the EML4 gene (Fig. 6g).

Discussion

Human neoplasms are highly heterogeneous, and clinical evidence indicates that pharmacogenetic relationships often involve oncogenic events occurring at low prevalence. Accordingly, precision oncology depends on the ability to functionally interrogate preclinical models capturing the molecular heterogeneity observed in patients. While collections of cancer cells derived from multiple tumor types have been previously

described [1, 2, 4], comprehensive databases of cells derived from single tumor lineages are much less common [16, 31]. To capture the clinical heterogeneity of CRCs, we assembled a collection of 151 cell lines. Notably, genetic finger printing and transcriptional profiling highlighted that several CRC lines previously thought to be unrelated are in fact derived from the same individual. This finding should be helpful in designing future studies and may facilitate the interpretation of previous analysis involving 'redundant' cell models.

The clinical relevance of the cell database was confirmed by showing that common oncogenic events (*KRAS, NRAS, BRAF*, *PIK3CA* and *PTEN* alterations) are present in the collection at rates similar to those found in CRC patients. Previous studies evaluated sensitivity to EGFR blockade (with the anti EGFR antibody cetuximab) in a fraction of the CRC lines described in this work [32]. We found a strong concordance among measurements obtained in independent laboratories, indicating that sensitivity to EGFR inhibition is a stable phenotype of colorectal tumors. While the molecular determinants of EGFR sensitivity are presently unknown, our data suggest a prominent intrinsic (cell autonomous) component in colorectal tumors. We conclude that genetic and pharmacological profiling CRC lines can successfully nominate clinically relevant molecular determinants of response to targeted therapies.

In addition to the mutation-based categories described above, CRC can be subdivided in up to 6 transcriptional subtypes with distinct molecular and clinical features [8, 9, 10, 11, 12], which we found to be robustly maintained in cells grown *in vitro*. This observation, along with the evidence that independent cell lines derived from the same patient consistently co-cluster transcriptionally, suggests that the transcriptional profile is stable and predominantly controlled by the genome in the absence of environmental inputs. These results further support the functional correspondence of CRC cell panel described here

with CRC specimens. Previous studies have also suggested that the transcriptional CRC subtypes may be associated with drug response. Cell models recapitulated the association between response to EGFR inhibitors and enrichment for the TA transcriptional class, which was identified in CRC clinical samples [12]. Along this line, the classification of cells reported in our work will facilitate future screenings with chemical or RNA interference libraries aimed at identifying novel pharmacogenomic relationships in specific transcriptional subtypes.

We previously reported that a subset of colorectal tumors, in which resistance to EGFR blockade occurs in a *RAS/RAF* wild type background, display overexpression of two tyrosine kinase genes, HER2 or MET [33, 34]. In 41/151 lines, which were analyzed in this study, the mechanism of resistance to cetuximab is unaccounted for. We hypothesized that in some of these cells resistance might likewise be driven by overexpression of a kinase gene. We therefore employed an outlier analysis based on expression levels of TK genes to identify additional kinase (and therefore actionable) targets in the subset of WT *RAS/BRAF* CRC cells intrinsically resistant to EGFR therapies.

Seven of the TKs retrieved by this approach (ALK, FGFR2, KIT, NTRK1, NTRK2, RET and PDGFRA) are the target of drugs undergoing clinical testing or approved. In some instances (FGFR2, ALK or NTRK1), overexpression of the TK gene was associated with amplification or translocation of the corresponding locus. The same genetic alterations have been previously reported in CRC cells and patients [26, 27, 28], thus validating our approach. Our data further suggest that overexpression of TK sustains primary resistance to EGFR blockade, and could be used to identify patients unlikely to respond to cetuximab or panitumumab. Of note, with the exception of RET and FGFR2, which coexisted in NCIH716 cells, outlier expression of TKs was mutually exclusive.

Our results have broad implications for the use of kinase inhibitors in CRCs and potentially other cancer types. First, companion diagnostic assays are essential to identify those (rare) patients with tumors bearing aberrantly expressed kinases. Since the molecular events leading to protein overexpression can be multiple (gene amplification, gene fusions and translocations with different partners and breakpoints), we propose that IHC should be better suited than other methods to serve this purpose. As a test case, we demonstrated the feasibility of this approach to identify CRC patients carrying overexpressed and genetically altered Trk-A and ALK kinases. Additional work is warranted to develop and validate IHC assays for the other outlier TK genes. Importantly, given the low prevalence of these outlier events, strategies aimed at concomitantly screen for multiple molecular alterations should be implemented. Eventually, it is likely that next-generation sequencing will become the most effective approach to detect molecularly deregulated kinase genes as it allows concomitant assessment of single nucleotide variants as well as a gene copy number alterations and translocations.

Our results also highlight the relevance of clinical trials involving CRC patients carrying molecularly altered tyrosine kinase genes, which can be readily identified using FFPE specimens. Our data suggest that these patients are unlikely to benefit from anti EGFR antibodies but likely to profit from currently available drugs such as ALK and TRK inhibitors. It is important to note that in some instances (i.e. KIT), blockade of the kinase activity with a specific kinase inhibitor did not impair growth, while TK suppression with siRNA revealed functional dependency and triggered apoptosis. This suggests that down regulation of the candidate protein, for example through specific antibodies, might be necessary in some instances. Our data further suggest that tyrosine kinase inhibitors effectively impair tumor growth in cell lines in which transcriptional deregulation is

accompanied by genetic activation of the kinase (TRK, ALK, FGFR2), but not in the cell lines where no genetic activation was identified (PDGFR, KIT).

In conclusion our results describe a powerful preclinical resource capturing the heterogeneity of colorectal cancer patients, which can be used to define multidimensional information of clinical relevance and applicability.

METHODS

Cell lines and reagents

A collection of 152 cell lines of intestinal origin was assembled from different worldwide cell line banks or academic laboratories as indicated in Supplementary Data 1. All cell lines were maintained in their original culturing conditions according with supplier guidelines. Cells were ordinarily supplemented with FBS at different concentrations, 2mM L-glutamine, antibiotics (100U/mL penicillin and 100 mg/mL streptomycin) and grown in a 37°C and 5% CO_2 air incubator. Cells were routinely screened for absence of Mycoplasma contamination using the Venor® GeM Classic kit (Minerva biolabs). The identity of each cell line was checked by Cell ID™ System and by Gene Print® 10 System (Promega), throught 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 analyzed using GeneMapperID software from Life Technologies. Resulting cell line STR profiles were cross-compared and matched with the available STR from ATCC, DSMZ, JCRB, Korean Cell Line Bank, ECCAC, and CellBank

Australia repositories online databases. Results of the analyzed loci for each cell line are provided in Supplementary Data 1.

Cetuximab was obtained from the Pharmacy at Niguarda Ca' Granda Hospital in Milan, Italy; Crizotinib, AZ4547, imatinib, ponatinib and nilotinib were purchased from Selleck Chemicals. CEP701 was purchased from Sigma Aldrich.

DNA analysis

Genomic DNA samples were extracted by Wizard® SV Genomic DNA Purification System (Promega). The MSI status has been evaluated by mean of MSI Analysis System kit (Promega). 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 (3730 DNA Analyzer, ABI capillary electrophoresis system (Applied Biosystems). The results were analyzed using GeneMarker V2.2.0 software. As cell lines do not have a corresponding normal tissue available for comparison, their profile has been analyzed with an MSS cell line of control, K562. Samples with instability in one, two or more markers (mononucleotide repeat) are defined as MS-instable (MSI). Samples with no detectable alterations are MS-stable (MSS). The same samples of genomic DNAs were used for PCR amplification to check 16 hot spot CRC mutations in all cells (reported in Supplementary Data 1). The mutational status of KRAS (exons 2, 3, 4), NRAS (exons 2, 3), HRAS (exons 2 and 3), BRAF (exon 15) and PIK3CA (exons 9 and 20) was determined by Sanger sequencing (primers' sequences are listed in Supplementary Data 3).

RNA extraction and gene expression profiling

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). Synthesis of cDNA and biotinylated cRNA was performed using the IlluminaTotalPrep RNA Amplification Kit (Ambion), according to the manufacturer's protocol using 500ng of total RNA. Quality assessment and quantification of cRNAs were performed with Agilent RNA kits on Bioanalyzer 2100. Hybridization of cRNAs (750 ng) was carried out using Illumina Human 48k gene chips (Human HT-12 V4 BeadChip). Array washing was performed using Illumina High Temp Wash Buffer for 10' at 55°C, followed by staining using streptavidin-Cy3 dyes (Amersham Biosciences). Probe intensity data were obtained using the Illumina Genome Studio software (Genome Studio V2011.1). Raw data were Loess normalized with the Lumi R package and further processed with Excel software..

Tissue of origin analysis

Thirty panels of genes with tissue-specific expression were retrieved from the TIGER portal[13] (http://bioinfo.wilmer.jhu.edu/tiger/) and filtered to remove gene symbol redundancies. Similarly, to avoid probe redundancy in array data, when multiple probes matched the same gene we selected the one with the highest standard deviation. For the tissue of origin analysis, we assumed that the vast majority of the cell lines present in this dataset were of colorectal origin, while only a very small fraction could possibly derive from other tissues. Therefore, to increase sensitivity of the colon-specific gene panel, genes not expressed (L2S<9) in at least 50 of the 152 CRC cells were removed from the panel. To increase specificity of other tissue-specific gene panels, genes from these panels expressed (L2S>9) in more than three CRC cell lines (2%) were also removed. For each gene of a tissue-specific panel, the TIGER database provided an "EST Enrichment" (EE) score, proportional to the extent of tissue specificity. To give more weight to genes with

high EE score in the estimation of the tissue of origin, EE scores were squared and summed. Then, the squared EE score of each gene was divided by the sum, to obtain a "Scaled EE score" representing what fraction of the tissue specificity of the gene panel is brought by each individual gene. Finally, a "tissue score" was calculated for each gene panel by summing the products of gene expression and scaled EE score, and using the resulting value as exponential of 2 to obtain linear values.

Unsupervised expression clustering

After removing the COLO741 melanoma cell line from the dataset, we assessed global transcriptional correlations across the remaining 151 cell lines by unsupervised clustering A detection filter was applied to the 151 cell line dataset by requesting that the reported detection p-value reported in GSE59857 was zero in at least one of the 151 samples. After detection filtering, when multiple probes matched the same gene we selected the one with the highest standard deviation. This led to identification of 19,828 unique genes detected in at least one cell line. Before clustering,, we applied a further filter based on standard deviation (SD) across samples (SD of Log_2 Signal > 0.8). We then employed the Log_2 signal of 3,132 genes passing this filter for hierarchical clustering with maximum linkage based on Pearson correlation, using the GEDAS software [35].

Drug proliferation assay

CRC cell lines were seeded at different densities (2-4 $x10^3$ cells/well) in 100µl 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 (in case of cetuximab) or DMSO-only (for all the other drugs) treated cells were included as controls. Plates were incubated at 37°C in 5% CO_2 for 4 or 5 days, after which cell viability was assessed by measuring ATP content through Cell Titer-Glo® Luminescent

Cell Viability assay (Promega). Luminescence was measured by Perkin Elmer Victor X4. Based on cell viability after 4 days of treatment with clinical relevant concentration of 10μg/ml of cetuximab and on the shape of the drug screening-cell response growth curves, CRC cells were defined sensitive when cell viability was ≤40% at 10μg/ml of cetuximab. We then calculated the area under the curve index (AUC) for each cell line of the panel and selected the threshold of AUC index ≥13000 based on the values of AUC obtained in the subgroup of sensitive cell line. The area under the concentration–inhibition curve (Index AUC) was computed using Excel [36]. The "cetuximab effect arbitrary index" was calculated by subtracting the threshold value of 13000 from the AUC index of each cell line to highlight the subpopulation of sensitive cell lines (those with a negative index).

Classification of CRC cell lines into intrinsic subtypes

To classify cell lines into CRC subtypes according to previously published transcriptional classification systems, gene lists for the five classifiers were obtained from the supplementary information of the relevant publications. We then employed the NTP algorithm [23] on the 151 CRC lines expression dataset of 19,828 genes described above. NTP is a nearest neighbor classifier developed to classify samples with defined gene signatures also when generated in a different microarray platform, and to provide an estimate of classification robustness by computing false discovery rate and p-value. NTP analysis was conducted using the GenePattern Bioportal[37] (http://www.broadinstitute.org/cancer/software/genepattern/).

For NTP implementation, we selected genes positively and specifically associated to one subtype. The CRC subtype classifiers were generated by different procedures. To perform reciprocal comparisons in an independent dataset, based on a different microarray platform, we had to adapt them to a common classification strategy. The thresholds for gene selection described below were chosen to remove genes with low class

discrimination power (to improve sensitivity and specificity) while at the same time maintaining a reasonable number of classifying genes (for the stability and platform independence of the classifier).

De Sousa E Melo [9] and Sadanandam [12] gene signatures have originally been defined using the same approach, i.e. Significance Analysis of Microarrays [38]followed by Prediction analysis for microarrays (PAM) [39], to build lists of genes positively or negatively associated to each molecular subtype. The extent and sign of association is represented by the PAM score. To select only genes with a strong positive association to one class, we associated each gene to the class corresponding to the maximum PAM value, and only if the difference between the maximum and second highest value was sufficiently high (greater than 0.1).

Subtype signatures of the Marisa[10] classifier were defined by filtering the list of 1108 genes provided in Table S10 of their work, using the associated log fold changes and adjusted p-values for each gene in each of the six CRC subtypes. In particular, we assigned a gene to a class signature if two conditions were met: (i) its fold change in the class was the highest among its fold changes in all the classes; (ii) the adjusted p-value in the same class was the minimum between the adjusted p-values of all the other classes with a positive fold change. As done for De Sousa E Melo and Sadanandam classifiers, in order to select only genes with a strong association to a single subtype, we selected only those genes having a difference between the highest and second highest log fold change greater than 0.2. The Roepman [11] signatures were defined on the basis of positive genes provided in Table 2 of the original work. The Budinska [8] gene signatures were provided on Table 1 of the original work as metagenes, each with statistical assessment of differential expression across subtypes. Adaptation to NTP required two steps: (i) selection of the metagenes specifically upregulated in each subtype, and (ii) extraction of the lists of genes belonging to the selected metagenes. A metagene was associated with a subtype if it was

significantly upregulated in that subtype, had the highest positive fold change (Log2 ratio vs median) in that subtype and if the difference against the second highest subtype was greater than a log2 ratio of 0.2. Then, for each subtype, the lists of genes belonging to the specifically upregulated metagenes were merged to build the subtype signatures. To compare the results obtained by the five different classifiers employed, we computed hypergeometric tests performing a paired-sample enrichment analysis, with Bonferroni correction of the resulting p-values. Associations between CRC subtypes and MSI status, *BRAF/KRAS* mutations and cetuximab sensitivity were evaluated by hypergeometric distribution analysis and Fisher's exact test.

Outlier expression analysis

Outlier analysis was performed starting from the 151 CRC lines expression dataset. Expression of a gene in a cell line was considered outlier if it was both >5 standard deviations and >5-fold greater than its median expression across all cells. Kinase-encoding genes were obtained from (http://kinase.com/human/kinome/) [25], after annotation with updated gene symbols and mapped on the detection-filtered expression dataset of 19,828 unique genes described above.

Gene copy number analysis

CRC cell lines were trypsinized, washed with PBS and centrifuged; pellets were lysed and DNA was extracted as described above. 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 [40].

siRNA screening

The siRNA targeting reagents were purchased from Dharmacon, as a SMARTpool of four distinct siRNA species targeting different sequences of the target transcript. Cell lines were grown and transfected with SMARTpool siRNAs using RNAiMAX (Invitrogen) transfection reagents following manufacturer's instructions. Briefly RNAi screening conditions were as follow: on day one siRNA were distributed in each well of a 96-wellplate at final concentration of 20 nmol/L. Transfection reagents were diluted in OptiMEM and aliquoted at 10µl/well; after 20 minutes of incubation, 70µl of cells in media without antibiotics were added to each well. After five days, cell viability was estimated with a luminescent assay measuring cellular ATP levels with CellTiter-Glo® Luminescent Assay (Promega). Each plate included the following controls: mock control (transfection lipid only), AllStars (Qiagen) as negative control; Polo-like Kinase 1 (PLK1) (Dharmacon) as positive control[41]. siRNA sequences are listed in Supplementary Data 4.

RNA seq

Total RNA extracted from C10, KM12, SNU503, HCA24, CACO2, and, NCIH716 cells was processed for RNAseq analysis with Ion TotRNA Seq kit v2® following manifacturer's instruction and sequenced with the Ion Proton™ System (Ambion - Life Technology). Each fastq file was realigned using Tophat2 v2.0.11[42]. Version hg19 of the genome was used and Gencode v19 as reference transcriptome database. The average value of the reads aligned by Tophat2 is 61%. The Fragments Per Kilobase of exon per Million fragments mapped (FPKM) was computed on the aligned reads using Cufflinks software v2.2.1 [43]. Moreover, each fastq file was analyzed by FusionMap software v 7.0.1.25 [44] obtaining the list of fusion transcripts for each cell line.

Immunohistochemistry.

FFPE colorectal cancer samples were available through the clinical database of Ospedale Niguarda and Istituto Nazionale Tumori Regina Elena. Additional CRC specimens tissue microarrays were purchased from Abcam (Cat. ab128127). Trk-A and ALK protein expression was evaluated by immunohistochemistry (IHC) performed on 3-4µm thick tissue sections, using specific antibodies In order to assess Trk-A expression, the monoclonal antibody (Clone ID EP1058Y, Epitomics, dilution 1:200), which recognizes the C-terminal portion of the protein was used and the analysis was performed with the automated system BenchMark Ultra (Ventana Medical System, Inc., Roche) according to the manufacturer‘s instructions, with minimum modifications. Samples were incubated with the primary antibody for 1h at 42°C, after which a chromogenic reaction was performed using the ultraView Universal Alkaline Phosphatase Red Detection kit (Ventana, Roche). ALK protein espression was detected using the anti ALK mouse monoclonal antibody (clone 5A4) which recognizes the intracellular region of ALK. The analysis was performed manually, using *N*-Histofine ALK Detection Kit, according to the manufacturer's nstructions (Nichirei Biosciences Inc). Samples were incubated with the primary antibody for 30 min at room temperature and DAB was used as chromogen. For both proteins, specimens were considered positive when the expression was strong and cytoplasmic and present in at least 10 % of cells. Healthy tissue, i.e, normal colon mucosa, was used as internal negative control. Images were captured with the AxiovisionLe software (Zeiss) using an Axio Zeiss Imager 2 microscope (Zeiss).

Fluorescent in situ hybridisation analysis.

Tissue sections and cyto-embedded cell lines for Fluorecent in situ hybridization (FISH) experiment were prepared according to the manufacturer‘s instructions of Histology FISH Accessory kit (Dako). For both types of samples (tissue and cell lines) the last steps before

hybridization were: dehydration in ethanol series (70%, 90%, 100%), 3 washes (5' each) and air drying. To identify possible NTRK1 gene rearrangements, a dual colour break apart probe has been designed a contiguous of two BAC (Bacterial Artificial Chromosome) (http://genome.ucsc.) probes, RP11-349I17(1q23.1) and RP11-1038N13 (1q23.1), all together spanning an approximately 335 kb region encompassing the NTRK1 gene. A dual colour FISH analysis was performed using respectively: 10 µl mix- probe made up by 1.5µl BAC genomic probe RP11-349I17 (1q23.1) labelled in Spectrum Orange (Empire Genomics), 1,5 µl BAC genomic probe RP11-1038N13 (1q23.1) labelled in SpectrumGreen, and 7 µl LSI-WCP hybridisation buffer (Vysis) for each slide. In order to identify possible *ALK* gene rearrangements, an LSI ALK (2p23) Dual Color, Break Apart (Vysis) probe was used. Probes and target DNA of specimens were co-denatured in HYBRite System (DakoGlostrup) respectively at 75°C for 5 min using NTRK1 probe and at 73° for 5 min for ALK probe, therefore hybridized overnight at 37° C. Slides were washed with post-hybridisation buffer (DakoGlostrup) at 73°C for 5 min and counterstained wit 4,6-diamino-2phenylindole (DAPI II; Vysis, Downers Grove, IL. USA). Fluorescent *in situ* hybridisation signals were evaluated with a Zeiss Axioscope Imager. Z1 (Zeiss) equipped with single and triple band pass filters. Images for documentation were captured with CCD camera and processed using the MetaSystems Isis software. Samples were scored as positive either for *NTRK1* or *ALK* gene's rearrangement when a split signal of the probes was observed, in at least 15% of 100 cells analyzed in 10 different fields. Healthy tissue, i.e, normal colon mucosa, was used as internal negative control.

Western blotting analysis

Prior to biochemical analysis, all cells were grown in their specific media supplemented with 10% FBS. Total cellular proteins were extracted by solubilizing the cells in EB buffer (50 mMHepes pH 7.4, 150 mMNaCl, 1% Triton X-100, 10%

glycerol, 5 mM EDTA, 2 mM EGTA; all reagents were from Sigma-Aldrich, except for Triton X-100 from Fluka) in the presence of 1 mM sodium orthovanadate, 100 mM sodium fluoride and a mixture of protease inhibitors. Extracts were clarified by centrifugation, normalized with the 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-FGFR2 (1:1000) ; anti-ALK; anti-Trk-A (Santa Cruz) (1:500); anti-Trk (Santa Cruz) (1:500); anti-cKIT (1:1000); anti-PDGFRA (1:1000); anti-phospho-p44/42 ERK (Thr202/Tyr204) (1:1000); anti-p44/42 ERK (1:1000); anti-phospho-MEK1/2 (Ser217/221) (1:1000); anti-MEK1/2 (1:1000); anti-phospho AKT (Ser473) (1:1000); anti-AKT (1:1000); anti-PARP (1:1000); anti-actin (Santa Cruz) (1:3000); anti-vinculin (Millipore) (1:5000); anti-PTEN (1:1000). Anti-RET antibody[45] (1:500) was kindly provided by Prof. Massimo Santoro (Federico II, Naples). The full-length western blot membranes are shown in Supplementary Figs 9–13.

References

1. Barretina J*, et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607 (2012).

2. Garnett MJ*, et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570-575 (2012).

3. Park ES*, et al.* Integrative analysis of proteomic signatures, mutations, and drug responsiveness in the NCI 60 cancer cell line set. *Mol Cancer Ther* **9**, 257-267 (2010).

4. Basu A*, et al.* An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* **154**, 1151-1161 (2013).

5. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337 (2012).

6. Smith CG, *et al.* Somatic profiling of the epidermal growth factor receptor pathway in tumors from patients with advanced colorectal cancer treated with chemotherapy ± cetuximab. *Clin Cancer Res* **19**, 4104-4113 (2013).

7. Roth AD, *et al.* Integrated analysis of molecular and clinical prognostic factors in stage II/III colon cancer. *J Natl Cancer Inst* **104**, 1635-1646 (2012).

8. Budinska E, *et al.* Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J Pathol* **231**, 63-76 (2013).

9. De Sousa E. Melo F, *et al.* Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat Med* **19**, 614-618 (2013).

10. Marisa L, *et al.* Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* **10**, e1001453 (2013).

11. Roepman P, *et al.* Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. *Int J Cancer* **134**, 552-562 (2014).

12. Sadanandam A, *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* **19**, 619-625 (2013).

13. Liu X, Yu X, Zack DJ, Zhu H, Qian J. TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics* **9**, 271 (2008).

14. Vécsey-Semjén B, *et al.* Novel colon cancer cell lines leading to better understanding of the diversity of respective primary cancers. *Oncogene* **21**, 4646-4662 (2002).

15. Wilding JL, Bodmer WF. Cancer cell lines for drug discovery and development. *Cancer Res* **74**, 2377-2384 (2014).

16. Mouradov D, *et al.* Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res*, (2014).

17. Barault L, *et al.* Mutations in the RAS-MAPK, PI(3)K (phosphatidylinositol-3-OH kinase) signaling network correlate with poor survival in a population-based series of colon cancers. *Int J Cancer* **122**, 2255-2259 (2008).

