

# The Oncology Biomarker Discovery framework reveals cetuximab and bevacizumab response patterns in metastatic colorectal cancer

---

Received: 4 March 2022

---

Accepted: 17 August 2023







---

Published online: 04 September 2023

---

 Check for updates

---

Alexander J. Ohnmacht<sup>1,2,10</sup>, Arndt Stahler<sup>3,10</sup>, Sebastian Stintzing <sup>3,4,10</sup>, Dominik P. Modest <sup>3</sup>, Julian W. Holch<sup>4,5</sup>, C. Benedikt Westphalen <sup>5</sup>, Linus Hölzel<sup>1</sup>, Marisa K. Schübel<sup>1,2</sup>, Ana Galhoz<sup>1,2</sup>, Ali Farnoud<sup>1</sup>, Minhaz Ud-Dean<sup>1</sup>, Ursula Vehling-Kaiser<sup>6</sup>, Thomas Decker<sup>7</sup>, Markus Moehler<sup>8</sup>, Matthias Heinig<sup>1</sup>, Volker Heinemann<sup>5</sup>  & Michael P. Menden <sup>1,2,9</sup> 

Precision medicine has revolutionised cancer treatments; however, actionable biomarkers remain scarce. To address this, we develop the Oncology Biomarker Discovery (OncoBird) framework for analysing the molecular and biomarker landscape of randomised controlled clinical trials. OncoBird identifies biomarkers based on single genes or mutually exclusive genetic alterations in isolation or in the context of tumour subtypes, and finally, assesses predictive components by their treatment interactions. Here, we utilise the open-label, randomised phase III trial (FIRE-3, AIO KRK-0306) in metastatic colorectal carcinoma patients, who received either cetuximab or bevacizumab in combination with 5-fluorouracil, folinic acid and irinotecan (FOLFIRI). We systematically identify five biomarkers with predictive components, e.g., patients with tumours that carry chr20q amplifications or lack mutually exclusive ERK signalling mutations benefited from cetuximab compared to bevacizumab. In summary, OncoBird characterises the molecular landscape and outlines actionable biomarkers, which generalises to any molecularly characterised randomised controlled trial.

Precision medicine aims to tailor therapeutic interventions to specific patient subgroups defined by predictive biomarkers detected in tumours. Accordingly, strategies are required to identify such patient subgroups systematically<sup>1</sup>. For performing subgroup analysis and

exploratory biomarker discovery, the European Medicines Agency (EMA) has provided specific guidelines<sup>2</sup>. According to these, biological knowledge should underpin subgroup definitions, and subgroup-specific effects in late-stage clinical trials should still be interpreted

---

<sup>1</sup>Computational Health Center, Helmholtz Munich, 85764 Neuherberg, Germany. <sup>2</sup>Department of Biology, Ludwig-Maximilians University Munich, 82152 Martinsried, Germany. <sup>3</sup>Charité Universitätsmedizin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Hematology, Oncology, and Cancer Immunology, Charitéplatz 1, 10117 Berlin, Germany. <sup>4</sup>German Cancer Consortium (DKTK), partner sites Berlin and Munich, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. <sup>5</sup>Department of Medicine III and Comprehensive Cancer Center Munich, University Hospital, Ludwig-Maximilians University Munich, 81377 Munich, Germany. <sup>6</sup>Oncological Practice, 84028 Landshut, Germany. <sup>7</sup>Oncological Practice, 88212 Ravensburg, Germany. <sup>8</sup>Department of Medicine I and Research Center for Immunotherapy (FZI), Johannes Gutenberg-University Clinic, 55131 Mainz, Germany. <sup>9</sup>Department of Biochemistry and Pharmacology, University of Melbourne, Victoria 3010, Australia. <sup>10</sup>These authors contributed equally: Alexander J. Ohnmacht, Arndt Stahler, Sebastian Stintzing. ✉ e-mail: [volker.heinemann@med.uni-muenchen.de](mailto:volker.heinemann@med.uni-muenchen.de); [michael.menden@helmholtz-munich.de](mailto:michael.menden@helmholtz-munich.de)

with caution owing to the exploratory and retrospective nature of the analyses. For this purpose, a large number of computational methods have been proposed and discussed<sup>3–5</sup>, e.g., tree-based methods using recursive partitioning<sup>6–8</sup>, virtual twins<sup>9</sup>, outcome weighted methods<sup>10,11</sup>, causal forests<sup>12</sup> and metalearners for estimating heterogeneous treatment effects<sup>13</sup>. However, most of these computational methods neglect cancer biology, i.e., exploiting the molecular landscape of a clinical trial and customising models to cancer subtypes and mutational patterns.

Clinical outcomes of patients with metastatic colorectal cancer (mCRC) significantly improved upon the introduction of targeted treatments, including anti-EGFR and anti-VEGF directed monoclonal antibodies such as cetuximab and bevacizumab, respectively<sup>14</sup>. Tumours of colorectal cancer patients were shown to exhibit, for instance, either *KRAS* or *NRAS* mutations (referred to as *RAS* mutations) with a rate of about 50%, which tend to occur mutually exclusive<sup>15,16</sup>. These *RAS* mutations are clinically approved predictive biomarkers of resistance against anti-EGFR directed monoclonal antibodies such as cetuximab<sup>17</sup>. Bevacizumab has been reported to improve progression-free survival in first-line mCRC trials<sup>18</sup>; however, no comparable biomarker has been depicted yet.

In this study, we focused on the open-label randomised phase III clinical trial FIRE-3. Here, patients with *KRAS* exon 2 wild-type mCRC were randomised to receive either cetuximab or bevacizumab in combination with 5-fluorouracil, leucovorin and irinotecan (FOLFIRI) as a first-line regimen. Several retrospective subgroup analyses revealed potential prognostic and predictive biomarkers based on tumour DNA and clinical characteristics, such as the relevance of the molecular status, i.e., alterations other than *KRAS* exon 2, such as *KRAS* exon 3-4, *NRAS* exon 2-4 and BRAF V600E, or primary tumour sidedness<sup>19–23</sup>. For example, targeting EGFR in *RAS* wild-type mCRC tumours located in the left hemicolon (left-sided) was shown to be beneficial, whilst *RAS* wild-type tumours located in the right colon (right-sided) were less likely to respond<sup>24</sup>. Additionally, in the more recent FIRE-4.5 study, it was demonstrated that patients with BRAF V600E mutant tumours may benefit from the treatment with 5-fluorouracil, oxaliplatin, leucovorin and irinotecan (FOLFOXIRI) backbone plus bevacizumab<sup>25</sup>, whereas in contrast, these patients lacked benefits from cetuximab<sup>26,27</sup>. This hints towards tumour subtype-specific interactions and alternative mechanisms to acquire EGFR inhibitor resistance<sup>28</sup>.

