bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genome-wide chemical mutagenesis screens allow unbiased saturation of the cancer genome and identification of drug resistance mutations

Jonathan Brammeld[1], Mia Petljak[1], Luz Garcia Alonso[2], Alba Dalmases[3], Beatriz Bellosillo[3], Carla Daniela Robles-Espinoza[1], Stacey Price[1], Syd Barthorpe[1], Patrick Tarpey[1], Constantine Alifrangis[1], Graham Bignell[1], Joana Vidal[4], Jamie Young[1], Lucy Stebbings[1], Kathryn Beal[1], Michael R Stratton[1], Julio Saez-Rodriguez[2,5], Mathew Garnett[1], Clara Montagut[4], Francesco Iorio[2,6*], Ultan McDermott[1,6*]

[1] Wellcome Trust Sanger Institute, Hinxton, UK

[2] European Molecular Biology Laboratory – European Bioinformatics Institute, Cambridge, UK

[3] Pathology Department, Hospital del Mar, Barcelona, Spain

[4] Cancer Research Program, FIMIM and Medical Oncology Department, Hospital del Mar, Barcelona, Spain

[5] WTH Aachen University Hospital, Aachen, Germany

[6] Co-senior author

* Co-corresponding author

Corresponding authors: Ultan McDermott: um1@sanger.ac.uk; Francesco Iorio: iorio@ebi.ac.uk

Keywords: mutagenesis, ENU, cancer, drug resistance

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Drug resistance is an almost inevitable consequence of cancer therapy and ultimately proves fatal for the majority of patients. In many cases this is the consequence of specific gene mutations that have the potential to be targeted and re-sensitize the tumor. The means therefore to saturate the genome with point mutations and that avoids chromosome or nucleotide sequence context bias would open the door to identify all possible drug resistance mutations in cancer models. Here we describe such a method for elucidating drug resistance mechanisms using genome-wide chemical mutagenesis allied to next-generation sequencing. We show that chemically mutagenizing the genome of cancer cells dramatically increases the number of drug-resistant clones and allows the detection of both known and novel drug resistance mutations. We have developed an efficient computational process that allows for the rapid identification of involved pathways and druggable targets. Such *a priori* knowledge would greatly empower serial monitoring strategies for drug resistance in the clinic as well as the development of trials for drug resistant patients.

Introduction

Despite an increasing array of new cancer therapies, drug resistance is an almost universal phenomenon that is likely due to the presence of rare subclonal populations that act as a reservoir for resistance mutations. The emergence of drug resistance ultimately proves fatal for the majority of patients, and therefore the early detection of resistance and the identification of novel re-sensitization strategies is a subject of intense activity.

Previously, the identification of drug resistance genes has relied on either re-biopsy of cancer patients following the development of resistance or the use of cancer cell lines made resistant by exposure *in vitro* to drug over many weeks. Both approaches can suffer from inherent biases. With respect to the former, biopsy of a single resistant site of disease may miss alternate resistance mechanisms in other metastatic sites (Van Allen et al. 2014). Equally, serial drug exposure in cancer cell lines will favor pre-existing drug resistant clones that are specific for that cell line and may not represent the entire spectrum of resistance mechanisms for that treatment.

For these reasons, there is considerable interest in the use of forward genetic screens capable of engineering into the cancer genome mutational events that can be tested for their ability to cause drug resistance in an unbiased fashion. Such screens, if sufficiently unbiased, could in theory capture the entire breadth of genetic resistance mechanisms for any drug. Recent studies have demonstrated the power of both genome-wide gain- and loss-of-function screens using CRISPR/Cas9, lentiviral shRNA and large-scale open-reading frame technologies to identify clinically relevant drug resistance mechanisms in cancer (Hu and Zhang 2016).

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

However, they all fail to capture a third important mechanism of drug resistance, namely that of point mutations. Point mutations account for resistance in large numbers of patients receiving targeted therapies in melanoma, colon and lung cancers and chronic myeloid leukemia (**Supplementary Table 1**) (Kobayashi et al. 2005; Katayama et al. 2012; Montagut et al. 2012; Ohashi et al. 2012; Bettegowda 2014; Long et al. 2014; Van Allen et al. 2014; Wagle et al. 2014; Arena et al. 2015; Russo et al. 2015; Siravegna et al. 2015; Thress et al. 2015).

N-Ethyl-N-nitrosourea (ENU) has been used as a potent mutagen in mouse models of development for over four decades (Acevedo-Arozena et al. 2008). Exposure results in the efficient generation of random point mutations throughout the cell genome (Tokunaga et al. 2014). We therefore tested whether, in cancer cell line models, ENU could be used to mutagenize the genome and enable expansion of drug resistant cells following the application of a targeted agent. As proof of concept, we chose to investigate whether this approach could identify all clinically demonstrated resistance mutations in colorectal cancer patients treated with the EGFR monoclonal antibody Cetuximab (Van Cutsem et al. 2009). We developed a sequencing and informatics approach to detect novel resistance mutations from next-generation sequence data and to detect statistical enrichment for mutually exclusive mutations in specific signalling pathways comprising more than 8,000 genes at the sample population level (**Figure 1a**). Our mutagenesis screen was able to successfully identify all drug resistance mutations to Cetuximab previously observed in the clinic as well as a novel mutation that we subsequently identified in a colorectal cancer patient. We suggest that this approach is a powerful and facile means to draw the landscape of point mutations that confer resistance to targeted therapies. Such knowledge could be used to discover therapeutic strategies to re-sensitize resistant tumors as well as identify which genes should be prioritised for non-invasive monitoring during treatment using plasma DNA sequencing.

Results

ENU exposure confers stable resistance to Cetuximab in colon cancer cells

We screened 51 colorectal cancer cell lines with a concentration range of the EGFR monoclonal antibody Cetuximab and assessed viability after 6 days (**Figure 1b**). In keeping with clinical experience of the genetic factors that underpin response to this drug, those cell lines wild-type for *KRAS/NRAS/BRAF* (green bars) exhibited heightened sensitivity to Cetuximab (Douillard et al. 2013). We therefore chose two of these lines, CCK-81 and NCI-H508, to use in the ENU resistance experiment. Both cell lines additionally demonstrated Cetuximab sensitivity in long-term clonogenic survival assays (**Supplementary Figure 1a, b**). Moreover, CCK-81 has features of microsatellite instability (MSI) whereas NCI-H508 is microsatellite stable (MSS). MSI

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is detected in 16% of colorectal cancers and is associated with a different phenotype and clinical outcome compared to MSS cancers. The CCK-81 cell line was exposed to a dose range of ENU (0.01-1mg/ml) for 24 hours, following which the mutagenized cells were treated with Cetuximab (10μg/ml) for 8 consecutive weeks. The number of drug resistant colonies was counted at the end of the experiment. We observed no drug resistant colonies in the absence of ENU (**Figure 1c**). With increasing ENU concentration we observed a linear increase in both the number of drug resistant colonies (left y-axis, blue bars) as well as their mutational burden (right y axis, green triangles). Subsequently, NCI-H508 cells were also treated with ENU (0.1mg/ml) for 24 hours followed by weekly Cetuximab treatment for 8 weeks. Drug resistant colonies were picked, expanded in culture and submitted for whole exome Illumina sequencing (total of 14 CCK-81 and 58 NCI-H508 colonies). Data was analysed for substitutions and insertions/deletions to enable an estimation of the number of ENU-associated mutations per Mb of exome and to detect novel (and putative drug resistance) mutations (**Supplementary Table 2**). We then performed clonogenic survival assays on a subset of resistant clones and confirmed robust and stable resistance to Cetuximab (**Figure 1d**).

The spectrum of ENU-induced mutations

The ability of any mutagenesis screen to capture a particular phenotype is strongly dependent on its ability to evenly saturate the genome with all 6 possible classes of base substitution type (expressed as the pyrimidine of a mutated Watson–Crick base pair, C>A, C>G, C>T, T>A, T>C, T>G). On average, we detected 451 novel mutations per exome in each clone (mean 550 and 199 in CCK-81 and NCI-H508 clones respectively), for a total of 33,857 (**Supplementary Table 2**). The mutations were almost exclusively composed of base substitutions (96% of total). A third of such mutations were non-synonymous variants within the coding exon of a gene where resistance mutations are more likely to occur (**Figure 2a**). Only 4% were potential loss-of-function truncating mutations (frameshift indels or nonsense mutations). The remaining mutations were either silent or intronic. Significantly, analysis of exome sequence data across all 72 clones (regardless of whether MSS or MSI) revealed that of the 6 possible classes of base substitution, only C>G substitutions are less well represented (3% of all substitution base changes) (**Figure 2a**). The mutation spectrum in ENU-derived clones was similar regardless of whether the cells came from a MSI or MSS background (**Supplementary Figure 2**). There was no evidence of a significant bias towards mutations in coding genes in any particular chromosome or indeed any specific region within a chromosome (**Figure 2b and Supplementary Figure 3**). However, given these 33,857 mutations are a mixture of those caused by ENU, background mutational processes and private subclonal

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

variants, we elected to use a mathematical approach to specifically extract the ENU signature from the data and more accurately determine the mutation spectrum of ENU mutations.