18. De Roock W, *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* **11**, 753-762 (2010).

19. Sartore-Bianchi A, *et al.* Multi-determinants analysis of molecular alterations for predicting clinical benefit to EGFR-targeted monoclonal antibodies in colorectal cancer. *PLoS ONE* **4**, e7287 (2009).

20. Cunningham D, *et al.* Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med* **351**, 337-345 (2004).

21. De Roock W, De Vriendt V, Normanno N, Ciardiello F, Tejpar S. KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer. *Lancet Oncol* **12**, 594-603 (2011).

22. Tabernero J*, et al.* Cetuximab administered once every second week to patients with metastatic colorectal cancer: a two-part pharmacokinetic/pharmacodynamic phase I dose-escalation study. *Ann Oncol* **21**, 1537-1545 (2010).

23. Hoshida Y. Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment. *PLoS One* **5**, e15543 (2010).

24. Kothari V*, et al.* Outlier kinase expression by RNA sequencing as targets for precision therapy. *Cancer Discov* **3**, 280-293 (2013).

25. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science* **298**, 1912-1934 (2002).

26. Lin E*, et al.* Exon array profiling detects EML4-ALK fusion in breast, colorectal, and non-small cell lung cancers. *Mol Cancer Res* **7**, 1466-1476 (2009).

27. Ardini E*, et al.* The TPM3-NTRK1 rearrangement is a recurring event in colorectal carcinoma and is associated with tumor sensitivity to TRKA kinase inhibition. *Mol Oncol*, (2014).

28. Mathur A, Ware C, Davis L, Gazdar A, Pan BS, Lutterbach B. FGFR2 Is Amplified in the NCI-H716 Colorectal Cancer Cell Line and Is Required for Growth and Survival. *PLoS One* **9**, e98515 (2014).

29. Gao J*, et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).

30. Cerami E*, et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404 (2012).

31. Beaufort CM*, et al.* Ovarian Cancer Cell Line Panel (OCCP): Clinical Importance of In Vitro Morphological Subtypes. *PLoS One* **9**, e103988 (2014).

32. Ashraf SQ, Nicholls AM, Wilding JL, Ntouroupi TG, Mortensen NJ, Bodmer WF. Direct and immune mediated antibody targeting of ERBB receptors in a colorectal cancer cell-line panel. *Proc Natl Acad Sci U S A* **109**, 21046-21051 (2012).

33. Bertotti A*, 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* **1**, 508-523 (2011).

34. Bardelli A*, et al.* Amplification of the MET receptor drives resistance to anti-EGFR therapies in colorectal cancer. *Cancer Discov* **3**, 658-673 (2013).

35. Fu L, Medico E. FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. *BMC Bioinformatics* **8**, 3 (2007).

36. Whitehouse PA*, et al.* Heterogeneity of chemosensitivity of colorectal adenocarcinoma determined by a modified ex vivo ATP-tumor chemosensitivity assay (ATP-TCA). *Anticancer Drugs* **14**, 369-375 (2003).

37. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet* **38**, 500-501 (2006).

38. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-5121 (2001).

39. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* **99**, 6567-6572 (2002).

40. Roig AI*, et al.* Immortalized epithelial cells derived from human colon biopsies express stem cell markers and differentiate in vitro. *Gastroenterology* **138**, 1012-1021.e1011-1015 (2010).

41. Brough R*, et al.* Functional viability profiles of breast cancer. *Cancer Discov* **1**, 260-273 (2011).

42. Trapnell C*, et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515 (2010).

43. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).

44. Ge H, Liu K, Juan T, Fang F, Newman M, Hoeck W. FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics* **27**, 1922-1928 (2011).

45. De Falco V*, et al.* Ponatinib (AP24534) is a novel potent inhibitor of oncogenic RET mutants associated with thyroid cancer. *J Clin Endocrinol Metab* **98**, E811-819 (2013).

46. Valtorta E*, et al.* KRAS gene amplification in colorectal cancer and impact on response to EGFR-targeted therapy. *Int J Cancer* **133**, 1259-1265 (2013).

Author Contributions

A.B., E.M., F.D.N., M.R. and G.P. designed the study. M.R., G.P., C.C., E.V., M.B. S.L., B.M.S.V., E.M. and F.D.N performed experiments and analyzed the data. G.C., C.I., M.B., M.M. and F.C., contributed to the analysis of data. M.L. established HROC panel of CRC

cells. A.B., F.D.N., E.M., M.R. and G.P. wrote the manuscript. A.B., F.D.N. and E.M. supervised the study.

Acknowledgments

We thank Lazzari L., Lallo A., Misale S., Arena S. and Rossi T. for helping in the cetuximab pharmacological screening; Cantarella D. and Porporato R. for microarray expression profiling. We thank Prof. Massimo Santoro for supplying RET antibody. We are grateful to the researchers named in Table S1 for the kind gift of the cell lines. Supported by the European Community's Seventh Framework Programme under grant agreement no. 602901, MErCuRIC (A.B.); Associazione Italiana per la Ricerca sul Cancro (AIRC) IG grant no. 12812 (A.B.) and no. 12944 (E.M.); AIRC MFAG no. 11349 (F.D.N.); grant "Farmacogenomica"—5 per mille 2009 MIUR—Fondazione Piemontese per la Ricerca sul Cancro—ONLUS (F.D.N.); AIRC 2010 Special Program Molecular Clinical Oncology 5 per mille, project no. 9970 (A.B. and E.M.); FPRC 5 per mille 2010 and 2011 Ministero della Salute (A.B., F.D.N. and E.M.); Ministero dell'Istruzione, dell'Università e della Ricerca, progetto PRIN (A.B.); Progetti di Ateneo-2011, Università di Torino (ORTO11RKTW to A.B.).

Accession codes

Gene expression microarray data have been deposited in the GEO database with accession number GSE59857

Competing financial interests

Authors declare NO competing financial interests

Figure Legends

Figure 1. Expression clustering of 151 CRC cell lines.

Unsupervised hierarchical clustering of 151 CRC cell lines based on their global expression profile. Cell lines derived from the same individual (same STR) are highlighted by the same color. Names of cell lines with identical STR are reported below the dendrogram in the order by which they occur in the cluster, from left to right.

Figure 2. Cetuximab screening on CRC cell lines. The indicated cell lines were treated with increasing concentrations of cetuximab for 4 days and cell viability was assessed by measuring ATP content. Bars represent an arbitrary index of cetuximab effect on each cell line as detailed in the methods. Cell lines sensitive to cetuximab treatment are shown with a negative index. Red bars represent *KRAS* altered lines; yellow bars indicate *NRAS* mutated cells; blue bars indicate genetic alterations affecting codon V600 of *BRAF*; black bars indicate *RAS/BRAF* wild type cells. NCIH630 cells are KRAS amplified [46].

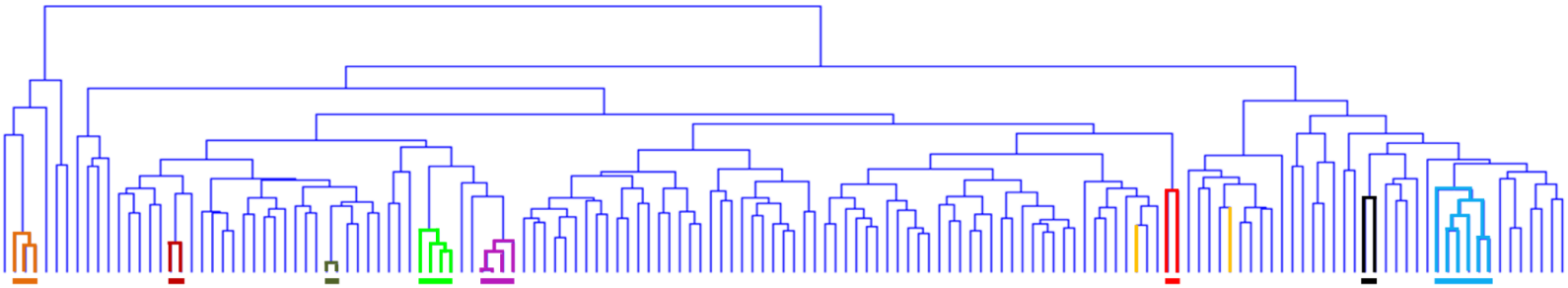
Figure 3. CRC transcriptional subtypes are maintained in cell lines. CRC cell lines are classified according to the "CRC-assigner" intrinsic transcriptional subtypes and the colon crypt location previously described[12]. MSI status, *RAS/BRAF* mutations and cetuximab sensitivity are indicated. FDR: false discovery rate. The Fig. includes only CRC cell lines that were assigned to a subtype with FDR < 0.2. Among cell lines sharing the same STR profile we selected, the original cell line in case of groups of cells including also derived clones, and in the absence of such information, the one classified with the highest confidence by the NTP algorithm..RAS_M: *RAS* mutated cells. $BRAF_M$: *BRAF* mutated cells. Cetux: cetuximab.

Figure 4. Identification of outlier kinase genes in 151 CRC cell lines. (a) Scatter plot representation of transcriptional outlier kinases in 151 CRC cells. Colored circles represent outlier kinases with >5 standard deviations and >5-fold differential expression compared to median expression. Asterisk labels the KM12 family of cells: KM12, KKM12C, KM12SM, KM12L4. **(b)** RNA seq analysis unveiled an in-frame gene fusion event between exon 20 of the ALK gene and the exon13 of the EML4 gene on chromosome 2 in the C10 CRC cell line. **(c)** Break apart FISH analysis of EML4-ALK in C10 cells shows clear separation of green (5') and red (3') signals corresponding to the ALK gene (original 63x magnification ,scale bar=10μm). Scale bars, 10μm **(d)** PCR on the cDNA of C10 cells confirmed the EML4-ALK translocation **(e)** Western blot analysis confirmed overexpression of ALK protein in C10 cells, at a molecular weight compatible with an EML4-ALK genetic rearrangement. **(f)** RNAi knockdown of ALK inhibits cell proliferation of C10 cell line harboring the EML4-ALK fusion gene. C10 cells were analyzed by ATP lite proliferation assay 5 days after transfection with ALK-specific pooled siRNAs, scrambled siRNA, or transfection reagent (mock). Data are expressed as average ± s.d. of three independent experiments. si-RNA mediated downregulation of ALK levels were verified by western blot analysis. **(g)** C10 cells, resistant to the EGFR inhibitor cetuximab, are sensitive to the ALK inhibitor crizotinib. Cell viability was assessed by measuring ATP content after 5 days of treatment. Data are expressed as average ± s.d. of three independent experiments. **(h)** Crizotinib treatment downregulates MAPK and PI3K pathways in C10 cells harboring the ALK genetic rearrangement. C10 cell lines were treated with ALK inhibitor 1μM crizotinib for 12h, after which whole-cell extracts were subjected to western blot analysis.

Figure 5. Overexpressed kinase genes are therapeutic targets in CRC. (a) RNA seq analysis of the indicated cell lines shows overexpression of TK outlier originally identified by microarray data analysis. FPKM: Fragments Per Kilobase of exon per Million fragments

mapped. **(b)** Drug profiling with kinase inhibitors. The indicated cell lines were treated with crizotinib (ALK inhibitor) 240nM, CEP701 (TRK inhibitor) 100nM, AZD4547 (FGFR inhibitor) 100nM, ponatinib[45] (RET inhibitor) 100Nm or nilotinib (KIT and PDGFRA inhibitor) 240nM for 5 days. Viability was then assessed by measuring ATP content.

Fig. 6. NTRK1 and ALK molecular alterations in CRC specimen. (a) Scatter plot representation of transcriptional outlier kinases in 352 CRC tissue samples from the TCGA database. Red circles highlight samples with *Z-scores* values>6. **(b-c)** Immunohistochemistry based identification of FFPE CRC displaying Trk-A protein overexpression. 1/742 samples was markedly positive for Trk-A overexpression. In panel **b** original 4x magnification , scale bar=500μm; in panel **c** original 20x magnification, scale bar=100μm.**(d)** Break-apart FISH analysis revealed *NTRK1* gene rearrangement in the FFPE sample positive by IHC (original 63x magnification, scale bar=10μm). **(e-f)** Immunohistochemistry based identification of FFPE CRC displaying ALK protein overexpression. 1/742 samples was markedly positive for ALK overexpression. In panel **e** original 4x magnification, scale bar=500μm; in panel **f** original 20x magnification, scale bar=100μm) **(g)** Break-apart FISH analysis revealed *ALK* gene rearrangement in the FFPE sample positive by IHC (original 63x magnification, scale bar=10μm).



COLO320DM
COLO320HSR
COLO320

LS174T
LS180

GP5D
GP2D

CX1
WIDR
HT29
KM20

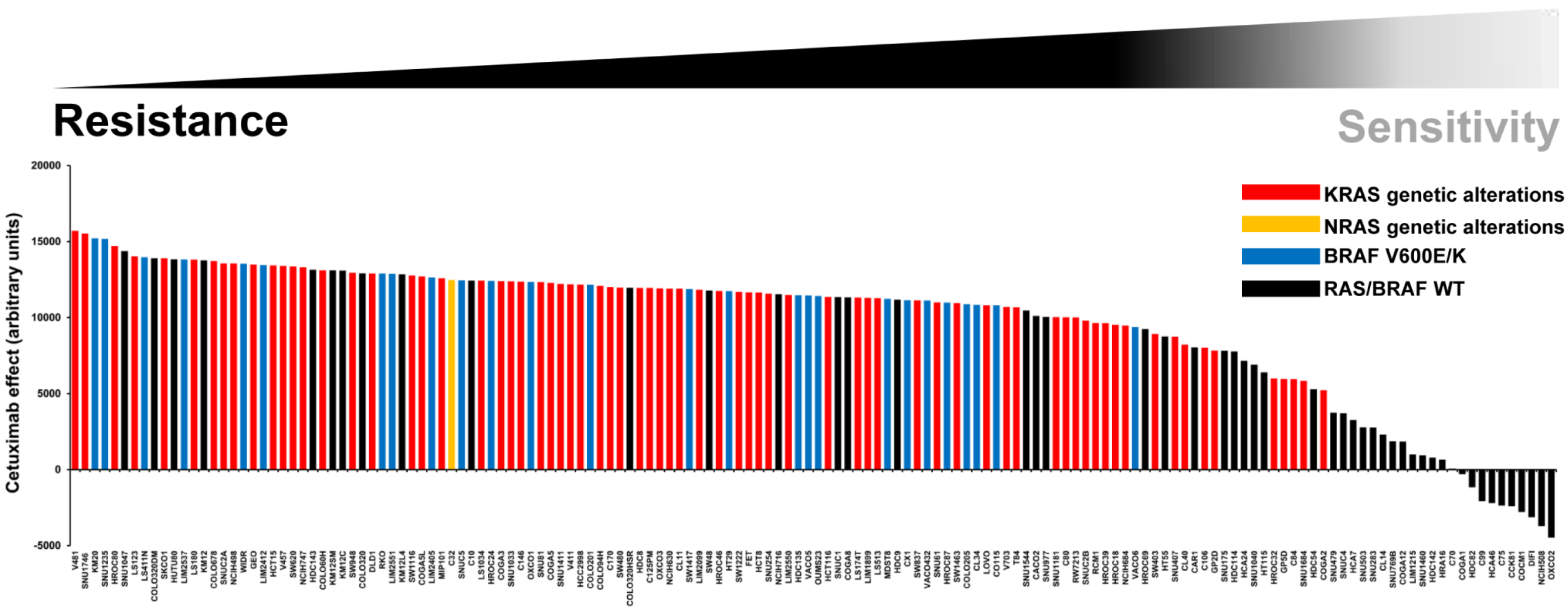
KM12SM
KM12L4
KM12C
KM12

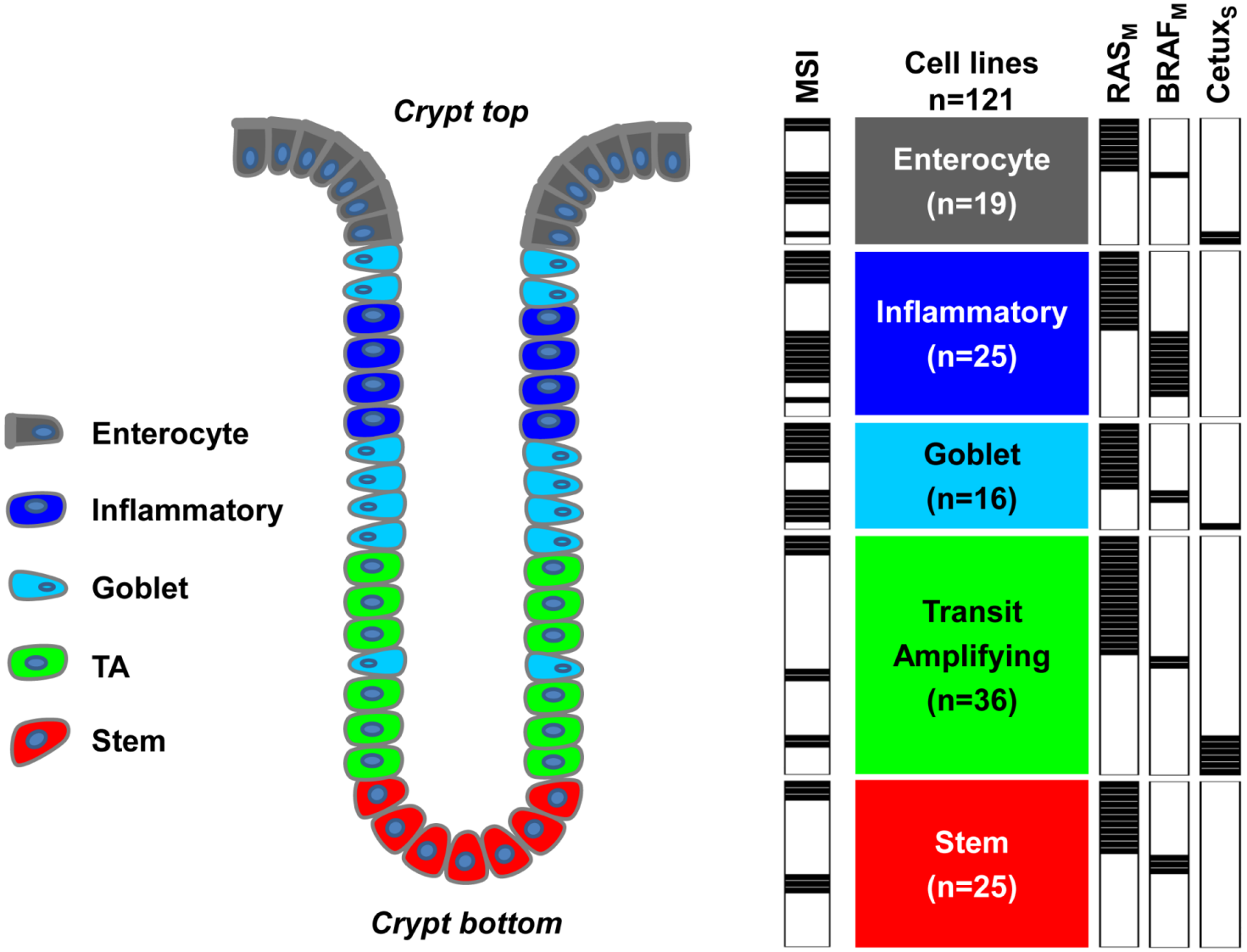
COGA5L
COGA5

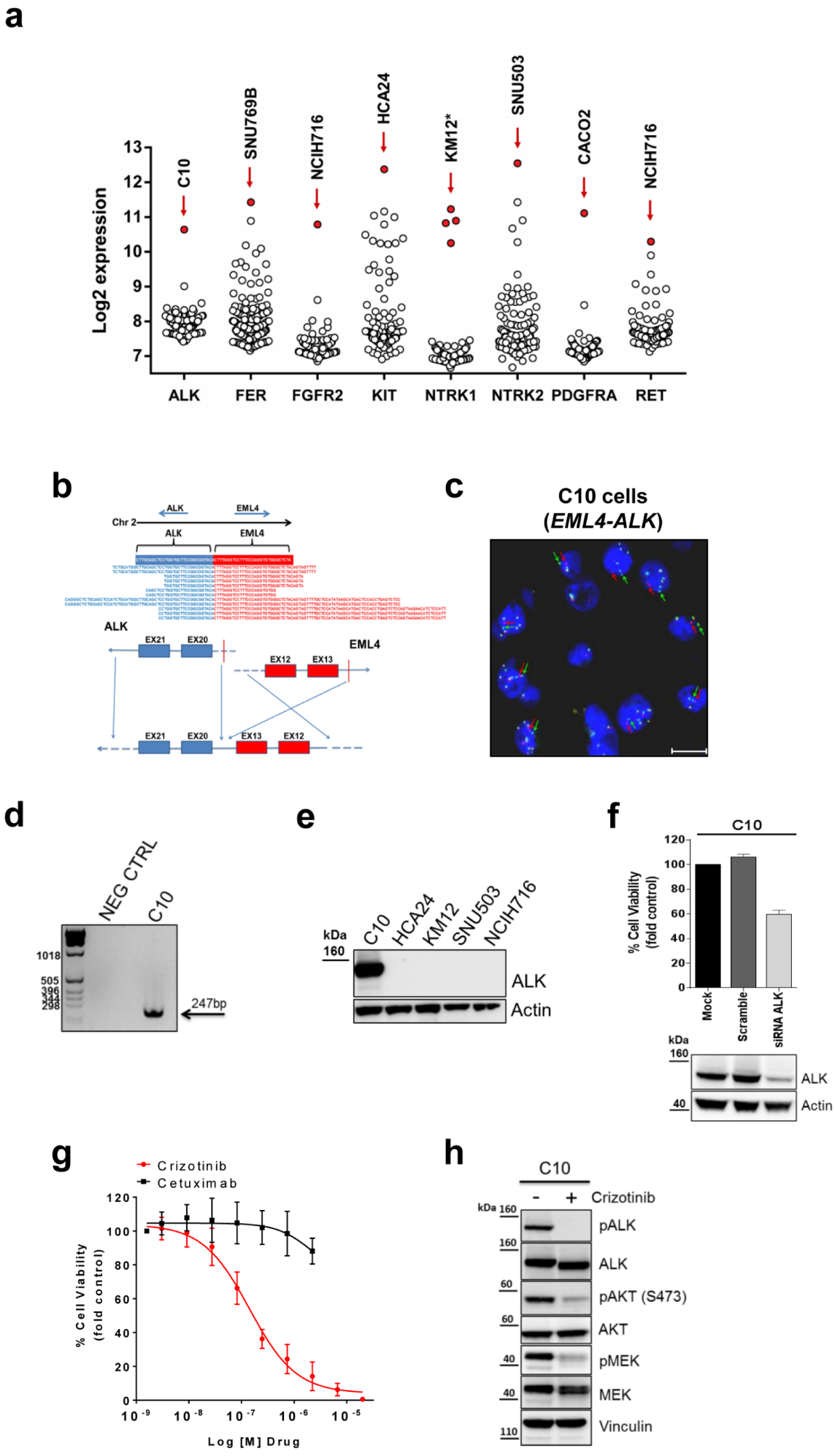
COLO201
COLO205

SW620
SW480

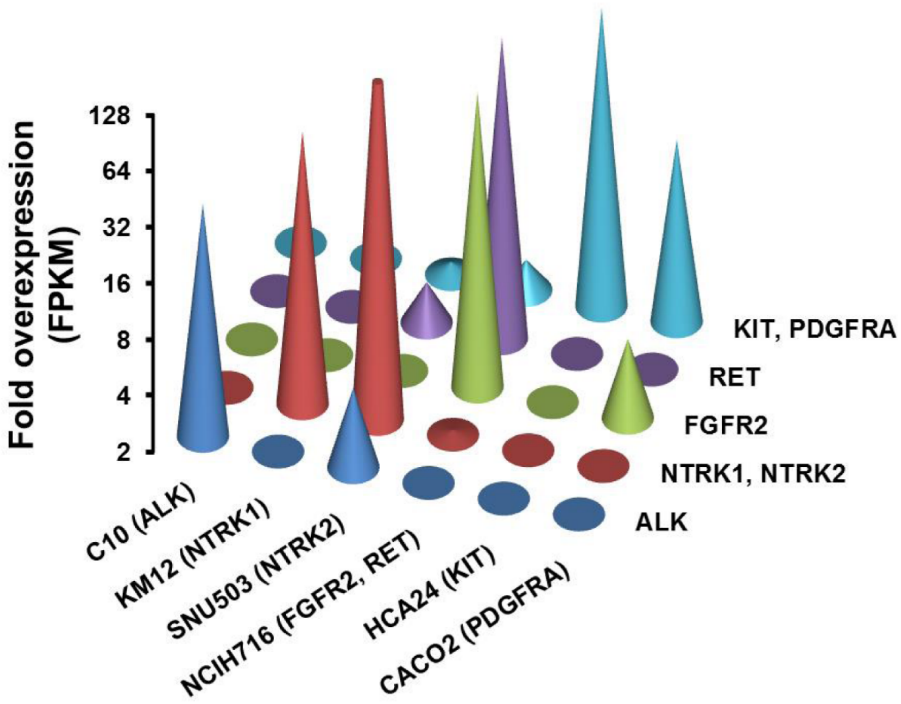
HCT8
HCT15
DLD1
MIP101
C146
C170


Resistance
Sensitivity
Cetuximab effect (arbitrary units)
20000
15000
10000
5000
0
-5000
KRAS genetic alterations
NRAS genetic alterations
BRAF V600E/K
RAS/BRAF WT

