Title: Longitudinal liquid biopsy and mathematical modelling of clonal evolution forecast waiting time to treatment failure in the PROSPECT-C phase II colorectal cancer clinical trial.

Running Title: Exploiting ctDNA to predict timing of treatment failure

Authors: Khurum H Khan[1,2], David Cunningham[1], Benjamin Werner[3], Georgios Vlachogiannis[2], Inmaculada Spiteri[3], Timon Heide[3], Javier Fernandez Mateos[2,3], Alexandra Vatsiou[3], Andrea Lampis[2], Mahnaz Darvish Damavandi[2], Hazel Lote[2], Ian S. Huntingford[2], Somaieh Hedayat[2], Ian Chau[1], Nina Tunariu[4], Giulia Mentrasti[2], Francesco Trevisani[2], Sheela Rao[1], Gayathri Anandappa[1,2], David Watkins[1], Naureen Starling[1], Janet Thomas[1], Clare Peckitt[1], Nasir Khan[4], Massimo Rugge[5], Ruwaida Begum[1], Blanka Hezelova[1], Annette Bryant[1], Thomas Jones[6], Paula Proszek[6], Matteo Fassan[6], Jens C Hahne[2], Michael Hubank[6], Chiara Braconi[1,7], Andrea Sottoriva[3,†] & Nicola Valeri[1,2,†]

Affiliations:

[1] Department of Medicine, The Royal Marsden NHS Trust, London and Sutton. United Kingdom.

[2] Division of Molecular Pathology, The Institute of Cancer Research London and Sutton, United Kingdom.

[3] Centre for Evolution and Cancer, The Institute of Cancer Research, London, United Kingdom.

[4] Department of Radiology, The Royal Marsden NHS Trust, London and Sutton. United Kingdom.

[5] Department of Medicine and Surgical Pathology, University of Padua, Padua, Italy.

[6] Clinical Genomics, The Centre for Molecular Pathology, The Royal Marsden NHS Trust, London and Sutton. United Kingdom.

[7] Division of Cancer Therapeutics, The Institute of Cancer Research, London and Sutton, United Kingdom

† **Correspondence should be addressed to:**

Dr Andrea Sottoriva

Centre for Evolution and Cancer, The Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK

Email: andrea.sottoriva@icr.ac.uk

Dr Nicola Valeri

Centre for Molecular Pathology, The Institute of Cancer Research & The Royal Marsden Hospital

Sutton, Surrey, SM2 5PT, UK

Email: nicola.valeri@icr.ac.uk

Keywords: treatment resistance, liquid biopsy, colorectal cancer, mathematical modelling, tumour evolution, precision medicine.

Funding: This work was supported by Cancer Research UK (grant number CEA A18052), the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research (grant numbers A62, A100, A101, A159) and the European Union FP7 (grant number CIG 334261) to N.V. A.S. is supported by the Wellcome Trust (202778/B/16/Z), Cancer Research UK (A22909), and The Chris Rokos Fellowship in Evolution and Cancer. B.W. is supported by the Geoffrey W. Lewis Post-Doctoral Training fellowship. The authors acknowledge support from the National Institute for Health Research Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research. This work was also

supported by Wellcome Trust funding to the Centre for Evolution and Cancer (105104/Z/14/Z).

Conflict of interest: D.C. received research funding from: Roche, Amgen, Celgene, Sanofi, Merck Serono, Novartis, AstraZeneca, Bayer, Merrimack and MedImmune. I.C. has had advisory roles with Merck Serono, Roche, Sanofi Oncology, Bristol Myers Squibb, Eli-Lilly, Novartis, Gilead Science. He has received research funding from Merck-Serono, Novartis, Roche and Sanofi Oncology, and honoraria from Roche, Sanofi-Oncology, Eli-Lilly, Taiho, Bayer and Prime Therapeutics. DW has received honoraria from Amgen. KK has advisory role with Bayer Oncology group. NV received honoraria from Merck Serono, Bayer and Eli-Lilly. All other authors declare no conflict of interest.

Word count: 6003

Total number of Figures: 5

Total number of Tables: 1

ABSTRACT

Sequential profiling of plasma cell-free (cf)DNA holds immense promise for early detection of patient relapse. However, how to exploit the predictive power of cfDNA as a liquid biopsy in the clinic remains unclear. RAS pathway aberrations can be tracked in cfDNA to monitor resistance to anti-epidermal growth factor receptor (EGFR) monoclonal antibodies in metastatic colorectal cancer patients. In this prospective phase II clinical trial of single-agent cetuximab in *RAS* wild-type patients, we combine genomic profiling of serial cfDNA and matched sequential tissue biopsies with imaging and mathematical modelling of cancer evolution. We show that a significant proportion of patients defined as *RAS* wild-type based on diagnostic tissue analysis harbour aberrations in RAS pathway in pre-treatment cfDNA and, in fact, do not benefit from EGFR inhibition. We demonstrate that primary and acquired resistance to cetuximab are often of polyclonal nature, and these dynamics can be observed in tissue and plasma. Furthermore, evolutionary modelling combined with frequent serial sampling of cfDNA allowed predicting the expected waiting time to treatment failure in individual patients. This study demonstrates how integrating frequently-sampled longitudinal liquid biopsies with a mathematical framework of tumour evolution allows individualised quantitative forecasting of relapse, providing novel opportunities for adaptive personalised therapies.

STATEMENT OF SIGNIFICANCE

Liquid biopsies capture spatial and temporal heterogeneity underpinning resistance to anti-EGFR monoclonal antibodies in colorectal cancer. Dense serial sampling is needed to predict the waiting time to treatment failure and generate a window of opportunity for intervention.

INTRODUCTION

Colorectal cancer (CRC) is one of the commonest cancers worldwide and accounts for 8-9% of cancer-related mortality[1], with poor 5–year survival for stage IV disease[2,3]. Genetic alterations in RAS pathway are responsible for primary and acquired resistance to anti-epidermal growth factor receptor (EGFR) monoclonal antibodies[4-12]. However, despite tailored patient selection based on genetic screening for somatic *RAS* mutations, 65-70% of patients progress within 3-12 months after starting therapy.

In many patients, *RAS* mutations occur early during colorectal carcinogenesis, manifesting as a clonal (truncal) alteration in primary and metastatic lesions[13,14]; these patients derive no benefit from anti-EGFR therapies. In seminal studies, plasma cell-free (cf)DNA analysis of patients with no detectable *RAS* mutations at baseline has demonstrated that under the selective pressure of therapy, small and undetectable *RAS* mutant subpopulations at baseline undergo clonal expansion, ultimately leading to acquired treatment resistance[15-17]. Hence, whilst resistance to anti-EGFR therapies can be polyclonal, it often converges on biochemical signalling pathways downstream of EGFR[18]. Since treatment resistance is driven by intra-tumour heterogeneity (ITH) that generates the substrate of variation necessary for evolution, multiple biopsies in time and space are critical to understanding the complexity of an evolving malignancy[19,20].

Although cfDNA holds immense promise for patient management, the predictive power of liquid biopsies in metastatic colorectal cancer (mCRC) has not been demonstrated within a prospective clinical trial, and data on concordance between liquid (cfDNA) and solid (tissue) biopsies remain sparse at this stage. Most of the previously reported cohorts of anti-EGFR resistant mCRC include patients treated

with anti-EGFR therapy in combination with chemotherapy, representing a potential confounding factor in understanding genomic patterns. Moreover, although seminal studies in other cancer types such as breast[21] and non-small cell lung cancer (NSCLC)[22] have shown prospectively that cfDNA positivity precedes clinical progression, a major challenge to the use of liquid biopsies in the clinic is the extreme variability between patients, with waiting times to progression ranging from 1 to 400 days in breast cancer, from 10 to 346 days in NSCLC, and unknown in mCRC. This variability prevents patient-specific clinical predictions, representing an obstacle for personalised medicine.

We have recently shown that combining functional genomics, molecular pathology and radiology in the context of well annotated clinical trials can help defining novel biomarkers to optimise patient selection[23,24]. In this prospective phase II trial of patients with *RAS* wild type (wt) mCRC treated with single agent anti-EGFR monoclonal antibodies we aimed to: (1) test the value of profiling sub-clonal mutations in *RAS* pathway in cfDNA to predict response to anti-EGFR therapy; (2) assess the mutational concordance between liquid and solid biopsies; (3) estimate waiting time to treatment failure in each individual patient using mathematical modelling of cancer evolution.

RESULTS

Trial design and patient characteristics

The Prospect-C trial (ClinicalTrials.gov identifier: NCT02994888), is a study of biomarkers of response and resistance to anti-EGFR therapies in *KRAS*/*NRAS* wt

chemo-refractory mCRC. The trial recruited forty-seven patients between November 2012 and December 2016. The objectives of the study were to track and validate the known mechanisms of resistance/response to anti-EGFR therapies and to identify novel mechanisms of response/resistance to such therapies.

Patients in the study were subjected to tissue sampling from metastatic deposits at predefined time-points including pre- [baseline (BL)] and post-treatment [disease progression (PD)]; optional partial response (PR) biopsies were conducted where clinically and technically feasible and archival material (primary cancer or original diagnostic biopsies) was used when available. Plasma was collected every 4 weeks until disease progression (**Table 1** and **Fig. 1A**; **Supplementary Table S1**). Median duration of cetuximab treatment was 10.7 months (IQR: 2.0-137.3); 20.0%, 24.4% and 46.7% of the patients experienced PR, stable disease (SD), and PD (indicative of primary resistance) by RECIST v1.1 respectively. Median Progression Free Survival (PFS) and Overall Survival (OS) were 2.6 months (95% CI: 1.9 – 4.2) and 8.2 months (95% CI: 4.2 - 12.0) respectively (**Supplementary Table S2**). These results were found to be consistent with previously published literature[25].

A consort diagram describing the study is reported in **Fig. 1B**. Patients were tested for *RAS* pathway aberrations in cfDNA and tissue biopsies according to different methods as described below:

1) in the initial 22 consecutive eligible patients (cohort 1) cfDNA was tested by digital-droplet (dd)-PCR using a tiered approach based on frequency of RAS pathway aberrations previously reported as being associated with primary and acquired resistance to anti-EGFR treatment[18]. Patients in this cohort also had tissue biopsies sequenced: a total of 84 cores from 22/22 (44 cores), 4/4 (8 cores) and

16/22 (32 cores) biopsies at BL, PR and PD respectively were tested; six patients did not have progression biopsies (2 deceased prior to assessment, 2 were considered clinically unfit and 2 declined biopsy). Twenty available archival samples from eight patients were also sequenced (**Supplementary Table S1** and **Supplementary Figure S1**).

2) From the remaining 23 eligible patients on the trial, PD and/or BL plasma samples from patients (n=18, cohort 2) with primary resistance [PFS ≤3 months (11 patients, 12 samples)] and long term benefit [PFS>6 months (4 patients, 13 samples)] were subjected to targeted sequencing with the Roche AVENIO cfDNA Expanded Kit, covering 77 cancer related genes. In the 2 long term responders for whom material was available from the first cohort (4 samples), a similar approach was adapated to validate our findings by sequencing BL and PD cfDNA (**Supplementary Figure S1**).