The non-negative matrix factorization algorithm has previously been used to detect the presence of mutational signatures in human cancers, including from defects in DNA mismatch repair and altered activity of the error-prone polymerase *POLE* (http://cancer.sanger.ac.uk/cosmic/signatures) (Alexandrov et al. 2013b). It extracts signatures based on a 96-mutation classification that incorporates the 6 base substitution types described above as well as the immediate flanking sequence context of the mutated base (four possible 5' and four possible 3' bases). In our data it revealed a distinct and unique signature that was represented across almost all trinucleotide contexts in both CCK-81 and NCI-H508 clones and not previously detected in prior tumor studies, including a panel of 51 colorectal cancer cell lines (data not shown) (**Figure 2c**, **Supplementary Figure 4**). This signature ('Signature A') is likely one of ENU exposure. Reassuringly, the pattern of base substitutions that comprise this signature was almost identical to that seen across the entire set of substitutions detected in the ENU-derived clones, with again only C>G substitutions seen at lower frequency (**Figure 2c**). Thus, using this approach it should be feasible to generate the majority of theoretical coding point mutations for drug resistance across the entire genome.

As expected we detected a signature of MSI ('Signature B') in the combined mutational catalogue for the CCK-81 clones but surprisingly also in the NCI-H508 clones (**Supplementary Figure 4**) (Alexandrov et al. 2013a). On closer examination, this signature was the result of two hypermutator clones in the NCI-H508 mutational catalogue (red arrows) (clones NCI-H508_26 and NCI-H508_40) (**Supplementary Figure 5**). These clones had mutation rates as high as any of the MSI CCK-81 clones and increased numbers of small insertions and deletions. This would be in keeping with a defect in the mismatch repair pathway (**Supplementary Table 2**). Clone NCI-H508_26 was found to harbor a novel nonsense (stop-gained) mutation in the mismatch repair gene *MLH1* and clone NCI-H508_40 a nonsense mutation in the DNA repair gene *EXO1*. Two other clones (black arrows) also have elevated mutation rates that may be the result of gaining nonsense mutations in *POLQ*, a gene involved in DNA damage repair. These gave rise to the third signature of unknown origin detected in NCI-H508 clones ('Signature C') (**Supplementary Figure 4**).

ENU mutagenesis identifies clinically relevant resistance mutations and pathways

A challenge in the identification of drug resistance mutations in ENU-derived clones is that each clone harbors many hundreds of 'passenger' mutations in addition to that conferring resistance. We hypothesised that with a sufficient population of individual resistant clones it might become feasible to use statistical enrichment for non-synonymous coding mutations in

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

specific pathways to help identify drug resistance mutations. We therefore developed a statistical framework (SLAP-Enrich) to identify whether genetic alterations observed in multiple samples are enriched within a specific pathway in a statistically significant manner using a network of 8,056 unique genes (https://github.com/francescojm/SLAPenrich). Once significantly enriched pathways are identified SLAP-Enrich applies a final filter based on the tendency for genes in a positively selected pathway to be mutated in a mutually exclusive manner. When applying this method to the set of ENU-mutations across the 72 Cetuximab-resistant clones, we found several statistically enriched pathways (false discovery rate (FDR) < 5%) (**Supplementary Table 3**). The pathway most significantly enriched with mutations, 'Signalling to P38 via RIT and RIN', contains many of the key genes of the canonical MAP kinase pathway (**Figure 3a**). In total, we were able to identify credible resistance mutations in 42 of the 72 resistant clones (59%) (**Supplementary Table 4**). We detected credible resistance mutations in all of the genes found clinically to confer resistance to EGFR therapy in colorectal cancer (**Supplementary Table 1**). *EGFR, KRAS, NRAS, BRAF* and *MAP2K1 (MEK1)* were each found to be mutated in ≥3 clones and in a mutually exclusive manner. There was no clear difference in the frequency of specific mutations between NCI-H508 and CCK-81. Furthermore, 38/42 (90%) of these putative resistance mutations have previously been identified in colorectal patients developing resistance to Cetuximab. The most frequently observed ENU resistance mutation was that of *BRAF* p.V600E (13/42 clones), followed by *NRAS* p.Q61K (8/42) and *KRAS* p.G12C (4/42) (**Figure 3b,c and Supplementary Figure 6**). These mutations are all canonical driver mutations in tumorigenesis and known to activate oncogenic signalling and confer resistance to Cetuximab both experimentally and clinically (Diaz et al. 2012; Misale et al. 2012). We also detected *EGFR* mutations in 3 of the resistant clones. The *EGFR* I491K substitution has been shown to induce structural changes to the extracellular domain of EGFR such as to prevent Cetuximab binding and confer resistance (Montagut et al. 2012). Of the remaining 30 clones where SLAP-Enrich was unable to detect statistical enrichment of mutations in specific pathways, there are two plausible explanations – one is that these mutations do not cluster into previously characterized pathways and are therefore not detectable by SLAP-Enrich, the other is that the observed resistance is the result of mutations outside of the coding exome (for example, in enhancer/promoter or untranslated (UTR) regions) and therefore not amenable to detection using this whole exome capture approach.

ENU mutagenesis identifies novel resistance mutations in MAP2K1

Recently, two studies of plasma DNA sequencing in colorectal cancer patients undergoing treatment with EGFR monoclonal antibodies jointly identified the first *MAP2K1* (*MEK1*) codon K57 resistance mutations (p.K57T and p.K57N) (Russo et al. 2015; Siravegna et

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

al. 2015). In our study we also identified *MAP2K1* mutations at the K57 codon (p.K57N, p.K57E) as well as at two sites not previously reported (p.F53L, p.C121S) (**Supplementary Table 4**). We therefore sequenced these *MAP2K1* loci (together with additional mutation hotspots in 34 other genes) in a series of plasma DNA samples collected from 22 colorectal cancer patients who acquired resistance to treatment with EGFR therapies (either Cetuximab or Panitumumab) after an initial response. In addition to many of the known canonical resistance mutations (in *KRAS, NRAS* and *BRAF*), we detected in one such patient a novel p.F53L *MAP2K1* mutation predicted by our screen to be a resistance mutation (**Table 1, Figure 4a**). As previously reported, we detected in a number of these patients more than one likely resistance mutation, in keeping with different metastatic sites evolving different resistance mechanisms.

To functionally validate the resistance effects of these *MAP2K1* mutations, we treated CCK-81 cells expressing the novel p.F53L and p.C121S mutations as well as the previously identified p.K57N mutation with Cetuximab (alongside empty vector and wild-type *MAP2K1* controls). We found that all of our candidate resistance mutations induced resistance to Cetuximab and the strength of the resistance effect for the mutations was comparable to that conferred by overexpression of the MET receptor tyrosine kinase, a previously identified resistance mechanism (**Figure 4b, left panel**) (Bardelli et al. 2013). Long-term growth inhibition assays similarly showed robust and durable resistance to Cetuximab in the *MAP2K1* mutant cells (**Figure 4b, right panel**). Immunoblot analysis demonstrated failure to completely suppress pERK1/2 expression following Cetuximab treatment in all of the MAP2K1 clones (**Figure 4c**).

Rational targeting of pathways can re-sensitize drug resistant mutants to Cetuximab

Identification of the key signalling pathways that underpin drug resistance opens up the possibility of rationally targeting key components of such pathways and thus re-sensitizing resistant cells. The creation of mutagenized resistant cell lines, either through the ENU screen or through the deliberate genetic modification of the parental cell line for specific mutations, allowed us the opportunity for such experiments to be carried out *in vitro*. The pathway most frequently mutated in the drug resistant CCK-81 and NCI-H508 ENU clones converges towards MAPK family members and targeting these nodes might be expected to overcome resistance (**Figure 3c**). A Cetuximab-resistant CCK-81 *BRAF* V600E mutant clone (ENU-10) was re-sensitized when the monoclonal antibody was combined with the BRAF inhibitor Dabrafenib (**Supplementary Figure 7a**). Cetuximab-resistant clones harbouring mutations (in *KRAS, NRAS* and *MAP2K1*) that would be predicted to activate MAPK signalling were re-sensitized when a MEK inhibitor (Trametinib) was combined with Cetuximab (**Supplementary Figure 7a,b**). Similarly, combining Cetuximab with Trametinib almost completely re-sensitized the resistant

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MAP2K1 mutant CCK-81 cells (**Figure 5a, b**). Indeed, such a combination has already been suggested as putative therapeutic strategy for colon cancer patients (Misale et al. 2014).

Discussion

Drug resistance is an almost inevitable feature of anticancer treatment and is ultimately fatal in the majority of patients. Subclonal mutations, present in a small subset of tumor cells prior to drug treatment, can serve as a reservoir for the emergence of drug resistance. The population of resistant cells is likely to be at very low frequency because resistance comes with a fitness cost. Such cells are naturally subject to a process of Darwinian selection following the treatment of any cancer cell population, such that we only detect the emergence of resistant clones following treatment. The emergence of plasma DNA sequencing to allow longitudinal samples to be assayed during cancer treatments makes it ideal to detect such cancer evolution (Russo et al. 2015). In theory, therefore, this approach would avoid the bias of single organ biopsy and would be eminently more feasible to execute across large clinical cohorts.

Regardless of what technology is used to detect resistance mutations, some *a priori* knowledge of the likely drug resistance candidates greatly increases the sensitivity of such assays. Identifying the complete catalogue of drug resistance effectors to any drug requires *in vitro* studies that model resistance in the relevant tissue and genetic background. In the past such *in vitro* studies featured cell lines that had undergone serial passage in the presence of the candidate drug in order to force the emergence of resistant clones. Although such studies have successfully identified clinically relevant mechanisms of drug resistance in some instances, they are biased towards mainly selecting for those pre-existing resistant subclones that are particular to that specific cell line.