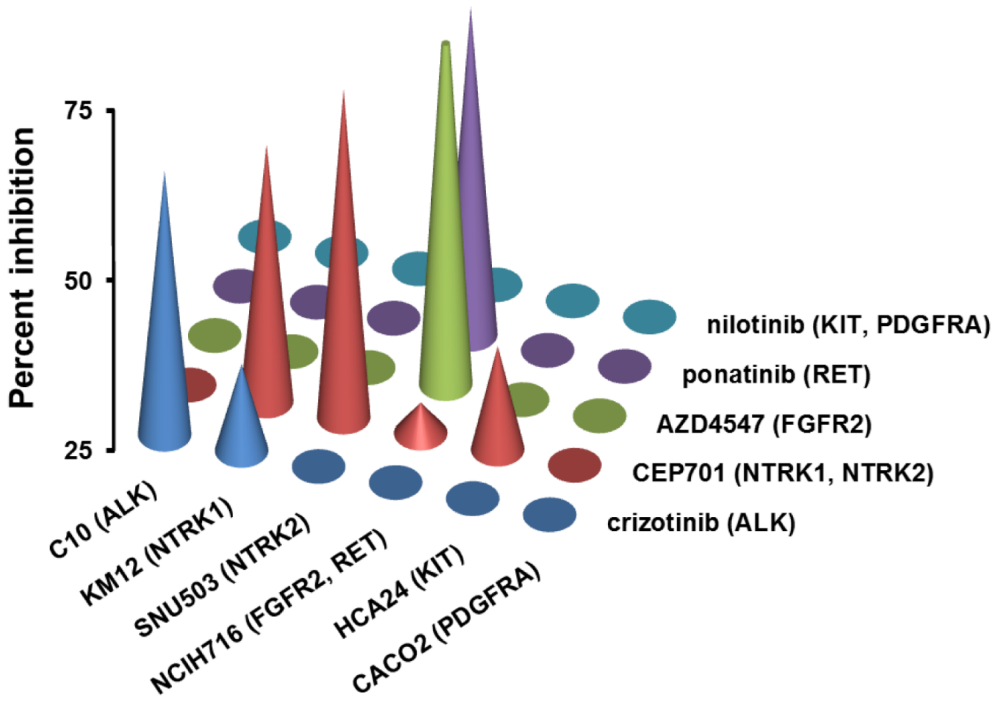

Crypt top
Crypt bottom
Enterocyte
Inflammatory
Goblet
TA
Stem
MSI
Cell lines
n=121
RAS$_M$
BRAF$_M$
Cetux$_S$
Enterocyte
(n=19)
Inflammatory
(n=25)
Goblet
(n=16)
Transit
Amplifying
(n=36)
Stem
(n=25)

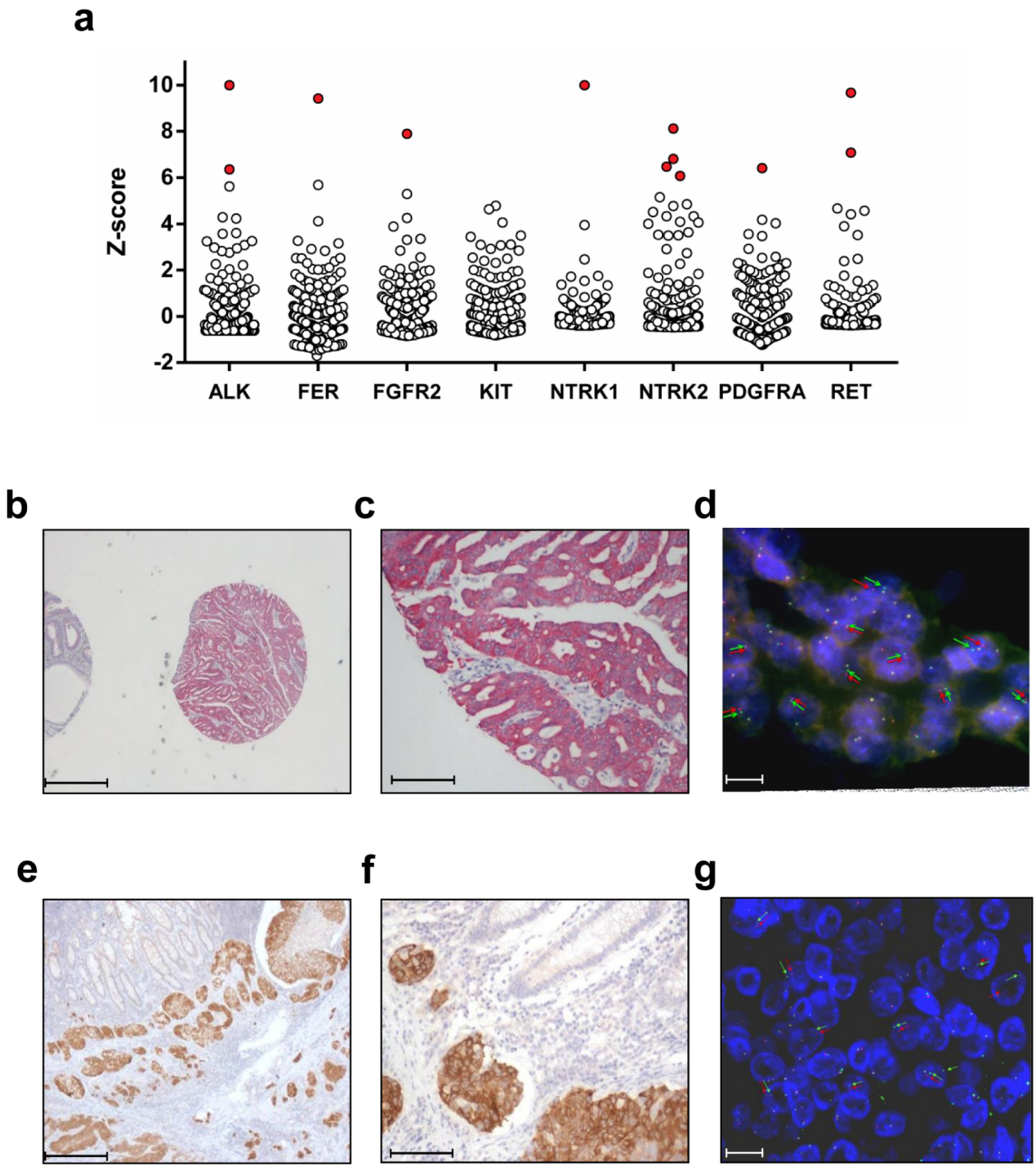

a
Log2 expression
13
12
11
10
9
8
7
C10
SNU769B
NCIH716
HCA24
KM12*
SNU503
CACO2
NCIH716
ALK
FER
FGFR2
KIT
NTRK1
NTRK2
PDGFRA
RET
b
Chr 2
ALK
EML4
EX21
EX20
EX12
EX13
c
C10 cells
(EML4-ALK)
d
NEG CTRL
C10
1018
505
396
344
298
247bp
e
kDa
160
C10
HCA24
KM12
SNU503
NCIH716
ALK
Actin
f
C10
% Cell Viability
(fold control)
120
100
80
60
40
20
0
Mock
Scramble
siRNA ALK
kDa
160
40
ALK
Actin
g
Crizotinib
Cetuximab
% Cell Viability
(fold control)
120
100
80
60
40
20
0
10 -9
10 -8
10 -7
10 -6
10 -5
Log [M] Drug
h
C10
-
+
Crizotinib
kDa
160
160
60
60
40
40
110
pALK
ALK
pAKT (S473)
AKT
pMEK
MEK
Vinculin

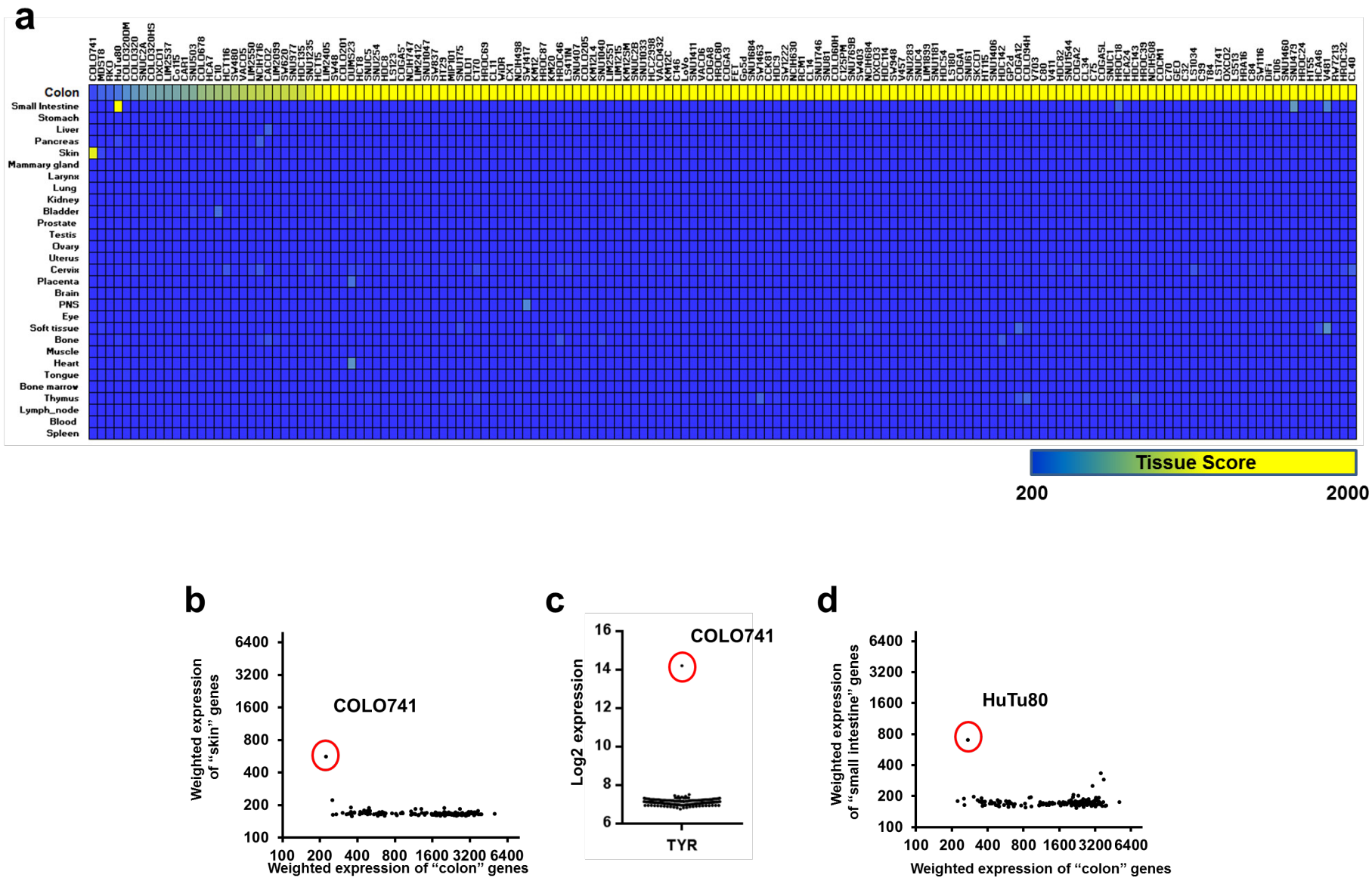
a


Fold overexpression (FPKM)
128
64
32
16
8
4
2
KIT, PDGFRA
RET
FGFR2
NTRK1, NTRK2
ALK
C10 (ALK)
KM12 (NTRK1)
SNU503 (NTRK2)
NCIH716 (FGFR2, RET)
HCA24 (KIT)
CACO2 (PDGFRA)

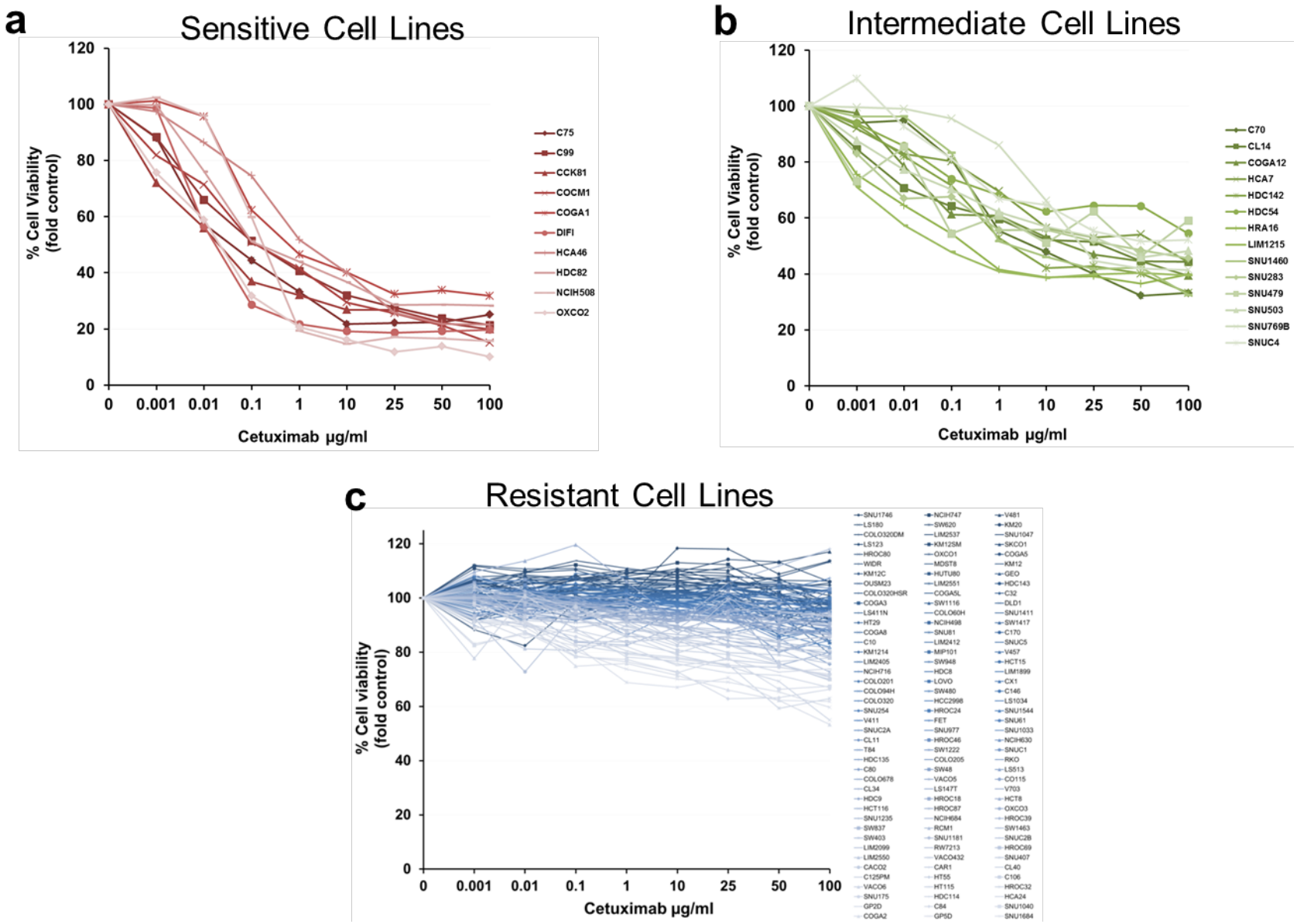
b


Percent inhibition
75
50
25
nilotinib (KIT, PDGFRA)
ponatinib (RET)
AZD4547 (FGFR2)
CEP701 (NTRK1, NTRK2)
crizotinib (ALK)
C10 (ALK)
KM12 (NTRK1)
SNU503 (NTRK2)
NCIH716 (FGFR2, RET)
HCA24 (KIT)
CACO2 (PDGFRA)

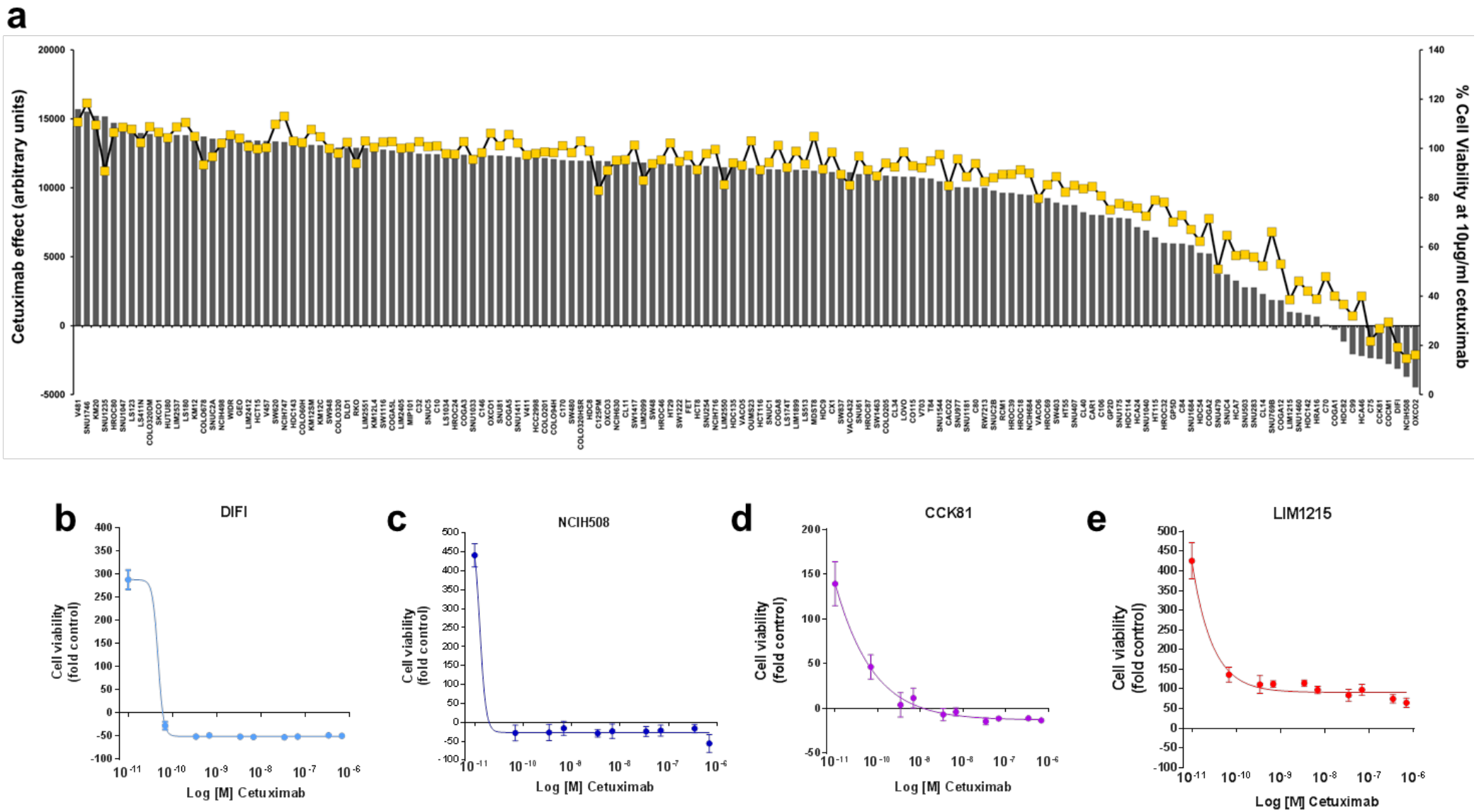

a
Z-score
10
8
6
4
2
0
-2
ALK
FER
FGFR2
KIT
NTRK1
NTRK2
PDGFRA
RET
b
c
d
e
f
g



Supplementary Figure 1: Tissue of Origin analysis on 152 cell lines. (a) Heatmap representation of the 30 "Tissue scores" for the 152 cell lines. The scores summarize the global expression of the tissue specific genes for each one of the 30 tissues analyzed. **(b)** Plot representation of the weighed expression of the colon genes versus skin genes. Circle highlights COLO741 cells showing high "skin" score. **(c)** Log2 expression of the skin-specific tyrosinase (*TYR*) gene in 152 cell lines. Circle highlights COLO741 cells showing an outlier expression profile for the *TYR* gene. **(d)** Plot representation of the weighed expression of the colon genes versus small intestine genes. Circle highlights Hutu80 cells displaying high "small intestine" score.



Supplementary Figure 2. Cetuximab response curves in 150 CRC cell lines. 150 cell lines of colorectal origin were treated with increasing concentrations of cetuximab for 4 days and cell viability was assessed by measuring ATP content. Data are expressed as percent viability compared with medium-only treated cells. Based on their response cells can be classified in sensitive (red lines); partially sensitive (green lines) and resistant (blue lines) to cetuximab treatment.

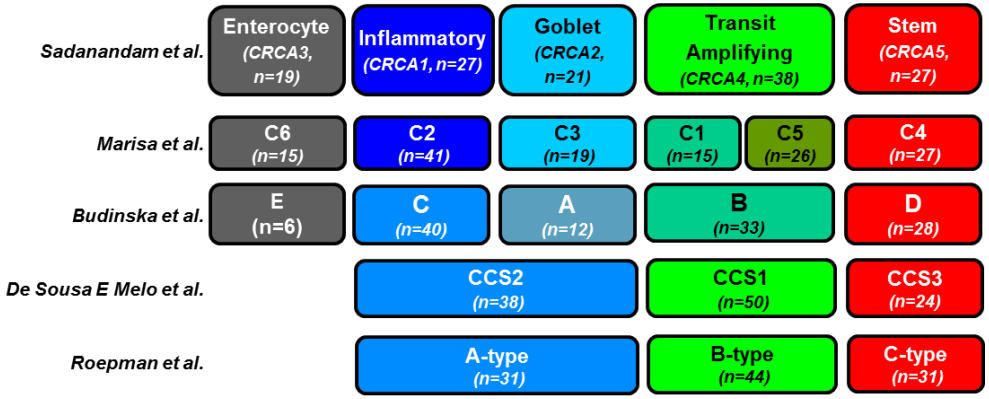


Supplementary Figure 3. Cetuximab effect on CRC cell lines. (a) Cell viability of each CRC cell line at clinical relevant concentration of 10µg/ml of cetuximab after 4 days of treatment (indicated by yellow dots) closely parallel AUC values. **(b-e)** Analysis of cell viability after 4-5 days of cetuximab treatment indicates net loss of cells compared to the number of cells at time "0", in three representative 'sensitive' cell models, even at clinical relevant dose of 10 µg/ml of cetuximab **(b-d)**; while cetuximab exerts mainly a cytostatic effect (reduction of growth rate) on the intermediate cell lines LIM1215 **(e)**. Data are expressed as percent viability compared to the number of cells at day "0".