Patients with RAS pathway aberrations in baseline cfDNA do not respond to cetuximab

We initially investigated cfDNA concentration (**Supplementary Table S3**) and *RAS* pathway hotspot mutations in 143 plasma samples using ddPCR (10 samples could not be tested because of haemolysis; **Supplementary Figure S1**). Eleven patients (50%) had *RAS* pathway aberrations in their baseline cfDNA: 6 patients with *KRAS/NRAS* mutations, 2 patients with *BRAF* V600E mutations, 1 patient with *PIK3CA* E545K mutation and 2 with *ERBB2* amplification (**Supplementary Figure S2** and **Supplementary Table S4**). Detection of RAS pathway aberrations in baseline cfDNA was significantly associated with inferior PFS (HR=3.41; CI=1.24-9.37; p=0.02), worst OS (HR=2.78; CI=1.09-7.11; p=0.03), and showed also a trend

towards poor Response Rate (RR) (0% vs. 36.4%; p=0.09) compared to wt patients (**Fig. 2A-D**). In order to validate these findings, we tested the prevalence of RAS pathway aberrations in all the patients that showed primary resistance to cetuximab in the second cohort. Considering that hotspot-based methods such as ddPCR allow testing only a limited number of known potential drivers, thus perhaps underestimating the biological and clinical relevance of polyclonal resistance, we subjected all baseline samples in the second cohort to next-generation sequencing (NGS) of a broad panel of 77 cancer related genes. Interestingly, in keeping with data from the first set, known RAS pathway aberration were found in 6 out of 11 (54.5%) cases (**Supplementary Figure S3** and **Supplementary Table S5**). In several cases, resistance could be attributed to multiple drivers (**Fig. 2E**). In the same cohort, two patients with no common RAS pathway aberrations displayed mutations in *EGFR* T739P, *EGFR* C326R, *FGFR2* R203C and *KRAS* A18D. Among these, the *KRAS* A18D is a gain-of-function mutation with transforming activity that lies within the GTP binding region of the KRAS protein[26] and might impair response to anti-EFGR treatment thus deserving further validation (**Supplementary Table S5**). Interestingly in either of the two cohorts, no significant differences in cfDNA concentration were observed between patients with or without baseline aberrations (**Supplementary Figure S4**).

Emerging sub-clonal RAS pathway aberrations as drivers of acquired resistance to cetuximab

Next we screened for drivers of acquired resistance in the first cohort of patients. Nineteen patients (86.3%) displayed RAS pathway aberrations at progression. As

expected, the majority of patients with *KRAS* mutations had amino acid substitutions in codons 12 or 13 and 61 (**Supplementary Figure S2**). Interestingly, in 75% of the patients who achieved PR, aberrations were detectable in cfDNA in advance of clinical-radiological progression: in the first patient *c-MET* amplification was detected in cfDNA 2 months in advance of the radiological progression. In the second patient, a *KRAS* Q61H A-T mutation was detected one month prior to PD. In the last patient who experienced a remarkable PR with PFS of 10 months, two mutations were detected: a *KRAS* (Q61H A-T) mutation was found as early as eight months ahead of PD while another *KRAS* (G12D) mutation emerged four months prior to radiological progression (**Fig. 2F**). Polyclonal resistance due to multiple independent *RAS* mutant sub-clones was observed in primary and secondary resistance cases and was supported by the slope of the *APC* clonal mutation that was used as a reference to infer sub-clonal composition (**Fig. 2G** and **Supplementary Figure S5**). Specifically, polyclonal resistance could be determined because (a) the sum of the mutational frequencies of two distinct sub-clones cannot be larger than the whole, indicated by the frequency of a clonal mutation like *APC* (pigeonhole principle [27]), and (b) the trajectories over time of the two mutant sub-clones do not travel together and one can even overtake another, showing they are independent (**Fig. 2G**). The occurrence of polyclonal resistance in patients with acquired resistance (PFS>6 months) was further confirmed by NGS of progression, best response and baseline cfDNA of four patients with long-term clinical benefit in the second cohort and two patients in the first cohort for whom no mutations were initially detected by ddPCR (**Fig. 2H** and **Supplementary Table S5**). Multiple drivers of resistance emerged over treatment in 83.3% of cases and were confirmed to be sub-clonal by comparing

their variant allele frequency to the one of truncal mutations in genes such as *APC* and or *TP53* (**Supplementary Figure S5**).

In keeping with previous evidence, *RAS* mutant clones emerged during cetuximab treatment and faded away once treatment was interrupted[17]. Intriguingly, despite a period of 12 months without any treatment and radiological evidence of PD, these mutations remained undetectable when the patient entered fourth line treatment supporting the rationale for potential re-challenge with anti-EGFR treatment (**Supplementary Figure S6**)[28].

In two patients who showed a sustained benefit from cetuximab lasting 33 and 16 months respectively, we tested whether the durable response observed was due to persistent EGFR pathway inhibition due to a phenomenon of oncogenic addiction. Consistent with this hypothesis, an *EGFR* amplification was observed in the BL liver biopsy of one of these two patients (**Fig. 3A**). Treatment was halted in this patient after 16 months due to symptomatic progression of a non-target lesion assessed by RECIST V.1.1 criteria. Interestingly, after cetuximab was withdrawn a rapid progression of the *EGFR*-amplified metastatic liver lesion that was biopsied at BL along with development of new liver deposits were observed within 6 weeks. On the contrary, the soft tissue pelvic mass biopsied at time of PD did not show any significant change in volume. Consistent with this, no *EGFR* amplification was found in the latter lesion suggesting that this metastasis was not dependent on EGFR signalling (**Fig. 3A**). A rise in *APC* mutant clones was observed synchronously with the increase in size of the non-target metastasis that led to treatment discontinuation. CEA lagged behind and no RAS pathway mutant clones were detected at any time point (**Fig. 3B**). While these data confirm the correlation between tumour burden and cfDNA and highlight the higher sensitivity of cfDNA

compared to CEA, a critical review of this case makes us wonder whether, in absence of circulating RAS pathway aberrations, this patient should have received radiotherapy to the single progressing metastatic deposit while continuing cetuximab treatment.

Comparison of liquid versus matched solid biopsies provides insights on the subclonal architecture of cetuximab resistant CRC

Based on the data gathered from our cfDNA analysis, we performed ddPCR validation and amplicon based ultra-deep sequencing of sequential tissue biopsies (**Supplementary Figures S1** and **S7**) collected at clinically relevant time-points to dissect the structure of *RAS* pathway aberrations and to test if the evolutionary patterns observed in cfDNA were represented in tissues. We started our analysis by comparing mutations detected in cfDNA with mutational data obtained by ddPCR in tissues. Twenty-four mutations were detected in cfDNA of 14 patients for whom at least one tissue sample was available for cfDNA/tissue comparison: 79% of cfDNA mutations were present in at least one tissue biopsy of all 14 patients. In patients with paired pre-treatment biopsies and archival tissues (diagnostic material obtained prior to any treatment) available (6 patients, 9 mutations), 6 mutations from 5 patients were detected in both samples suggesting that these mutations pre-existed cetuximab treatment (**Fig. 4A** and **Supplementary Table S6**). In order to validate these findings we ran amplicon based ultra-deep sequencing of the same tissue biopsies using a custom library covering the most frequently mutated codons in *KRAS*, *NRAS* and *BRAF,* plus the *APC* gene. NGS analysis showed agreement among different cores from the same lesion and was in line with ddPCR analysis,

although many mutations could not be detected with statistical significance due to the limits of sensitivity of NGS versus ddPCR (**Fig. 4A** and **Supplementary Table S7**). Concordance in VAF between the two methods was good (R^2=0.5; p<0.0001) (**Supplementary Figure S8**). Interestingly, ultra-deep sequencing showed the presence of a number of RAS pathway mutations in the samples analysed (**Supplementary Figure S9)**. An average of 6.4 mutations (range 3-9) was observed in different *RAS* hotspots with proven contribution to anti-EGFR treatment supporting the notion that cetuximab resistance might often be polyclonal. In keeping with this observation, and in line with cfDNA data, purity adjustment of variants using clonal *APC* mutations confirmed that most of these mutations were sub-clonal **Supplementary Figure S10**). More importantly, the allele frequency of these abnormalities was below the detection threshold of clinically approved methods such as COBAS (**Fig. 4B**) or clinically validated targeted NGS at moderate depth (**Supplementary Figure S11**) explaining why these patients were initially classified as *RAS* wt.

The ability to study multi-region and/or sequential tissue biopsies also allowed dissecting spatial and temporal heterogeneity in response to anti-EGFR treatments. Multi-region sequencing of two different cores of two liver metastases (segment II and segment IV) and primary cancer resected after neo-adjuvant chemotherapy in one of our patients (1009) showed the presence of 7 mutations in *KRAS* and *NRAS* hotspots: among the mutations previously tested and detected in cfDNA of this patient, an *NRAS* G12C mutation appeared as clonal and common to all the metastases while the other 3 mutations (*KRAS* Q61R and *KRAS* G12S and *NRAS* G12D) appeared as sub-clonal, emerged during cetuximab treatment and, as stated above, faded away once treatment was halted (**Fig. 4C** and **Supplementary Figure**

S6) suggesting that truncal and private mutations in the same pathway might coexist and contribute to cetuximab resistance. Two patients had *ERBB2* amplifications detected in cfDNA; the amplification was confirmed by *Chromogenic In Situ Hybridisation* in BL and PD biopsies (Fig. 4D) supporting the concept that *ERBB2* amplifications are clonal and present in approximately 10% of CRC[29]. A *c-MET* amplification was observed in cfDNA of a patient with PR two months in advance of radiological progression and was supported by a synchronous increase in *APC* mutant alleles (Fig. 4E top panel). Fluorescent *In Situ Hybridisation* revealed patchy areas of amplification in both the PR and PD biopsies of this patient (Fig. 4E lower panel). Furthermore NGS analysis of his sequential biopsies including two different areas of the primary resected CRC along with two different cores at BL, PR and PD revealed the emergence of an *NRAS* G12C mutation suggesting possible polyclonal resistance to cetuximab in this patient (Fig. 4F left panel). Interestingly, the emergence of (*c-MET* and *NRAS*) resistant clones in the PD biopsy of this patient appeared to be associated with a decay in other sub-clones, consistent with the presence of an evolutionary bottleneck leading to the survival of the fittest, treatment resistant, clone (Fig. 4F cartoon).

Frequent serial cfDNA sampling and evolutionary modelling predict waiting time to relapse in individual patients

Our longitudinal dataset offered a unique opportunity to study the dynamics of treatment resistance quantitatively because of the frequent (4-week) sampling.