Forward genetic screens enable an unbiased approach to be taken in the definition of genetic events that cause drug resistance. At the forefront of such screens have been large-scale lentiviral shRNA or CRISPR/Cas9 libraries that repress gene expression (Berns et al. 2004; Konermann et al. 2015) or conversely open reading-frame libraries that overexpress full length gene transcripts (Yang et al. 2011). Such screens have been recently used to define the resistance effectors to ALK inhibition in lung cancer (Wilson et al. 2015) and to BRAF inhibition in melanoma (Shalem et al. 2014; Wilson et al. 2015). An advantage of such screens is the identification/prioritization of candidate genes for subsequent investigation in clinical samples (either tissue biopsies or plasma DNA) and the ability to model the means to re-sensitize resistant cells using combinatorial drug strategies (Misale et al. 2014). However, although such unbiased screens present us with an extraordinary opportunity to map the entire landscape of genes that when repressed or overexpressed confer drug resistance, they fail to capture a third important class of drug resistance drivers, namely point mutations. In almost every analysis of

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

drug resistant disease across a range of tumor types, including lung cancer, melanoma, colon cancer and chronic myeloid leukemia, the landscape of resistance effectors includes significant numbers of genes that confer resistance exclusively through the acquisition of point mutations (**Supplementary Table 1**). Indeed, in colorectal cancer, it is increasingly evident that *KRAS*/*NRAS*/*BRAF*/*EGFR* point mutations collectively account for the majority (up to 90%) of both acquired and intrinsic resistance to EGFR therapies (Bettegowda 2014) (Douillard et al. 2013).

Genome-wide ENU chemical mutagenesis and subsequent phenotype-driven screening has been pivotal to a complete understanding of how complex biological processes operate in classical model organisms including yeast (Forsburg 2001), flies (St Johnston 2002), zebrafish (Patton and Zon 2001), and, perhaps most extensively, mice (Kile and Hilton 2005). The alkylating agent *N*-ethyl-*N*-nitrosourea (ENU) can introduce a high rate of point mutations into the genome and has two distinct advantages over previously used mutagens. First, it is very efficient, inducing a point mutation every 1 to 2 Mb throughout the genome in mouse models (~100-fold higher than the spontaneous mutation rate and 3-fold higher than X-irradiation) (Concepcion et al. 2004). Second, unlike irradiation, which induces multi-locus deletions, ENU is a point mutagen and affects single loci. ENU functions by the transfer of its ethyl group to Oxygen or Nitrogen atoms in DNA, resulting in misidentification of these ethylated bases during replication. If the mismatch is not repaired a base-pair substitution results (Justice 2000). To date, the use of ENU to define drug resistance mechanisms in cancer has been focused on specific genes in non-cancer cell line models rather than to interrogate the entire coding genome (Tiedt et al. 2011; Zhang et al. 2011; Ercan et al. 2015).

Here we establish a model for the use of genome-wide chemical mutagenesis screens to capture the diversity of clinically relevant drug resistance mutations. As proof of concept, we employed this screen in the setting of an EGFR therapy and colorectal cancer, a disease in which response to such therapy is invariably followed by the acquisition of resistance. Such resistance mechanisms are dominated by point mutations in the MAP kinase signalling pathway and have been extensively validated in patient cohorts (**Supplementary Table 1**) (Yonesaka et al. 2011; Diaz et al. 2012; Misale et al. 2012; Montagut et al. 2012; Bardelli et al. 2013; Bettegowda 2014). We are able to identify all clinically detected resistance mutations to Cetuximab treatment in colorectal cancer, and in addition potential therapeutic avenues to re-sensitize resistant cells. We propose that ENU mutagenesis should be incorporated alongside newer CRISPR gene editing technologies in the systematic interrogation of drug resistance given the prevalence (and potential for therapeutic targeting) of point mutations as mediators of resistance in cancer.

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Materials.

All cell culture was performed in either RPMI or DMEM/F12 medium (according to the supplier's recommendations) and supplemented with 5% FBS and penicillin/streptavidin. Cells were maintained at 37°C and 5% CO_2 during culture. The identity of all cell lines used in this paper was confirmed using a panel of 95 single nucleotide polymorphisms (SNPs) used previously for cell line authentication (Fluidigm, San Francesco, CA).

Immunoblotting.

Differential phosphorylation of proteins in signalling pathways were analysed by western blot. Cells were plated 24 hours prior to drug treatment and incubated for indicated times and concentrations. Adherent cells were then washed with PBS and collected after indicated incubation time with drug using lysis buffer containing 5% β-mercaptoethanol, 150mM NaCl, 50nM Tris pH 7.5, 2nM EDTA pH 8, 25nM NaF, 1% NP-40, protease inhibitors (Roche), phosphatase inhibitors (Roche). Lysates were then normalised after bicinchoninic acid (BCA) assay using lysis buffer. Protein lysates were resolved using SDS page electrophoresis in pre-cast Invitrogen 4-12% Bis-Tris gels and transferred for 12 hours. Primary antibodies: p44/42 MAPK, Phospho-p44/42 MAPK (Thr202/Tyr204) and Akt were sourced from Cell Signalling and phospho-Akt (pS473) was sourced from Invitrogen. Monoclonal β-tubulin was sourced from Sigma (USA).

Drug sensitivity assays.

Cells were seeded in 96-well plates for 6-day assays and 6-well plates for clonogenic assays. Cells were incubated in drug free media to allow for adherence for 24 hours before the addition of drug at indicated concentrations. Each cell line was seeded to achieve approx. 70% confluency at the end of the assay. Cetuximab was obtained from the Addenbrookes' Hospital Pharmacy, Cambridge (UK). Trametinib (GSK1120212) and Dabrafenib were obtained from Selleckchem (USA).

ENU mutagenesis of cell lines.

Cells were incubated in ENU at the indicated concentration for 24 hours before being washed 3 times with PBS and incubated in media for a further 24 hours. Cells were then selected with 10μg/ml Cetuximab 48 hours post-ENU exposure for 8 weeks. Clones were then picked using Scienceware small cloning cylinders (Wayne, NJ) and either transferred to 96 well plates or expanded into large flasks for drug sensitivity assays. DNA was extracted in 96-well plate format using the Agencourt DNAdvance Genomic DNA Isolation kit.

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Whole Exome Sequencing.

Exome sequencing was carried out using the Agilent SureSelectXT Human All Exon 50Mb bait set. 72 clones were DNA extracted and underwent library construction, flow cell preparation and cluster generation according to the Illumina library preparation protocol. We performed 75-base paired-end Illumina sequencing. Read alignment to the reference human genome (GRCh37) was performed using the Burrows-Wheeler Aligner (BWA) (http://bio-bwa.sourceforge.net/). Unmapped reads were excluded from the analysis. The average coverage across CCK-81- and NCI-508-derived clones was 65X and 62X respectively. The matched parental cell lines were sequenced at greater depth (158X in CCK-81 and 144X in NCI-H508).

Variant Detection.

Single nucleotide substitutions were called using the CaVEMan C (Cancer Variants through Expectations Maximisation) algorithm and insertions/deletions were called using split-read mapping implemented in the Pindel algorithm (https://github.com/cancerit). Variants were identified by comparison to a reference single matched sample consisting of a high sequence coverage contemporary parental cell line control.

Data Filtering To Remove Pre-existing Subclonal Variants.

A number of clones shared mutations which were present in a small percentage of reads in their corresponding contemporary parental cell line sequence. These subclonal mutations could confound subsequent pathway analysis by causing enrichment in a pathway due to mutations which were present before ENU treatment but were not called due to their low representation. To overcome this problem, variants were filtered against the deep sequenced contemporary parental control after mutation calling via Caveman and Pindel. The Samtools mpileup algorithm was used to remove any mutations which were present in 0.5% or more reads in the high coverage parental cell line control (http://samtools.sourceforge.net/). The final set of mutations were used to generate an event matrix for all 72 clones (**Supp. Table 5**) and used as the input file for SLAP-Enrich analysis described below.

Deciphering mutational signatures branding exome sequences of clones exposed to ENU.

The immediate 5′ and 3′ sequence context of base substitutions identified across Cetuximab-resistant clones was extracted using the ENSEMBL Core APIs for human genome build GRCh37 and was used to generate mutational catalogues for the downstream analysis. The mutational catalogue of CCK-81 Cetuximab-resistant clones contained 7,198 substitutions, while the NCI-H508 clones contained a total of 23,862 substitutions. Mutational signatures were deciphered

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

separately across both catalogues of mutations, using a previously developed computational framework (Alexandrov et al. 2013b). Briefly, the algorithm identifies a minimal set of mutational signatures that optimally explains the proportions of mutation types found across a given mutational catalogue (i.e. across all substitutions identified in CCK-81 and NCI-H508 clones; **Supplementary Figure 4**); and then estimates the contribution of each identified signature to a mutation spectra of each sample included in analysis (i.e. to a mutation spectra of each individual clone; see for NCI-H508 clones - **Supplementary Figure 5**).

Sample Level Analysis of Pathway Enrichments (SLAP-Enrich).