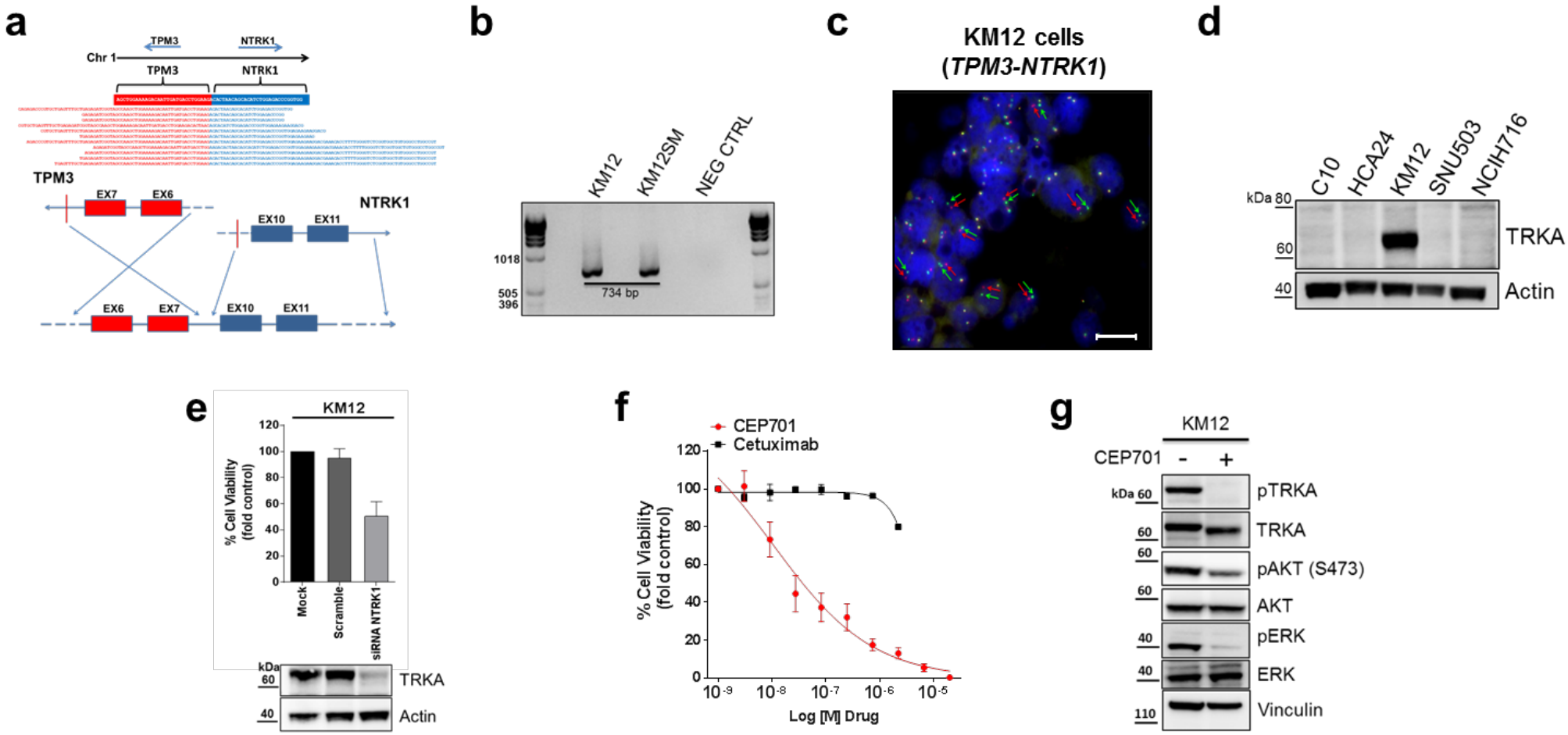
a

		Sadanandam et al.				
		Enterocyte	Inflammatory	Goblet	TA	Stem
Marisa et al.	C6	3,06E-06	21,959	9,001	14,511	3,761
	C2	14,339	1,02E-04	5,338	3,875	6,971
	C3	17,441	1,925	2,16E-08	8,854	1,925
	C1	16,782	18,249	12,326	1,67E-04	18,249
	C5	22,491	19,428	2,050	6,94E-06	3,073
	C4	3,226	14,779	1,727	0,046	9,08E-17
Budinska et al.	Subtype E	7,296	33,981	42,886	32,523	17,982
	Subtype C	18,906	1,21E-09	0,993	2,663	1,643
	Subtype A	4,817	16,797	1,47E-02	24,530	9,817
	Subtype B	16,474	0,499	16,267	8,21E-15	0,089
	Subtype D	18,379	8,072	3,959	0,682	1,72E-17
De Sousa E Melo et al.	CCS2	20,814	6,17E-05	8,92E-03	6,272	7,253
	CCS1	0,135	10,647	21,772	6,48E-14	0,072
	CCS3	11,871	21,929	14,085	0,153	7,93E-18
Roepman et al.	A-Type	4,331	5,78E-04	2,73E-02	2,743	8,924
	B-Type	3,869	10,139	18,358	2,99E-16	0,595
	C-Type	22,691	19,054	29,743	0,294	1,37E-13

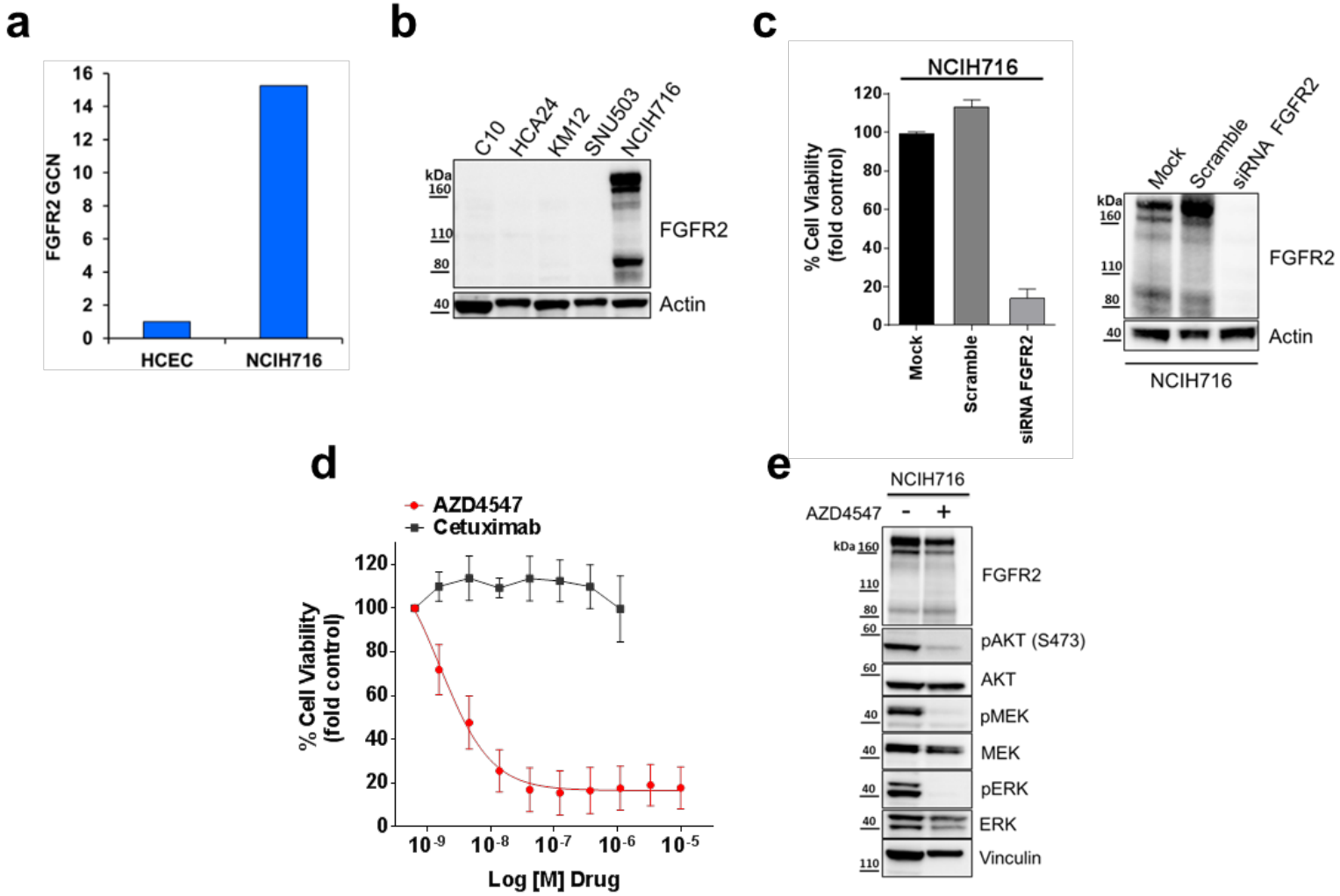
b



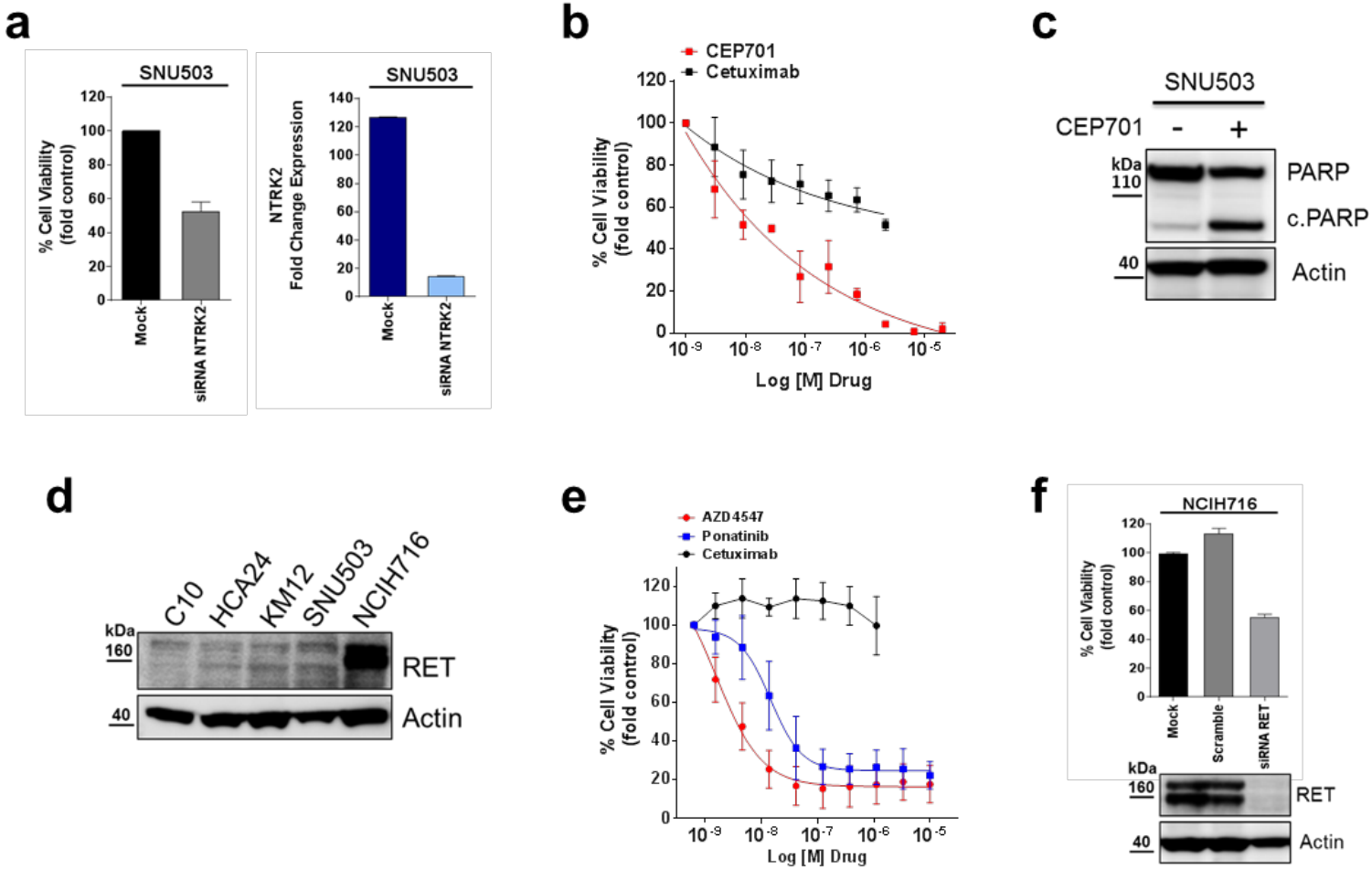
Supplementary Figure 4: Expression based classification of CRC cell lines. (a) CRCassigner (CRCA) is the reference classifier. Enrichment matrix (Bonferroni-corrected hypergeometric p-values) comparing members of CRCA subtype versus others expression-based CRC classifiers. Orange highlighting indicates enrichment p-values< 0.05, green highlighting indicates depletion p-values< 0.05. **(b)** Graphic representation of the consensus across the intrinsic subtypes identified in cell lines by CRCassigner (CRCA) by Sadanandam et al. and other four different classification systems. The number of cell lines significantly assigned to each subtype (FDR<0.2) is reported.



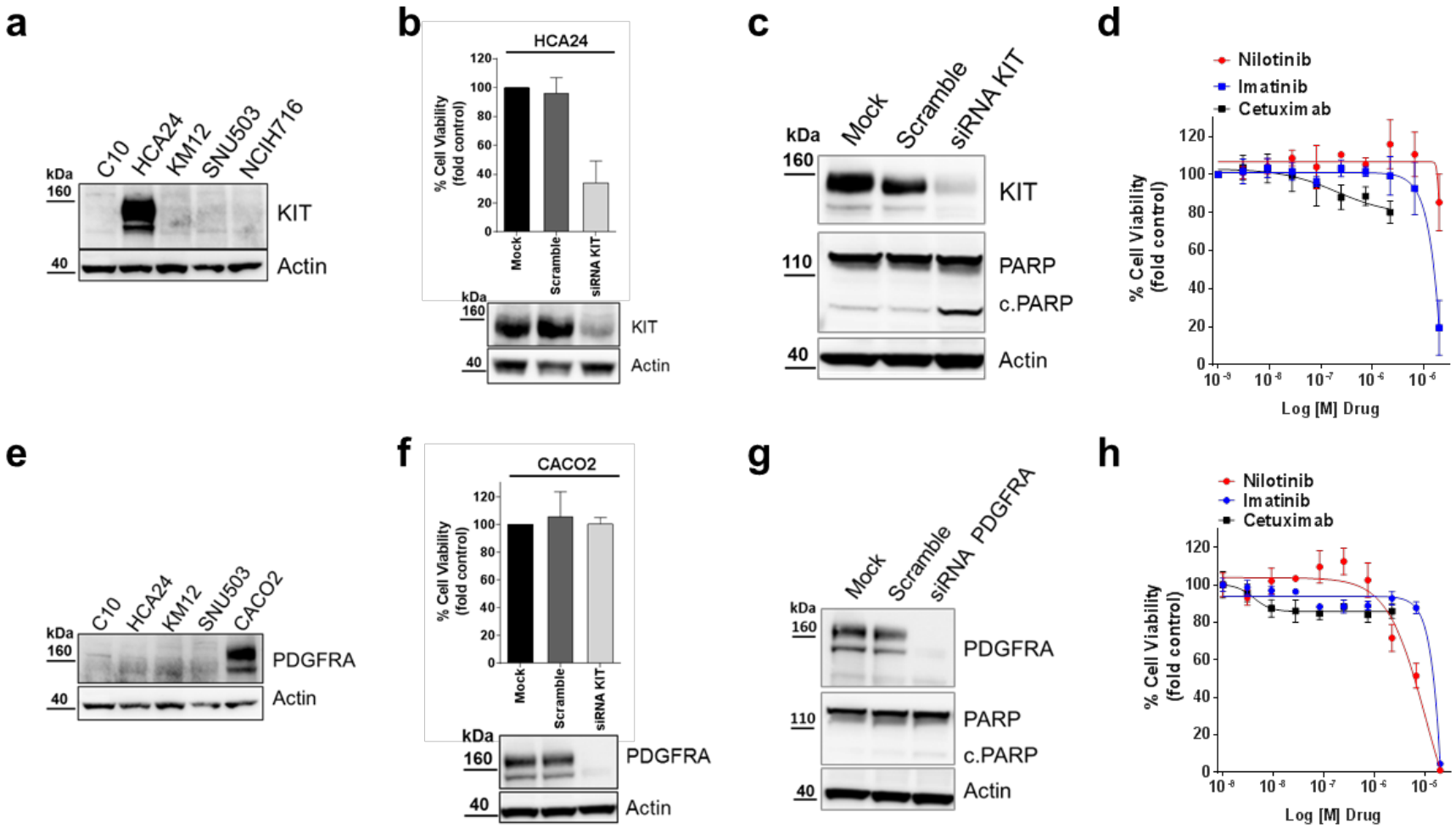
Supplementary Figure 5. Characterization of oncogenic *TPM3-NTRK1* gene fusion in KM12 cells. **(a)** RNA seq analysis identified gene fusion event between exon 7 of the *TPM3* gene and the exon10 of *NTRK1* gene in the KM12 CRC cell line. **(b)** PCR on the cDNA of KM12 cell lines confirmed the *TPM3-NTRK1* gene rearrangements. KM12SM cells, a derivate of the KM12 cells, were used as positive control. **(c)** Break apart FISH analysis of *TPM3-NTRK1* in KM12 cells shows clear separation of green (5') and red (3') signals corresponding to the *NTRK1* gene. Scale bars, 10µm. **(d)** Western blot analysis confirmed overexpression of the TRKA protein (codified by *NTRK1* gene) in KM12 cells, at a molecular weight compatible with *TPM3-NTRK1* translocation. **(e)** RNAi knockdown of NTRK1 inhibits cell proliferation of KM12 cells harboring the*TPM3-NTRK1* fusion gene. KM12 cells were analyzed by ATP lite proliferation assay 5 days after transfection with NTRK1-specific pooled siRNAs, scrambled siRNA, or transfection reagent (mock). Data are expressed as average ± s.d. of three independent experiments. si-RNA mediated downregulation of NTRK1 levels were verified by western blot. **(f)** KM12 cells, resistant to the EGFR inhibitor cetuximab, are sensitive to the TRK inhibitor CEP701. Cell viability was assessed by measuring ATP content after 5 days of treatment. Data are expressed as average ± s.d. of three independent experiments. **(g)** CEP701 treatment downregulates MAPK and PI3K pathways in KM12 cells harboring *NTRK1* genetic rearrangment. KM12 cells were treated with 1µM CEP701 for 12h, after which whole-cell extracts were subjected to western blot analysis.



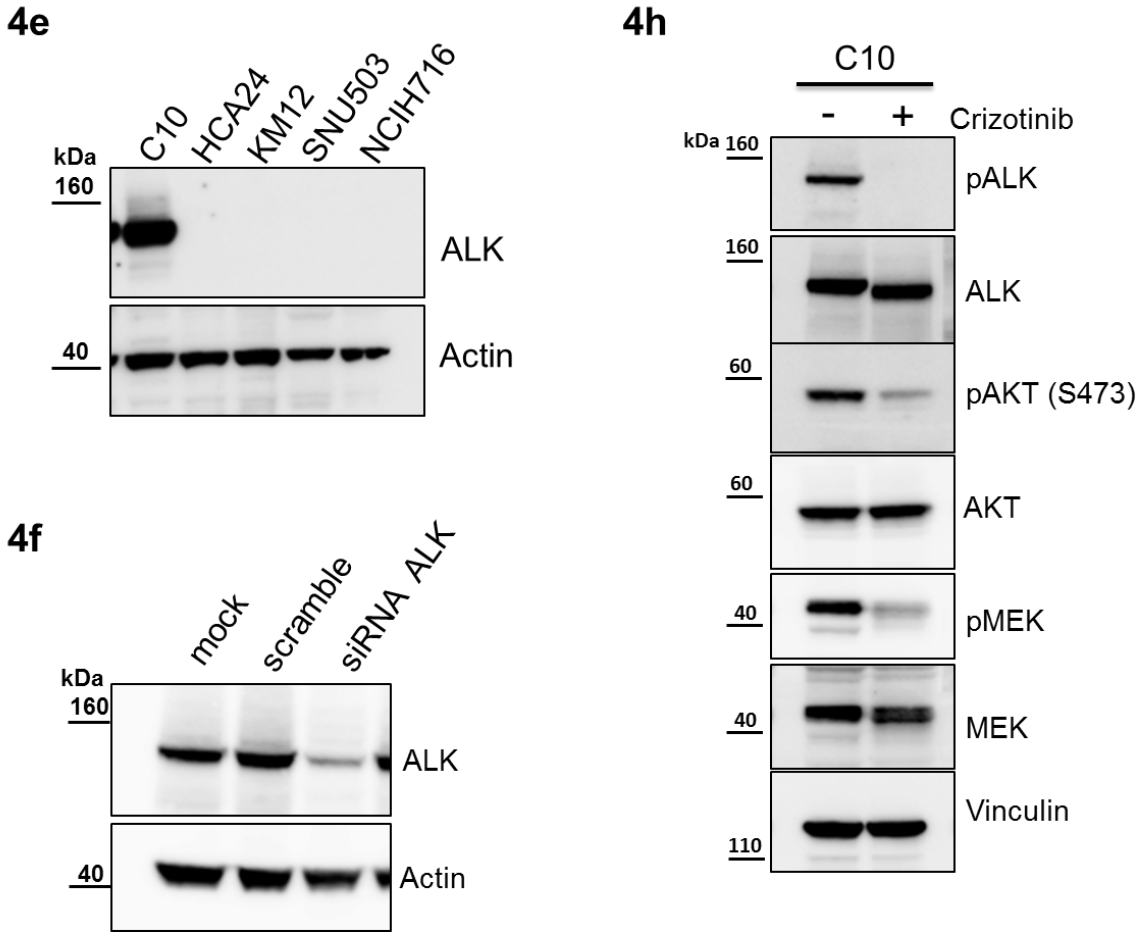
Supplementary Figure 6. Oncogenic addiction to *FGFR2* genetic amplification. (a) gDNA RTPCR identified *FGFR2* gene copy number increase in the NCIH716 CRC cell line. **(b)** Western blot analysis confirmed overexpression of FGFR2 protein in the NCIH716 cells. **(c)** RNAi knockdown of FGFR2 inhibits NCIH716 cell proliferation overexpressing the tyrosine kinase receptor. Cells were analyzed by ATP lite proliferation assay 5 days after transfection with FGFR2-specific pooled siRNAs, scrambled siRNA, or transfection reagent (mock). Data are expressed as average ± s.d. of three independent experiments. si-RNA mediated downregulation of FGFR2 levels were verified by western blot analysis. **(d)** NCIH716 cell lines, resistant to EGFR inhibitor cetuximab, are sensitive to FGFR inhibitor AZD4547. Cell viability was assessed by measuring ATP content after 5 days of treatment. Data are expressed as average ± s.d. of three independent experiments. **(e)** AZD4547 treatment downregulates MAPK and PI3K pathways in NCIH716 cells harboring *FGFR2* gene copy number increase. NCIH716 cells were treated with 0.5μM AZD4547 for 12h, after which whole-cell extracts were subjected to western blot analysis.