Previous seminal studies showed that time series treatment resistance datasets could be interpreted using mathematical modelling of tumour evolution[30]. Similar

types of modelling were also applied to longitudinal liquid biopsies to demonstrate that resistant sub-clones are pre-existing to anti-EGFR treatment[15]. Although it is likely that resistant sub-clones are present at diagnosis, their extremely low prevalence in the treatment-naïve cancer cell population prevents the detection by standard assays such as COBAS, as exemplified in **Fig. 4B.**

Here we sought to use mathematical modelling to jointly analyse both CEA (tumour burden) and cfDNA from individual patients, with the aim of exploring the predictive power of evolutionary principals when applied to a prospective clinical trial cohort. Finally, we validated our predictions using RECIST v1.1 measurements from radiological imaging data.

In our model, at baseline a tumour consists of a total of N cells. These N cells are divided into two distinct subpopulations: a population of treatment sensitive cells of size n_s and a population of treatment resistant cells of size n_r, with $n_s + n_r = N$. Sensitive cells die under treatment at rate λ_s, whereas resistant cells continue to grow under treatment at rate λ_r. This leads to the following equation for the change of cancer cell population over time during treatment:

$$N(t) = n_s e^{-\lambda_s t} + n_r e^{\lambda_r t}. \quad (1)$$

Equation (1) models the exponential decay of sensitive cells and the exponential growth of resistant cells (**Fig. 5A**) and predicts a typical U-shape for the dynamical response to treatment[30] (**Fig. 5B**). As CEA is proportional to the total tumour burden $N(t)$, Equation (1) can be applied to CEA values over time from a patient. **Fig. 5C** illustrates the U-shape dynamics for the cetuximab therapy schedule in patient 1014. Model fits to the CEA dynamics under cetuximab treatment showed remarkably high goodness of fit (average R^2=0.995) in patients with a sufficient

number of measurements (≥3 time points, n=32) and allowed estimating the response rate λ_s and the relapse rate λ_r for each case (see Methods for details and **Supplementary Figure S12**). Response rates varied between 0 (non-responders) and 0.58 per week with a median response rate of 0.2 (excluding non-responders) (**Fig. 5D**). Relapse rates ranged from 0.03 to 0.38 per week with a median value of 0.165 (**Fig. 5E** and **Supplementary Figure S13**). Additionally, the model allowed estimating the initial frequency of the resistant population at treatment initiation. Responders had an initially small or a slow growing resistant subclone, or both. On the other hand, most non-responders had a nearly dominant (~100%) resistant population pre-existing at baseline (**Fig. 5F**). Interestingly, in two cases (1002 and 1045) the resistant population at baseline was not dominant (8.3% and 4.1% respectively), although not as low as the responders (<1%), but the growth rate of the resistant sub-clone was extremely high, leading to relapse even before the first CT scan (**Fig. 5F**, grey dots, bottom right of the plot). This led the patient to be labelled as progressor although our CEA analysis predicted that there was an initial response that remained clinically undetected.

Importantly, cfDNA can independently inform on the dynamics of the resistant population, modelled by the second part of Equation (1):

$$R(t) = n_r e^{\lambda_r t}. \quad (2)$$

We applied Equation (1) and Equation (2) to CEA and cfDNA mutant frequencies respectively for those patients in our cohort for which enough time points were available for both measurements (≥3 time points, n=11). We found that the model described the dynamics extremely well (CEA mean R^2=0.996, cfDNA mean R^2=0.979, **Supplementary Figure S14**), highlighting responders in which rise of

cfDNA preceded CEA (**Fig. 5G**) versus non-responders in which both cfDNA and CEA raised at the same time (**Fig. 5H**)**.** As discussed previously, in patient 1007 multiple mutant clones rising in the plasma indicated polyclonal resistance, a phenomenon previously reported[18]. In this case, our model allowed estimating the relapse rate for both sub-clones independently (*KRAS* G12D=0.303/week, *KRAS* Q61H=0.122/week). The radically different growth rates of these resistant sub-clones, as well as the fact that they crossed over each other, confirmed that these variants were indeed in two different cancer cell populations, corroborating polyclonal resistance (rather than two mutations in the same subclone or in nested subclones).

Importantly, the application of the model to both CEA and cfDNA allowed estimating the relapse rate from two independent measurements, allowing determining how accurate the dynamics observed in the plasma would recapitulate the dynamics observed in CEA later on. Remarkably, we found that despite the limited cohort, relapse rate measured by CEA with Equation (1) and relapse rate measured by cfDNA with Equation (2) were significantly correlated (**Fig. 5I**). This not only confirmed that the mutant sub-clones detected in plasma were indeed major contributors to resistance, but also indicated that cfDNA profiling allows to *quantitatively* predict the waiting time to progression in those patients that initially responded. Since in responders cfDNA dynamics precede CEA by several weeks (**Fig. 5G**, the green line preceding the black line) and that the growth rates of these two curves correlate, it becomes possible to use plasma to forecast the time when we expect to observe clinical relapse by RECIST v1.1 measurements. Given these measurements are considered proportional to tumour burden (RECIST v1.1 standards: 20% increase in lesions diameter, 72.8% increase in volume), we were

able to test our model against the RECIST v1.1 data to prove if predictions can be made based on either cfDNA or CEA. Specifically, plugged in the resistance parameters estimated for each patient using CEA and cfDNA (**Fig. 5F**), back into equation [1] and ask the following question: if the population at baseline is $N(t=0)$, at what time t do we expect $N(t)$ to be 72.8% larger than $N(t=0)$? This simply requires solving Equation [1] for *t*. We found that measurements of relapse rate from CEA and cfDNA using our model (**Fig. 5F**) were considerably precise in predicting when progression was identified through RECIST (**Fig. 5J**), particularly given the low accuracy of RECIST v1.1 measurements (a manual assessment of a CT-scan every three months). In the case of patient 1007, both CEA and cfDNA *KRAS* G12D predicted the time of relapse with reasonable accuracy, whereas the secondary resistant sub-clone *KRAS* Q61H, which was slow growing, was not accurate at all because of the dominant effects of G12D that anticipated relapse of several months (**Fig. 5J**). It is important to note that our method allows predicting which sub-clone will dominate the dynamics of relapse because of the estimation of each sub-clone's relapse rate independently. The estimated parameters can then be used to calculate which sub-clone would lead to relapse first, as well as the combined effects of multiple sub-clones. In the case of 1007, the dynamics of relapse were driven by the fastest growing sub-clone *KRAS* G12D. Hence, as expected, accuracy in predicting relapse from cfDNA relies on identifying the dominant resistant sub-clones. The predictions of the waiting time to relapse for all the other responders is reported in **Fig. 5K**. Percent error in these estimations with respect to the time from baseline to RECIST v1.1 recurrence is reported in **Supplementary Figure S15**). We note that our current predictions have two limitations that are not dependent on our model: (a) they are tested against RECIST v1.1 measurements, which may not respresent and

entirely accurate estimations of the exact time of relapse because they are manual assessments of CT-scans that are only performed every three months (hence relapse could have happened up to three months minus one day before), and (b) the predictions are based on the assumption that the detected mutants in the cfDNA are responsible for the majority of the resistance. If there are other undetected resistant sub-clones, or a component of non-genetic resistance, our model needs such information to perform an accurate prediction. Indeed, as expected, when resistance is polyclonal but not all sub-clones are detected or profiled using cfDNA, the errors in the predictions are higher, as for the case of patient 1014 subdominant *cMET* amplification. Interestingly however, our model allows quantifying *a posteriori* the magnitude of the 'unexplained' resistance and determine the growth rate of the undetected resistant sub-clones, the 'dark matter' of resistance, depending on the actual waiting time observed with RECIST v1.1. Despite some limitations, in several cases our predicted waiting time to recurrence was remarkably accurate. Especially considering the extensive inter-patient variability of clinical response and the extraordinary underlying complexity of the disease these dyanamic can be captured by a relatively simple and applicable model with relatively high accuracy.

In **Fig. 5L** we illustrate the clinical impact of predictive modelling for different dynamics of resistance in each patient and for different sensitivity of cfDNA detection. Assuming an example of initial resistant frequency of 0.03% (1 every 3,300 cells, the median estimated in our cohort) and the range of relapse rates in our cohort, with the model we can calculate the expected waiting time to progression according to RECIST v1.1 criteria for each rate. Initially the resistant population will be undetectable in the cfDNA because of biological and technical limitations (white area). Assuming weekly blood profiling, at some point cfDNA will become positive

for resistance variants (blue) depending on the accuracy of the detection method (different accuracies are reported in different panels) and after enough time points, we will be able to fit the model and infer the crucial parameters (frequency and relapse rate) that will predict recurrence (red). The sooner we can detect the mutant alleles in the blood, the earlier we can forecast relapse, thus creating a larger window of opportunity (yellow area) to take clinical decisions for a specific patient, such as changing treatment or adjusting treatment dynamically. We extended this illustrative analysis to the case of using CEA alone to predict resistance (**Supplementary Figure S16**), as well as for different time intervals of blood sampling (**Supplementary Figure S17**). The latter indicates that, when possible, at least a biweekly blood sampling provides significantly greater predictive power with respect to four weeks, and that beyond the four weeks interval the predictive power is limited. This may help the design of future studies.

DISCUSSION

The ability to design optimal personalised treatments relies strongly on the possibility of predicting the course of the disease in individual patients. Cancer evolution is the fundamental paradigm to understand how tumours change over time[31] and evolutionary mathematical modelling of tumour growth[32,33] provides the theoretical framework to construct predictive models in cancer.

For the first time within a prospective phase II study, we demonstrated that the combination of longitudinal plasma biopsies and solid tissue biopsies can be coupled with mathematical modelling of tumour evolution to anticipate tumour relapse quantitatively, thus impacting on future clinical decisions. We validated previous

retrospective and pre-clinical observations[16-18,34,35] supporting the notion that RAS pathway aberrations clonally expand during anti-EGFR treatment. More importantly, we showed that approximately 50% of mCRC patients considered *KRAS* wt, and as such eligible for anti-EGFR treatment, in fact present *RAS* aberrations and do not benefit from cetuximab. Our data might, at least in part, explain the observation that even in pan-*RAS* wt patients response and benefit from cetuximab are limited[36].

Whereas 2/3rd of all the abnormalities detected on baseline bloods prior to cetuximab treatment were observed in *RAS* genes, one third were found in genes not routinely tested in clinical practice, such as *PIK3CA* and *HER2*, proved to be involved in primary resistance[37] suggesting that extending genomic testing beyond the *RAS* genes might be useful for patient selection. The discrepancy in *RAS* status between archival material (usually primary cancer resections or biopsies) and baseline bloods prior to anti-EGFR treatment can be easily explained by a number of causes including ITH or sampling/technical errors. Whether previous chemotherapy treatments[38] or evolutionary dynamics during the metastatic progression[39,40] might have had a role in priming *RAS* mutant sub-clones remains unknown, however, it is important to point out that our data, as well as previous observations[41-43] suggest that these clones pre-existed treatment. While further studies should address if evolutionary bottlenecks induced by previous lines of chemotherapy might have had a role in selecting *RAS* mutant clones, it is intuitive to think about blood-based *RAS* genotyping as a rational strategy to overcome hurdles in patients selection in keeping with similar approaches used for *EGFR* testing in NSCLC[44].