As a first step, SLAP-Enrich estimates the probability of observing at least one gene belonging to a given pathway mutated in a given sample, based on the length of the total exon blocks of the genes in that pathway, and the sample mutation burden. Once this probability has been estimated for each individual sample, SLAP-Enrich models the likelihood of observing a given number of samples with mutations in the pathway under consideration through a Poisson binomial distribution. This is the discrete distribution of a sum of Bernoulli trials in which the probability of success is not constant. It is used by SLAP-Enrich to compute the deviance of the number of observed samples with mutations in a given pathway from its expectation through a corresponding p-value assignment.

Pathway gene-sets collection and post-processing.

A collection of pathway gene sets was downloaded from the Pathway Commons data portal (v4-201311) (http://www.pathwaycommons.org/archives/PC2/v4-201311/) and used in SLAP-Enrich. This contained an initial catalogue of 2,893 gene sets (one for each pathway) assembled from multiple public available resources, such as Reactome, Panther, HumanCyc and PID, and covering 8,148 unique genes. From this catalogue, gene sets containing less than 4 genes were discarded.

In order to remove redundancies those gene sets 1) corresponding to the same pathway across different resources or 2) with a large overlap (Jaccard index (JI) > 0.8, as detailed below) were merged together by intersecting them. The gene sets resulting from this compression were then added to the collection (with a joint pathway label) and those participating in at least one merging were discarded. The final collection resulting from this pre-processing was composed by 1,636 gene sets, for a total amount of 8,056 unique genes. Given two gene sets P_1 and P_2 the corresponding JI is defined as:

$$JI(P_1, P_2) = \frac{|P_1 \cap P_2|}{|P_1 \cup P_2|}$$

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Statistical framework.

Let $\boldsymbol{G} = \{\boldsymbol{g_1}, \boldsymbol{g_2}, \cdots, \boldsymbol{g_n}\}$ be a list including all the genes whose mutational status across a set of samples $\boldsymbol{S} = \{\boldsymbol{s_1}, \boldsymbol{s_2}, \cdots, \boldsymbol{s_m}\}$ has been determined by whole exome sequencing profiling, and $\boldsymbol{f}: \boldsymbol{G} \times \boldsymbol{S} \rightarrow \{\boldsymbol{0}, \boldsymbol{1}\}$ a function defined as
$\boldsymbol{f}(\boldsymbol{g}, \boldsymbol{s_i}) = \{\boldsymbol{1}$ if $\boldsymbol{g}$ is altered in sample $\boldsymbol{s_i}$ and 0 otherwise$\}$.

Given a pathway gene set P, the aim is to assess whether there is a tendency for that pathway to be recurrently genetically altered across the samples in S. In what follows we will refer to P and the pathway corresponding to P interchangeably.
We assume that the pathway P is altered in sample s_i if $\exists g \in P$ such that $g \in G$ and $f(g, s_i) = 1$, i.e. at least one gene in the pathway P is altered in the i-th sample. First of all we quantify how likely it is to observe at least one gene belonging to P altered in sample s_i. To quantify this probability, let us introduce the variable $X_i = |\{g \in P: g \in G \text{ and } f(g, s_i) = 1\}|$, whose value is equal to the number of genes in P that are altered in sample s_i.

Under the assumption of both a gene-wise and sample-wise statistical independence, the probability of X_i assuming a value greater or equal than 1 is given by:

Equation 1

$$p_i = \Pr(X_i \geq 1) = 1 - \exp(-\rho\, k),$$

where $k = \sum_{g \in P} \mathcal{L}(g)$ is the sum of the exonic block length of all the genes in pathway P, and ρ the background mutation rate, which can be computed from analysed dataset directly or set to established estimated values (typically 10^{-6}/nucleotide).

These p_i can be considered as the success probability of a set of Bernoulli trials $\{i\}$ (with $i = 1, \ldots, M$) and summing them across all the elements of S gives the expected number of samples harbouring a mutation in at least one gene belonging to the pathway P:

Equation 2

$$E(P) = \sum_{i=1}^{M} p_i.$$

On the other hand, if we consider a function ϕ on the domain of the X variables, defined as $\phi(X) = \{1 \text{ if } X \geq 1 \text{ and } 0 \text{ otherwise}\}$, then summing the values assumed by this function across

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

all the samples S gives the observed number of samples harbouring a mutation in at least one gene belonging to P:

$$O(P) = \sum_{i=1}^{M} \phi(X_i).$$

A pathway alteration index quantifying the deviance of $O(P)$ from its expectation can be then computed as:

$$\mathcal{A}(P) = \log_{10} \frac{O(P)}{E(P)}.$$

To assess the significance of such deviation, let us observe that the probability of the event $O(P) = y$, with $y \leq M$, (i.e. the probability of observing exactly y samples in which the pathway P is altered) distributes as a Poisson binomial B (a discrete probability distribution modeling the sum of a set of independent Bernoulli trials that are not identically distributed). In our case, the i-th Bernoulli trial accounts for the event "the pathway P is altered in the sample s_i" and its probability of success is given by p_i introduced above. The parameters of such $\mathcal{B}$ distribution are then the probabilities $\{p_i\}$, and its mean is given by Equation 2.

The probability of the event $O(P) = y$ can be written then as:

$$\Pr(O(P) = y) = \mathcal{B}(\rho, y) = \sum_{A \in F_y} \prod_{i \in A} p_i \prod_{j \in A^c} (i - p_j)$$

where F_y is the set of all the possible subsets of y that can be selected from $\{1, 2, \dots, M\}$ (for example if $M = 3$, then $F_2 = \{\{1,2\},\{1,3\},\{2,3\}\}$), and A^c is the complement of A (i.e. $\{1, 2, \dots, M\} \backslash A$).

Hence a p-value can be computed against the null hypothesis that $O(P)$ is drawn from a Poisson binomial distribution parametrised through the vector of probabilities $\{p_i\}$. Such p-value can be derived for an observation $O(P) = z$, with $z \leq M$, as:

$$\Pr(O(P) \geq z) = \sum_{i=z}^{M} \Pr(O(P) = i) = \sum_{i=z}^{M} \mathcal{B}(\{p_i\}, i).$$

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mutual Exclusivity Filter

After correcting the p-values yielded by testing all the pathways in the collection with the Benjamini-Hockberg method (http://www.math.tau.ac.il/~ybenja/MyPapers/benjamini_hochberg1995.pdf), SLAP-Enrich further filters the pathways whose enrichment false discovery rate (FDR) is below a user defined threshold (in this study 5%) based on mutual exclusivity criteria as a further evidence of positive selection. Particularly for a given enriched pathway $\boldsymbol{P}$, an *exclusive coverage* score $\boldsymbol{M}(\boldsymbol{P})$ is computed as

$$M(P) = 100\,\frac{\dot{O}(P)}{O(P)}$$

where $O(P)$ is the number of samples in which at least one gene belonging to the pathway gene-set P is mutated, and $\dot{O}(P)$ is the number of samples in which exactly one gene belonging to the pathway gene-set P is mutated. In this study, all the pathways P with $M(P) \geq 50$ pass this final filter.

Site-directed mutagenesis of MAP2K1 expression vectors.

In order to validate candidate there four drug resistance mutations from the ENU-based forward genetic screen we sought to create mutated vectors to express within Cetuximab sensitive colorectal cell lines. Wild-type construct for MAP2K1 was ordered from Dharmacon (Lafayette, CO) and taken forward for *in vitro* site-directed mutagenesis reactions using the GENEART® Site-Directed Mutagenesis System from Thermo Fisher Scientific (Waltham, MA). To achieve this, two complementary mutagenic oligonucleotide primers were designed (obtained from Sigma-Aldrich (St Louis, MO) and used to generate gene cDNA expression constructs with desired mutations. Mutations were confirmed using Sanger Sequencing, before being delivered into cells using lentiviral infection.

Plasma DNA sequencing.

DNA extraction was performed with QIAmp DNA Mini kit (Qiagen, Hilden, Germany). Library preparation was done with the Oncomine™ Focus Assay (Thermofisher Scientific, Waltham, MA USA) following the manufacturer's instructions. After barcoding, libraries were equalized to 100pM. The sequencing template was prepared using the IonPGMSequencing 200 Kit v2 and sequenced in an Ion Select 318 chip using the PGM Sequencing 200 Kit v2 with 500 flows. Hotspot mutations in 35 genes were targeted using the Oncomine Focus Assay (ThermoFisher) (**Supplementary Table 6**). Variant Caller v4.0.r73742 was used for variant calling with the Ion

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reporter Software. All filtered variants were also analyzed with the Integrative Genomic Viewer (IGV v2.3) software.

Data Access

All raw sequence data has been deposited in the European Genome-Phenome Archive (EGAS00001001743, EGAS00001001744, EGAS00001001745). SLAP-Enrich is implemented as an R-Package and public available on GitHub at https://github.com/francescojm/SLAPenrich together with a pipeline and detailed instructions to reproduce the results presented in this manuscript

Acknowledgements

UM was supported by a Cancer Research UK Clinician Scientist Fellowship. FI was supported by the European Bioinformatics Institute and Wellcome Trust Sanger Institute. JB, SP and JY are supported by a ERC Synergy Grant.

Author Contributions: J.B. performed the majority of experiments, analysed the data and contributed to writing the manuscript. M.P., L.G.A and C.M. analysed the data. S.B. and M.G. designed and ran the combination screen. C.M., A.D. and B.B. carried out and analysed the plasma sequencing. F.I. designed the computational methods, analysed the data and contributed to writing the manuscript. J.S.R supervised the analysis. U.M. conceived the project, analysed the data and contributed to writing the manuscript.