Previously proposed tumour subtypes in colorectal adenocarcinoma are based on the gene expression-derived consensus molecular subtypes (CMS) and could identify subtypes that reflected distinct tumour biology<sup>29</sup>. Recently, the prognostic value of CMS has been confirmed in the FIRE-3, CALGB/SWOG 80405 and AGITG MAX clinical trials for FOLFIRI combined with either cetuximab or bevacizumab<sup>21,30,31</sup>. In particular, CMS4 patients with *RAS* wild-type have shown a significantly longer overall survival when treated with cetuximab compared to bevacizumab in metastatic disease<sup>21</sup>. However, the clinical translation of the CMS classification of colorectal cancer is still in its infancy and is further investigated in multiple clinical trials<sup>32</sup>. These sparse results have illustrated that modelling interactions between somatic alterations and tumour subtypes can yield insights into complex biomarkers and highlight the urgent need for computational frameworks to systematically decipher the molecular landscape, tumour subtypes and biomarkers. Thus, we hypothesised that predictive response biomarkers may be revealed by systematically deconvoluting cancer genetic events and tumour subtypes within a clinical trial.

Here, we present the Oncology Biomarker Discovery (OncoBird) framework, which empowers the systematic identification of actionable biomarkers for clinical trials in oncology. OncoBird is publicly available as a software package at <https://github.com/MendenLab/OncoBird> and a demo run is available at [https://codeocean.com](https://codeocean.com/capsule/9911222/tree/v1)

[capsule/9911222/tree/v1](https://codeocean.com/capsule/9911222/tree/v1). Furthermore, users can run a graphical user interface within a docker container (Supplementary Fig. 1).

The OncoBird workflow is divided into five distinct steps: it systematically (1) investigates the molecular landscape of a clinical trial, i.e., copy number alterations, somatic mutations, mutually exclusive patterns and predefined tumour subtypes; (2) identifies biomarkers within a treatment arm based on genetic alterations, and (3) in relation to the predefined tumour subtypes; consecutively, (4) evaluates their predictive component across treatment arms; and finally, (5) it comprehensively corrects for multiple hypothesis testing and adjusts treatment effects of biomarkers based on resampling methods. To enhance the biological signal, this analysis integrates the molecular and biomarker landscape of cancer clinical trials by customising models to established cancer subtypes and mutational patterns. In essence, OncoBird yields subtype-specific biomarkers with treatment benefits in an interpretable and transparent manner and therefore operates complementary to existing methods. The utility of OncoBird is exemplified by the application to the FIRE-3 clinical trial, generalises to the ADJUVANT clinical trial<sup>33–35</sup>, and in fact, would generalise to any molecularly characterised randomised controlled trial (RCT) in oncology.