By comparing matched tissue and liquid biopsies, we demonstrated that cfDNA captures the overall evolutionary dynamics of the disease remarkably well. However, we note that sparse cfDNA sampling, such as sampling plasma every 2-3

months or even more, does not provide the sufficient predictive power to forecast treatment resistance in individual patients due to the inherent inter-patient variability of malignant evolution. On the other hand, collecting frequent longitudinal cfDNA samples from each individual patient (every 4 weeks or even less), was uniquely amenable to mathematical modelling of cancer evolutionary dynamics. Combining sequential mutant frequency data, tumour burden (CEA), and evolutionary modelling allowed us to measure the dynamics of resistance in each patient, and then predict the estimated waiting time of progression. This represents a fundamental step in the clinical translation of liquid biopsies for patient care as it allows overcoming the striking variability between patients and make patient-specific predictions.

Even though our study clearly demonstrates the superiority of liquid versus tissues biopsies in providing clinically relevant information and, in keeping with other studies[45], highlights limitations of tissues biopsies in capturing spatial ITH, it also offered a unique opportunity to track temporal ITH upon cetuximab treatment concomitantly in plasma and tissues. This study also confirms that polyclonal resistance is a common feature in anti-EGFR refractory patients and suggest the presence of a complex ecosystem ruling the emergence of “dominant” resistant clones[46]. Genomic bottlenecking has been observed upon response to chemotherapy in gastro-oesophageal cancers[47,48]; here we observed evolutionary bottlenecks in *RAS* pathway aberrations at time of disease progression suggesting a hierarchical structure in the selection of the “fittest” resistant clone(s)[35]. We acknowledge that our study was limited to the *RAS* pathway and few other cancer related genes as such other genetic[37] or non-genetic determinants[49] are likely to also play a role in this selection process and therefore in our predicted waiting time to relapse. We also recognise that the ability to predict time to disease progression

using our mathematical model will have to be prospectively validated in future trials. In this context, ongoing trials such as CHRONOS [Rechallenge With Panitumumab Driven by RAS Dynamic of Resistance (ClinicalTrials.gov Identifier: NCT03227926) will offer the opportunity to couple theoretic modelling with therapeutic intervention in order to define the validity and clinical utility of “drug holidays” and windows of opportunity.

In conclusion, we show that combining liquid biopsies with mathematical modelling of tumour evolution allows quantitative anticipation of tumour relapse, informing clinicians about timing of clinical decisions and future treatment strategies, facilitating the application of precision medicine with significant health economic benefits for patients and health systems.

MATERIAL AND METHODS

Clinical Trial Design

PROSPECT-C trial [clinical trials.gov number (NCT02994888)] is a phase II, open label, non-randomised study of anti-EGFR monoclonal antibodies in patients with *RAS* wt, refractory mCRC. Patients who were at least 18 years old and had a World Health Organisation (WHO) performance status (PS) of 0-2 were considered eligible for this study if they fulfilled all the following criteria: i) chemo-refractory (at least two lines of chemotherapy) metastatic disease; ii) *KRAS/NRAS* WT (on archival material according to hospital policy); iii) measurable disease; and iv) metastatic sites amenable to biopsy. Patients received cetuximab/panitumumab through the Cancer Drug Fund (CDF). Written informed consent was obtained from all patients. The study was carried out in accordance with the Declaration of Helsinki and approved

by National Institutional review boards [National Research Ethics Service (NRES): 12/LO/0914]. All participants were required to have mandatory pre-treatment biopsies (2 cores), biopsies at 3 months [if partial response (PR) by Response Evaluation Criteria In Solid Tumors (RECIST) v1.1 criteria (2 cores)] and at the time of progressive disease (PD) (2 cores from two suitable progressing metastatic sites). Treatment consisted of cetuximab 500mg/m^2 once every 2 weeks until progression or intolerable side effects. All but one patient received the cetuximab mAb and was not entirely anti-EGFR naïve at the time of trial entry; indeed this patient was switched to panitumumab due to a Common Toxicity Criteria for Adverse Events (CTCAE) 3.0 Grade II allergic reaction after the first dose of cetuximab and had previously received 3 cycles of fluorouracil, oxaliplatin and cetuximab combination with PR as neo-adjuvant chemotherapy for liver resection in the context of the New-EPOC trial[50] 13 months before entering the PROSPECT-C trial.

Isolation of cfDNA

cfDNA was extracted from EDTA anti-coagulated blood within 1 h after collection, plasma was separated from the cells by centrifugation (1500*g* for 15 min at 4 °C) followed by a second centrifugation of the supernatant at 1500*g* for 10 min at 4 °C to remove all cell debris. If not used immediately, plasma was frozen at -80 °C until further processing. cfDNA from 4 ml of plasma was isolated by the use of Qiagen blood mini kit (Qiagen, Hilden, Germany) according to manufacturer's protocol.

Digital Droplet (dd)PCR

The QX200 ddPCR system (Bio-Rad, Berkeley, California) was used and all reactions were prepared using the ddPCR Supermix with no dUTTP for Probes. All

PCR reactions were performed as duplex PCR using the relevant digital PCR assays for the wild-type and the mutation in question. Droplets were generated starting from 8ul of cfDNA template using the QX200 droplet generator according to the manufacturer's protocols. The PCR reaction was performed in a C1000 Touch Thermo Cycler (Bio-Rad) using the following protocol: 95°C for 10 min followed by 40 cycles of 94°C for 30 sec and 55°C for 1 min, then 98°C for 10 min. Droplets were read in the QX200 droplet reader and analyzed using the Quantasoft software version 1.6.6.0320 (Bio-Rad). Fractional Abundance (FA) was defined as follows: F.A. % = (N*mut*/(N*mut* + N*wt*)) × 100), where N*mut* is the number of mutant events and N*wt* is the number of WT events per reaction. The number of positive and negative droplets was used to calculate the concentration of the target and reference DNA sequences and their Poisson-based 95% confidence intervals. ddPCR analysis of normal control plasma DNA (from cell lines) and no DNA template controls were always included. Samples with very low positive events were repeated at least twice in independent experiments to validate the obtained results as previously described[17].

Targeted deep sequencing analysis

The targeted panel was sequenced on a HiSeq 2500. Residual adapter sequences were trimmed with Skewer[51] and reads aligned with Burrows-Wheeler Aligner (BWA). Base level (base quality ≥ 25) coverage of amplified genomic regions were extracted and used to call variants with a bayesian beta-binomial model (shearwater algorithm) implemented in the deepSNV package for R. Locus specific error rates were estimated from a composite set of buffy coats. The beta-binomial model, which includes a global dispersion factor in addition to a site-specific error rate, was

chosen, since a simpler binomial model did not reflected the variability (overdispersion) observed in the normal samples (buffy coats). Mutations were only called when i) coverage was at least 20,000x and ii) the posterior probability of the mutation being a false-positive was less than 5%. Overall FFPE specific background rates were low in buffy coats and FFPE samples, indicating sufficient removal/repair of FFPE related DNA damages during sample preparation.

cfDNA sequencing using Avenio Panel

The Avenio panel was run in the clinically accredited Molecular Diagnostic Laboratory in the Centre for Molecular Pathology at the Royal Marsden Hopsital. DNA sequencing libraries were prepared using the Avenio ctDNA Analysis kit (Roche), starting with 25ng DNA, following the manufacturer's instructions, and hybridised to the Avenio Expanded capture kit to enrich for a panel of 77 target genes and regions. Libraries were quality checked on an Agilent TapeStation. Sequencing was performed (150 bases, paired end) on an Illumina Nextseq 500 (High Output), eight samples per run (approx. 100 million PE reads/sample). Data were analysed with the Roche Avenio Custom App via a locally installed Roche server to generate variant allele frequency and unique allele depth data (mean fold depth, 15,008 ± 2955; Unique depth 4710 ± 2319). The Roche analysis pipeline supports VAF detection of SNVs to 0.5%, targeted Indels and fusions to 1%, and CNVs over 2.3 fold with sensitivities of >99%. Variants are detectable to 0.1% VAF.

REFERENCES

1. DeSantis, C.E.*, et al.* Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin* (2014).
2. Khan, K., Cunningham, D. & Chau, I. Targeting Angiogenic Pathways in Colorectal Cancer: Complexities, Challenges and Future Directions. *Curr Drug Targets* **18**, 56-71 (2017).
3. Lieberman, D. Colorectal cancer screening: practice guidelines. *Dig Dis* **30 Suppl 2**, 34-38 (2012).
4. Amado, R.G.*, et al.* Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **26**, 1626-1634 (2008).
5. Bardelli, A.*, et al.* Amplification of the MET receptor drives resistance to anti-EGFR therapies in colorectal cancer. *Cancer Discov* **3**, 658-673 (2013).
6. Engelman, J.A.*, et al.* MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316**, 1039-1043 (2007).
7. Inno, A.*, et al.* Is there a role for IGF1R and c-MET pathways in resistance to cetuximab in metastatic colorectal cancer? *Clin Colorectal Cancer* **10**, 325-332 (2011).
8. Karadima, M.L.*, et al.* The Prognostic Influence of BRAF Mutation and other Molecular, Clinical and Laboratory Parameters in Stage IV Colorectal Cancer. *Pathol Oncol Res* **22**, 707-714 (2016).
9. Montagut, C.*, et al.* 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 (2012).
10. Rowland, A.*, et al.* Meta-analysis comparing the efficacy of anti-EGFR monoclonal antibody therapy between KRAS G13D and other KRAS mutant metastatic colorectal cancer tumours. *Eur J Cancer* **55**, 122-130 (2016).
11. Turke, A.B.*, et al.* Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell* **17**, 77-88 (2010).
12. Yonesaka, K.*, et al.* Activation of ERBB2 signaling causes resistance to the EGFR-directed therapeutic antibody cetuximab. *Sci Transl Med* **3**, 99ra86 (2011).
13. Knijn, N.*, et al.* KRAS mutation analysis: a comparison between primary tumours and matched liver metastases in 305 colorectal cancer patients. *British journal of cancer* **104**, 1020-1026 (2011).
14. Vogelstein, B.*, et al.* Genetic alterations during colorectal-tumor development. *The New England journal of medicine* **319**, 525-532 (1988).
15. Diaz, L.A., Jr.*, et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537-540 (2012).
16. Misale, S.*, et al.* Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532-536 (2012).
17. Siravegna, G.*, et al.* Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nature medicine* **21**, 827 (2015).
18. Misale, S., Di Nicolantonio, F., Sartore-Bianchi, A., Siena, S. & Bardelli, A. Resistance to anti-EGFR therapy in colorectal cancer: from heterogeneity to convergent evolution. *Cancer Discov* **4**, 1269-1280 (2014).
19. Greaves, M. & Maley, C.C. Clonal evolution in cancer. *Nature* **481**, 306-313 (2012).
20. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613-628 (2017).
21. Garcia-Murillas, I.*, et al.* Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med* **7**, 302ra133 (2015).
22. Abbosh, C.*, et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446-451 (2017).