Disclosure Declaration

Competing financial interests - none of the authors hold any competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Acevedo-Arozena A, Wells S, Potter P, Kelly M, Cox RD, Brown SD. 2008. ENU mutagenesis, a way forward to understand gene function. *Annu Rev Genomics Hum Genet* **9**: 49-69.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL et al. 2013a. Signatures of mutational processes in human cancer. *Nature* **500**: 415-421.

Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. 2013b. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* **3**: 246-259.

Arena S, Bellosillo B, Siravegna G, Martinez A, Canadas I, Lazzari L, Ferruz N, Russo M, Misale S, Gonzalez I et al. 2015. Emergence of Multiple EGFR Extracellular Mutations during Cetuximab Treatment in Colorectal Cancer. *Clin Cancer Res* **21**: 2157-2166.

Bardelli A, Corso S, Bertotti A, Hobor S, Valtorta E, Siravegna G, Sartore-Bianchi A, Scala E, Cassingena A, Zecchin D et al. 2013. Amplification of the MET Receptor Drives Resistance to Anti-EGFR Therapies in Colorectal Cancer. *Cancer Discovery* **3**: 658-673.

Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B et al. 2004. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**: 431-437.

Bettegowda C. 2014. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Science Translational Medicine* **6**.

Concepcion D, Seburn KL, Wen G, Frankel WN, Hamilton BA. 2004. Mutation rate and predicted phenotypic target sizes in ethylnitrosourea-treated mice. *Genetics* **168**: 953-959.

Diaz LA, Jr., Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, Allen B, Bozic I, Reiter JG, Nowak MA et al. 2012. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**: 537-540.

Douillard J-Y, Oliner KS, Siena S, Tabernero J, Burkes R, Barugel M, Humblet Y, Bodoky G, Cunningham D, Jassem J et al. 2013. Panitumumab-FOLFOX4 Treatment and RAS Mutations in Colorectal Cancer. *New England Journal of Medicine* **369**: 1023-1034.

Ercan D, Choi HG, Yun CH, Capelletti M, Xie T, Eck MJ, Gray NS, Janne PA. 2015. EGFR Mutations and Resistance to Irreversible Pyrimidine-Based EGFR Inhibitors. *Clin Cancer Res* **21**: 3913-3923.

Forsburg SL. 2001. The art and design of genetic screens: Yeast. *Nature Reviews Genetics* **2**: 659-668.

Hu X, Zhang Z. 2016. Understanding the Genetic Mechanisms of Cancer Drug Resistance Using Genomic Approaches. *Trends Genet* **32**: 127-137.

Justice MJ. 2000. Capitalizing on large-scale mouse mutagenesis screens. *Nature Reviews Genetics* **1**: 109-115.

Katayama R, Shaw AT, Khan TM, Mino-Kenudson M, Solomon BJ, Halmos B, Jessop NA, Wain JC, Yeo AT, Benes C et al. 2012. Mechanisms of acquired crizotinib resistance in ALK-rearranged lung Cancers. *Sci Transl Med* **4**: 120ra117.

Kile BT, Hilton DJ. 2005. The art and design of genetic screens: Mouse. *Nature Reviews Genetics* **6**: 557-567.

Kobayashi S, Boggon TJ, Dayaram T, Janne PA, Kocher O, Meyerson M, Johnson BE, Eck MJ, Tenen DG, Halmos B. 2005. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* **352**: 786-792.

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H et al. 2015. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**: 583-U332.

Long GV, Fung C, Menzies AM, Pupo GM, Carlino MS, Hyman J, Shahheydari H, Tembe V, Thompson JF, Saw RP et al. 2014. Increased MAPK reactivation in early resistance to dabrafenib/trametinib combination therapy of BRAF-mutant metastatic melanoma. *Nat Commun* **5**: 5694.

Misale S, Arena S, Lamba S, Siravegna G, Lallo A, Hobor S, Russo M, Buscarino M, Lazzari L, Sartore-Bianchi A et al. 2014. Blockade of EGFR and MEK intercepts heterogeneous

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mechanisms of acquired resistance to anti-EGFR therapies in colorectal cancer. *Sci Transl Med* **6**: 224ra226.

Misale S, Yaeger R, Hobor S, Scala E, Janakiraman M, Liska D, Valtorta E, Schiavo R, Buscarino M, Siravegna G et al. 2012. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**: 532-U131.

Montagut C, Dalmases A, Bellosillo B, Crespo M, Pairet S, Iglesias M, Salido M, Gallen M, Marsters S, Tsai SP et al. 2012. Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. *Nature Medicine* **18**: 221-223.

Ohashi K, Sequist LV, Arcila ME, Moran T, Chmielecki J, Lin YL, Pan Y, Wang L, de Stanchina E, Shien K et al. 2012. Lung cancers with acquired resistance to EGFR inhibitors occasionally harbor BRAF gene mutations but lack mutations in KRAS, NRAS, or MEK1. *Proc Natl Acad Sci U S A* **109**: E2127-2133.

Patton EE, Zon LI. 2001. The art and design of genetic screens: Zebrafish. *Nature Reviews Genetics* **2**: 956-966.

Russo M, Siravegna G, Blaszkowsky LS, Corti G, Crisafulli G, Ahronian LG, Mussolin B, Kwak EL, Buscarino M, Lazzari L et al. 2015. Tumor heterogeneity and lesion-specific response to targeted therapy in colorectal cancer. *Cancer Discov* doi:10.1158/2159-8290.CD-15-1283.

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG et al. 2014. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**: 84-87.

Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G, Ponzetti A, Cremolini C, Amatu A, Lauricella C et al. 2015. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med* **21**: 795-801.

St Johnston D. 2002. The art and design of genetic screens: Drosophila melanogaster. *Nature Reviews Genetics* **3**: 176-188.

Thress KS, Paweletz CP, Felip E, Cho BC, Stetson D, Dougherty B, Lai Z, Markovets A, Vivancos A, Kuang Y et al. 2015. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat Med* **21**: 560-562.

Tiedt R, Degenkolbe E, Furet P, Appleton BA, Wagner S, Schoepfer J, Buck E, Ruddy DA, Monahan JE, Jones MD et al. 2011. A Drug Resistance Screen Using a Selective MET Inhibitor Reveals a Spectrum of Mutations That Partially Overlap with Activating Mutations Found in Cancer Patients. *Cancer Research* **71**: 5255-5264.

Tokunaga M, Kokubu C, Maeda Y, Sese J, Horie K, Sugimoto N, Kinoshita T, Yusa K, Takeda J. 2014. Simulation and estimation of gene number in a biological pathway using almost complete saturation mutagenesis screening of haploid mouse cells. *BMC Genomics* **15**: 1016.

Van Allen EM, Wagle N, Sucker A, Treacy DJ, Johannessen CM, Goetz EM, Place CS, Taylor-Weiner A, Whittaker S, Kryukov GV et al. 2014. The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discov* **4**: 94-109.

Van Cutsem E, Kohne CH, Hitre E, Zaluski J, Chang Chien CR, Makhson A, D'Haens G, Pinter T, Lim R, Bodoky G et al. 2009. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med* **360**: 1408-1417.

Wagle N, Van Allen EM, Treacy DJ, Frederick DT, Cooper ZA, Taylor-Weiner A, Rosenberg M, Goetz EM, Sullivan RJ, Farlow DN et al. 2014. MAP kinase pathway alterations in BRAF-mutant melanoma patients with acquired resistance to combined RAF/MEK inhibition. *Cancer Discov* **4**: 61-68.

Wilson FH, Johannessen CM, Piccioni F, Tamayo P, Kim JW, Van Allen EM, Corsello SM, Capelletti M, Calles A, Butaney M et al. 2015. A Functional Landscape of Resistance to ALK Inhibition in Lung Cancer. *Cancer Cell* **27**: 397-408.

Yang X, Boehm JS, Yang X, Salehi-Ashtiani K, Hao T, Shen Y, Lubonja R, Thomas SR, Alkan O, Bhimdi T et al. 2011. A public genome-scale lentiviral expression library of human ORFs. *Nature Methods* **8**: 659-U680.

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Yonesaka K, Zejnullahu K, Okamoto I, Satoh T, Cappuzzo F, Souglakos J, Ercan D, Rogers A, Roncalli M, Takeda M et al. 2011. Activation of ERBB2 Signaling Causes Resistance to the EGFR-Directed Therapeutic Antibody Cetuximab. *Science Translational Medicine* **3**.

Zhang S, Wang F, Keats J, Zhu X, Ning Y, Wardwell SD, Moran L, Mohemmad QK, Anjum R, Wang Y et al. 2011. Crizotinib-Resistant Mutants of EML4-ALK Identified Through an Accelerated Mutagenesis Screen. *Chemical Biology & Drug Design* **78**: 999-1005.