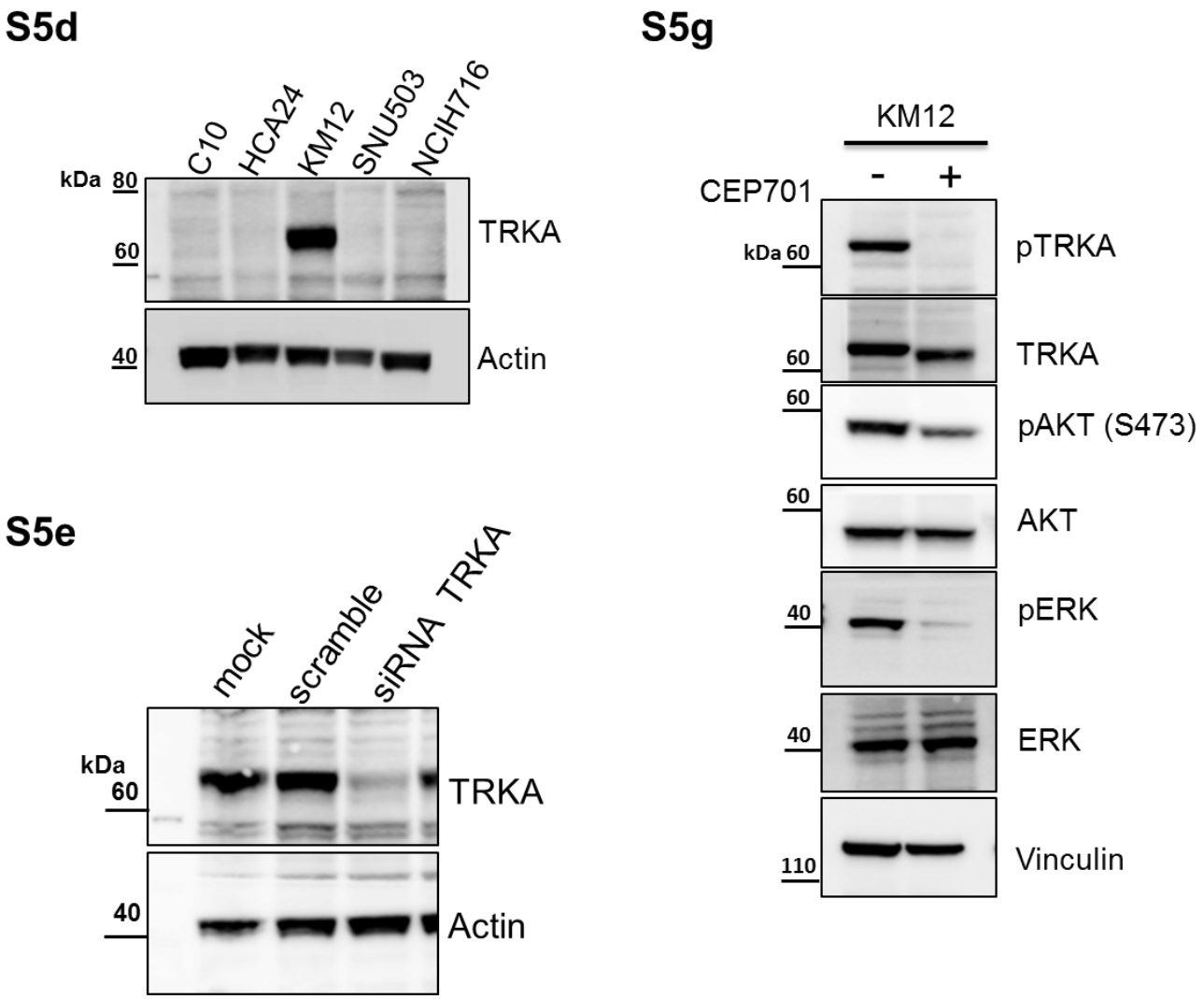
Supplementary Figure 7. Oncogenic addiction to overexpressed NTRK2 and RET kinases. (a) RNAi knockdown of NTRK2 inhibits cell proliferation of SNU503 cell line. SNU503 cells were analyzed by ATP lite 5 days after transfection with NTRK2-specific pooled siRNAs, or transfection reagent (mock). Data are expressed as average ± s.d. of three independent experiments. si-RNA mediated downregulation of NTRK2 levels were verified by cDNA RTPCR. **(b)** SNU503 cells, resistant to the EGFR inhibitor cetuximab, are sensitive to the TRK inhibitor CEP701. Cell viability was assessed by measuring ATP content after 5 days of treatment. Data are expressed as average ± s.d. of three independent experiments. **(c)** Treatment with TRK inhibitor CEP701 initiates apoptosis in NTRK2-dependent SNU503 CRC cells. Protein lysates were analyzed by western blot after 3 days of treatment with 1μM CEP701. **(d)** Western blot analysis confirmed overexpression of RET protein in NCIH716 cells. **(e)** NCIH716 cells, resistant to the EGFR inhibitor cetuximab, are sensitive to the RET inhibitor ponatinib and to the FGFR inhibitor AZD4547. Cell viability was assessed by measuring ATP content after 5 days of treatment. Data are expressed as average ± s.d. of three independent experiments. **(f)** RNAi knockdown of RET inhibits cell proliferation in the NCIH716 cells. NCIH716 cells were analyzed by ATP lite proliferation assay 5 days after transfection with RET-specific pooled siRNAs, scrambled siRNA, or transfection reagent (mock). Data are expressed as average ± s.d. of three independent experiments. si-RNA mediated downregulation of RET levels were verified by western blot analysis.



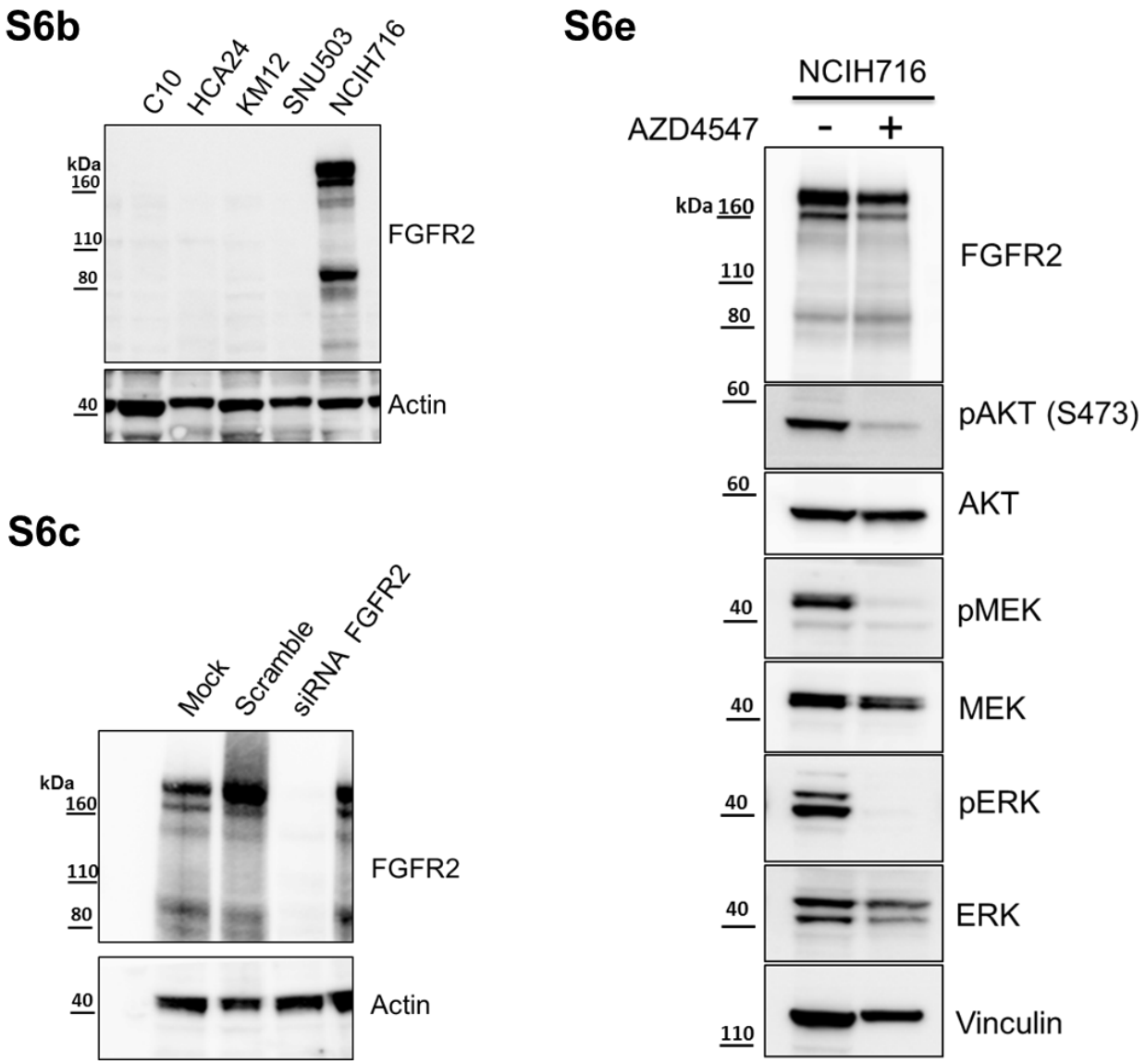
Supplementary Figure 8. Functional validation of KIT and PDGFRA overexpression. **(a)** Western blot analysis confirmed overexpression of KIT protein in HCA24 cell lines. **(b)** RNAi knockdown of KIT inhibits cell proliferation in HCA24 cells. HCA24 cells were analyzed by ATP lite 5 days after transfection with KIT-specific pooled siRNAs, scrambled siRNA, or transfection reagent (mock). Data are expressed as average ± s.d. of three independent experiments. si-RNA mediated downregulation of KIT levels were verified by western blot. **(c)** RNAi knockdown of KIT initiates apoptosis in KIT-dependent HCA24 CRC cell lines. Protein lysates were analyzed by western blot 3 days after transfection with KIT-specific pooled siRNAs, scrambled siRNA, or transfection reagent (mock). **(d)** HCA24 cells, resistant to the EGFR inhibitor cetuximab, are resistant also to the KIT kinase inhibitors nilotinib or imatinib. Cell viability was assessed by measuring ATP content after 5 days of treatment. Data are expressed as average ± s.d. of three independent experiments. **(e)** Western blot analysis confirmed overexpression of PDGFRA protein in CACO2 cell lines. **(f-g)** CACO2 cells are not addicted to the PDGFRA overexpression. RNAi knockdown of PDGFRA failed to induce apoptosis. Protein lysates were analyzed by western blot 3 days after transfection with PDGFRA-specific pooled siRNAs, scrambled siRNA, or transfection reagent (mock). **(h)** CACO2 cell lines, resistant to the EGFR inhibitor cetuximab, are resistant also to the PDGFRA kinase inhibitors nilotinib or imatinib. Cell viability was assessed by measuring ATP content after 5 days of treatment. Data are expressed as average ± s.d. of three independent experiments.



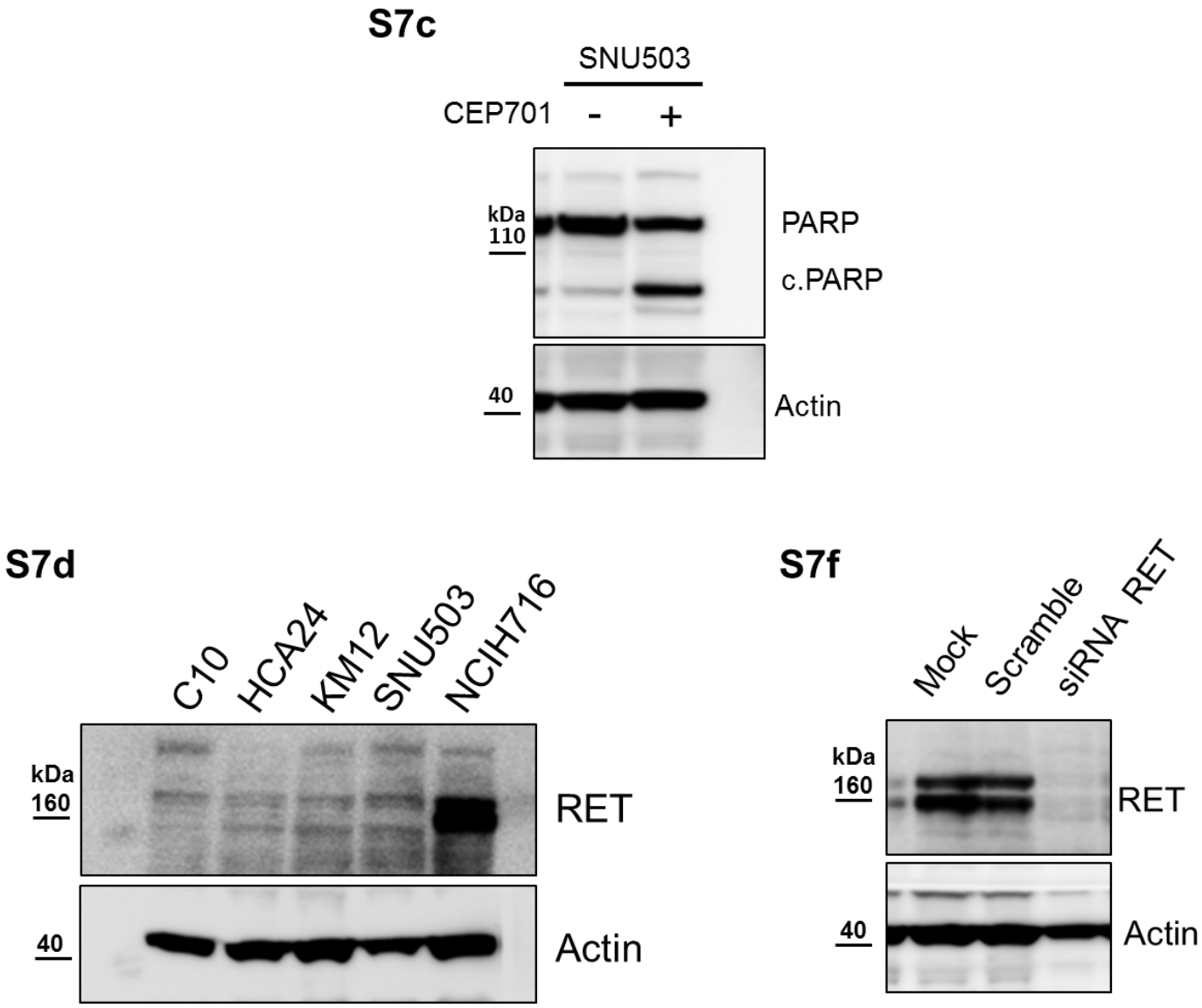
Supplementary Figure 9. Full lengths western blot of Figure 4. (4e) C10 CRC cells overexpress ALK protein. **(4f)** RNAi knockdown of ALK inhibits cell proliferation of C10 cell line harboring the EML4-ALK fusion gene. **(4h)** Crizotinib treatment downregulates MAPK and PI3K pathways in C10 cells.



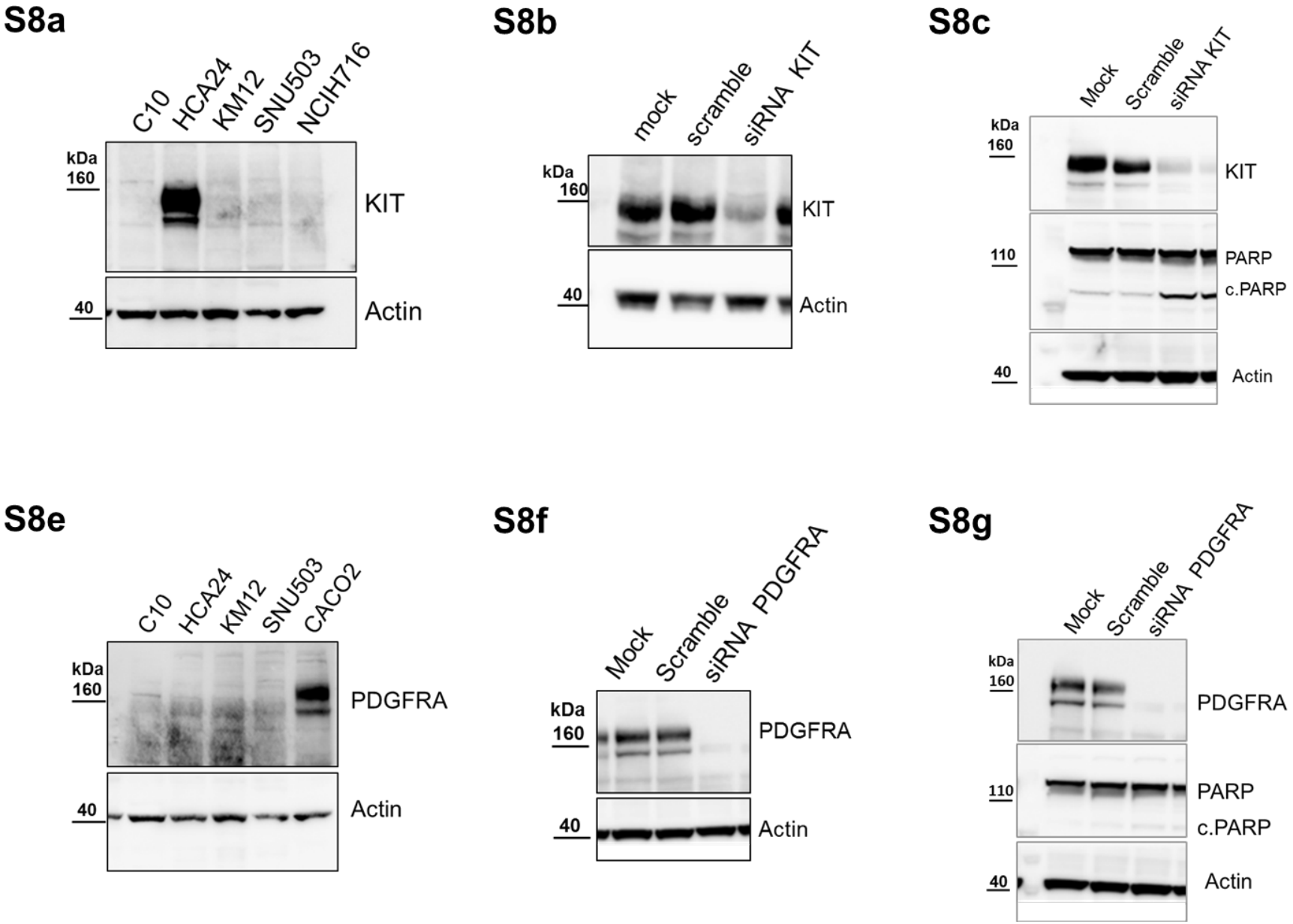
Supplementary Figure 10. Full lengths western blot of Supplementary Figure 5. (S5d) KM12 CRC cells overexpress TRKA protein. **(S5e)** RNAi knockdown of TRKA inhibits cell proliferation of KM12 cell line harboring the TPM3-NTRK1 fusion gene. **(S5g)** CEP701 treatment downregulates MAPK and PI3K pathways in KM12 cells.



Supplementary Figure 11. Full lengths western blot of Supplementary Figure 6. (S6b) NCIH716 CRC cells overexpress FGFR2 protein. **(S6c)** RNAi knockdown of FGFR2 inhibits cell proliferation of NCIH716 cell line harboring the FGFR2 gene amplification. **(S6e)** AZD4547 treatment downregulates MAPK and PI3K pathways in NCIH716 cells.



Supplementary Figure 12. Full lengths western blot of Supplementary Figure 7. (S7c) Treatment with TRK inhibitor CEP701 initiates apoptosis in NTRK2-dependent SNU503 CRC cells. **(S7d)** NCIH716 CRC cells overexpress RET protein. **(S7f)** RNAi knockdown of RET inhibits cell proliferation of NCIH716 cell line harboring the RET overexpression



Supplementary Figure 13. Full lengths western blot of Supplementary Figure 8. (S8a) HCA24 CRC cells overexpress KIT protein. **(S8b)** RNAi knockdown of KIT inhibits cell proliferation of HCA24 cells. **(S8c)** RNAi knockdown of KIT initiates apoptosis in KIT-dependent HCA24 CRC cells. **(S8e)** CACO2 CRC cells overexpress PDGFRA protein. **(S8bf)** RNAi knockdown of PDGFRA does not affect cell proliferation of CACO2 cells. **(S8g)** RNAi knockdown of PDGFRA does not induce apoptosis in CACO2 CRC cells.

Supplmentary Data 1. Molecular and pharmacological annotation of 151 CRC cell lines. This table is subdivided in four sections, reporting: (i) identification name and acquisition sources, (ii) STR profiles, (iii) MSI and mutational annotations and (iv) pharmacological annotations. On top of the sections, a note provides details and a brief description of the section content. Footnotes list contact details of scientists who provided the indicated cell lines.

CRC cell lines identity was checked by Short Tandem Repeats (STR) profiles through simultaneously amplification of ten STR loci (D5S818, D13S317, D7S820, D16S539, D21S11, vWA, TH01, TPOX, CSF1PO and amelogenin). Resulting cell line STR profiles were cross-compared and matched with the available ATCC STR, DSMZ, JCRB, Korean Cell Line Bank, ECCAC, and CellBank Australia repositories online databases. Cell lines derived from the same individual (similar STR) are highlighted by the same color.

List of the genetic status of the 151 CRC cell lines for the HRAS, NRAS, KRAS, BRAF and PIK3CA reported hotspots. MSI/MSS status are reported. PTEN protein expression, evalueted by western blot analysis, is also indicated. Positive: PTEN protein detected. Null: PTEN protein absent.