23. Khan, K.*, et al.* Functional imaging and circulating biomarkers of response to regorafenib in treatment-refractory metastatic colorectal cancer patients in a prospective phase II study. *Gut* (2017).
24. Vlachogiannis, G.*, et al.* Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science (New York, N.Y.)* **359**, 920-926 (2018).
25. Lee, M.S. & Kopetz, S. Current and Future Approaches to Target the Epidermal Growth Factor Receptor and Its Downstream Signaling in Metastatic Colorectal Cancer. *Clin Colorectal Cancer* **14**, 203-218 (2015).
26. Scholl, C.*, et al.* Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. *Cell* **137**, 821-834 (2009).
27. Nik-Zainal, S.*, et al.* The life history of 21 breast cancers. *Cell* **149**, 994-1007 (2012).
28. Santini, D.*, et al.* Cetuximab rechallenge in metastatic colorectal cancer patients: how to come away from acquired resistance? *Ann Oncol* (2017).
29. Sartore-Bianchi, A.*, et al.* Dual-targeted therapy with trastuzumab and lapatinib in treatment-refractory, KRAS codon 12/13 wild-type, HER2-positive metastatic colorectal cancer (HERACLES): a proof-of-concept, multicentre, open-label, phase 2 trial. *Lancet Oncol* **17**, 738-746 (2016).
30. Misale, S.*, et al.* Vertical suppression of the EGFR pathway prevents onset of resistance in colorectal cancers. *Nat Commun* **6**, 8305 (2015).
31. Gerlinger, M.*, et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England journal of medicine* **366**, 883-892 (2012).
32. Anderson, A.R. & Quaranta, V. Integrative mathematical oncology. *Nat Rev Cancer* **8**, 227-234 (2008).
33. Beerenwinkel, N., Schwarz, R.F., Gerstung, M. & Markowetz, F. Cancer evolution: mathematical models and computational inference. *Syst Biol* **64**, e1-25 (2015).
34. Russo, M.*, et al.* Tumor Heterogeneity and Lesion-Specific Response to Targeted Therapy in Colorectal Cancer. *Cancer Discov* **6**, 147-153 (2016).
35. Van Emburgh, B.O.*, et al.* Acquired RAS or EGFR mutations and duration of response to EGFR blockade in colorectal cancer. *Nat Commun* **7**, 13665 (2016).
36. Allegra, C.J.*, et al.* Extended RAS Gene Mutation Testing in Metastatic Colorectal Carcinoma to Predict Response to Anti-Epidermal Growth Factor Receptor Monoclonal Antibody Therapy: American Society of Clinical Oncology Provisional Clinical Opinion Update 2015. *J Clin Oncol* **34**, 179-185 (2016).
37. Bertotti, A.*, et al.* The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature* **526**, 263-267 (2015).
38. Kreso, A.*, et al.* Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science (New York, N.Y.)* **339**, 543-548 (2013).
39. Naxerova, K.*, et al.* Origins of lymphatic and distant metastases in human colorectal cancer. *Science* **357**, 55-60 (2017).
40. Lote, H.*, et al.* Carbon dating cancer: defining the chronology of metastatic progression in colorectal cancer. *Ann Oncol* **28**, 1243-1249 (2017).
41. Normanno, N.*, et al.* RAS testing of liquid biopsy correlates with the outcome of metastatic colorectal cancer patients treated with first-line FOLFIRI plus cetuximab in the CAPRI-GOIM trial. *Annals of oncology : official journal of the European Society for Medical Oncology* **29**, 112-118 (2018).
42. Thierry, A.R.*, et al.* Circulating DNA Demonstrates Convergent Evolution and Common Resistance Mechanisms during Treatment of Colorectal Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **23**, 4578-4591 (2017).
43. Grasselli, J.*, et al.* Concordance of blood- and tumor-based detection of RAS mutations to guide anti-EGFR therapy in metastatic colorectal cancer. *Annals of oncology : official journal of the European Society for Medical Oncology* **28**, 1294-1301 (2017).

44. Kwapisz, D. The first liquid biopsy test approved. Is it a new era of mutation testing for non-small cell lung cancer? *Ann Transl Med* **5**, 46 (2017).
45. Bardelli, A. & Pantel, K. Liquid Biopsies, What We Do Not Know (Yet). *Cancer Cell* **31**, 172-179 (2017).
46. Tabassum, D.P. & Polyak, K. Tumorigenesis: it takes a village. *Nat Rev Cancer* **15**, 473-483 (2015).
47. Findlay, J.M.*, et al.* Differential clonal evolution in oesophageal cancers in response to neo-adjuvant chemotherapy. *Nat Commun* **7**, 11111 (2016).
48. Noorani, A.*, et al.* A comparative analysis of whole genome sequencing of esophageal adenocarcinoma pre- and post-chemotherapy. *Genome Res* **27**, 902-912 (2017).
49. Hobor, S.*, et al.* TGFalpha and amphiregulin paracrine network promotes resistance to EGFR blockade in colorectal cancer cells. *Clinical cancer research : an official journal of the American Association for Cancer Research* **20**, 6429-6438 (2014).
50. Primrose, J.*, et al.* Systemic chemotherapy with or without cetuximab in patients with resectable colorectal liver metastasis: the New EPOC randomised controlled trial. *Lancet Oncol* **15**, 601-611 (2014).
51. Jiang, H., Lei, R., Ding, S.W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182 (2014).

FIGURE LEGENDS

Fig 1. Overview of trial related biopsies and CONSORT diagram describing the PROSPECT-C Trial. (**A**) Patients (pts) meeting all the inclusion and no exclusion criteria were required to have pre-treatment CT scan. All pts were also required to have pre-treatment mandatory core biopsy, followed by a core biopsy at 3 months if they had PR. Pts were monitored by CT scans every 3 months until the time of PD and if clinically feasible, they had biopsy of 1 or 2 progressing lesions from PD sites. Plasma samples were collected every 4 weeks until the time of PD. (**B**) Out of 47 pts initially consented, two were excluded from analysis: one (1015) was found to harbour an *NRAS* mutation on archival material during the screening process while the other one (1031) rapidly progressed before commencing cetuximab. Two pts in the first cohort (1005-1018) for whom no mutations were detected by ddPCR were also tested by cfDNA NGS. 5 pts in the second cohort were not tested by cfDNA NGS as their progression free survival was between 3 and 6 months thus they were not considered either primary resistant (PFS ≤3 months) nor long-term responders (PFS >6 months).

cfDNA= cell-free DNA; ddPCR= digital-droplet PCR; PFS=Progression Free Survival; CT=computed tomography; PD=progressive disease; PR=partial response.

Fig. 2. Clinical efficacy outcomes according to mutations in cell-free (cf)DNA, in primary and acquired resistance to anti-EGFR therapy. (A) Waterfall plot demonstrating changes in tumour burden by RECIST v1.1, according to patients with detectable or undetectable mutations/amplifications in baseline plasma. Asterisks indicate patients who rapidly progressed and died prior to re-staging scan (3 months) **(B)** Spider plot showing depth and duration of tumor regressions, according to

presence or absence of mutations in cfDNA. Progression Free Survival **(C)** and Overall Survival **(D)** of patients according to detectable or undetectable baseline mutations/amplifications. **(E)** Avenio cfDNA NGS results in two patients with primary resistance to cetuximab shows polyclonal resistance. **(F)** An example of a patient who received 10 months of cetuximab treatment with initial PR (at 3 months) followed by PD on a subsequent CT scan; tracking of plasma mutations on a 4-weekly basis revealed two mutations *(KRAS G12D and KRAS Q61H)* preceding the changes in CEA **(G)** An example of a primary progression on the treatment with polyclonal resistance demonstrated by the presence and relative increase in frequency of *KRAS G12D* and *APC* mutations along with emergence and relative increase in frequency of *KRAS G13D* and *KRAS Q61H* mutations under the selective pressure of anti-EGFR therapy. (**H**) Heat-map showing results of cfDNA analysis using the Avenio panel in patients with acquired resistance in the PROSPECT-C Trial. Baseline (BL), intermediate [INT (cycle 3)] and progression (PD) cfDNA from patients with PFS >6 months was tested for a panel of 77 cancer related genes. Red boxes indicate presence of mutations in different genes.

CEA= carcinoembryonic antigen; CT= computed tomography; ddPCR= digital droplet polymerase chain reaction; RECIST= response evaluation criteria in solid tumours; EGFR= epidermal growth factor receptor; PR= partial response; PFS= Progression Free Survival; VAF= variant allele frequency.

Fig. 3. EGFR pathway addiction, and impact of cancer heterogeneity on clinical decisions. An *EGFR* amplification was observed in the baseline liver biopsy of a long term responder to cetuximab. Treatment was halted after 16 months

due to clinical and minor RECIST v1.1 progression of a non-target lesion (orange circle). Interestingly, after cetuximab was withdrawn a rapid progression of the *EGFR* amplified metastatic liver lesion originally biopsied was observed along with new liver deposits within 6 weeks (red circles). On the contrary, the metastasis biopsied at time of progression did not show any significant change in volume and showed no *EGFR* amplification suggesting that this metastasis was not dependent upon EGFR signalling. A rise in *APC* mutant clones was observed synchronously with the increase in size of the non-target metastasis that led to treatment discontinuation; CEA lagged behind and no RAS pathway mutant clones were detected at any time point.

Fig. 4. Concordance between liquid versus tissue biopsies. (A) Heat map demonstrating validation of mutations/amplifications detected in plasma by ddPCR, in tissue samples obtained at clinically relevant timepoints. For NGS variants detected with confidence are reported with a star (posterior probability < 0.05, see Material and Methods). **(B)** comparions between limit of detectability of clinically validated assays (i.e. COBAS) and ultra-deep sequencing used in the study showing that most of the *KRAS* sub-clonal mutations causing cetuximab resistance (solid symbols) were below the detection threshold of standard clinical assays. (**C**) Example of a patient with RAS pathway intra-tumour heterogeneity between resected primary and synchronous liver metastases. **(D)** Example of a patient with primary progression and detection of *ERBB2* amplification in plasma, further validated with IHC and CISH in the tissue obtained both at baseline and progressive disease. Of note, in this case of non-CEA secreting tumour, baseline biopsy of a supraclavicular lymph node and subsequently of a progressing peritoneal biopsy

were conducted and both demonstrated concordance in detection of a (likely clonal) *HER2* amplification. **(E)** A case of a patient with no baseline mutation/amplification and initial clinical benefit with PR to cetuximab but later emergence of *c-MET* amplification at 2 months (preceding CEA changes) and subsequent increase in fractional abundance at the time of progression. FISH analysis of the tissue at four different time points confirmed the emergence of foci of *cMET* at PR and PD. (**F**) NGS of tissue biopsies [archival (n=2), baseline (n=2), PR [n=2 (3 months)] and disease progression [n=1] revealed the emergence of an *NRAS* G12C mutation and decay in other mutant clones suggestive of a selection bottleneck (cartoon).