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1

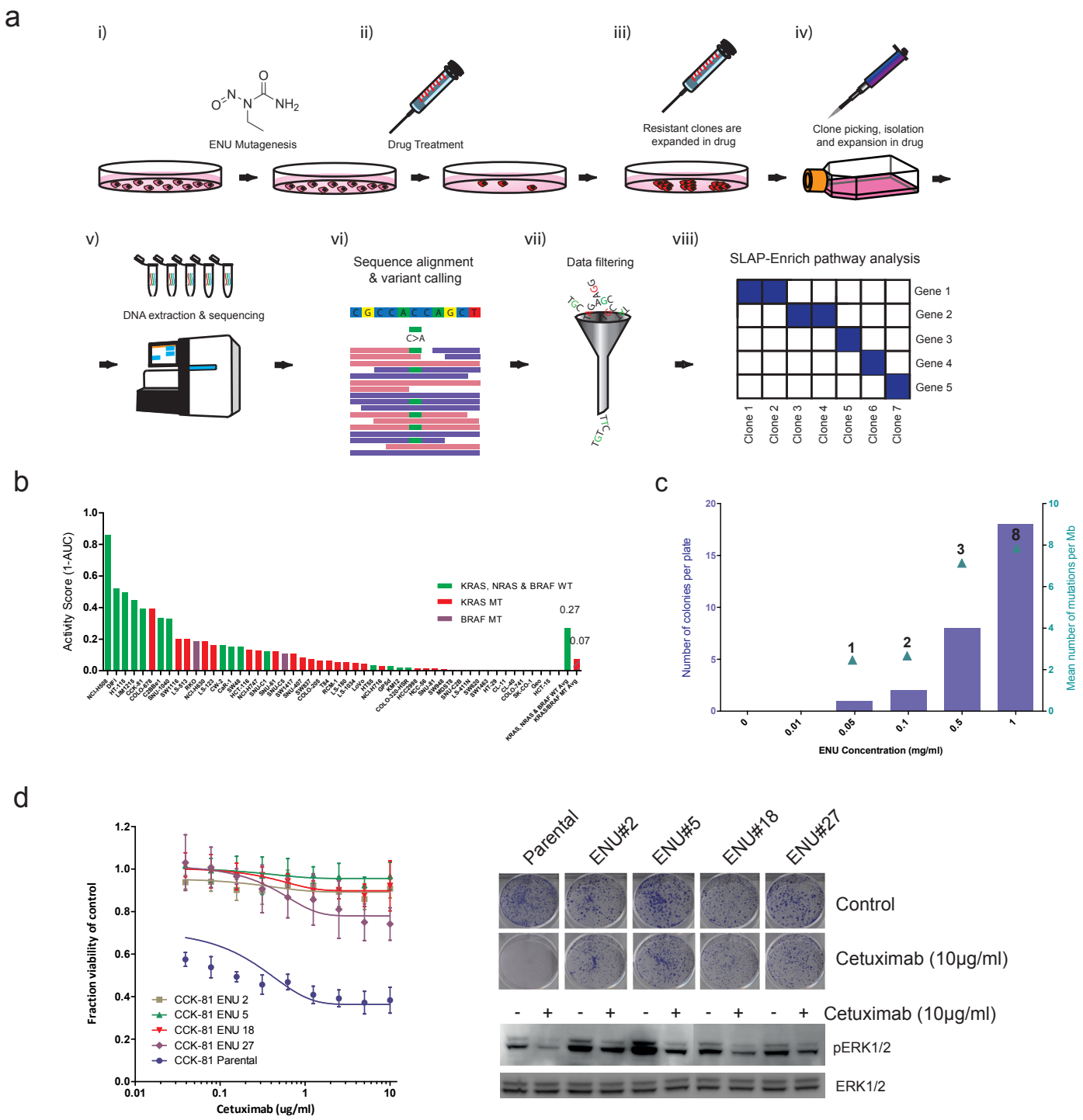

Figure 1: Genome-wide chemical mutagenesis screens to define pathways of drug resistance in cancer.
(a) Experimental design: (i) Cancer cells are mutagenized by 24 hour ENU exposure, (ii) the library of mutagenized cells is then drug selected, (iii) resistant clones are expanded in drug, (iv) clones are picked and expanded as separate cell lines in drug, (v) clones undergo whole exome sequencing together with the matched non-mutagenized parental cell line to high coverage, (vi) whole exome sequences undergo alignment and variant calling, (vii) filtering to remove confounding subclonal parental cell line variants, and (viii) filtered exome sequence data is finally analysed by SLAP-Enrich pathway analysis to identify pathways which are recurrently mutated across the population of resistant clones more than would be expected by chance considering both the exonic pathway size and the total number of mutations per clone.
(b) Cetuximab viability screen in colorectal cancer cell lines. 51 colorectal cancer cell lines were screened with a concentration range of the EGFR monoclonal antibody Cetuximab and viability (as measured by the Activity Score, 1-AUC) was measured after 6 days. The KRAS, NRAS and BRAF mutation status of each line is indicated. The mean Activity Score for those cell lines wild-type for all 3 genes (green bar) versus those with a canonical driver mutation in at least one gene (red bar) is indicated in the last two columns.
(c) Dose dependent effect of ENU on mutation burden and drug resistance. CCK-81 colorectal cancer cells were treated for 24 hours with increasing concentrations of ENU (x-axis, 0.01-1mg/ml) and then weekly with Cetuximab (10µg/ml) for 8 weeks to allow drug resistant colonies to develop. The number of resistant colonies per plate were counted for each ENU concentration (blue bars) and submitted for whole exome sequencing to calculate the mean number of mutations per Mb (green triangles). Numbers above each triangle indicate the number of clones sequenced at that concentration.
(d) ENU mutagenesis generates stably drug-resistant clones. A subset of Cetuximab-resistant clones generated in the CCK-81 cell line following exposure to ENU and subsequent serial weekly Cetuximab treatment were picked from the plate and expanded separately out of drug for 4 weeks. 21-day clonogenic survival assays in 4 clones (along with the parental CCK-81 cell line) treated with Cetuximab 10µg/ml confirmed that resistance to the EGFR mAb had been maintained and was stable. A 6-day viability assay of cells treated with a concentration range of Cetuximab (right panel) demonstrates resistance of ENU clones at all concentrations. Immunoblot analysis of effect of Cetuximab treatment (10µg/ml) for 6 hours confirms persistence of MAPK signalling in ENU clones.

Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

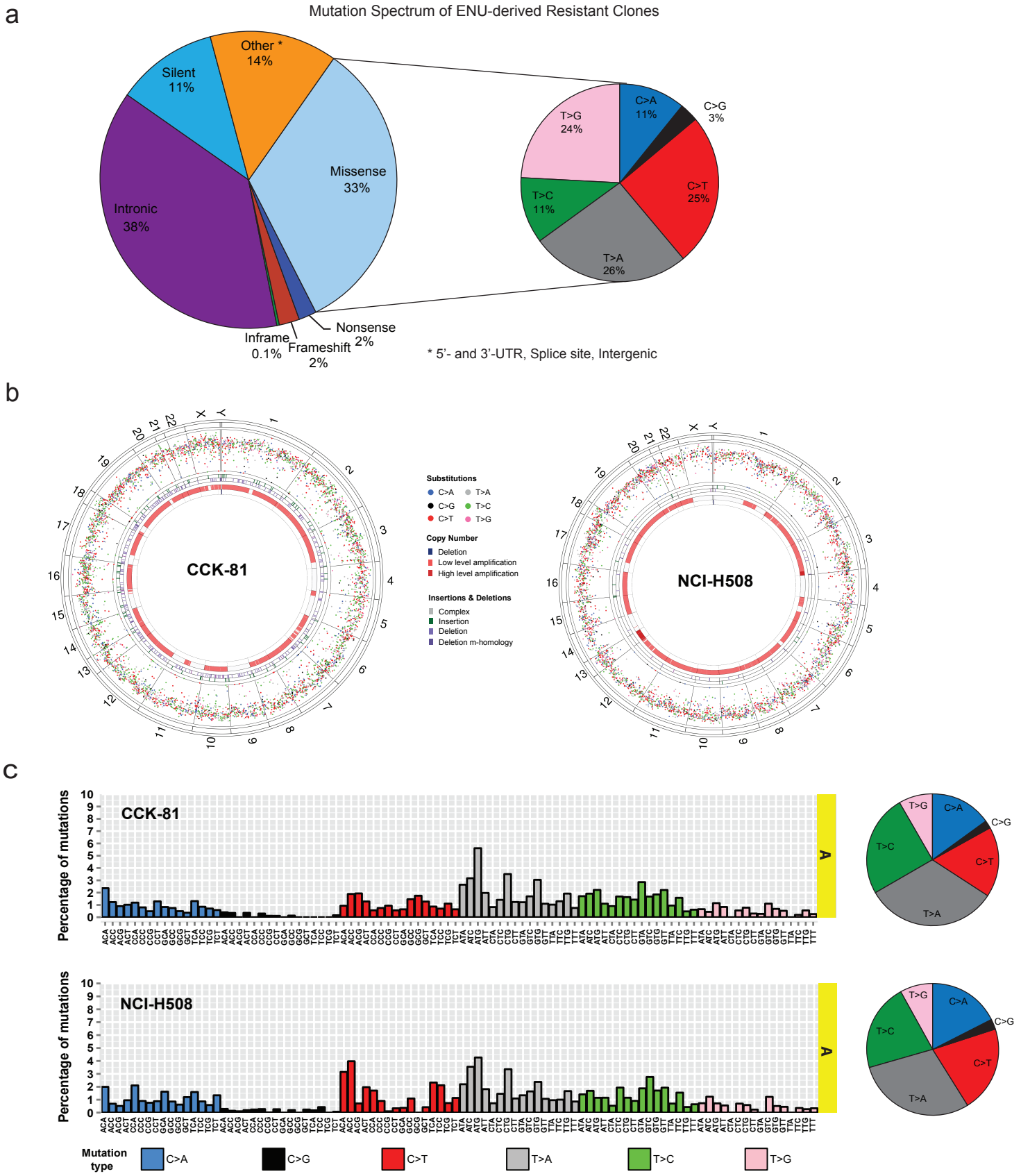

Figure 2: The spectrum of ENU-induced mutations.