CRC cells were treated with indicated concentrations of cetuximab for 4 days, after that cell viability was assessed by ATP content. For each cell line the area under the concentration–inhibition curve (Index AUC) (see Methods) is reported. For each drug concentration the percentage (%) of growth inhibition at indicated concentrations is listed. Cetux: cetuximab

Section 1. Cell lines names, sources and acquisition details / **Section 2. Short Tandem Repeats (STR) profiles**

CRC Cell Line Name	Supplier ID	Cat ID	Date of Arrival	Locus AMEL		Locus CSF1PO				Locus D5S818				Locus D7S820				Locus D13S317				Locus D16S539			Locus D21S11			Locus TH01			Locus TPOX			Locus vWA			
C10	ECACC	12022901	January 2013	X	Y	11				12				10	13			13				11			30	31.2		6	9.3		8	9		14	15		
C106	ECACC	12022906	January 2013	X		10	11			12				8	10			8	14			12	13		32.2			6			10	11		18			
C125PM	ECACC	12022909	January 2013	X		11	14			9				8	10			11	12			9			28			8			8			17	19		
C146	Lindy Durrant[a]	UK	February 2012	X		12				13				10	12			8	11			12	13		29	32.2		7	9.3		8	11		14	18	19	
C170	Lindy Durrant[a]	UK	February 2012	X	Y	12				13				10	12			8	11			12	13		29	32.2		7	9.3		8	11		18	19		
C32	ECACC	12022908	January 2013	X		10	11			11				9				8	11			9			28	30		8	9			9		14	18		
C70	ECACC	12022903	January 2013	X		12				10				9				12				11	12		29	31.2		6	9.3		8	11		17	18		
C75	ECACC	12022907	January 2013	X	Y	12				8	12			9	11			10	13			8	11		31			8	9.3		10			17			
C80	ECACC	12022904	January 2013	X	Y	10	12			11	12			10	11			8	11			9	10		28	29		6	7		8	12		14	18		
C84	ECACC	12022902	January 2013	X		11	12			12				11				11	12			13			33.2			7	9.3		8	9		14	15		
C99	ECACC	12022905	January 2013	X		10	11			11	12			9	11			8	14			11	13		28	30		8	9.3		8	10		16	17		
CACO2	ATCC	HTB-37	Before 2010	X		9	11			12	13			11	12			11	13	14		12	13		30	32		6			9	11		16	18		
CAR1	JCRB	JCRB0207	April 2010	X		11				10				8	12			9	12			9			32.2			6			8			14	18		
CCK81	JCRB	JCRB0208	March 2012	X		10	13			11	12	13		8	8.1	10.3	12	9	12	13		9	10		28	31		6	8		9	11		14	18		
CL11	DSMZ	ACC467	October 2010	X	Y	11				12				9	10			13	14			11	13		31	32.2		6	9		8			16	17		
CL14	DSMZ	ACC504	October 2010	X		12				11				11	12			8	12			11	13		28	29		6	8		8			16	18		
CL34	DSMZ	ACC520	October 2010	X		11	16			9	11			10	15			11	13			9	11		30			6	7		8	10		15	18		
CL40	DSMZ	ACC535	October 2010	X		10	12			11				8	9			8	11			11	12		30	32.2		6	9		8	9		16	17		
CO115	Dr. Hamelin[b]	France	July 2011	X		11	12	13	14	10	12			7	10.3			12	16			10	11		34.2	35		9	9.3		8	10		15	20		
COCM1	JCRB	JCRB0257	April 2010	X		12				11	12			9	11			10				12	13		29	31		7	9		8			16	17		
COGA1	Prof. Huber[c]	Austria	November 2011	X		10	11			11	12			7.3	10			8	11			9	13		27	34		9	9.3		9	11		16	18		
COGA12	Prof. Huber[c]	Austria	November 2011	X		9	11			11	15			12	13			14	15			11	12		30	30.2		9.3			10	13		14	16		
COGA2	Prof. Huber[c]	Austria	November 2011	X		10	11			12	16			11	13			11	12			11	12	13	29	36.2		8	9.3		8	10		18	20	21	
COGA3	Prof. Huber[c]	Austria	November 2011	X		10	13			12				8	11	11.1		10	12	13		10	11		30.2	31.2		7	8		8	11		17	19		
COGA5	Prof. Huber[c]	Austria	November 2011	X	Y	11				13				10	11			13				11	12		29			6			11			18			
COGA5L	Prof. Huber[c]	Austria	November 2011	X		10	11			13				10	11			13				11	12		29			6			11			18	20		
COGA8	Prof. Huber[c]	Austria	November 2011	X	Y	11	12			10	12			8	11			11	12			9	11		29	31		9.3			8			17	18		
COLO201	ATCC	CCL-224	July 2011	X		11	12			10	13			9	10			10	12			12	13		30.2	33.2		8	9		11			15			
COLO205	NCI60	507320	Before 2010	X		11	12			10	13			9	10			10	12			12	13		30.2	33.2		8	9		11			15			
COLO320	ECACC	UK	November 2011	X		11				12				9	12			11				11	12		33.2			9			8	9		15	18		
COLO320DM	ATCC	CCL-220	July 2011	X		11				12				9	12			11				11	12		31	33.2		9			8	9		15	18		
COLO320HSR	ECACC	UK	February 2012	X		11				12				9	12			11				11	12		31	33.2		8	9		8	9		15	18		
COLO60H	CLS	Germany	December 2011	X	Y	11	15			9	16			7.3	10			11				9	13		29	33.2		6	9.3		7	10		15	16	17	19
COLO678	DSMZ	ACC194	October 2010	X		12				13	14			8	12			10	13			9			27	29		9	9.3		8			16	19		
COLO741	ATCC	USA	Before 2010	X		11				13				12				9	11				12		29	32.2		7	9		8	11		17	18		
COLO94H	CLS	Germany	December 2011	X	Y	11	14			12				8								13			27	28		7	9.3		8			15	19		
CX1	CLS	Germany	December 2011	X		11	12			11	12			10				11				11	12		29	30		6	9		8	9		17	19		
DIFI	Dr. Baselga[d]	Spain	Before 2010	X		10	11			11	12			10	12			8	11			12			32			7	9.3		8	9		17	18		
DLD1	ATCC	CCL-221	December 2012	X	Y	11	12			13				10	12			8	11			12	13		29	32.2		7	9.3		8	11		18	19		
FET	IEOM[e]	Italy	June 2012	X		12				14				8	11			11				12			28	32.2		7			8			16	17		
GEO	Prof. Tortora[f]	Italy	February 2012	X						12				10	11	11.1		8	12			12			29			8			8			15			
GP2D	ECACC	95090714	January 2012	X		9	12			11	12			9	11			8	12			11	13		29	30		7	9.3		8	11		17	18		
GP5D	ECACC	95090715	January 2012	X		9	12			12				9	11			9	12			11	13		29	30		7	9.3		8	11		17	18		
HCA24	ECACC	6061903	January 2012	X	Y	11				13				8	11.3			12	13			11	12		28	29		6	9.3		10	11		14	18		
HCA46	ECACC	7031601	October 2009	X		11	12			11	12			10	14			10	11			11	13		30			6	9.3		8	9		17			
HCA7	ECACC	6061902	Before 2010	X		11	14			9	11	12		11	12			9				10	11		26	28.2	29.2	6	7		8	10		14	16		
HCC2998	NCI60	507628	May 2011	X		11	13			12				10	11			11	12			9	12		29	30		7	9.3		8	11		18	19		
HCT116	NCI60	507326	May 2011	X		7	10			10	11			11	12			10	12			11	13		29	30		8	9	10	8			17	22		
HCT15	NCI60	507616	May 2011	X	Y	12				13				10	12			8	11			12	13		29	32.2		7	9.3		8	11		18	19		
HCT8	ATCC	CCL-244	January 2012	X	Y	12				13				10		12		8	11			12	13		29	32.2		7	9.3		8	11		18	19		
HDC114	DKFZ[g]	Germany	November 2012	X	Y	11				12				10				12				11	13		28	30		6	9.3		8			14	19		
HDC135	DKFZ[g]	Germany	November 2010	X		9	11			10	14			10	11			9				10	12		26	28		9	9.3		7	8		16	18		
HDC142	DKFZ[g]	Germany	November 2010	X		11	12			12	13			8				8				12	13		27	28		9	9.3		8	11		17			
HDC143	DKFZ[g]	Germany	November 2012	X		11	12			9	10			9	10			9	12			12	13		30	32.2		7	9		8			16			
HDC54	DKFZ[g]	Germany	November 2010	X	Y	11				12				8	9			11	12			11			30.2	32.2		9.3			9	10		17	19		
HDC8	DKFZ[g]	Germany	November 2010	X		12				12				8	10			11				13			31			9.3			8	10		15	19		
HDC82	DKFZ[g]	Germany	November 2010	X		13				11				8	12			9	11			13			29	30.2		7	8		8	9		17	18		
HDC9	DKFZ[g]	Germany	November 2010	X	Y	10	12		13	11	12			7	13			11	12			10	12		26	31.2		6	7		9			18	19		
HRA16	ECACC	6061905	January 2012	X	Y	11					12			11	12			9	13			10	13		28			9.3			8	12		16	18		
HROC18	Dr. Linnebacher[h]	Germany	January 2012	X		12				9	11			8	11			11				12	13		30	33.2		7	8		8			17			
HROC24	Dr. Linnebacher[h]	Germany	September 2011	X	Y	11	12			10	11			11	13			7	13			9	11		27	28		6	9		11			16	20		
HROC32	Dr. Linnebacher[h]	Germany	January 2012	X		14				12				8	11			12				11	12		31			8	9		8	11		19			
HROC39	Dr. Linnebacher[h]	Germany	January 2012	X		12	13			13				12				12	13			12			29			7	9.3		8	11		16	18		
HROC46	Dr. Linnebacher[h]	Germany	January 2012	X	Y	12				13				8	11			12				12	13		28	32.2		6	9.3		8	12		15	17		
HROC69	Dr. Linnebacher[h]	Germany	January 2012	X	Y	10	11			12	13			10	12			11	12	13		12	13		29	33.2		7	9.3		8	12		17	19		
HROC80	Dr. Linnebacher[h]	Germany	January 2012	X		12				12	13			9	11			8				10			30	32.2		6			8	11		17			
HROC87	Dr. Linnebacher[h]	Germany	September 2011	X		12				9	11			9				10	14			11	13		28	29		8			11	13		16	20		
HT115	ECACC	85061104	January 2012	X		7	12			12				11	12			13				10			30	32.2		6	9		8			18			
HT29	NCI60	507335	May 2011	X		11	12			11	12			10				11	12			11	12		29	30		6	9		8	9		17	19		
HT55	ECACC	85061105	January 2012	X		11				11	12			11	13			9	13			14			28			8	9.3		8	9		14	17		
HUTU80	ATCC	HTB-40	Before 2010	X	Y	11	13			12	13			9	11			8	11			10	11		31	32.2		7			9	11		16	18		
KM12	NCI60	507347	May 2011	X		10	12			10				8	8.3			12	15			11			27			9.3			11	12		17	19		
KM12C	KCLB	80015	November 2010	X		10				10				8	9			12	15			11			27	34.2		10			11	12		17			
KM12L4	KCLB	80014	November 2010	X		10	12			11				8	9			13	15			11	12		27	34		10			11	12		16	17		
KM12SM	KCLB	80016	November 2010	X		10	12			11				8	9			13	15			11	12		27	34.2		9.3			11	12		16	17		
KM20	KCLB	80017	November 2010	X		11	12			11	12			10				11				11	12		29	30		6			8	9		17	19		
LIM1215	Dr. Whitehead[i]	Australia	September 2009	X	Y	10				9	12			9	10.3			8	12			8, 9	11	12	29	31		9.3	10		8			15	16		
LIM1899	Dr. Walker[j]	Australia	December 2010	X	Y	10				11	14			8	9			9	12			11			29			6	9.3		9	11		17	20		
LIM2099	Dr. Walker[j]	Australia	December 2010	X	Y	11				11				13				8				11			29	30		9			8			19			
LIM2405	Dr. Walker[j]	Australia	December 2010	X	Y	14	15			12	13			7	8			8	11			14	15		29	30		9.3			8			14	18		
LIM2412	Dr. Walker[j]	Australia	December 2010	X		9				10				8	11			8	13			11	12		28	29	30	6			7	8	9	14	17		
LIM2537	Dr. Walker[j]	Australia	December 2010	X		9				10				8	11			8	14			12			28	29	30	6			8	9		14	17		
LIM2550	CBA	CBA-0169	September 2011	X	Y	12				12	13			9	10			12	15			9	14		29	31.2		8	9		8			14	19		
LIM2551	CBA	CBA-0170	September 2011	X	Y	10	11			12	16			9	11			8	10			11			28	33.2		6	9.3		8	11		17	19		
LOVO	ATCC	CCL-229	July 2011	X	Y	11	13		14	11	12	13		9.3	10	11		8	11			9	12		29	31.2		9.3			8	9		17	18		
LS1034	ATCC	CRL-2158	July 2011	X	Y	11	12			12	13			8	12			12				9	11		29	30		7	7		9	11		17			
LS123	ATCC	CCL-255	July 2011	X		10	11			11				9				9	13			13			30	31		7	8		9			16	17		
LS174T	ATCC	CL-188	January 2011	X		10	13		14	11	15			11				10				11	13		29	31		8	7		8	9		15	17		
LS180	ATCC	CL-187	January 2011	X		10	13			10	15			10	11			10				11	13		29	31		8	7		8	9		15	18		
LS411N	ATCC	CRL-2159	July 2011	X		10				9	16			7.3	9.3			12	13	14		11	12		29	32	34	6			8	9		16	17		
LS513	ATCC	CRL-2134	July 2011	X	Y	10	13			9	11			8	11			9	10			12	13		30	30		8			8			16	17		
MDST8	ECACC	99011801	January 2012	X		11				10				8	10			10	11			11			28	31		8	9.3		8			15	20		
MIP101	Prof. Jessup[m]	USA	June 2012	X		12				13				10	11.3			8	11			12	13		29	32.2		7	9.3		8	11		18	19		
NCIH498	ATCC	CCL-254	May 2012	X		12				13				8	11			9	11			9	11		29	31		7			8			15			
NCIH508	ATCC	CCL-253	February 2011	X		11	12			12				8	12			8	12			12			29	32.2		9	9.3		8	11		15	16		
NCIH630	KCLB	90630	November 2010	X	Y	10	11			12				8	11			11	12	13	14	11	12		30	33.2		7	9		8	9		15	16	17	18
NCIH684	KCLB	90684	November 2010	X	Y	10				11				10	11				12			11	12		31.2			9			8	9		16	17		
NCIH716	ATCC	CCL-251	July 2011	X		11				11				10	11			8	11			11	12		32.2	32.2		6	9.3		8	11		16	17		
NCIH747	ATCC	CCL-252	Before 2010	X	Y	11	12			11				10				11	12			9	12		29			6			8	11		16	19		
OUMS23	JCRB	JCRB1022	40269	X		12				11				10	13			12				11			28	30		7	9.3		11			16			
OXCO1	Prof. Cerundolo[n]	UK	November 2010	X		12				12				8	12			11				11	12		30			6	9.3		9	10		16	17		
OXCO2	Prof. Cerundolo[n]	UK	November 2010	X	Y	9	13			11	12			10	13			10	11			11	12		30			8			11			17	22		
OXCO3	Prof. Cerundolo[n]	UK	November 2010	X		12				11				10	11			11	12			9			28			9	9.3		8			16	18		
RCM1	HSRRB	JCRB0256	March 2012	X		10	12			11				11	12			10	12			11	12		29			9			8	11		20			
RKO	Prof. Bert Vogelstein[o]	USA	August 2010	X		8	11			11	12	13	14	8	10	11		8	11			11	12	13	28	30		6	10		9	10	11		16		
RW7213	Diego Arango[p]	Spain	October 2012	X	Y	11				11				10				10	12			8	10		30	31		6			11			18			
SKCO1	ATCC	HTB-39	July 2011	X	Y	13				9	11			8	10			14				13			28	30		8	9		8	11		18	19		
SNU1033	KCLB	719	November 2010	X		10				13				11	12			10				9			29	30		9			11			17	19		
SNU1040	KCLB	1140	November 2010	X	Y	12				10	13			11	13			12				9			28	29	32	6	9		8			17	19	20	
SNU1047	KCLB	1147	November 2010	X		9	11			8	12			8	10			10	11			9			29	32.2		9			8	11		14	21		
SNU1181	KCLB	1181	July 2012	X		12				10				10	11			10	11			10	12		29	32.2		9			8			14	18		
SNU1235	KCLB	1235	July 2012	X		12				12				8	12			13				12			30			9			8	9		17			
SNU1406	KCLB	1406	July 2012	X		11	12			11	12			8	11			8	9			11	12		30	33.2		6	9.3		8	11		15	18		
SNU1411	KCLB	1411	July 2012	X		10	12			11	13			11				10				12	13		30			9			8			20			
SNU1460	KCLB	1460	July 2012	X		12				11				8	11			9	12			10			30.2	31		6	7		8	11		14	18		
SNU1544	KCLB	1544	July 2012	X		11	12			12				11	12			9	11			8	14		29	30		7	9		8			14	20	21	
SNU1684	KCLB	1684	July 2012	X		9	10			10	11			9	13			8				9	10		28			9			8	10		14	17		
SNU1746	KCLB	1746	July 2012	X	Y	7	9			11	12	13	14	8	11			9				9	14		28	33		6			8	9		18			
SNU175	KCLB	175	November 2010	X		10	12			11				11				9	12			9	12		30	32.2		7	9		8	11		14	17		
SNU254	KCLB	254	July 2012	X		10	11			10	11			13				11				9			31			7			8	12		14	17		
SNU283	KCLB	638	November 2010	X		11	13			9				11				8	11			9	12		29	30		7	9		8	11		14	19		
SNU407	KCLB	407	October 2012	X	Y	10	11			10				8	10			8	10			11	12		28	29		6			7	8		14	16		
SNU479	KCLB	479	July 2012	X		11	12			13				9	12			9	11			9	11		32.2			6	7		9	12		16	17		
SNU503	KCLB	503	October 2012	X		11				10				10	11			10				11	12		30			9			8	11		14	17		
SNU61	KCLB	61	November 2010	X		12				11				8				8				12			29			9			8	9		16			
SNU769B	KCLB	668	November 2010	X	Y	11	14			14				11.3	13			9	12			12			29	31		7	8		8			18			
SNU81	KCLB	81	October 2012	X		11	13			10	11			12				9	10			9	12		30	31		7	9		8	11		16	17		
SNU977	KCLB	977	July 2012	X		11	12			11	12			10	12			11				10			29			7			11			16	17	18	
SNUC1	KCLB	0000C1	November 2010	X		12				10				10	11			9	11			9	11		30			7	9		11	12		16	18		
SNUC2A	KCLB	0000C2A	November 2010	X		11	14			8	9			13	16			9	12			8	9					9			8			15	19		
SNUC2B	KCLB	0000C2B	November 2010	X		11	14			8	9			13	16			9	10			8	9					9			8			15	18		
SNUC4	KCLB	0000C4	November 2010	X	Y	10				8	9			7.3		11	12	8	9			9	10		28	30		7	8		11			14	15		
SNUC5	KCLB	0000C5	November 2010	X		11	13			11	12			12				8				8	13		29	31.2		6	9		8			15			
SW1116	ATCC	CCL-233	May 2011	X	Y	10	11			11	12			12				11	14			9	12		28	29		6			8	11		14	19		
SW1222	Dr Herlyn[q]	USA	July 2011	X	Y	11				9				10	12			9				9	10		29			7			10			18			
SW1417	ATCC	CCL-238	May 2011	X		12				12	12			9	11			12				11	12		29	32.2	33.2	6	8		8	11		15	18		
SW1463	ATCC	CCL-234	May 2011	X		11	12			13	14			9				12	13			11			30	31.2		6	7		8	11		16			
SW403	ATCC	CCL-230	40634	X		10	13			11				8	9			13				10	12		28	29		6			8	9		14	18		
SW48	ATCC	CCL-231	March 2011	X		9	10			10	14			9	10			11	12			11	13		28	31		8			6	9.3		18	20		
SW480	ATCC	CCL-228	January 2011	X		13	14			13				8				12				13			30	30.2		8			11			16			
SW620	ATCC	507424	May 2011	X		13	14			13				8	9			12				9	13		30	30.2		8			11			16			
SW837	ATCC	CCL-235	May 2011	X		10				12				9	12			13				12			28	31.2		9.3			8	9		15	19		
SW948	ATCC	CCL-237	40634	X		12				11				9	11			10	11			11	12		25.2	29		6	9.3		8	11		16	18		
T84	ATCC	CCL-248	July 2011	X		10				12				8	10			10	11			10	11		31			6	9		8	11		17	18		
V411	Prof. Markowitz[r]	USA	March 2012	X		11	12			13				9	11			9	12			11	12		29	32.2		9			8			15			
V457	Prof. Markowitz[r]	USA	March 2012	X		11	14			11	16			12	13			11	12			11	12		29	32		7	9		8	9		18			
V481	Prof. Markowitz[r]	USA	March 2012	X		7	12			8	11			8				11	13			11	15		31			7			8	9		17	18	21	
V703	Prof. Markowitz[r]	USA	March 2012	X		9	12			13				11				9	12			9			29.2	31		9	9.3		7	10		18	21		
VACO432	Prof. Bert Vogelstein[o]	USA	August 2010	X		9	10			9	10	15	16	11	13			11	12			11	14		28	33.2		8	9.3		7	8		14	19		
VACO5	Prof. Markowitz[r]	USA	February 2012	X		11	13.1			13				10				9	13			9	13		24.2	29		9.3			9	11		14	18		
VACO6	Prof. Markowitz[r]	USA	February 2012	X	Y	11				11				7	10			11	14			10	12	13	29	31.2		9	9.3		8	10		18	20	22	
WIDR	Prof. Bernards[s]	The Netherland	September 2011	X		11	12			11	12			10	10			11	12			11	12		29	30		6	9		8	9		17	19		

Section 3. Microsatellite Instability (MSI/MSS) and mutational status / **Section 4. Pharmacological profile of response to cetuximab treatment**