CEA= carcinoembryonic antigen; CT= computed tomography; RECIST= response evaluation criteria in solid tumours; PD= progressive disease; PR= partial response. NGS= next generation sequencing.

Fig. 5. Forecasting waiting time to progression in responders using evolutionary modelling and frequent cfDNA sampling. **(A)** From the point of view of therapy, the tumour at baseline can be modelled as comprised of a treatment sensitive population with size n_s and a resistant population with size n_r, usually small. Under treatment, the sensitive decreases at rate λ_s, and the resistant increases at rate λ_r. These dynamics are capture by the equation of $N(t)$. **(B)** The sum the two populations $N(t)$ corresponds to tumour burden over time and has the characteristic U-shape curve of an initial response (tumour shrinks) followed by relapse (tumour comes back). **(C)** As CEA measurements are a surrogate for the tumour burden $N(t)$, in this illustrative example we applied the model to the CEA values over the course of cetuximab treatment for patient 1014. **(D)** Fitting the model to CEA values allowed measuring the n and λ parameters, thus estimating the

response and relapse rates in each individual patient. Here response rates λ_s are shown. Unsurprisingly, most progression and stable disease patients showed $\lambda_s = 0$ (no response), indicated by the absence of bars. **(E)** Relapse rates λ_r varied between patients, with stable disease patients showing a trend of slow relapse. Both progression patients and responders showed relatively high relapse. In responders with high relapse rates, the initial frequency of the resistant subpopulation was likely low. **(F)** The combination of relapse rate and initial frequency of the resistant subclone allowed the stratification of patients, where responders showed low initial frequency of the resistant clone or very slow relapse rates, or both. We note that even for moderately low frequencies of resistant subclones (e.g. 1-10%) rapid growth is sufficient to induce lack of response as the tumour grows back even before the first CT-scan. **(G)** We applied the total tumour burden model (Equation 1) and the resistant population model (Equation 2) to CEA and cfDNA respectively. In responders the cfDNA preceded tumour burden of several weeks, supporting the predictive value of cfDNAm whereas in non-responders **(H)** the two curves overlapped. **(I)** Strikingly, we found a significant correlation between the relapse rate measured from CEA and the relapse rate measured from cfDNA, indicating that measuring evolutionary dynamics from plasma can inform on the expected dynamics that will occur later on at the macro-scale (CEA and RECIST v1.1). **(J)** We measured the evolutionary parameters (n_r and λ_r) of the resistant population from CEA and cfDNA for responder 1007 and use the evolutionary framework to predict when we expect to observe relapse in the CT-scan under RECIST v1.1 criteria (+20% increase in target lesions diameter). Despite the limited precision of RECIST v1.1 and the infrequency of the CT-scan, both predictions based on cfDNA (the dominant *KRAS* G12D subclone) and CEA were strikingly accurate. **(K)** We verified the same

predictions for all responders in our cohort for which we had at least CEA and confirm the power of predicting waiting time to relapse using our approach was remarkable. We note that when resistance was polyclonal but only one clone was detected (e.g. patient 1014 – *cMET* amplified clone), our method allows measuring the contribution of *unobserved* resistant subclones by studying the difference between predicted time of relapse and RECIST. **(L)** The predictive power of the mathematical framework applied to cfDNA is illustrated here. In this example we consider a fix initial resistant population n_r=0.0003 (median n_r in our cohort) and vary the relapse rate λ_r between 0.01 and 0.41 (range in our cohort). The time when we observed the mutant alleles in plasma (blue) depends on the sensitivity of the assay. Assuming to profile cfDNA every week, with a sufficient number of positive time points (n=5), we can fit our model to the data and determine the evolutionary parameters, thus allowing for predicting the expected waiting time to RECIST v1.1 progression. The higher the sensitivity, the earlier progression can be predicted, thus creating a window of opportunity for clinicians to take patient-specific treatment decisions (yellow).

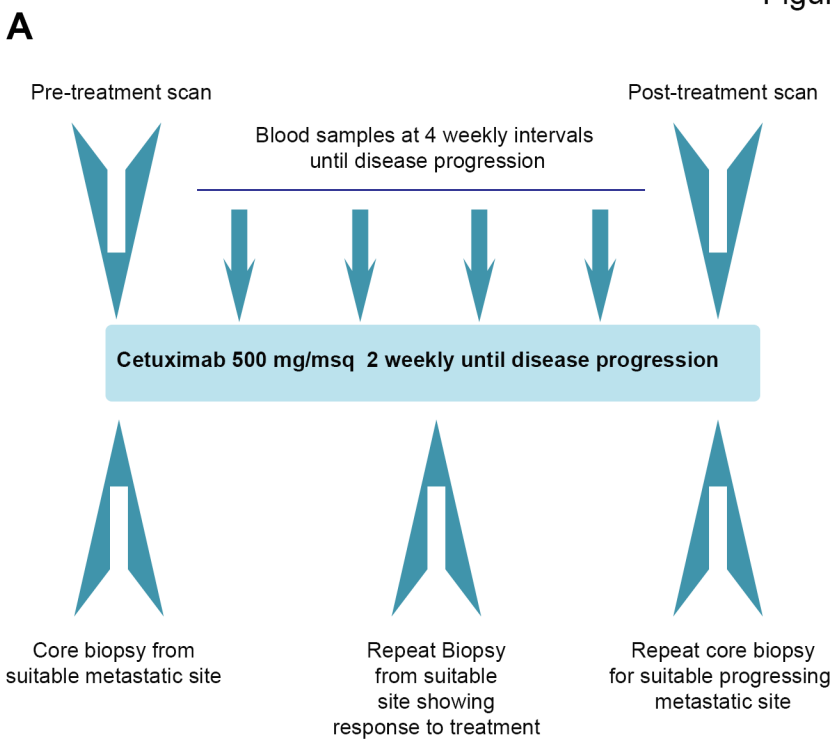
CEA= carcinoembryonic antigen; RECIST= response evaluation criteria in solid tumours; variant allele frequency=VAF

Table 1. Baseline patients characteristics (n=46*)

	N	%
Age (median & range)	66	33 – 84
Gender		
Male	29	63
Female	17	37
Primary tumour resected	31	67.4
Site of Primary		
Left Colon	15	32.6
Rectal	20	43.5
Right Colon	11	23.9
Histology		
Adenocarcinoma (mucinous)	11	23.9
Adenocarcinoma (non-mucinous)	35	76.1
Differentiation		
Poor	7	15.2
Moderate	38	82.6
Well	1	2.2
Dukes staging at diagnosis		
B	5	10.9
C	15	32.6
D	26	56.5
Extent of metastatic disease		
Liver	31	67.4
Lung	20	43.5
Omentum/Peritoneum	14	30.4
Distant Lymph Nodes	11	23.9
Other	11	23.9
Soft tissue	3	6.5
Radiotherapy to Primary	11	23.9
Lines of prior therapy		
1	3	6.5
2	22	47.8
3	16	34.8
4	3	6.5
6	1	2.2
9	1	2.2

* Overall 47 patients were screend for the study; one patient (1015) was excluded based on presence of an *NRAS* mutation before commencing cetuximab and didn't undergo any trial-related procedures. One patient (1031) had rapid deterioration prior to cetuximab and did not start treatment.

Figure 1


A
Pre-treatment scan
Post-treatment scan
Blood samples at 4 weekly intervals
until disease progression
Cetuximab 500 mg/msq 2 weekly until disease progression
Core biopsy from
suitable metastatic site
Repeat Biopsy
from suitable
site showing
response to treatment
Repeat core biopsy
for suitable progressing
metastatic site

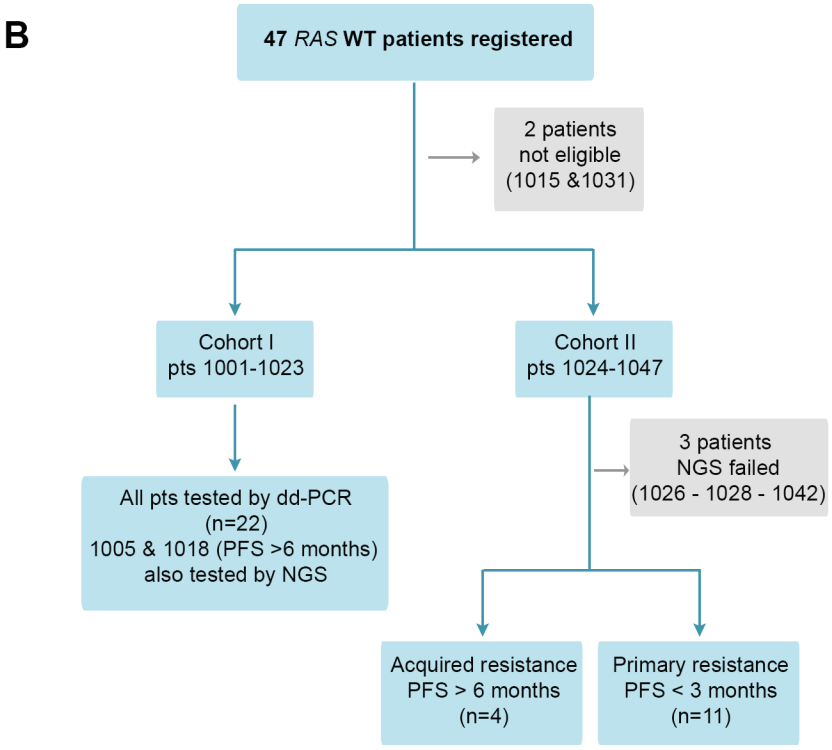

B
47 RAS WT patients registered
2 patients
not eligible
(1015 &1031)
Cohort I
pts 1001-1023
Cohort II
pts 1024-1047
All pts tested by dd-PCR
(n=22)
1005 & 1018 (PFS >6 months)
also tested by NGS
3 patients
NGS failed
(1026 - 1028 - 1042)
Acquired resistance
PFS > 6 months
(n=4)
Primary resistance
PFS < 3 months
(n=11)