(a) Spectrum of mutations in ENU-derived drug resistant clones. A pie chart representation of the proportion of 33,857 mutations detected in CCK-81 and NCI-H508 clones categorised according to mutation type. For the missense mutations, there is a further representation of the proportion of mutations falling within the 6 possible nucleotide base substitutions (C>A, C>G, C>T, T>A, T>C, T>G).

(b) Circos plots of mutations in CCK-81 and NCI-H508 clones. All substitution mutations in CCK-81 or NCI-H508 clones are represented by their intermutation distance per chromosome. Each chromosome is composed of all coding exons in genes found within that chromosome. The four circles in each plot, from outermost inwards, represent the human chromosomes, substitution mutations, insertion mutations, deletion mutations, copy number gains, and copy number losses. Low and high level amplification refers to 2-4 and 4-8 copies respectively.

(c) The mutational signature of ENU. The trinucleotide representation of ENU mutations in CCK-81 and NCI-H508 drug-resistant colonies is displayed as the distribution of mutations for all 96 possible combinations of mutations generated as a result of ENU exposure. Adjacent pie charts display the relative distribution of each of the six classes of possible base substitutions for the CCK-81 and NCI-H508 clones.

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

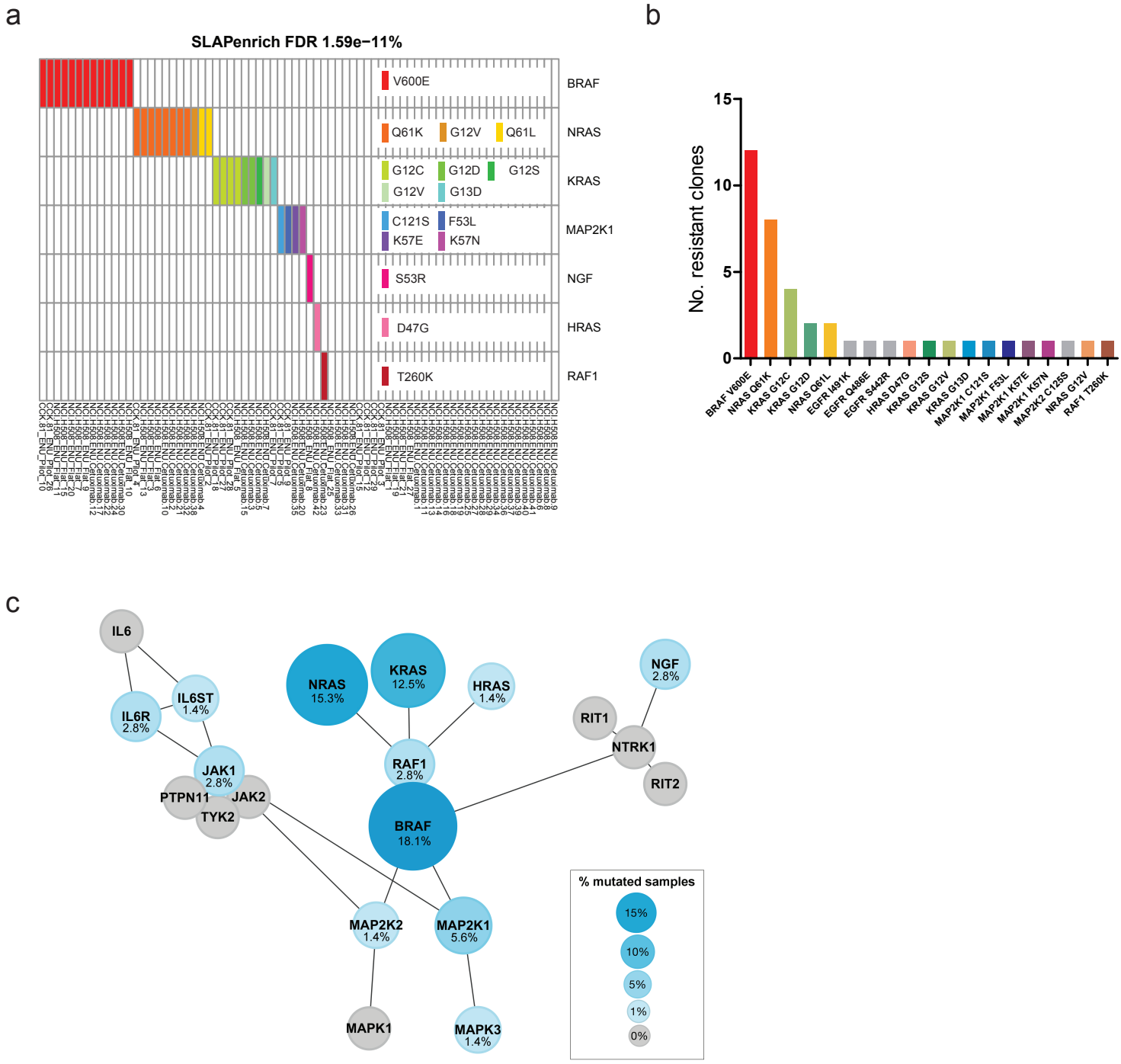

Figure 3: ENU mutagenesis identifies clinically relevant mutations and pathways.
(a) SLAP-Enrich schematic for Cetuximab drug resistance pathway. The pathway most significantly enriched with mutations, 'Signalling to P38 via RIT and RIN', contains many of the key genes of the canonical MAP kinase pathway and which are mutated in a mutually exclusive manner. Whole exome sequence from 72 Cetuximab resistant ENU clones was used to identify pathways using the SLAP-Enrich algorithm. Amino acid substitutions are labelled for the subset of genes most frequently mutated in the pathway and/or demonstrating mutual exclusivity with other mutations.
(b) Cetuximab resistance mutations. Frequency of likely ENU-derived drug resistance mutations across 42 resistant clones. The amino acid consequence of each mutation is indicated with the gene name.
(c) Rational targeting of resistance pathways. A visual representation of the mutated genes that comprise the 'Signalling to P38 via RIT and RIN' pathway indicating gene-gene interactions and the hierarchy of signalling.