CRC Cell Line Name	MSI / MSS status	AA change HRAS (codon 12, 13, 61)	AA change NRAS (codon 12, 13, 61)	AA change KRAS (codon 12, 13, 61, 117, 146)	AA change BRAF (codon 600)	AA change PIK3CA (codon 542, 545, 546, 1047)	PTEN expression	AUC index	NT	0.001µg/ml Cetux	0.01µg/ml Cetux	0.1µg/ml Cetux	1µg/ml Cetux	10µg/ml Cetux	25µg/ml Cetux	50µg/ml Cetux	100µg/ml Cetux
C10	MSI	WT	WT	WT	WT	WT	positive	562.44	0	-3.57	-0.96	-3.84	-1.41	-0.97	1.05	6.62	-0.71
C106	MSS	WT	WT	p.G12C	WT	WT	positive	4972.23	0	1.94	6.94	3.41	12.46	19.31	20.78	24.39	30.06
C125PM	MSS	WT	WT	p.K117N	WT	WT	positive	1048.79	0	-4.71	8.32	9.09	9.18	17.15	0.28	4.83	6.65
C146	MSI	WT	WT	p.G13D	WT	p.E545K	positive	634.22	0	0.14	4.84	0.45	-1.41	1.75	2.13	3.85	3.32
C170	MSI	WT	WT	p.G13D	WT	p.E545K	positive	986.56	0	-1.00	2.15	2.84	3.40	-1.09	3.35	5.29	6.90
C32	MSS	WT	p.G12D	WT	WT	WT	positive	524.23	0	2.48	-6.76	-6.76	-3.28	-2.73	-4.12	7.41	1.50
C70	MSS	WT	WT	WT	WT	WT	positive	12933.35	0	6.02	5.05	19.10	44.51	52.09	60.04	67.78	66.81
C75	MSS	WT	WT	WT	WT	WT	positive	15355.71	0	11.98	42.74	55.58	66.86	78.34	77.90	77.63	74.94
C80	MSS	WT	WT	p.A146V	WT	WT	positive	2986.99	0	-0.86	0.45	1.36	2.01	6.20	4.44	15.03	24.44
C84	MSS	WT	WT	p.G12A	WT	WT	positive	7041.12	0	8.79	10.71	21.44	23.47	27.20	37.18	36.78	33.64
C99	MSS	WT	WT	WT	WT	WT	positive	15079.23	0	11.68	34.14	48.70	59.43	68.08	72.40	76.18	78.64
CACO2	MSS	WT	WT	WT	WT	WT	positive	2887.05	0	4.35	12.51	19.99	13.94	15.20	11.60	15.94	14.05
CAR1	MSS	WT	WT	WT	WT	WT	positive	4954.64	0	-8.70	-2.74	5.06	10.52	15.53	22.83	24.39	29.19
CCK81	MSI	WT	WT	WT	WT	WT	positive	15413.17	0	28.03	44.06	63.09	67.99	73.21	73.29	77.69	80.15
CL11	MSS	WT	WT	p.Q61H	WT	WT	positive	1095.11	0	2.30	2.94	2.23	2.18	4.67	5.59	6.00	4.89
CL14	MSS	WT	WT	WT	WT	WT	positive	10695.69	0	15.39	29.32	35.77	40.43	47.75	48.51	55.50	55.72
CL34	MSI	WT	WT	WT	p.V600E	WT	positive	2162.36	0	0.56	0.05	-0.57	7.86	7.56	7.20	12.61	11.49
CL40	MSI	WT	WT	p.G12D	WT	WT	positive	4770.78	0	0.58	6.07	7.56	4.63	16.40	17.32	24.01	30.39
CO115	MSI	WT	WT	WT	p.V600E	WT	null	2196.48	0	2.75	1.98	4.47	5.34	7.11	8.63	11.02	13.60
COCM1	MSS	WT	WT	WT	WT	WT	positive	15779.12	0	18.12	28.69	49.17	58.39	70.60	74.39	78.79	84.87
COGA1	MSI	WT	WT	WT	WT	WT	positive	13301.17	0	-1.30	4.31	37.70	53.44	59.94	67.64	66.21	68.21
COGA12	MSI	WT	WT	WT	WT	WT	positive	11155.18	0	2.47	21.53	38.73	39.15	46.98	53.12	55.49	60.62
COGA2	MSS	WT	WT	p.G12S	WT	WT	positive	7769.69	0	11.19	8.38	25.13	24.15	28.59	33.95	37.75	46.69
COGA3	MSI	WT	WT	p.G13D	WT	WT	null	603.64	0	-0.51	-1.71	-1.80	-1.02	-2.69	0.21	4.79	3.70
COGA5	MSS	WT	WT	p.G13D	WT	WT	positive	713.58	0	-1.03	-7.82	-1.07	2.29	-5.62	-1.58	9.13	0.90
COGA5L	MSS	WT	WT	p.G13D	WT	WT	positive	290.48	0	-4.59	-2.84	-0.30	-2.05	-2.75	1.24	2.14	1.59
COGA8	MSS	WT	WT	WT	WT	WT	positive	1663.39	0	0.53	2.06	1.50	-2.73	-1.19	-0.01	5.60	20.61
COLO201	MSS	WT	WT	WT	p.V600E	WT	positive	839.14	0	1.14	2.45	1.76	2.40	1.45	2.37	6.04	3.38
COLO205	MSS	WT	WT	WT	p.V600E	WT	positive	2113.39	0	-1.14	6.75	8.14	6.59	6.03	3.82	12.52	13.49
COLO320	MSS	WT	WT	WT	WT	WT	positive	87.20	0	1.54	-1.70	-2.04	-1.27	2.04	-1.41	0.34	1.56
COLO320DM	MSS	WT	WT	WT	WT	WT	positive	-901.48	0	-0.98	1.96	-4.63	-9.16	-8.77	-4.25	-3.23	-5.58
COLO320HSR	MSS	WT	WT	WT	WT	WT	positive	1036.72	0	-0.99	-0.24	-1.66	-2.33	-2.90	-1.04	6.72	9.21
COLO60H	MSI	WT	WT	p.A146T	WT	WT	positive	-110.90	0	1.48	0.41	4.03	-0.40	-2.30	-1.90	-0.68	0.95
COLO678	MSS	WT	WT	p.G12D	WT	WT	positive	-718.07	0	0.23	-1.62	0.57	-1.66	6.69	-3.44	-5.15	-3.69
COLO741	--> Other analyses not done because of the melanoma origin of this cell line																
COLO94H	MSS	WT	WT	p.G12D	WT	p.E545K	positive	915.33	0	1.56	8.21	-0.03	2.43	1.69	4.05	5.71	3.95
CX1	MSS	WT	WT	WT	p.V600E	WT	positive	1858.22	0	-2.79	-1.88	-4.16	-1.64	1.63	4.91	7.89	16.46
DIFI	MSS	WT	WT	WT	WT	WT	positive	16133.06	0	1.39	43.45	71.45	78.26	80.82	81.32	80.82	80.33
DLD1	MSI	WT	WT	p.G13D	WT	p.E545K	positive	98.34	0	0.54	1.49	5.87	0.36	-2.43	0.65	1.34	-0.25
FET	MSS	WT	WT	p.G12D	WT	WT	positive	1348.64	0	-0.01	3.42	0.75	2.21	2.92	4.26	8.76	6.44
GEO	MSI	WT	WT	p.G12A	WT	WT	positive	-490.44	0	-6.78	-5.48	-4.79	-4.01	-4.06	-3.09	-4.47	1.41
GP2D	MSI	WT	WT	p.G12D	WT	p.H1047L	positive	5160.15	0	8.18	8.98	17.32	17.83	24.91	22.70	27.34	26.15
GP5D	MSI	WT	WT	p.G12D	WT	p.H1047L	positive	7031.60	0	2.13	9.11	19.64	26.25	29.89	31.03	34.64	40.35
HCA24	MSS	WT	WT	WT	WT	WT	positive	5854.93	0	8.69	9.51	12.79	19.46	24.27	27.63	28.52	33.03
HCA46	MSS	WT	WT	WT	WT	p.E542K	positive	15201.24	0	2.57	13.63	25.43	48.26	60.06	74.20	78.20	78.51
HCA7	MSI	WT	WT	WT	WT	WT	positive	9731.98	0	8.00	17.10	19.77	30.46	43.52	46.98	45.85	55.85
HCC2998	MSS	WT	WT	p.A146T	WT	WT	positive	825.45	0	-2.41	-2.75	-0.18	-1.20	2.07	0.48	5.55	5.14
HCT116	MSI	WT	WT	p.G13D	WT	p.H1047R	positive	1638.93	0	-0.48	7.54	9.02	9.09	8.75	6.93	7.40	10.16
HCT15	MSI	WT	WT	p.G13D	WT	p.E545K	positive	-434.33	0	1.82	2.54	0.30	-0.79	0.25	0.70	-3.88	-2.20
HCT8	MSI	WT	WT	p.G13D	WT	p.E545K	positive	1352.38	0	-9.65	-13.65	-19.62	-6.03	8.74	-2.35	8.38	10.64
HDC114	MSS	WT	WT	WT	WT	WT	positive	5233.58	0	7.48	14.76	11.30	21.71	23.35	25.44	25.38	28.71
HDC135	MSI	WT	WT	WT	p.V600E	WT	positive	1527.86	0	3.03	2.92	2.62	3.93	5.90	8.65	9.44	4.73
HDC142	MSS	WT	WT	WT	WT	WT	positive	12203.91	0	6.68	18.01	28.37	47.06	57.97	57.17	59.84	66.77
HDC143	MSI	WT	WT	WT	WT	WT	positive	-150.05	0	-0.20	-2.88	4.18	3.67	-2.94	-3.79	-1.53	2.92
HDC54	MSS	WT	WT	WT	WT	WT	positive	7716.15	0	6.14	14.31	26.04	31.80	37.72	35.53	35.71	45.55
HDC8	MSS	WT	WT	p.G12D	WT	WT	positive	1044.27	0	-1.76	0.38	0.51	0.40	1.06	0.75	5.61	8.82
HDC82	MSS	WT	WT	WT	WT	WT	positive	14156.82	0	0.31	23.99	48.85	56.07	63.37	71.55	71.34	71.69
HDC9	MSI	WT	WT	WT	WT	WT	positive	1813.42	0	-0.35	5.49	6.00	6.25	8.48	6.91	12.34	5.91
HRA16	MSS	WT	WT	WT	WT	WT	positive	12342.48	0	24.46	35.51	45.52	58.48	61.22	60.77	63.53	60.11
HROC18	MSS	WT	WT	p.A146V	WT	p.E545K	positive	3474.66	0	-0.05	-1.90	2.10	6.20	8.74	0.90	24.82	19.99
HROC24	MSI	WT	WT	WT	p.V600E	WT	positive	587.69	0	3.30	1.81	3.89	1.51	2.41	7.05	2.89	0.29
HROC32	MSS	WT	WT	p.G12A	WT	WT	positive	6995.80	0	3.12	8.50	17.37	20.26	21.87	24.55	40.69	37.10
HROC39	MSS	WT	WT	p.A146T	WT	WT	positive	3365.57	0	-5.31	4.48	16.26	17.42	10.53	15.73	19.85	14.50
HROC46	MSS	WT	WT	p.G12V	WT	WT	positive	1241.91	0	6.78	-4.77	4.01	13.69	4.71	12.81	9.58	-3.27
HROC69	MSS	WT	WT	WT	WT	WT	positive	3748.48	0	6.50	10.14	16.02	16.15	14.76	17.76	17.66	22.07
HROC80	MSS	WT	WT	p.G12V	WT	WT	positive	-1710.63	0	8.40	0.96	-2.41	-7.38	-6.47	-3.96	-7.83	-13.39
HROC87	MSI	WT	WT	WT	p.V600E	WT	positive	2018.73	0	-1.59	1.53	6.69	7.15	8.78	6.31	10.26	12.89
HT115	MSS	WT	WT	WT	WT	WT	positive	6596.29	0	-2.75	-0.59	7.29	13.60	21.04	25.20	30.73	45.06
HT29	MSS	WT	WT	WT	p.V600E	WT	positive	1255.35	0	-0.55	-3.56	0.33	-0.76	-2.13	0.83	8.72	8.47
HT55	MSS	WT	WT	WT	WT	WT	positive	4242.82	0	-3.85	-0.24	2.86	12.38	17.77	16.83	23.44	22.03
HUTU80	MSS	WT	WT	WT	WT	WT	positive	-836.36	0	1.11	-1.51	-2.84	-2.56	-4.30	-4.68	-5.35	-2.10
KM12	MSI	WT	WT	WT	WT	WT	null	-771.80	0	-1.55	-1.36	-2.53	-4.36	-4.80	-4.90	-4.05	-2.64
KM12C	MSI	WT	WT	WT	WT	WT	null	-95.10	0	-4.88	-5.91	-5.03	-5.25	-4.72	-5.11	0.42	2.48
KM12L4	MSI	WT	WT	WT	WT	WT	null	147.82	0	-7.19	-3.96	-3.65	-3.13	-0.41	-3.48	0.16	4.92
KM12SM	MSI	WT	WT	WT	WT	WT	null	-109.42	0	-6.71	-6.45	-7.05	-9.63	-7.66	-6.60	-2.08	7.77
KM20	MSS	WT	WT	WT	p.V600E	WT	positive	-2213.87	0	-10.84	-7.28	-8.22	-10.24	-9.51	-14.25	-13.25	-5.98
LIM1215	MSI	WT	WT	WT	WT	WT	positive	12002.99	0	28.86	42.71	52.16	59.05	61.42	60.12	59.67	60.42
LIM1899	MSI	WT	WT	p.G12A	WT	WT	positive	1690.11	0	-4.91	-6.70	-2.68	0.47	1.17	7.36	7.25	12.86
LIM2099	MSS	WT	WT	p.G12C	WT	WT	positive	1168.03	0	-0.15	-0.34	1.78	3.01	13.04	4.45	4.74	6.98
LIM2405	MSI	WT	WT	WT	p.V600E	WT	null	362.68	0	8.06	5.78	8.34	5.01	-0.03	1.18	2.70	1.25
LIM2412	MSI	WT	WT	WT	p.V600E	WT	null	-445.10	0	-2.97	5.11	1.32	3.12	-0.73	-3.69	-1.19	-3.22
LIM2537	MSI	WT	WT	WT	p.V600E	WT	null	-823.98	0	-5.89	-9.36	-13.78	-10.99	-8.65	-11.34	-6.00	4.91
LIM2550	MSI	WT	WT	p.Q61K	WT	WT	positive	1506.88	0	9.91	18.70	19.36	10.05	14.77	11.95	5.43	5.95
LIM2551	MSI	WT	WT	WT	p.V600E	p.H1047R	positive	113.22	0	8.58	2.51	7.83	0.45	-2.97	-0.88	-1.37	5.26
LOVO	MSI	WT	WT	p.G13D	WT	WT	positive	2196.41	0	-4.03	-2.08	-0.13	-2.45	1.56	6.62	14.27	11.55
LS1034	MSS	WT	WT	p.A146T	WT	WT	positive	568.39	0	0.75	-1.04	0.45	1.88	2.17	3.70	0.98	5.22
LS123	MSS	WT	WT	p.G12S	WT	WT	positive	-1033.79	0	-1.74	-9.35	-10.32	-4.66	-7.76	-5.77	-4.80	-4.74
LS174T	MSI	WT	WT	p.G12D	WT	p.H1047R	positive	1675.79	0	2.26	5.40	0.37	7.94	7.77	8.93	9.72	8.18
LS180	MSI	WT	WT	p.G12D	WT	p.H1047R	positive	-820.39	0	-6.64	-2.21	-8.36	0.20	-10.51	-5.57	-4.68	-0.90
LS411N	MSI	WT	WT	WT	p.V600E	WT	null	-976.45	0	-0.23	-2.18	1.09	-4.00	-2.32	-4.96	-6.12	-3.59
LS513	MSS	WT	WT	p.G12D	WT	WT	positive	1716.36	0	-1.13	1.77	2.30	4.57	6.39	7.28	8.41	10.37
MDST8	MSS	WT	WT	WT	p.V600K	WT	null	1770.32	0	-6.19	-1.71	-4.06	-6.12	-4.81	4.91	7.89	16.46
MIP101	MSI	WT	WT	p.G13D	WT	p.E545K	positive	416.29	0	-0.40	-0.90	-0.08	-0.64	-0.33	2.04	2.38	2.27
NCIH498	MSS	WT	WT	p.Q61K	WT	p.E545K	positive	-560.10	0	0.58	1.70	-1.96	-0.54	-2.01	0.48	-4.65	-2.56
NCIH508	MSS	WT	WT	WT	WT	p.E545K	positive	16710.76	0	-2.45	4.06	40.08	80.64	85.36	82.92	83.42	84.31
NCIH630	MSI	WT	WT	WT	WT	WT	positive	1094.06	0	-0.89	0.61	2.66	2.43	4.87	4.68	4.97	7.01
NCIH684	MSS	WT	WT	p.G12D	WT	WT	positive	3521.58	0	-1.06	1.13	1.90	7.65	10.15	15.84	17.88	20.40
NCIH716	MSS	WT	WT	WT	WT	WT	positive	1460.05	0	3.94	8.72	-9.14	-7.77	0.44	-7.13	9.69	15.72
NCIH747	MSI	WT	WT	p.G13D	WT	WT	positive	-313.64	0	-11.78	-9.76	-12.24	-9.20	-13.02	-12.32	-1.20	8.11
OUMS23	MSS	WT	WT	WT	p.V600E	WT	null	1577.07	0	6.76	2.84	3.77	1.07	-3.04	0.88	13.81	6.49
OXCO1	MSS	WT	WT	WT	p.V600E	WT	positive	654.10	0	-7.83	-8.53	-7.11	-6.98	-6.16	-3.71	-0.44	16.06
OXCO2	MSI	WT	WT	WT	WT	WT	positive	17470.11	0	24.42	41.14	68.27	79.30	83.85	88.16	86.17	89.89
OXCO3	MSS	WT	WT	p.Q61H	WT	WT	positive	1076.96	0	1.22	4.51	2.97	4.54	8.96	5.30	2.01	9.76
RCM1	MSS	WT	WT	p.G12V	WT	WT	positive	3360.92	0	1.45	4.40	2.19	6.01	10.58	12.99	16.51	21.62
RKO	MSI	WT	WT	WT	p.V600E	p.H1047R	positive	99.57	0	6.51	3.66	0.72	8.36	6.16	3.31	0.36	-2.68
RW7213	MSS	WT	WT	p.G12C	WT	WT	positive	2993.63	0	3.00	7.59	7.65	11.64	13.50	14.73	15.81	14.34
SKCO1	MSS	WT	WT	p.G12V	WT	WT	positive	-898.14	0	-12.08	-10.70	-10.89	-7.41	-6.52	-7.74	-7.30	2.51
SNU1033	MSS	WT	WT	p.G12D	WT	WT	positive	610.88	0	-3.84	-3.08	0.08	2.06	4.51	3.03	1.21	5.55
SNU1040	MSI	WT	WT	WT	WT	WT	null	6091.38	0	17.50	12.67	21.65	22.59	27.56	21.95	33.65	32.57
SNU1047	MSI	WT	WT	WT	WT	WT	null	-1377.24	0	-5.39	-3.12	-6.53	-7.13	-8.55	-8.17	-6.48	-6.25
SNU1181	MSS	WT	WT	p.G12S	WT	WT	positive	2977.82	0	7.44	27.19	5.35	8.30	11.50	7.23	19.44	14.39
SNU1235	MSS	WT	WT	WT	p.V600E	WT	null	-2184.11	0	-0.14	-2.45	-0.45	0.50	9.24	-5.64	-11.81	-18.04
SNU1406	MSS	WT	WT	WT	WT	WT	positive										
SNU1411	MSS	WT	WT	p.G12C	WT	WT	positive	783.28	0	-2.76	-2.04	-3.30	-4.55	-2.16	0.28	4.83	6.65
SNU1460	MSS	WT	WT	WT	WT	WT	positive	12054.06	0	3.81	3.72	16.60	48.33	53.98	58.27	57.70	67.51
SNU1544	MSI	WT	WT	WT	WT	p.H1047R	positive	2526.72	0	2.52	2.34	0.93	0.04	2.48	7.12	13.29	18.10
SNU1684	MSI	WT	WT	p.G13D	WT	WT	null	7154.92	0	16.91	13.36	17.71	31.19	32.95	29.42	37.77	38.14
SNU1746	MSI	WT	WT	p.G13D	WT	WT	positive	-2525.53	0	11.65	17.58	1.73	0.42	-18.35	-18.02	-8.84	-13.65
SNU175	MSI	WT	WT	WT	WT	WT	positive	5176.65	0	2.05	9.80	14.31	20.70	22.47	24.77	25.53	28.18
SNU254	MSS	WT	WT	p.G12D	WT	p.E545A	positive	1417.81	0	-8.03	-2.93	-5.29	-1.91	2.18	-0.66	8.25	12.10
SNU283	MSS	WT	WT	WT	WT	WT	positive	10221.76	0	16.99	33.07	32.46	44.43	44.14	48.30	51.53	54.37
SNU407	MSI	WT	WT	p.G12D	WT	p.H1047R	positive	4252.95	0	-0.38	1.67	7.61	14.23	15.01	20.15	20.97	24.13
SNU479	MSS	WT	WT	WT	WT	WT	null	9244.47	0	26.90	14.59	45.59	39.19	49.04	37.70	53.56	40.90
SNU503	MSS	WT	WT	WT	WT	WT	positive	10216.08	0	12.32	22.79	29.75	37.89	43.18	46.83	54.14	51.75
SNU61	MSI	WT	WT	p.G12D	WT	WT	positive	2005.72	0	-2.60	-0.48	0.86	5.64	3.17	5.59	11.37	12.80
SNU769B	MSI	WT	WT	WT	WT	WT	positive	11134.12	0	4.44	4.00	4.44	4.06	33.93	55.25	58.14	58.52
SNU81	MSS	WT	WT	p.A146T	WT	WT	positive	659.04	0	-0.92	2.35	-3.83	-2.86	-1.10	-1.19	6.26	3.11
SNU977	MSS	WT	WT	WT	WT	WT	positive	2968.09	0	-3.18	0.57	-2.33	2.09	4.32	6.72	19.90	15.63
SNUC1	MSS	WT	WT	WT	WT	WT	positive	1651.74	0	2.20	5.74	7.67	5.36	5.73	3.37	8.68	11.71
SNUC2A	MSI	WT	WT	p.G12D	WT	WT	positive	-560.84	0	0.20	4.97	3.98	1.16	3.41	1.31	-2.73	-7.28
SNUC2B	MSI	WT	WT	p.G12D	WT	WT	positive	3201.17	0	-0.41	1.39	5.80	4.65	11.92	13.14	16.61	18.35
SNUC4	MSI	WT	WT	WT	WT	p.E545G	null	9286.57	0	-9.74	7.31	18.89	33.14	35.35	44.69	48.32	47.80
SNUC5	MSI	WT	WT	WT	p.V600E	p.H1047R	positive	542.73	0	-0.40	-1.84	-5.16	-0.83	-0.73	-3.68	4.94	4.72
SW1116	MSS	WT	WT	p.G12A	WT	WT	positive	228.80	0	-5.54	-5.10	-7.71	-3.47	-2.57	-1.16	2.00	2.42
SW1222	MSS	WT	WT	p.A146V	WT	WT	null	1308.71	0	0.10	3.23	1.35	5.25	5.34	6.47	7.09	6.11
SW1417	MSS	WT	WT	WT	p.V600E	WT	positive	1133.22	0	-0.65	0.19	-0.61	-0.58	-1.27	-1.94	8.32	8.71
SW1463	MSS	WT	WT	p.G12C	WT	WT	positive	2042.16	0	4.81	0.27	4.27	1.68	11.23	-0.38	13.19	13.18
SW403	MSS	WT	WT	p.G12V	WT	p.Q546K	positive	4077.36	0	-3.28	-1.41	4.33	7.69	11.45	17.40	18.75	27.30
SW48	MSI	WT	WT	WT	WT	WT	positive	1219.28	0	-5.68	-6.77	6.37	0.50	6.23	7.12	5.32	6.66
SW480	MSS	WT	WT	p.G12V	WT	WT	positive	1025.71	0	-4.82	1.63	2.54	3.80	1.70	5.60	3.42	8.18
SW620	MSS	WT	WT	p.G12V	WT	WT	positive	-362.57	0	0.84	-3.28	-6.86	-8.96	-9.77	-9.12	-0.25	2.88
SW837	MSS	WT	WT	p.G12C	WT	WT	positive	1859.80	0	0.03	2.69	5.42	8.45	10.55	11.82	7.90	9.43
SW948	MSS	WT	WT	p.Q61L	WT	p.E542K	positive	51.08	0	-0.91	0.65	0.87	1.08	0.06	-0.92	0.07	1.38
T84	MSS	WT	WT	p.G13D	WT	p.E542K	positive	2326.70	0	-1.11	-0.07	2.24	5.29	5.14	6.96	15.59	10.60
V411	MSS	WT	WT	p.G13D	WT	WT	positive	807.71	0	1.16	0.18	2.50	-1.90	2.74	2.84	4.46	4.68
V457	MSI	WT	WT	p.G12D	WT	WT	null	-399.99	0	5.71	3.31	4.41	2.24	-0.29	-4.57	-0.38	-3.12
V481	MSI	WT	WT	p.G12C	WT	p.Q546P	null	-2706.95	0	-3.90	-4.88	-7.07	-8.04	-10.72	-10.22	-13.27	-17.06
V703	MSI	WT	WT	p.G12D	WT	WT	positive	2291.68	0	0.35	4.37	6.91	6.51	7.85	9.04	11.93	13.41
VACO432	MSI	WT	WT	WT	p.V600E	WT	positive	1876.61	0	0.56	4.25	10.97	8.55	15.00	14.11	9.25	5.09
VACO5	MSI	WT	WT	WT	p.V600E	p.H1047R	positive	1544.19	0	7.12	3.34	1.73	4.74	6.94	6.55	7.99	8.42
VACO6	MSI	WT	WT	WT	p.V600E	WT	positive	3625.00	0	22.28	-1.05	11.78	16.16	20.34	14.28	18.95	19.21
WIDR	MSS	WT	WT	WT	p.V600E	WT	positive	-547.13	0	-6.43	-8.00	-6.56	-6.12	-5.45	-6.32	-4.01	2.36

Supplementary Data 2. Classification of 151 CRC cell lines into molecular subtypes according to five transcriptonal classifiers.
CRC cell lines from GSE59857 were assigned by the NTP algorithm to a molecular subtype for each of five different classifiers. For each classifier, in addition to the subtype class, the table reports the Benjamini-Hockberg (BH)-corrected FDR. In the BH.FDR columns, grey shading indicates low classification confidence (FDR > 0.2).