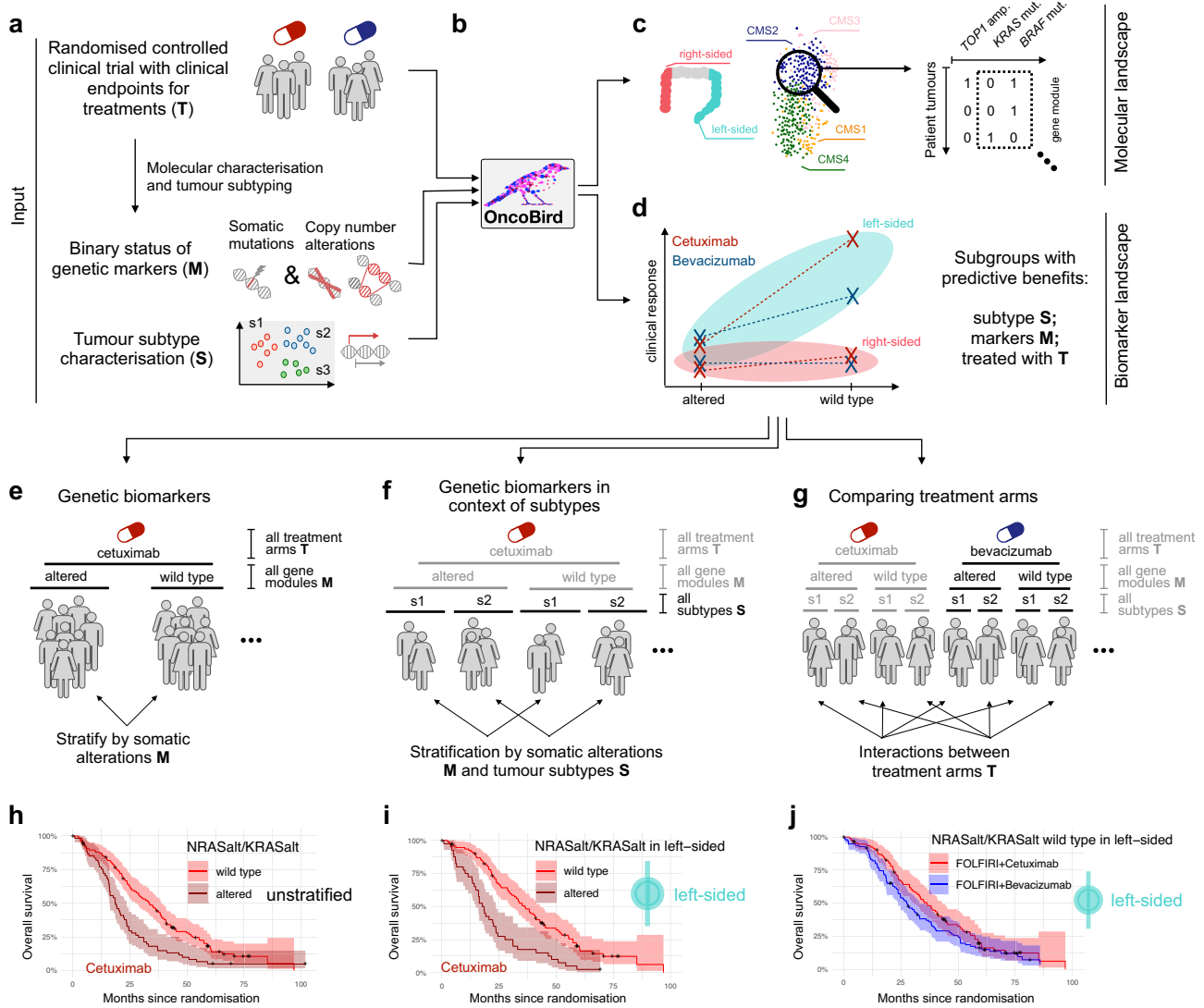
## Results

OncoBird is applicable to RCTs accompanied with molecular characteristics, including genetic sequencing panels which yield copy number alterations and somatic driver mutations (Fig. 1a, b). In addition, a second layer of stratification can be supplied in the form of predefined tumour subtypes (Fig. 1a). Then, OncoBird systematically assesses the genetic landscape in the context of tumour subtypes (Fig. 1c) and outlines the biomarker landscape across multiple clinical responses (Fig. 1d), i.e., time-to-event data (overall or progression-free survival; “Methods”), and binary variables capturing treatment success (objective response rate; “Methods”).

Here, we leveraged the FIRE-3 RCT, including 752 mCRC patients who have been treated with FOLFIRI and either cetuximab or bevacizumab. We defined tumour subtypes based on CMS<sup>21</sup>, and tumour sidedness, i.e., left- or right-sided mCRC. In addition, 373 tumours were genetically characterised, i.e., the mutational status of 277 frequently altered cancer genes. To reveal the biomarker landscape, we employed the following stratification and modelling strategies (Supplementary Data 1; “Methods”): We first investigated each alteration for stratifying patients by their prognosis within each treatment arm (Fig. 1e). Consecutively, we inspected alterations in tumour subtypes (Fig. 1f), revealing subtype-specific biomarkers. Finally, we tested for treatment interactions to reveal biomarkers with predictive effects (Fig. 1g). Importantly, subtypes and genetic alterations ought to be independent of the treatment assignment. The molecular landscape and individual treatment arm analysis could be applied to any trial design without limitations.

Exemplified with a well-established biomarker of cetuximab response<sup>17</sup>, *RAS* wild-type mCRC patients showed longer overall survival (Fig. 1h;  $p = 0.0002$ , HR = 0.53 [0.38–0.73]). Consistent with a previous study<sup>36</sup> and more recently defined treatment guidelines for mCRC<sup>37</sup>, the cetuximab overall survival (OS) benefit for patients with *RAS* wild-type tumours was conserved in left-sided tumours (Fig. 1i;  $p = 7.6 \times 10^{-5}$ , HR = 0.44 [0.29–0.66]). Furthermore, we observed interactions between *RAS* mutations and the treatment arm in left-sided tumours ( $p_{\text{int}} = 0.07$ ): Cetuximab remained superior to bevacizumab in *RAS* wild type and left-sided tumours (Fig. 1j;  $p = 0.05$ , HR = 0.73 [0.52–1.00]) in terms of OS, whilst bevacizumab and cetuximab achieved comparable OS for patients with *RAS* mutant and left-sided tumours (Supplementary Fig. 2;  $p = 0.32$ , HR = 1.22 [0.85–1.75]).

Whilst we particularly focused on the FIRE-3 trial in colorectal cancer, we also demonstrate the generalisability of OncoBird by applying it with the same default biomarker thresholds to the