Figure 2

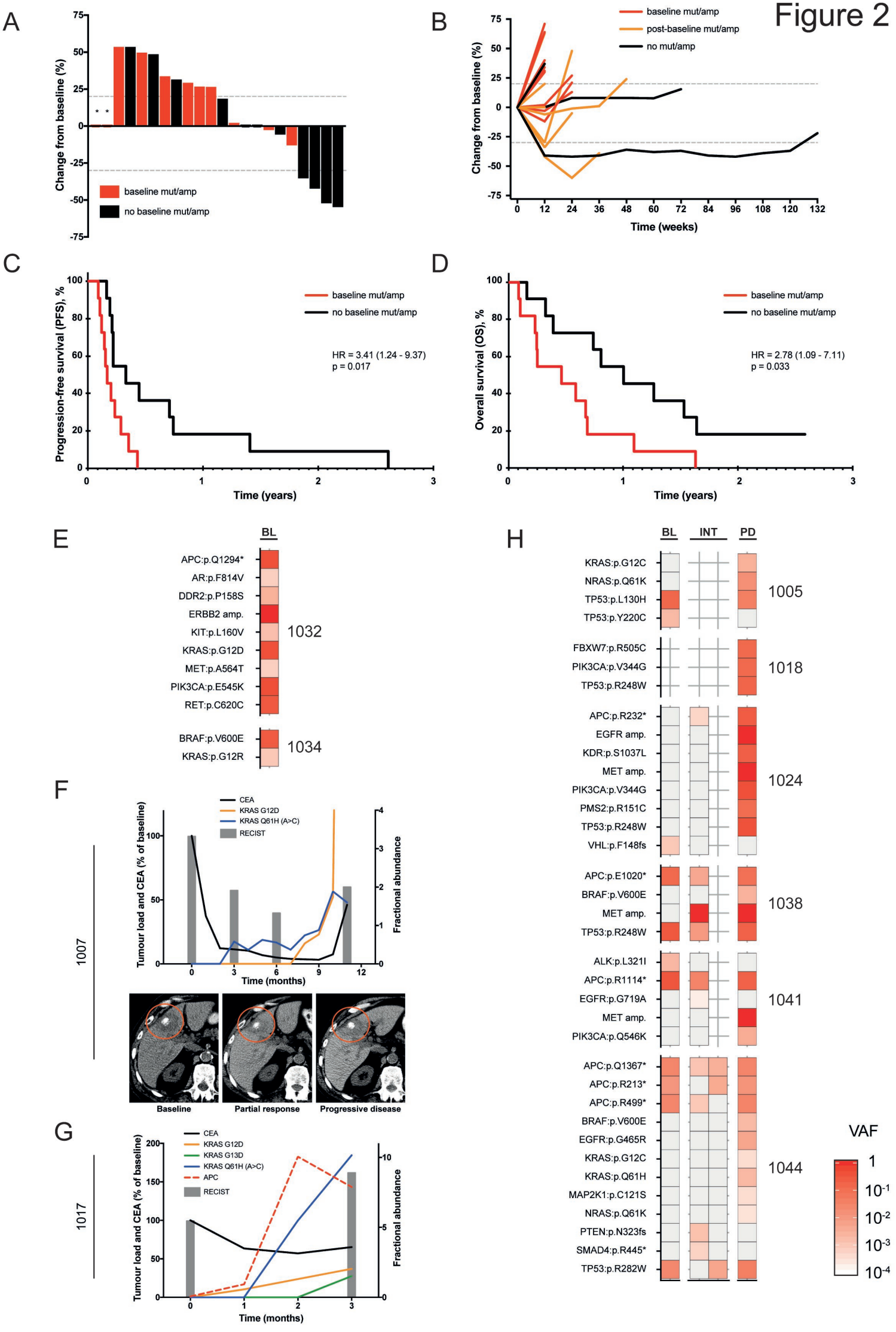
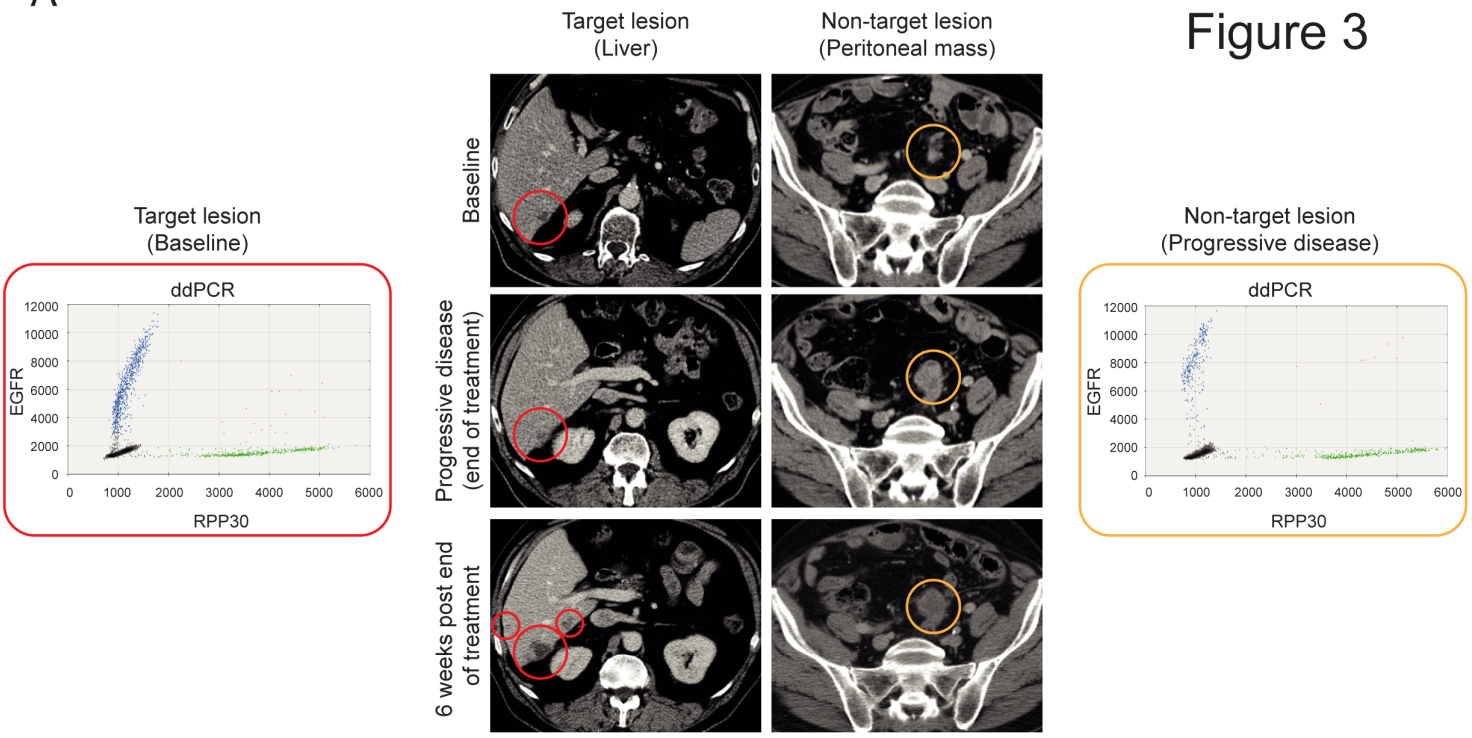

A
Change from baseline (%)
baseline mut/amp
no baseline mut/amp
B
baseline mut/amp
post-baseline mut/amp
no mut/amp
Change from baseline (%)
Time (weeks)
C
Progression-free survival (PFS), %
baseline mut/amp
no baseline mut/amp
HR = 3.41 (1.24 - 9.37)
p = 0.017
Time (years)
D
Overall survival (OS), %
baseline mut/amp
no baseline mut/amp
HR = 2.78 (1.09 - 7.11)
p = 0.033
Time (years)
E
BL
APC:p.Q1294*
AR:p.F814V
DDR2:p.P158S
ERBB2 amp.
KIT:p.L160V
KRAS:p.G12D
MET:p.A564T
PIK3CA:p.E545K
RET:p.C620C
1032
BRAF:p.V600E
KRAS:p.G12R
1034
F
1007
CEA
KRAS G12D
KRAS Q61H (A>C)
RECIST
Tumour load and CEA (% of baseline)
Fractional abundance
Time (months)
Baseline
Partial response
Progressive disease
G
1017
CEA
KRAS G12D
KRAS G13D
KRAS Q61H (A>C)
APC
RECIST
Tumour load and CEA (% of baseline)
Fractional abundance
Time (months)
H
BL
INT
PD
KRAS:p.G12C
NRAS:p.Q61K
TP53:p.L130H
TP53:p.Y220C
1005
FBXW7:p.R505C
PIK3CA:p.V344G
TP53:p.R248W
1018
APC:p.R232*
EGFR amp.
KDR:p.S1037L
MET amp.
PIK3CA:p.V344G
PMS2:p.R151C
TP53:p.R248W
VHL:p.F148fs
1024
APC:p.E1020*
BRAF:p.V600E
MET amp.
TP53:p.R248W
1038
ALK:p.L321I
APC:p.R1114*
EGFR:p.G719A
MET amp.
PIK3CA:p.Q546K
1041
APC:p.Q1367*
APC:p.R213*
APC:p.R499*
BRAF:p.V600E
EGFR:p.G465R
KRAS:p.G12C
KRAS:p.Q61H
MAP2K1:p.C121S
NRAS:p.Q61K
PTEN:p.N323fs
SMAD4:p.R445*
TP53:p.R282W
1044
VAF

Figure 3

A


Target lesion
(Liver)
Non-target lesion
(Peritoneal mass)
Baseline
Progressive disease
(end of treatment)
6 weeks post end
of treatment
Target lesion
(Baseline)
ddPCR
EGFR
RPP30
Non-target lesion
(Progressive disease)
ddPCR
EGFR
RPP30

B

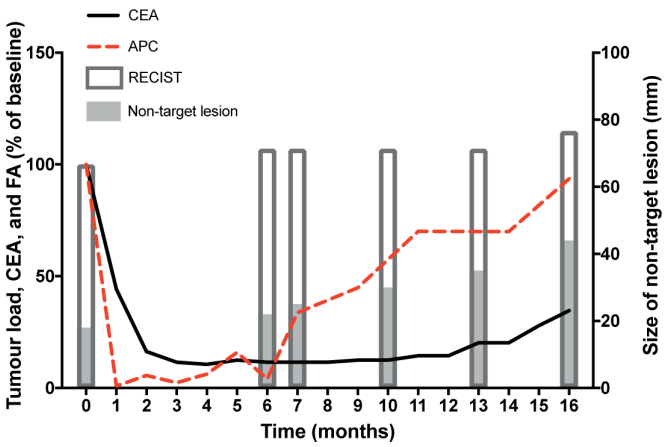

CEA
APC
RECIST
Non-target lesion
Tumour load, CEA, and FA (% of baseline)
Size of non-target lesion (mm)
Time (months)