GEO ID	Cell Line	Sadanandam class	BH.FDR	Marisa class	BH.FDR	Budinska Class	BH.FDR	De Sousa E Melo class	BH.FDR	Roepman Class	BH.FDR
GSM1448103	COGA12	Enterocyte	0.000990	C6	0.001235	Subtype B	0.001794	CCS1	0.030964	B-Type	0.020127
GSM1448130	HDC54	Enterocyte	0.000990	C6	0.000751	Subtype B	0.001776	CCS1	0.002580	B-Type	0.003871
GSM1448175	DIFI	Enterocyte	0.000990	C6	0.000751	Subtype A	0.001794	CCS1	0.002580	B-Type	0.003871
GSM1448185	HDC143	Enterocyte	0.000990	C6	0.000751	Subtype B	0.185241	CCS3	0.401371	B-Type	0.252024
GSM1448199	C32	Enterocyte	0.000990	C6	0.000751	Subtype E	0.005636	CCS1	0.002580	B-Type	0.003871
GSM1448208	SNU479	Enterocyte	0.000990	C6	0.000751	Subtype A	0.001776	CCS1	0.002580	B-Type	0.006136
GSM1448215	NCIH684	Enterocyte	0.001973	C5	0.003195	Subtype C	0.001776	CCS1	0.142241	B-Type	0.006136
GSM1448110	V481	Enterocyte	0.002293	C3	0.000751	Subtype A	0.001510	CCS2	0.392312	A-Type	0.169639
GSM1448221	V411	Enterocyte	0.004915	C6	0.002475	Subtype B	0.201545	CCS1	0.006773	B-Type	0.779002
GSM1448205	SNU1684	Enterocyte	0.007631	C6	0.024364	Subtype D	0.234064	CCS2	0.268919	C-Type	0.061305
GSM1448157	SNU1411	Enterocyte	0.013285	C3	0.010619	Subtype E	0.762087	CCS3	0.910059	C-Type	0.992003
GSM1448154	FET	Enterocyte	0.016416	C6	0.206853	Subtype C	0.146926	CCS1	0.637107	A-Type	0.036089
GSM1448189	SNU1040	Enterocyte	0.024998	C6	0.002706	Subtype D	0.004473	CCS3	0.279400	C-Type	0.127088
GSM1448220	VACO6	Enterocyte	0.025762	C3	0.016990	Subtype C	0.001510	CCS2	0.008386	A-Type	0.003871
GSM1448151	COGA5L	Enterocyte	0.042272	C5	0.032518	Subtype B	0.210863	CCS1	0.428267	B-Type	0.838494
GSM1448080	HROC69	Enterocyte	0.048097	C6	0.017273	Subtype B	0.217986	CCS1	0.191832	B-Type	0.159829
GSM1448127	SNUC4	Enterocyte	0.059593	C2	0.001235	Subtype B	0.001794	CCS1	0.004328	A-Type	0.232881
GSM1448099	COGA1	Enterocyte	0.079418	C2	0.012915	Subtype E	0.036904	CCS2	0.021943	A-Type	0.004645
GSM1448111	HT115	Enterocyte	0.149772	C2	0.340635	Subtype C	0.001776	CCS2	0.088469	A-Type	0.059678
GSM1448203	C84	Enterocyte	0.712161	C1	0.006564	Subtype E	0.325500	CCS1	0.213236	B-Type	0.207267
GSM1448202	C80	Goblet	0.000936	C3	0.014100	Subtype B	0.002157	CCS1	0.040515	B-Type	0.110592
GSM1448070	HDC9	Goblet	0.000990	C3	0.000751	Subtype A	0.002922	CCS2	0.057591	C-Type	0.174684
GSM1448071	HROC24	Goblet	0.000990	C3	0.000751	Subtype C	0.001510	CCS2	0.005134	A-Type	0.003871
GSM1448087	SNUC2B	Goblet	0.000990	C2	0.000751	Subtype C	0.006194	CCS2	0.048453	A-Type	0.054044
GSM1448164	LS174T	Goblet	0.000990	C3	0.000751	Subtype A	0.001776	CCS2	0.859147	A-Type	0.044358
GSM1448193	SKCO1	Goblet	0.000990	C3	0.000751	Subtype A	0.458161	CCS1	0.123550	C-Type	0.839294
GSM1448206	SNU1746	Goblet	0.000990	C3	0.000751	Subtype C	0.001776	CCS2	0.005920	A-Type	0.784579
GSM1448216	SNU175	Goblet	0.000990	C3	0.000751	Subtype C	0.001794	CCS2	0.229650	A-Type	0.053998
GSM1448141	LS513	Goblet	0.001267	C3	0.000751	Subtype A	0.090861	CCS1	0.300701	B-Type	0.873060
GSM1448119	CL34	Goblet	0.006785	C2	0.000751	Subtype C	0.008933	CCS2	0.009710	A-Type	0.293298
GSM1448139	LS180	Goblet	0.009365	C2	0.002249	Subtype A	0.185241	CCS2	0.501343	C-Type	0.796206
GSM1448188	KM12SM	Goblet	0.009535	C2	0.011895	Subtype A	0.231325	CCS2	0.011940	A-Type	0.272147
GSM1448120	CX1	Goblet	0.010379	C3	0.116078	Subtype C	0.072921	CCS2	0.450579	A-Type	0.484628
GSM1448073	KM12	Goblet	0.010895	C3	0.006564	Subtype A	0.242989	CCS2	0.078334	A-Type	0.212048
GSM1448186	KM12C	Goblet	0.012324	C3	0.005730	Subtype A	0.086067	CCS2	0.303117	C-Type	0.337246
GSM1448094	COLO60H	Goblet	0.013752	C2	0.001006	Subtype B	0.016922	CCS1	0.302302	A-Type	0.010960
GSM1448159	SNU81	Goblet	0.019742	C3	0.025757	Subtype C	0.001776	CCS3	0.764465	A-Type	0.036089
GSM1448187	KM12L4	Goblet	0.090582	C2	0.005921	Subtype A	0.637688	CCS2	0.600230	A-Type	0.747456
GSM1448204	C99	Goblet	0.146863	C3	0.185111	Subtype A	0.150970	CCS1	0.152360	B-Type	0.708839
GSM1448217	SNU407	Goblet	0.161949	C2	0.000751	Subtype C	0.028037	CCS2	0.162952	A-Type	0.635107
GSM1448136	LIM1899	Goblet	0.162815	C1	0.517844	Subtype A	0.273869	CCS2	0.446369	B-Type	0.826927
GSM1448218	SNU769B	Goblet	0.292993	C3	0.029487	Subtype C	0.001794	CCS1	0.105451	A-Type	0.006136
GSM1448081	V457	Goblet	0.332352	C2	0.001006	Subtype A	0.487231	CCS2	0.026638	A-Type	0.033065
GSM1448181	C125PM	Goblet	0.349946	C5	0.001006	Subtype B	0.002922	CCS1	0.673820	B-Type	0.339217
GSM1448172	COLO205	Goblet	0.350034	C3	0.386958	Subtype E	0.491730	CCS2	0.876460	C-Type	0.497533
GSM1448179	HT29	Goblet	0.474598	C1	0.516591	Subtype C	0.001510	CCS3	0.073678	A-Type	0.424097
GSM1448198	C146	Goblet	0.683481	C2	0.114711	Subtype E	0.827113	CCS3	0.900633	C-Type	0.438027
GSM1448169	HCC2998	Goblet	0.709348	C2	0.720713	Subtype B	0.771782	CCS1	0.970676	A-Type	0.881046
GSM1448124	VACO432	Goblet	0.817685	C2	0.000751	Subtype C	0.051803	CCS2	0.005134	A-Type	0.072679
GSM1448135	LIM2537	Inflammatory	0.000936	C2	0.000729	Subtype C	0.001776	CCS2	0.004328	A-Type	0.014506
GSM1448149	CL40	Inflammatory	0.000936	C6	0.000751	Subtype A	0.455554	CCS1	0.106044	C-Type	0.966413
GSM1448069	HDC135	Inflammatory	0.000990	C2	0.000751	Subtype C	0.001776	CCS2	0.037737	A-Type	0.004645
GSM1448072	HROC87	Inflammatory	0.000990	C2	0.000751	Subtype C	0.002922	CCS2	0.009710	A-Type	0.059054
GSM1448086	SNU61	Inflammatory	0.000990	C6	0.012015	Subtype C	0.001776	CCS2	0.214830	A-Type	0.126305
GSM1448090	T84	Inflammatory	0.000990	C2	0.013625	Subtype C	0.001776	CCS1	0.002580	B-Type	0.049663
GSM1448095	HROC18	Inflammatory	0.000990	C1	0.002475	Subtype C	0.001776	CCS1	0.002580	B-Type	0.054090
GSM1448108	V703	Inflammatory	0.000990	C2	0.000751	Subtype E	0.448469	CCS1	0.213236	B-Type	0.445882
GSM1448118	LIM1215	Inflammatory	0.000990	C2	0.000751	Subtype C	0.001776	CCS2	0.002580	A-Type	0.103811
GSM1448138	LS123	Inflammatory	0.000990	C4	0.000751	Subtype C	0.001776	CCS3	0.134671	C-Type	0.015615
GSM1448147	OXCO3	Inflammatory	0.000990	C5	0.007669	Subtype C	0.001776	CCS3	0.073293	A-Type	0.058604
GSM1448158	SNU1460	Inflammatory	0.000990	C6	0.000729	Subtype C	0.002581	CCS1	0.002580	B-Type	0.059054
GSM1448195	SW837	Inflammatory	0.002593	C5	0.019594	Subtype C	0.001776	CCS3	0.572890	C-Type	0.401727
GSM1448091	WIDR	Inflammatory	0.003274	C2	0.024351	Subtype E	0.807536	CCS3	0.399412	A-Type	0.437125
GSM1448150	COGA5	Inflammatory	0.003907	C5	0.003195	Subtype C	0.001776	CCS3	0.268213	C-Type	0.383461
GSM1448131	SNUC5	Inflammatory	0.006519	C2	0.000751	Subtype D	0.771897	CCS2	0.073293	C-Type	0.082238
GSM1448190	SNU1235	Inflammatory	0.007045	C4	0.073046	Subtype D	0.150440	CCS3	0.218617	C-Type	0.032815
GSM1448168	LIM2550	Inflammatory	0.009365	C2	0.005536	Subtype C	0.001776	CCS2	0.002580	A-Type	0.003659
GSM1448121	VACO5	Inflammatory	0.012810	C2	0.145403	Subtype C	0.006194	CCS2	0.002580	A-Type	0.006136
GSM1448145	NCIH747	Inflammatory	0.022430	C4	0.174720	Subtype D	0.209861	CCS3	0.789212	C-Type	0.078209
GSM1448163	LOVO	Inflammatory	0.024497	C2	0.001735	Subtype E	0.051419	CCS2	0.089148	A-Type	0.818343
GSM1448076	LIM2551	Inflammatory	0.024998	C2	0.000751	Subtype C	0.001776	CCS2	0.015055	A-Type	0.003871
GSM1448074	KM20	Inflammatory	0.044482	C5	0.005162	Subtype C	0.001776	CCS2	0.031017	A-Type	0.062896
GSM1448184	HDC114	Inflammatory	0.068717	C2	0.152224	Subtype C	0.018831	CCS2	0.151899	A-Type	0.822306
GSM1448107	C170	Inflammatory	0.095037	C2	0.000751	Subtype D	0.264851	CCS2	0.073403	A-Type	0.421550
GSM1448067	CO115	Inflammatory	0.137821	C2	0.000751	Subtype C	0.008933	CCS2	0.002580	A-Type	0.003871
GSM1448075	LIM2412	Inflammatory	0.150026	C2	0.036495	Subtype E	0.636429	CCS3	0.670887	C-Type	0.204893
GSM1448174	COLO678	Inflammatory	0.319792	C4	0.054298	Subtype D	0.099640	CCS3	0.471947	B-Type	0.097957
GSM1448132	RKO	Inflammatory	0.723828	C2	0.002249	Subtype E	0.713542	CCS2	0.002580	C-Type	0.361216
GSM1448165	SW48	Inflammatory	0.748892	C2	0.030819	Subtype C	0.061840	CCS2	0.114785	A-Type	0.028970
GSM1448178	HCT15	Inflammatory	0.880083	C2	0.012357	Subtype C	0.324066	CCS2	0.002580	A-Type	0.334894
GSM1448116	MIP101	Inflammatory	0.885287	C2	0.018819	Subtype B	0.304046	CCS2	0.008386	A-Type	0.013977
GSM1448140	LS411N	Inflammatory	0.912018	C2	0.001235	Subtype B	0.697417	CCS2	0.002425	A-Type	0.449418
GSM1448089	SW1417	TA	0.000936	C5	0.000751	Subtype C	0.001510	CCS1	0.116897	B-Type	0.004645
GSM1448197	C106	TA	0.000936	C1	0.000751	Subtype E	0.095568	CCS1	0.002580	B-Type	0.029072
GSM1448079	HROC32	TA	0.000990	C1	0.000751	Subtype B	0.001510	CCS1	0.002580	B-Type	0.003871
GSM1448082	HT55	TA	0.000990	C1	0.000751	Subtype B	0.001510	CCS1	0.002580	B-Type	0.003871
GSM1448083	HROC39	TA	0.000990	C5	0.001006	Subtype B	0.005139	CCS1	0.002580	B-Type	0.041595
GSM1448088	SW1222	TA	0.000990	C1	0.001006	Subtype B	0.001776	CCS1	0.002580	B-Type	0.003871
GSM1448101	COGA3	TA	0.000990	C2	0.013625	Subtype B	0.156176	CCS1	0.002580	B-Type	0.003659
GSM1448109	HCA24	TA	0.000990	C3	0.009407	Subtype B	0.018051	CCS1	0.002425	A-Type	0.205840
GSM1448112	LS1034	TA	0.000990	C1	0.000751	Subtype B	0.173422	CCS1	0.002580	B-Type	0.006136
GSM1448113	SW1463	TA	0.000990	C5	0.000751	Subtype B	0.001510	CCS1	0.048406	B-Type	0.003871
GSM1448115	SNUC1	TA	0.000990	C5	0.000751	Subtype B	0.004473	CCS1	0.002580	B-Type	0.003871
GSM1448117	HRA16	TA	0.000990	C5	0.000751	Subtype B	0.006710	CCS1	0.002580	B-Type	0.003871
GSM1448128	HDC82	TA	0.000990	C5	0.000729	Subtype B	0.004473	CCS1	0.002580	B-Type	0.003871
GSM1448129	HDC142	TA	0.000990	C5	0.000751	Subtype B	0.002157	CCS1	0.002580	B-Type	0.003871
GSM1448143	NCIH630	TA	0.000990	C1	0.001471	Subtype B	0.006194	CCS1	0.003354	B-Type	0.003871
GSM1448146	OXCO2	TA	0.000990	C6	0.118519	Subtype B	0.001510	CCS1	0.002580	B-Type	0.003871
GSM1448153	COLO94H	TA	0.000990	C5	0.000751	Subtype B	0.001776	CCS1	0.002425	B-Type	0.004645
GSM1448156	RW7213	TA	0.000990	C3	0.000751	Subtype A	0.001776	CCS1	0.002580	B-Type	0.205028
GSM1448167	COCM1	TA	0.000990	C6	0.000751	Subtype B	0.001776	CCS1	0.002580	B-Type	0.003871
GSM1448171	CL14	TA	0.000990	C5	0.000729	Subtype A	0.380716	CCS1	0.002580	B-Type	0.005514
GSM1448176	GEO	TA	0.000990	C3	0.207590	Subtype B	0.005636	CCS1	0.027711	B-Type	0.053501
GSM1448177	HCA46	TA	0.000990	C1	0.000751	Subtype A	0.009479	CCS1	0.002580	B-Type	0.003871

GSM1448191	SNU1406	TA	0.000990	C5	0.000751	Subtype B	0.001776	CCS1	0.009710	B-Type	0.003871
GSM1448194	SW403	TA	0.000990	C5	0.000751	Subtype B	0.001776	CCS1	0.002580	B-Type	0.003871
GSM1448200	C70	TA	0.000990	C1	0.051278	Subtype B	0.001794	CCS1	0.002580	B-Type	0.017056
GSM1448201	C75	TA	0.000990	C1	0.231460	Subtype B	0.001794	CCS1	0.002580	B-Type	0.007575
GSM1448213	SW948	TA	0.000990	C5	0.000751	Subtype B	0.006416	CCS1	0.116036	B-Type	0.013291
GSM1448106	Gp5D	TA	0.001267	C2	0.010018	Subtype B	0.002922	CCS1	0.002580	A-Type	0.159254
GSM1448114	SW1116	TA	0.001267	C5	0.000751	Subtype B	0.149920	CCS1	0.002580	B-Type	0.005514
GSM1448219	SNU1181	TA	0.001267	C5	0.000751	Subtype E	0.872426	CCS1	0.447968	A-Type	0.330043
GSM1448123	SNU283	TA	0.001610	C1	0.019548	Subtype C	0.001510	CCS1	0.002580	B-Type	0.010375
GSM1448100	COGA2	TA	0.002293	C5	0.023526	Subtype C	0.022248	CCS2	0.143282	B-Type	0.023586
GSM1448097	CCK81	TA	0.003595	C1	0.018872	Subtype B	0.001776	CCS2	0.002425	B-Type	0.677054
GSM1448214	NCIH498	TA	0.008861	C5	0.000751	Subtype B	0.748810	CCS1	0.513663	B-Type	0.003871
GSM1448098	RCM1	TA	0.075741	C1	0.000751	Subtype E	0.400633	CCS1	0.270296	B-Type	0.025545
GSM1448085	SNU1047	TA	0.096473	C2	0.000751	Subtype D	0.020879	CCS2	0.030964	C-Type	0.034726
GSM1448068	COLO201	TA	0.149543	C5	0.000751	Subtype C	0.001776	CCS2	0.053385	C-Type	0.339531
GSM1448105	GP2D	TA	0.153310	C2	0.001471	Subtype B	0.013420	CCS1	0.013418	A-Type	0.158788
GSM1448142	NCIH508	TA	0.219985	C1	0.121106	Subtype A	0.005414	CCS1	0.002580	B-Type	0.791861
GSM1448212	SW620	TA	0.280308	C5	0.000751	Subtype D	0.005414	CCS3	0.460303	C-Type	0.022363
GSM1448084	HROC46	Stem	0.000936	C5	0.000751	Subtype D	0.060681	CCS1	0.160733	B-Type	0.010375
GSM1448160	CACO2	Stem	0.000936	C4	0.002002	Subtype D	0.002922	CCS1	0.270967	B-Type	0.003659
GSM1448166	SW480	Stem	0.000936	C4	0.000751	Subtype D	0.001510	CCS3	0.021943	C-Type	0.007575
GSM1448077	OUMS23	Stem	0.000990	C4	0.000751	Subtype D	0.001510	CCS3	0.008386	C-Type	0.003871
GSM1448078	SNU1033	Stem	0.000990	C4	0.008071	Subtype D	0.007686	CCS3	0.015055	C-Type	0.006709
GSM1448096	SNUC2A	Stem	0.000990	C4	0.000751	Subtype D	0.005139	CCS3	0.014727	C-Type	0.020454
GSM1448122	MDST8	Stem	0.000990	C4	0.000751	Subtype D	0.001776	CCS3	0.002580	C-Type	0.003871
GSM1448134	OXCO1	Stem	0.000990	C4	0.000751	Subtype D	0.001776	CCS3	0.002580	C-Type	0.006136
GSM1448137	LIM2099	Stem	0.000990	C4	0.000751	Subtype D	0.001776	CCS3	0.002425	C-Type	0.003871
GSM1448170	CAR1	Stem	0.000990	C4	0.000751	Subtype E	0.001794	CCS3	0.029471	C-Type	0.004645
GSM1448180	HUTU80	Stem	0.000990	C4	0.000751	Subtype E	0.011042	CCS3	0.086656	C-Type	0.006136
GSM1448196	C10	Stem	0.000990	C4	0.000751	Subtype D	0.001776	CCS3	0.002425	C-Type	0.020454
GSM1448207	SNU254	Stem	0.000990	C4	0.000751	Subtype D	0.005636	CCS3	0.047736	C-Type	0.010960
GSM1448209	SNU503	Stem	0.000990	C4	0.000751	Subtype D	0.002581	CCS3	0.079715	C-Type	0.015007
GSM1448211	SNU977	Stem	0.000990	C4	0.001006	Subtype D	0.003252	CCS3	0.116734	C-Type	0.014506
GSM1448102	COGA8	Stem	0.001267	C1	0.001235	Subtype D	0.011042	CCS3	0.048430	C-Type	0.029072
GSM1448173	COLO320DM	Stem	0.001267	C4	0.000751	Subtype D	0.001510	CCS3	0.005134	C-Type	0.291489
GSM1448144	NCIH716	Stem	0.001973	C4	0.005337	Subtype D	0.002157	CCS3	0.113874	C-Type	0.158788
GSM1448152	COLO320	Stem	0.002293	C4	0.000751	Subtype D	0.001776	CCS3	0.014727	C-Type	0.051607
GSM1448125	HCA7	Stem	0.002593	C2	0.000751	Subtype D	0.022645	CCS2	0.032419	C-Type	0.011612
GSM1448148	CL11	Stem	0.006247	C4	0.000729	Subtype D	0.007686	CCS3	0.032419	C-Type	0.038077
GSM1448104	HDC8	Stem	0.015689	C4	0.001006	Subtype C	0.132201	CCS3	0.270032	A-Type	0.159419
GSM1448093	HROC80	Stem	0.042435	C2	0.024357	Subtype C	0.023968	CCS3	0.462838	A-Type	0.204528
GSM1448162	HCT8	Stem	0.069052	C4	0.003407	Subtype D	0.011535	CCS2	0.189027	C-Type	0.003659
GSM1448182	COLO320HSR	Stem	0.072001	C4	0.000751	Subtype D	0.004473	CCS3	0.017030	C-Type	0.074093
GSM1448192	SNU1544	Stem	0.080769	C2	0.001235	Subtype D	0.301940	CCS2	0.301107	A-Type	0.164496
GSM1448092	HCT116	Stem	0.162033	C4	0.005162	Subtype D	0.067002	CCS3	0.115157	C-Type	0.004645
GSM1448161	DLD1	Stem	0.311486	C4	0.072692	Subtype D	0.445976	CCS3	0.138248	C-Type	0.450929
GSM1448126	LIM2405	Stem	0.370679	C2	0.001235	Subtype D	0.051330	CCS2	0.189831	C-Type	0.068408

Supplementary Data 3. Sanger sequencing primers's sequencees.
Table lists primers used to sequence 16 hot spot mutations in KRAS, NRAS, HRAS, BRAF and PIK3CA.

Gene/Exon	Primers (5'-> 3')	
KRAS EX2	FORW	GGTGGAGTATTTGATAGTGTATTAACC
KRAS EX2	REV	AGAATGGTCCTGCACCAGTAA
KRAS EX2	SEQ FOR	TCATTATTTTTATTATAAGGCCTGCTG
KRAS EX3	FORW	AAAGGTGCACTGTAATAATCCAGAC
KRAS EX3	REV	ATGCATGGCATTAGCAAAGA
KRAS EX3	SEQ FOR	CCAGACTGTGTTTCTCCCTTC
KRAS EX4	FORW + SEQ	TGGACAGGTTTTGAAAGATATTTG
KRAS EX4	REV	ATTAAGAAGCAATGCCCTCTCAAG
NRAS exon2	FORW + SEQ	GTACTGTAGATGTGGCTCGC
NRAS exon2	REV	AGAGACAGGATCAGGTCAGC
NRAS exon3	FOR	CTTATTTAACCTTGGCAATAGCA
NRAS exon3	REV + SEQ	GATTCAGAACACAAAGATCATCC
HRAS exon2	FORW	GGCAGGAGACCCTGTAGGA
HRAS exon2	REV + SEQ	AGCCCTATCCTGGCTGTGT
HRAS exon3	FORW + SEQ	AGAGGCTGGCTGTGTGAACT
HRAS exon3	REV	ATGCGCAGAGAGGACAGGA
BRAF ex15	FORW	TGCTTGCTCTGATAGGAAAATG
BRAF ex15	REV	AGCATCTCAGGGCCAAAAAT
BRAF ex15	SEQ FOR	TGTTTTCCTTTACTTACTACACCTCA
PIK3CA EX9	FORW	GGGAAAAATATGACAAAGAAAGC
PIK3CA EX9	REV	CTGCTTTATTTATTCCAATAGGTATGG
PIK3CA EX9	SEQ FOR	TAGCTAGAGACAATGAATTAAGGGAAA
PIK3CA EX20	FORW	CTCAATGATGCTTGGCTCTG
PIK3CA EX20	REV	TGGAATCCAGAGTGAGCTTTC
PIK3CA EX20	SEQ FOR	TTGATGACATTGCATACATTCG

Supplementary Data 4. siRNA sequencees.
For each target gene a pool of 4 siRNA was used. The table lists the sequencees of each siRNA included in the corresponding pool.

siRNA Target	siRna Sequence (5'->3')
ALK	GACAAGAUCCUGCAGAAUA
ALK	GGAAGAGUCUGGCAGUUGA
ALK	GCACGUGGCUCGGGACAUU
ALK	GGUCAUAGCUCCUUGGAAU
FGFR2	CCAAAUCUCUCAACCAGAA
FGFR2	GAACAGUAUUCACCUAGUU
FGFR2	GGCCAACACUGUCAAGUUU
FGFR2	GUGAAGAUGUUGAAAGAUG
KIT	AAACACGGCUUAAGCAAUU
KIT	GAACAGAACCUUCACUGAU
KIT	GGGAAGCCCUCAUGUCUGA
KIT	GCAAUUCCAUUUAUGUGUU
NTRK1	GGACAACCCUUUCGAGUUC
NTRK1	CCAGUGACCUCAACAGGAA
NTRK1	CCACAAUACUUCAGUGAUG
NTRK1	GAAGAGUGGUCUCCGUUUC
NTRK2	GAACAGAAGUAAUGAAAUC
NTRK2	GUAAUGCUGUUUCUGCUUA
NTRK2	GCAAGACACUCCAAGUUUG
NTRK2	GAAAGUCUAUCACAUUAUC
PDGFRA	CAUCAGAGCUGGAUCUAGA
PDGFRA	GGCCUUACUUUAUUGGAUU
PDGFRA	GAGCUUCACCUAUCAAGUU
PDGFRA	CCUCUAGGAAUGACGGAUU
RET	GCAAAGACCUGGAGAAGAU
RET	GCACACGGCUGCAUGAGAA
RET	UUAAAUGGAUGGCAAUUGA
RET	GCGAGGAGAUGUACCGCCU