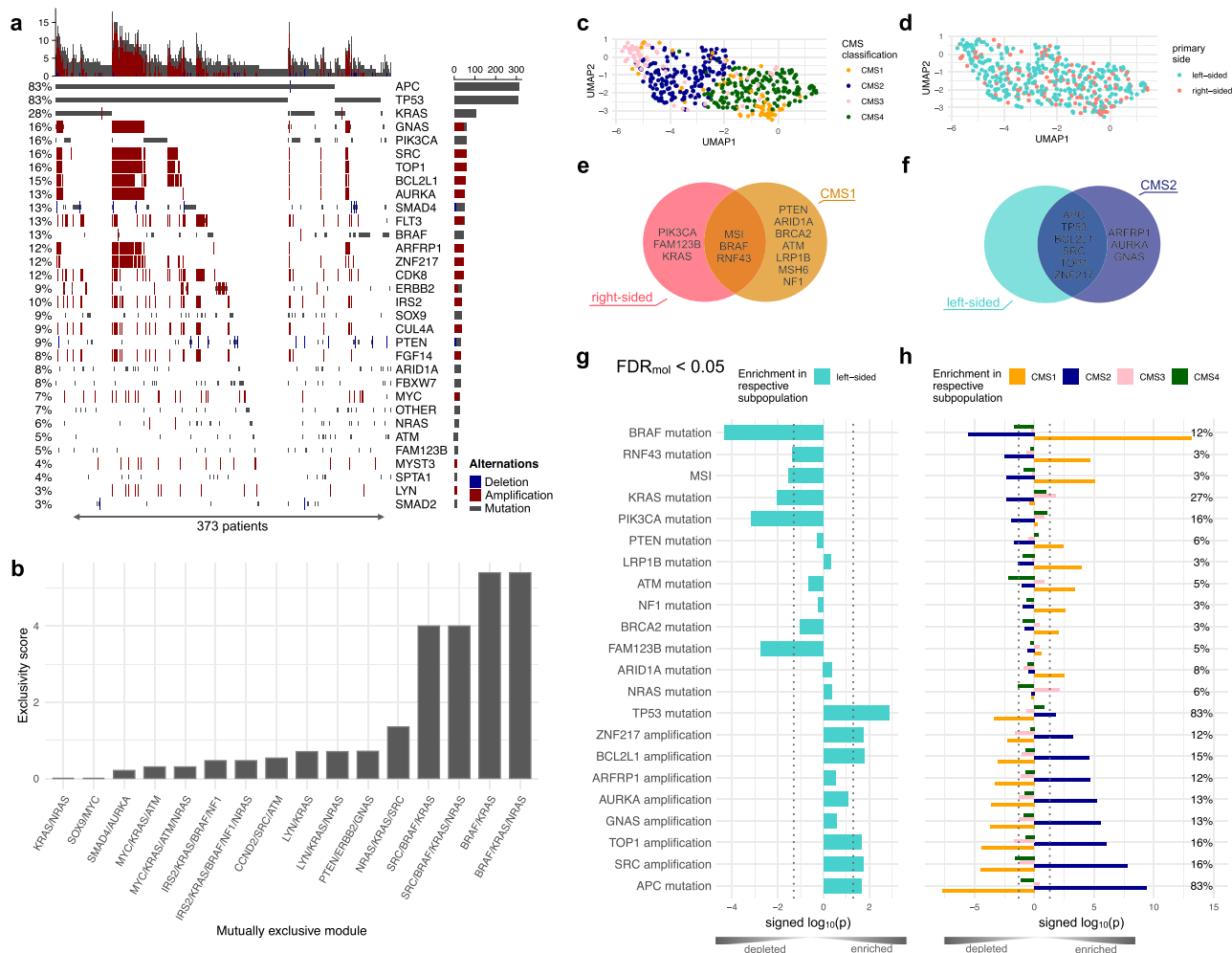


**Fig. 1 | The Oncology Biomarker Discovery (OncoBird) workflow.** **a** Patients in clinical trials were treated with (T) two treatment regimens with measured clinical endpoints. Subsequently, their tumours are characterised according to (M) tumour genetic alterations (somatic mutations and copy number alterations) and (S) tumour subtypes. **b** With this input, OncoBird outlines **c** the molecular landscape and **d** the biomarker landscape. For the latter, **e** somatic alterations are explored for a differential patient prognosis for each treatment arm. **f** Consecutively, for each treatment arm, subtype-specific biomarkers are derived. **g** Finally, interactions

between treatment arms are examined. The grey shadings indicate the data included in the previous analysis step. Here, this is exemplified in the FIRE-3 clinical trial using Kaplan–Meier plots, including 95% confidence intervals (CI) and summary statistics of the Cox regression models. **h** *RAS* mutations are established biomarkers of cetuximab resistance. **i** Patients with *RAS* wild-type tumours showed a better prognosis when treated with cetuximab within left-sided tumours compared to right-sided tumours. In addition, **j** the *RAS* wild-type subpopulation in left-sided tumours showed benefits when treated with cetuximab compared to bevacizumab.

ADJUVANT clinical trial (“Methods”), which explored gefitinib in non-small cell lung cancer (NSCLC)<sup>33–35</sup>. The ADJUVANT study reported predictive components of five alterations, i.e., *TP53* mutations, *RBI* alterations and copy number amplifications of *NKX2-1*, *CDK4* and *MYC*<sup>35</sup>. Four out of five biomarkers were concordantly identified for disease-free survival with OncoBird (FDR<sub>int</sub> < 0.2; Supplementary Data 2; Supplementary Fig. 3–6). In addition, OncoBird suggests that the mutual exclusivity patterns play a role in the biomarker landscape of NSCLC (Supplementary Fig. 3c, d). In more detail, we observed gefitinib benefits in tumours that were characterised by mutations in either *TP53*, *SMAD4* or *CDK4* amplifications ( $p = 0.0002$ , HR = 0.37 [0.21–0.63]; Supplementary Data 2; Supplementary Figs. 5c and 6a), for which the resampling-based adjustment of the conditional average treatment effect yielded  $p_{adj} = 0.001$  with HR = 0.32 [0.14–0.86] (Supplementary Data 2; “Methods”). These findings highlight the accessibility, reproducibility and interoperability of OncoBird.

**The molecular landscape of the FIRE-3 clinical trial**  
Leveraging OncoBird, we assessed the genetic landscape of patient tumours in the FIRE-3 clinical trial. In total, 373 tumours were genetically characterised, including 31 frequently altered cancer genes observed in at least 12 patients (Fig. 2a). We observed amplifications in chromosome arm 20q (chr20q) in 74/373 tumours (19.8%), which includes *SRC*, *TOP1*, *BCL2L1*, *ZNF217*, *AURKA*, *GNAS* and *ARFRP1* (Fig. 2a). Indeed, chr20q amplifications have been reported to define a distinct subtype of left-sided colon cancers<sup>38</sup>. In addition, we identified 39 mutually exclusive somatic alterations (gene modules) using the Mutex algorithm (Fig. 2b; “Methods”)<sup>39</sup>, thus grouping low frequent but functionally similar somatic events within a signalling pathway. We could confirm that chr20q amplifications were mutually exclusive to somatic mutations in the ERK signalling pathway (*KRAS*, *NRAS* or *BRAF*;  $p = 0.0002$ , Fisher’s exact test).



**Fig. 2 | Molecular landscape of the FIRE-3 clinical trial. a** OncoPrint of 373 mCRC tumours, including mutations and copy number alterations detected in more than 12 tumours. **b** The mutually exclusive alteration patterns were derived with the Mutex algorithm. Gene expression profiles of 451 mCRC tumours are annotated by **c** the consensus molecular subtypes (CMS) and **d** the primary tumour side. **e** Venn

diagram showing all enriched somatic alterations for CMS1 and right-sided tumours, and **f** enriched somatic alterations for CMS2 and left-sided tumours. **g** Frequently altered cancer genes tested for enrichment in left- or right-sided tumours, and **h** tested against CMS subtypes using one-sided hypergeometric tests. Source data for the figure panels are provided as Source Data file.

In addition, we analysed 451 gene expression profiles and showed consistency with their derived CMS subtypes (Fig. 2c), whilst the primary tumour side displayed a heterogeneous gene expression pattern (Fig. 2d). Right-sided tumours were particularly enriched in CMS1 tumours ( $p = 0.009$ , hypergeometric test; Supplementary Fig. 7) and depleted in CMS2 tumours ( $p = 0.007$ , hypergeometric test; Supplementary Fig. 7).

The concordance between right-sided tumours and CMS1 (Fig. 2e) was reflected by genetic alterations that were enriched in both tumour subtypes. Microsatellite instabilities (MSI) and somatic mutations in *BRAF* and *RNF43* were enriched in both CMS1 and right-sided tumours ( $FDR_{mol} < 0.05$ , hypergeometric test). Additionally, mutations in *PIK3CA*, *FAM123B* and *KRAS* were only associated with right-sided tumours (Fig. 2e;  $FDR_{mol} < 0.05$ , hypergeometric test). In contrast, the similarity of left-sided tumours and CMS2 (Fig. 2f) was characterised by mutations in *APC*, *TP53* and chr20q amplifications (*SRC*, *TOP1*, *BCL2L1*, *ZNF217*), which were all significantly enriched in both left-sided and CMS2 tumours (Fig. 2g, h;  $FDR_{mol} < 0.05$ , hypergeometric test). Somatic mutations in *PTEN*, *ARID1A*, *ATM*, *LRP1B*, *BRCA2* and *NF1* did not show a preference for a particular primary tumour side, but were enriched in CMS1 tumours (Fig. 2h), and were associated with an increased tumour mutational burden ( $p = 0.008$ ,  $p = 0.002$ ,  $p = 0.017$ ,  $p = 0.0001$ ,  $p = 0.010$  and  $p = 0.051$ , respectively, Fisher's exact test).