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

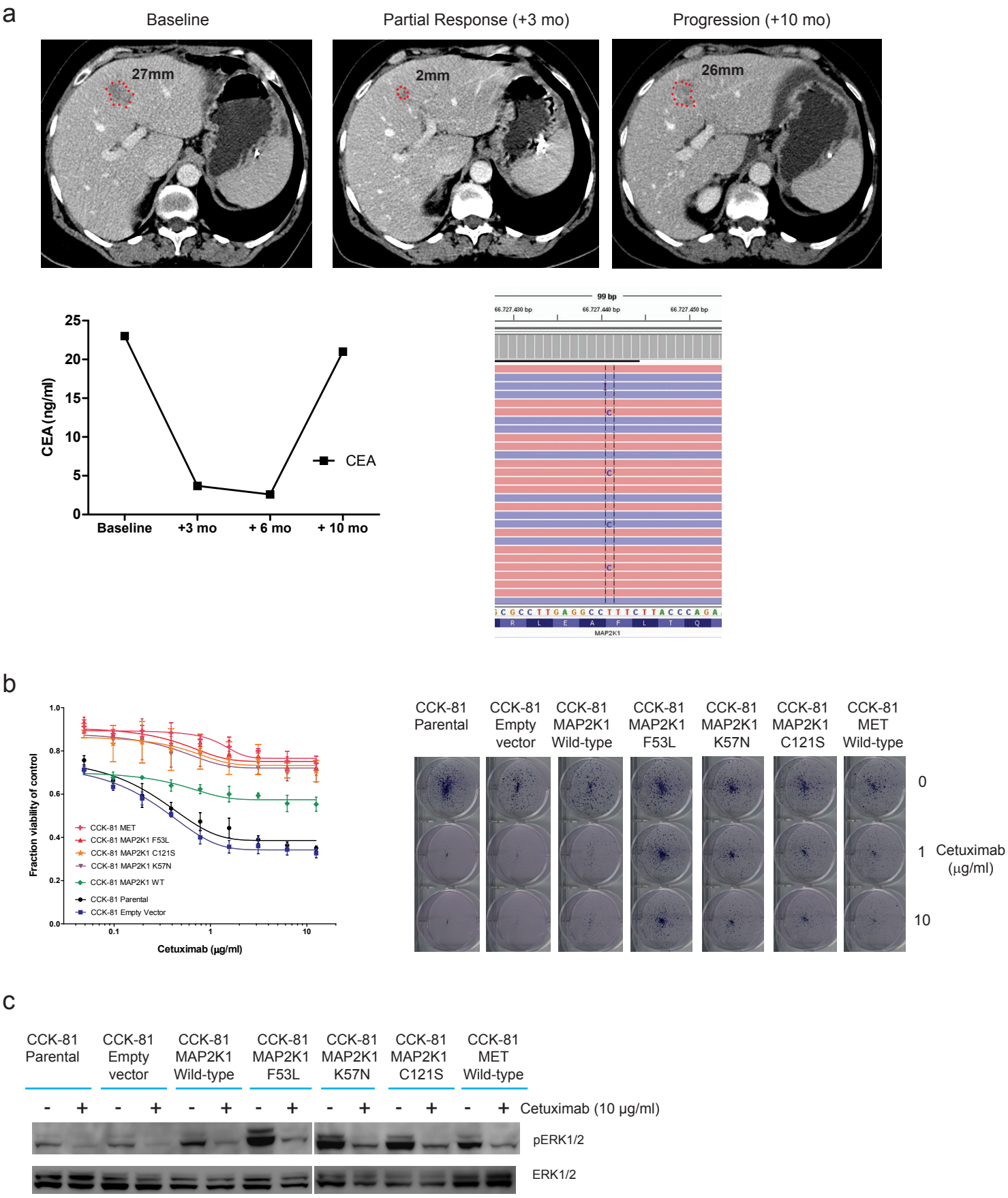

Figure 4. ENU mutagenesis identifies a novel MAP2K1 resistance mutation that is detected in a colorectal cancer patient after an initial response to Cetuximab.
(a) A colorectal cancer patient with inoperable liver metastases was treated with the combination of FOLFIRI chemotherapy and Cetuximab. Initial response to treatment was demonstrated radiologically at 3 months and by falling CEA marker levels. Subsequent measurements at 10 months confirmed progressive disease and sequencing of plasma DNA detected a novel mutation in MAP2K1 (MEK1), p.F53L.
(b) CCK-81 cells were transduced with the following lentiviral MAP2K1 vectors: empty vector (mock), wild-type (WT), K57N, C121S and F53L mutations. MET was also overexpressed as a positive control. Cells were seeded in 96-well plates and treated for 1 week with increasing concentrations of Cetuximab (0-12.5µg/ml) (left panel) or assayed using a clonogenic survival assay at 20 days at 1 or 10 µg/ml of Cetuximab (right panel).
(c) Immunoblot of the effect of Cetuximab (10 µg/ml) at 6 hours on pERK in MAP2K1 mutant CCK-81 cells (as b).

Figure 5

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

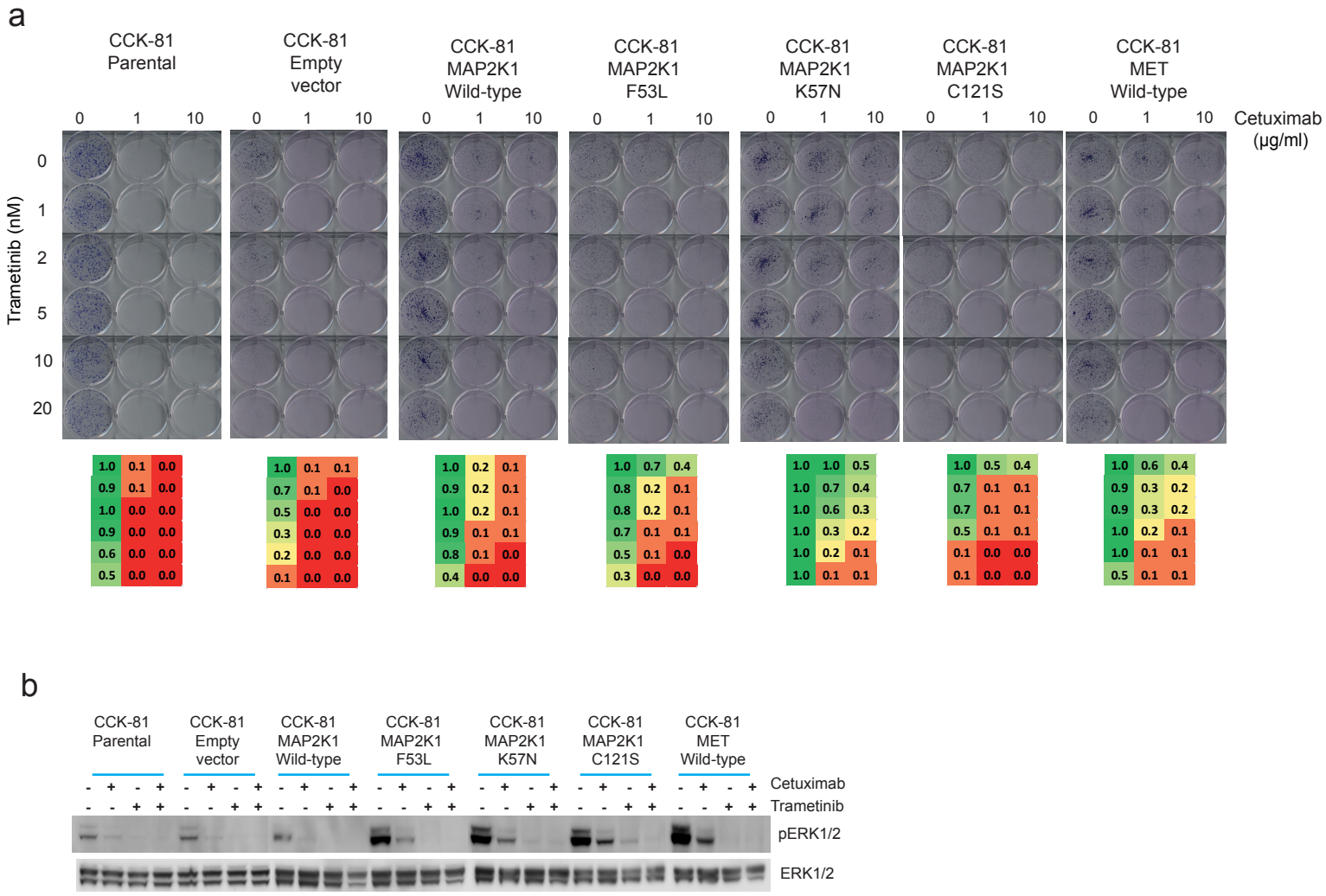

Figure 5. Rational targeting of resistance pathways to re-sensitize drug resistant mutants.
(a) Immunoblot of the effect of combing Cetuximab and Trametinib (MEK inhibitor) on MAPK signalling in MAP2K1 mutant CCK-81 cells. Cells were treated with Cetuximab (10 µg/ml), Trametinib (5nM) or both for 6 hours. MET was also overexpressed as a positive control.
(b) Clonogenic survival assay of MAK2K1 mutant CCK-81 cells when treated with Cetuximab, Trametinib or a combination of the two drugs.

bioRxiv preprint doi: https://doi.org/10.1101/066555; this version posted July 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Patient ID	Therapy	Resistance	Resistance mutation	Mutant Allele Frequency (%)	Oncogenic alteration in COSMIC database	Druggable
1630423	FOLFIRI + Cetuximab	Acquired	KRAS p.Q61H	47%	YES	NO
1155170	FOLFIRI + Cetuximab	Acquired	ERBB2 amplification	-	YES	YES
1532256	FOLFIRI + Cetuximab	Acquired	KRAS p.G12V	36%	YES	NO
			MAP2K1 p.F53L	14%	YES	YES
621581	Irinotecan+Cetuximab	Acquired	ND	-	-	-
222822	Irinotecan+Cetuximab	Acquired	KRAS p.Q61H	61%	YES	NO
880618	Irinotecan+Cetuximab	Acquired	ND	-	-	-
205149	FOLFOX + Cetuximab	Acquired	EGFR p.S492R	1%	-	YES
279889	FOLFOX + Cetuximab	Acquired	KRAS p.G12S	14%	YES	NO
1595011	Irinotecan + Cetuximab	Acquired	KRAS p.G12A	6%	YES	NO
			NRAS p.Q61H	11%	YES	NO
595082	Irinotecan + Cetuximab	Acquired	KRAS p.A146T	26%	YES	NO
1111065	FOLFOX+ Cetuximab	Acquired	NRAS p.Q61L	31%	YES	NO
			EGFR p.S492R	6%	-	YES
1617443	Irinotecan + Cetuximab	Acquired	ND	-	-	-
1510839	Irinotecan + Cetuximab	Acquired	BRAF p.V600E	14%	YES	YES
1114672	Irinotecan + Cetuximab	Acquired	NRAS p.Q61K	35%	YES	NO
1305689	Irinotecan + Cetuximab	Acquired	EGFR p.R451C	6%	-	-
392067	Irinotecan + Cetuximab	Acquired	ND	13%	YES	YES
1682384	Panitumumab	Acquired	ND	-	-	-
337548	Irinotecan + Cetuximab	Acquired	EGFR p.S464L	-	-	YES
750169	FOLFIRI + Cetuximab	Acquired	KRAS amplification	-	YES	-
1717528	FOLFIRI + Cetuximab	Acquired	ND	-	-	-
1284241	Cetuximab	Acquired	BRAF p.V600E	6%	YES	YES
1268932	Irinotecan+Cetuximab	Acquired	EGFR p.K467T	15%	-	-

Table 1. Identification of genetic alterations associated with resistance to anti-EGFR antibodies in plasma samples from 22 colorectal cancer patients.
The table lists putative genetic mechanisms of acquired resistance to anti-EGFR therapies that were identified in ctDNA of 22 patients. The novel *MAP2K1* mutation is indicated in red. FOLFIRI = folinic acid, 5-fluorouracil and irinotecan; FOLFOX = folinic acid, 5-fluorouracil and oxaliplatin. ND = no resistance mutation detected.