Figure 4

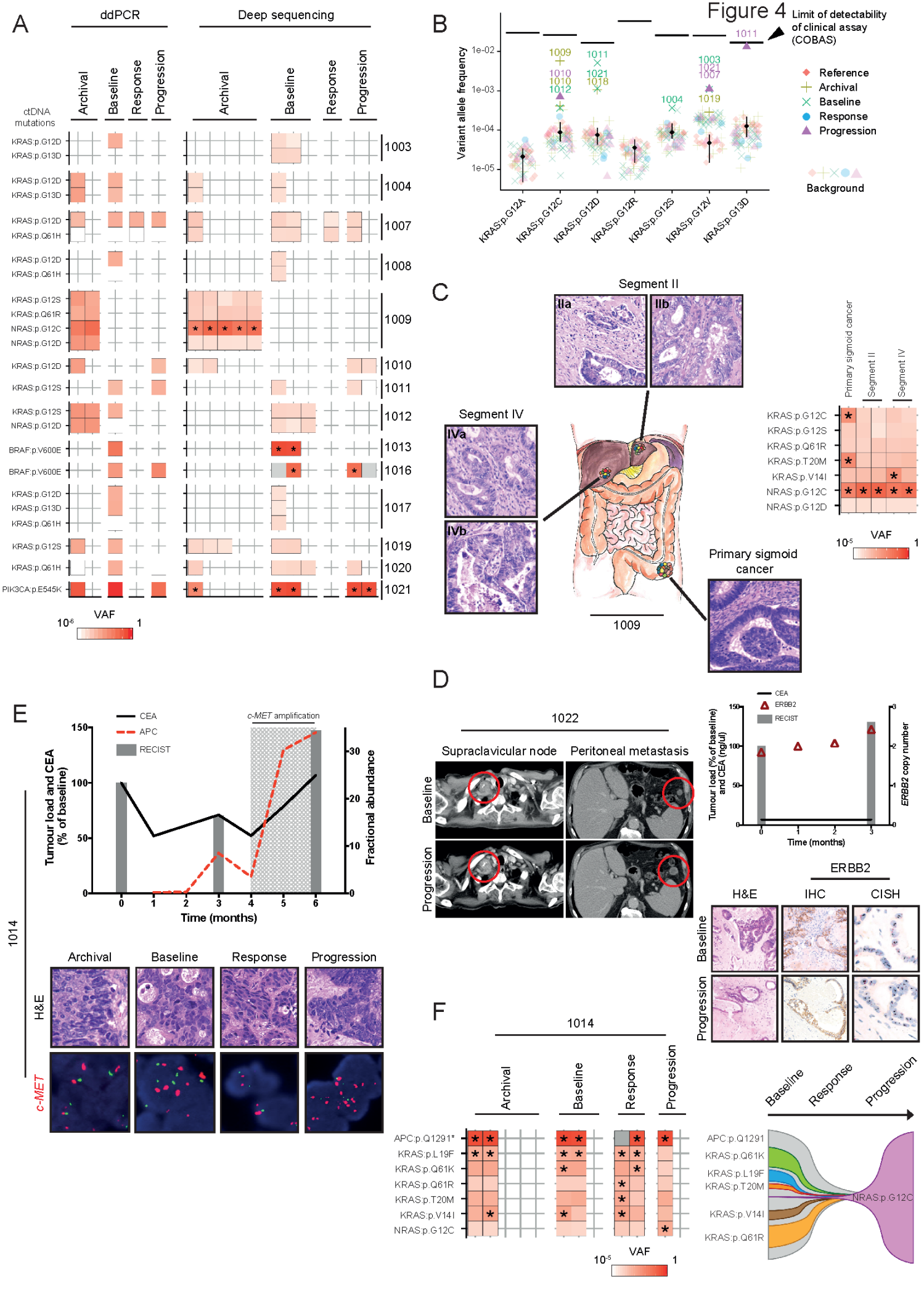
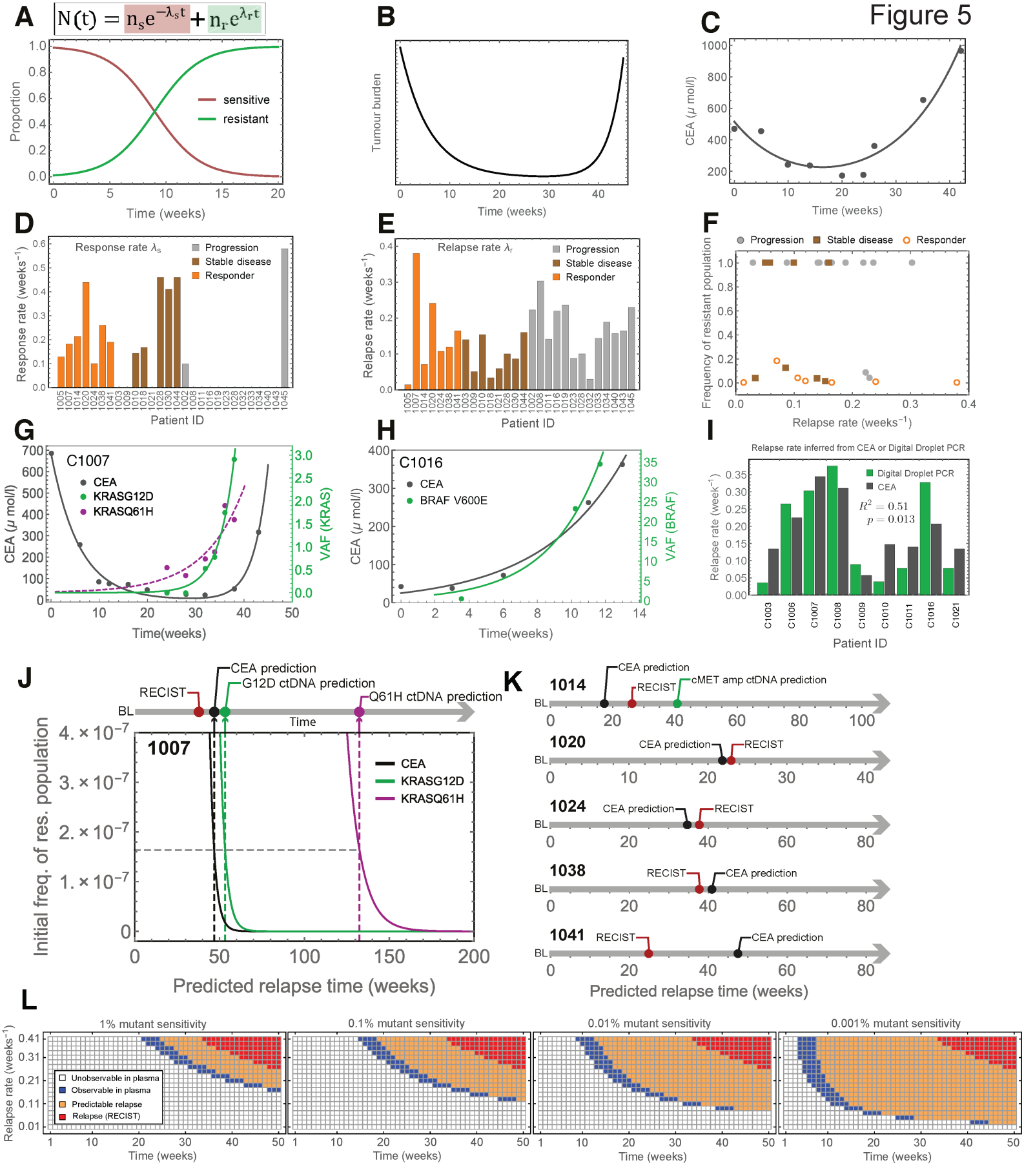

A
ddPCR
Deep sequencing
Archival
Baseline
Response
Progression
ctDNA mutations
KRAS:p.G12D
KRAS:p.G13D
KRAS:p.Q61H
KRAS:p.G12S
KRAS:p.Q61R
NRAS:p.G12C
NRAS:p.G12D
BRAF:p.V600E
KRAS:p.T20M
KRAS:p.V14I
PIK3CA:p.E545K
1003
1004
1007
1008
1009
1010
1011
1012
1013
1016
1017
1019
1020
1021
VAF
10-6
1
B
Limit of detectability of clinical assay (COBAS)
Variant allele frequency
1e-02
1e-03
1e-04
1e-05
Reference
Archival
Baseline
Response
Progression
Background
KRAS:p.G12A
KRAS:p.G12C
KRAS:p.G12D
KRAS:p.G12R
KRAS:p.G12S
KRAS:p.G12V
KRAS:p.G13D
C
Segment II
IIa
IIb
Segment IV
IVa
IVb
1009
Primary sigmoid cancer
Primary sigmoid cancer
Segment II
Segment IV
KRAS:p.G12C
KRAS:p.G12S
KRAS:p.Q61R
KRAS:p.T20M
KRAS:p.V14I
NRAS:p.G12C
NRAS:p.G12D
VAF
10-5
1
D
1022
Supraclavicular node
Peritoneal metastasis
Baseline
Progression
CEA
ERBB2
RECIST
Tumour load (% of baseline) and CEA (ng/ul)
ERBB2 copy number
Time (months)
ERBB2
H&E
IHC
CISH
Baseline
Progression
E
1014
CEA
APC
RECIST
c-MET amplification
Tumour load and CEA (% of baseline)
Fractional abundance
Time (months)
Archival
Baseline
Response
Progression
H&E
c-MET
F
1014
Archival
Baseline
Response
Progression
APC:p.Q1291*
KRAS:p.L19F
KRAS:p.Q61K
KRAS:p.Q61R
KRAS:p.T20M
KRAS:p.V14I
NRAS:p.G12C
VAF
10-5
1
Baseline
Response
Progression
APC:p.Q1291
KRAS:p.Q61K
KRAS:p.L19F
KRAS:p.T20M
NRAS:p.G12C
KRAS:p.V14I
KRAS:p.Q61R

Figure 5


A
$N(t) = n_s e^{-\lambda_s t} + n_r e^{\lambda_r t}$
Proportion
sensitive
resistant
Time (weeks)
B
Tumour burden
Time (weeks)
C
CEA (μ mol/l)
Time (weeks)
D
Response rate λ_s
Progression
Stable disease
Responder
Response rate (weeks^{-1})
Patient ID
E
Relapse rate λ_r
Progression
Stable disease
Responder
Relapse rate (weeks^{-1})
Patient ID
F
Progression
Stable disease
Responder
Frequency of resistant population
Relapse rate (weeks^{-1})
G
C1007
CEA
KRASG12D
KRASQ61H
CEA (μ mol/l)
VAF (KRAS)
Time(weeks)
H
C1016
CEA
BRAF V600E
CEA (μ mol/l)
VAF (BRAF)
Time(weeks)
I
Relapse rate inferred from CEA or Digital Droplet PCR
Digital Droplet PCR
CEA
$R^2 = 0.51$
$p = 0.013$
Relapse rate (week^{-1})
C1003
C1006
C1007
C1008
C1009
C1010
C1011
C1016
C1021
Patient ID
J
CEA prediction
RECIST
G12D ctDNA prediction
Q61H ctDNA prediction
BL
Time
1007
CEA
KRASG12D
KRASQ61H
Initial freq. of res. population
Predicted relapse time (weeks)
K
1014
CEA prediction
RECIST
cMET amp ctDNA prediction
1020
CEA prediction
RECIST
1024
CEA prediction
RECIST
1038
RECIST
CEA prediction
1041
RECIST
CEA prediction
Predicted relapse time (weeks)
L
1% mutant sensitivity
0.1% mutant sensitivity
0.01% mutant sensitivity
0.001% mutant sensitivity
Unobservable in plasma
Observable in plasma
Predictable relapse
Relapse (RECIST)
Relapse rate (weeks^{-1})
Time (weeks)