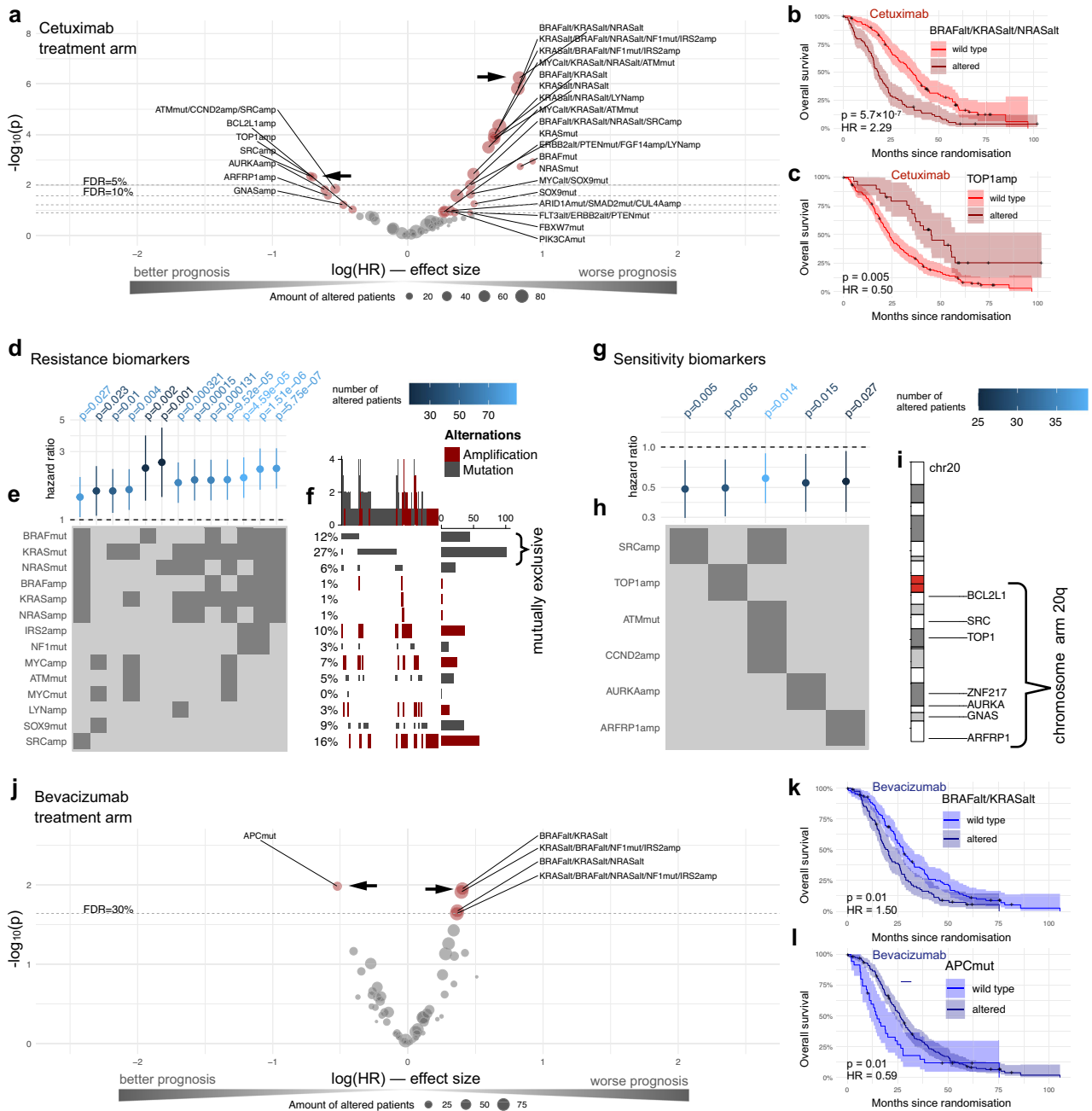
In summary, leveraging OncoBird and investigating patterns of genetic events in tumour subtypes revealed meaningful tumour biology. For example, mutations of either *BRAF* or *KRAS* promote ERK signalling and therefore occur mutually exclusive. *BRAF* mutations were predominantly found in CMS1, but nevertheless, 27 out of 53 *BRAF* mutant tumours were distributed among CMS2-4. Therefore, it is of utmost importance to gain an enhanced understanding of the molecular landscape of mCRC prior to the interpretation of biomarkers, which is further empowered by OncoBird.

### Genetic biomarkers of cetuximab

First, independent of tumour subtypes, we assessed single genes and mutually exclusive gene modules (Fig. 2a, b) as biomarkers for cetuximab. For this, we leveraged Cox proportional hazards regression and logistic regression models ("Methods"), considering overall survival (OS; Fig. 3a–h), progression-free survival (PFS; Supplementary Fig. 8) and the objective response rate (ORR; Supplementary Fig. 9). We quantified effect sizes by hazard ratios (HR) for survival data and odds ratios (OR) for binary data including 95% confidence intervals (Supplementary Data 3).

The clinically established resistance biomarkers of cetuximab were recovered, i.e., mutations in *RAS* (either *KRAS* or *NRAS*) referred to a poorer OS in the cetuximab treatment arm (Fig. 1h;  $p = 0.0002$ ,





**Fig. 3 | Identification of genetic biomarkers for FOLFIRI plus cetuximab or bevacizumab.** **a** Volcano plot for genetic biomarkers of cetuximab in the form of mutually exclusive gene modules or single gene mutations. Each point shows the effect of a particular group of alterations summarised by its hazard ratio derived by the Cox regression models and its raw  $p$ -value derived by a Wald test. Exemplifying the most significant associations, Kaplan–Meier plots, including 95% confidence intervals (CI) and summary statistics of the Cox regression models, are shown for **b** the mutually exclusive module consisting of *RAS* and *BRAF* mutations, and **c** the amplification of *TOP1* treated with cetuximab. For investigating the biomarker composition, we focus on **d** resistance biomarkers of FOLFIRI plus cetuximab with  $FDR_{cet} < 0.1$ , showing their hazard ratios and 95% CIs. For these, **e** the composition of mutually exclusive genes is indicated by dark grey colour, and **f** an oncoprint

highlighting mutational frequencies of biomarker combinations is shown. In like manner, **g** cetuximab sensitivity biomarkers, shown by their hazard ratios and 95% CIs, and **h** their composition are summarised. **i** Karyoplot showing transcription start sites of co-amplified genes on chromosome 20q. **j** Volcano plot of the genetic biomarkers of bevacizumab with  $FDR_{bev} < 0.3$ , shown in brown colour by their hazard ratios derived by the Cox regression models and their raw  $p$ -values derived by a Wald test. Kaplan–Meier plot including 95% CIs and summary statistics of the Cox regression models of **k** mutations in *KRAS* or *BRAF* and **l** *APC* mutations treated with bevacizumab. The compositions of bevacizumab biomarkers are shown in Supplementary Fig. 10a, b. A Source Data file is provided, which contains the source data for the figure panels and the sample sizes of the conducted statistical tests.

$HR = 1.90$  [1.36–2.65],  $FDR_{cet} < 0.1$ ). In addition, we confirmed that *BRAF* mutations are mutually exclusive to *RAS* mutations (Fig. 2b;  $p = 0.0008$ , Fisher’s exact test), and both contributed to a poor OS when treated with cetuximab (Fig. 3b;  $p = 5.7 \times 10^{-7}$ ,  $HR = 2.29$  [1.65–3.16],  $FDR_{cet} < 0.1$ ), which has been consistently observed in an independent cohort<sup>40</sup>.

Most resistance biomarker modules grouped mutations in *KRAS* and *BRAF* ( $FDR_{cet} < 0.1$ ). In addition, we found a gene module including mutations in *SOX9* and *MYC* amplifications, for which mutant tumours displayed a worse prognosis based on OS (Fig. 3d, e;  $p = 0.02$ ,  $HR = 1.50$  [1.07–2.37],  $FDR_{cet} < 0.1$ ). By inspecting their oncoprint (Fig. 3f) 27/59 tumours harboured mutations in either *SOX9* or *MYC* and were wild-

type in either *BRAF*, *KRAS* or *NRAS*, hinting towards an alternative cetuximab resistance mechanism.

In addition, we found *TOPI* amplifications to be a strong predictor of a prolonged OS for treatment with cetuximab (Fig. 3c;  $p = 0.005$ , HR = 0.50 [0.30–0.81],  $FDR_{cet} < 0.1$ ). In fact, we could identify multiple co-amplifications that showed prognostic value for the cetuximab treatment arm, which are located on chromosome 20q. Among the most predictive amplifications for a longer OS were *SRC*, *TOPI*, *AURKA* and *ARFRP1* (Fig. 3g–i; Supplementary Data 3). Consistent trends were observed with *SRC* amplifications in PFS ( $p = 0.10$ , HR = 0.69 [0.44–1.07], median PFS wild-type tumours 9.6 months vs mutants 11.1 months) and ORR ( $p = 0.18$ , OR = 0.45 [0.14–1.45], ratio ORR wild-type 0.66 vs mutant tumours 0.83).

### Genetic biomarkers of bevacizumab

Analogously to the cetuximab biomarker analysis, for the bevacizumab treatment arm, we also built Cox proportional hazards regression models (“Methods”) applied to OS (Fig. 3j–l; Supplementary Fig. 10) and PFS (Supplementary Fig. 11), and logistic regression models for ORR (Supplementary Fig. 12). For exploring bevacizumab biomarker trends, we employed a lenient threshold of  $FDR_{bev} < 0.3$ , which deviates from the default setting (“Methods”). The mutually exclusive module of *KRAS* and *BRAF* mutations showed lower OS (Fig. 3j, k;  $p = 0.01$ , HR = 1.50 [1.10–2.04],  $FDR_{bev} < 0.3$ ), which is consistent with literature reports<sup>41,42</sup>. A better predictor for poor OS was the *APC* wild-

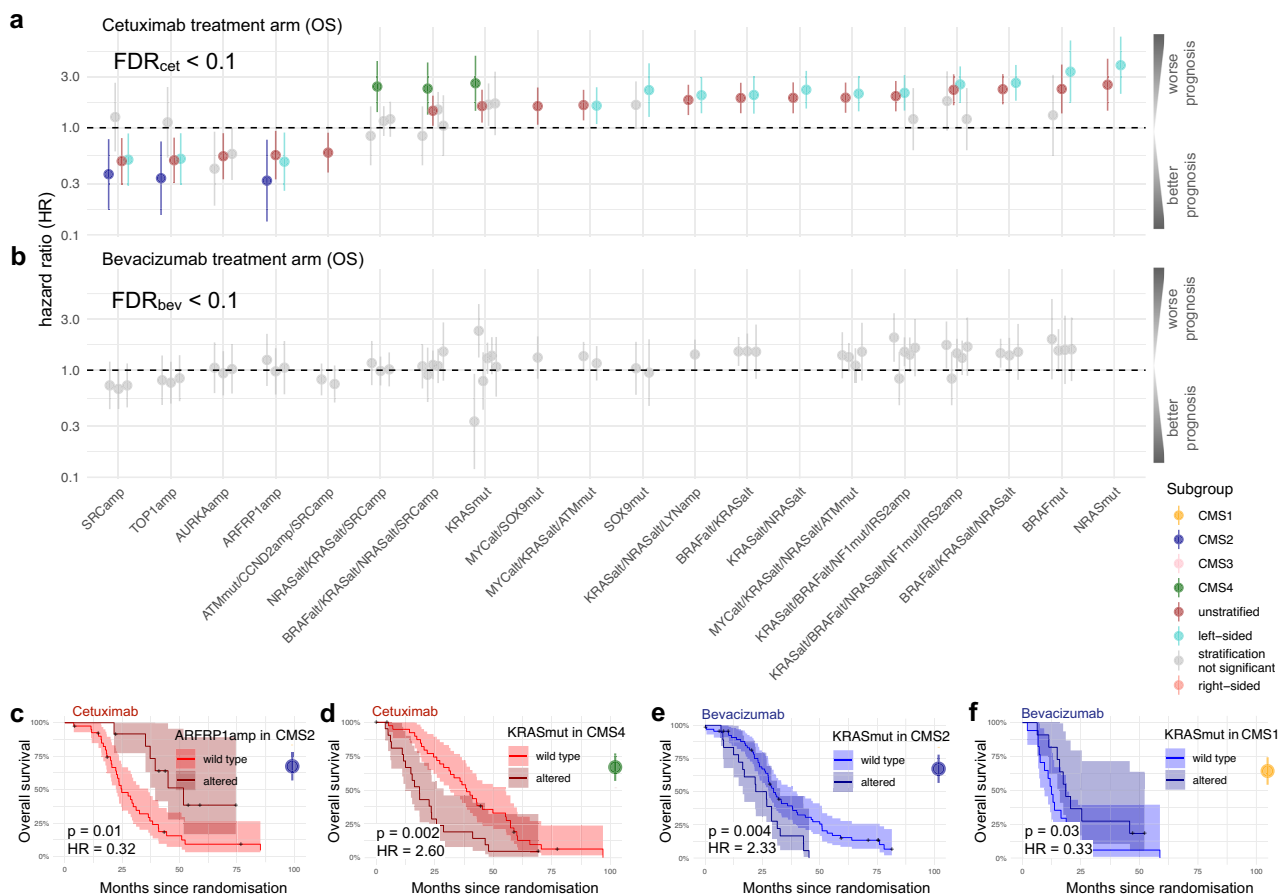
type status for tumours treated with FOLFIRI plus bevacizumab (Fig. 3j, l;  $p = 0.01$ , HR = 1.69 [1.14–2.50],  $FDR_{bev} < 0.3$ ).

### Subtype-specific biomarkers of cetuximab and bevacizumab

The previous analyses focused on genetic biomarkers in isolation, whilst here, we investigated them within the context of tumour subtypes (“Methods”). In FIRE-3, tumour subtypes are defined as either left- or right-sided tumours, or alternatively, classified according to the consensus molecular subtypes, i.e., CMS1–4 (“Methods”)29. Here, we tested stratifications based on each single gene or gene module within tumour subtypes for OS (Fig. 4a, b), PFS (Supplementary Fig. 13) and ORR (Supplementary Fig. 14).

In total, we found 38 subtype-specific biomarkers of cetuximab for OS ( $FDR_{cet} < 0.1$ ; “Methods”). In particular, we recovered favourable OS of CMS2 patients treated with cetuximab (Fig. 4a), if their tumours additionally carried chr20q amplifications, i.e., *ARFRP1* (Fig. 4c;  $p = 0.01$ , HR = 0.32 [0.13–0.77],  $FDR_{cet} < 0.1$ ), *TOPI* (Supplementary Fig. 15a;  $p = 0.01$ , HR = 0.34 [0.15–0.74],  $FDR_{cet} < 0.1$ ) and *SRC* (Supplementary Fig. 15b;  $p = 0.01$ , HR = 0.37 [0.17–0.78],  $FDR_{cet} < 0.1$ ). Additionally, CMS4 *KRAS* mutant tumours treated with cetuximab showed worse OS (Fig. 4d;  $p = 0.002$ , HR = 2.60 [1.44–4.70],  $FDR_{cet} < 0.1$ ) and PFS (Supplementary Fig. 13a, c).

For reporting bevacizumab biomarker trends, we employed a lenient false discovery rate ( $FDR_{bev} < 0.3$ ), which deviates from the conservative OncoBird default setting (“Methods”). Tumours with



**Fig. 4 | Identification of subtype-specific genetic biomarkers for FOLFIRI plus cetuximab or bevacizumab.** Subtype-specific genetic biomarkers for OS of **a** cetuximab and **b** bevacizumab using hazard ratios including 95% confidence intervals (CI) derived from single Cox regression models. Subtypes are defined by either the primary tumour side, CMS or unstratified (reference model). Kaplan–Meier plots including 95% CIs, hazard ratios and raw  $p$ -values derived by

Wald tests from the Cox regression models of subtype-specific genetic biomarkers for **c** *ARFRP1* in CMS2, **d** *KRAS* in CMS4, **e** *KRAS* in CMS2 and **f** *KRAS* in CMS1 in either the cetuximab or bevacizumab treatment arm. A Source Data file is provided, which contains the source data for the figure panels and the sample sizes of the conducted statistical tests.

*KRAS* mutations classified as CMS2 tended to show worse OS when treated with bevacizumab (Fig. 4e;  $p = 0.004$ , HR = 2.33 [1.31–4.15],  $FDR_{bev} < 0.3$ ). In contrast, *KRAS* mutated tumours classified as CMS1 tended to show a longer OS compared to wild-type tumours when treated with bevacizumab (Fig. 4f;  $p = 0.03$ , HR = 0.33 [0.12–0.93],  $FDR_{bev} < 0.3$ ).

### Predictive components of biomarkers

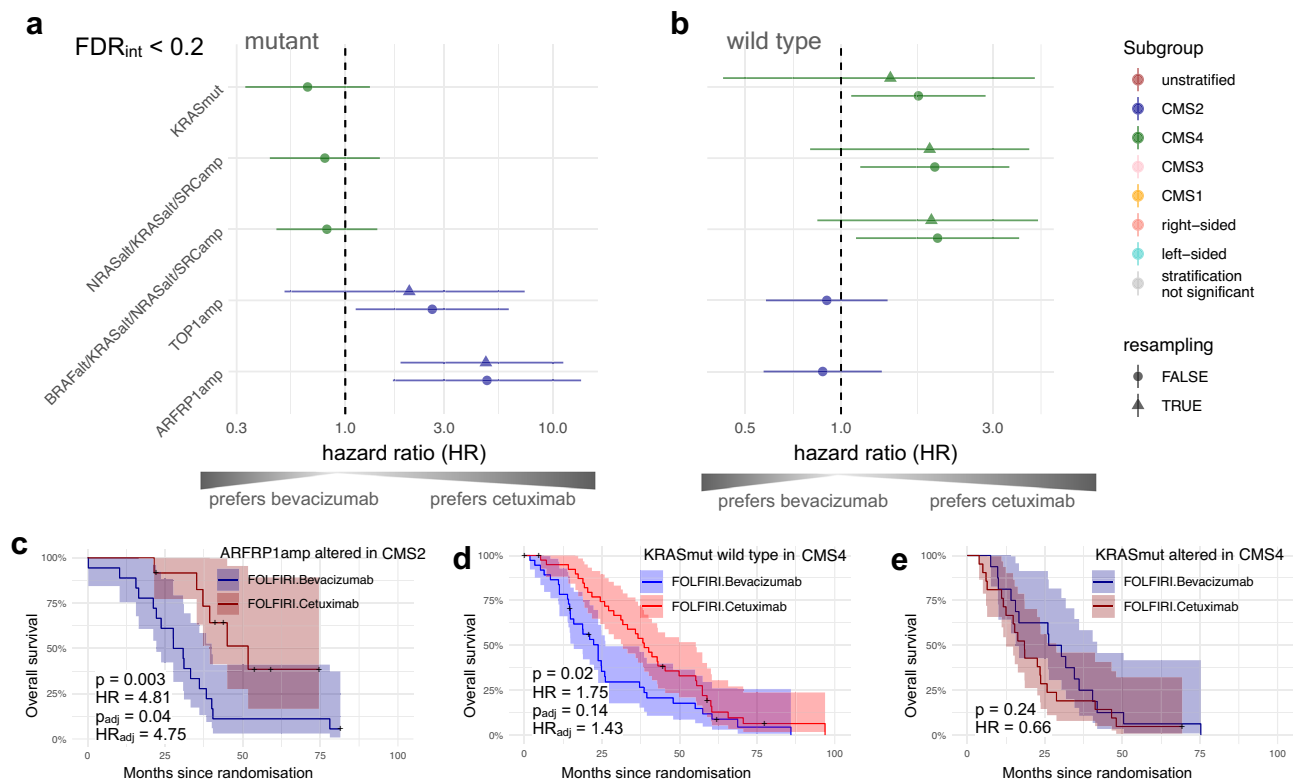
For assessing the predictive component of response biomarkers, here, we compared the cetuximab and bevacizumab treatment arms against each other by focusing on interactions between genetic alterations in the context of tumour subtypes (“Methods”). Subsequently, we compared the prognosis of both inhibitors for each subgroup according to the interaction biomarkers, thus assessing potential treatment benefits. In addition, we corrected the conditional average treatment effects in the identified subgroups using resampling methods to obtain multiplicity-adjusted  $p$ -values and bias-corrected confidence intervals (“Methods”). The results were summarised for OS (Fig. 5a, b) and PFS (Supplementary Fig. 16), whereas no significant interactions were detected for ORR. In total, we found five putative interactions (Supplementary Data 4;  $FDR_{int} < 0.2$ ; “Methods”). For reporting other biomarker trends, we also included summary statistics of 57 subgroups with a lenient threshold of  $FDR_{int} < 0.6$ , which deviates from the default setting (Supplementary Data 3).

For example, we found predictive value of chr20q amplifications in CMS2 tumours treated with FOLFIRI plus cetuximab (Fig. 5a, b), which is evident by the significant interactions of *TOP1* ( $p_{int} = 0.07$ ,  $FDR_{int} < 0.2$ ) and *ARFRP1* ( $p_{int} = 0.01$ ,  $FDR_{int} < 0.2$ ). *ARFRP1* amplifications showed the largest predictive component among the chr20q

amplifications. Accordingly, we observed longer OS in the cetuximab treatment arm compared to bevacizumab in CMS2 (Fig. 5a, c; *ARFRP1*:  $p = 0.003$ , HR = 0.21 [0.07–0.59],  $FDR_{int} < 0.2$ ; Supplementary Data 3). The resampling-based adjusted treatment effect confirmed this observation and yielded a hazard ratio in this subgroup of HR = 0.21 [0.09–0.54] with  $p_{adj} = 0.04$  (Fig. 5a, c). Previous reports have indicated a prognostic value of chr20q amplifications in colorectal cancer patients<sup>38,43</sup>, whilst OncoBird yielded additional evidence that they harbour a predictive component.

Another interaction example was tumours with *KRAS* mutations that showed CMS-specific responses. In CMS4, patients with *KRAS* wild-type tumours responded better to cetuximab compared to patients treated with bevacizumab (Fig. 5b, d; *KRAS* wild types:  $p = 0.02$ , HR = 0.57 [0.35–0.93];  $p_{int} = 0.02$ ,  $FDR_{int} < 0.2$ ), for which the resampling-based adjusted treatment effect yielded HR = 0.70 [0.25–2.35] with  $p_{adj} = 0.14$  (Fig. 5b, d). Our results suggest a predictive role of *KRAS* mutations in CMS4 for cetuximab, which we also identified for PFS (Supplementary Fig. 16c, d). Notably, modules containing alterations in *NRAS*, *BRAF* and *SRC* showed similar statistics since only four, eight and twelve mutant tumours were present in CMS4. Insignificant but numerically longer OS was observed for patients with *KRAS* mutated tumours classified as CMS4 treated with bevacizumab (Fig. 5e, *KRAS* mutants:  $p = 0.24$ , HR = 0.66 [0.33–1.31]), with a median OS 28.3 months compared to 18.4 months when treated with cetuximab.

In order to assess the ability of OncoBird to discover the same biomarkers for different datasets, we applied 5-fold cross-validation repeated five times and extracted the ten most significant biomarkers for OS across each of the 25 models (Fig. 6a). Consistent with our



**Fig. 5 | Predictive biomarkers in the context of tumour subtypes.** Overview of interaction biomarkers ( $FDR_{int} < 0.2$ ) focusing on **a** mutant and **b** wild-type tumours when comparing cetuximab and bevacizumab treatment, using hazard ratios including 95% confidence intervals (CI) derived from single Cox regression models fitted on OS. Triangle points and confidence intervals were obtained from the bootstrap-based bias-correction of treatment effects. For the conducted statistical

tests, the sample sizes are given in Supplementary Data 3. Here exemplified, Kaplan–Meier plots including 95% CIs, hazard ratios and raw  $p$ -values derived by Wald tests from the Cox regression models compare treatments in subgroups for **c** *ARFRP1* amplifications in CMS2 and **d, e** *KRAS* mutations in CMS4. Source data for the figure panels are provided as Source Data file.

previous findings, gene modules containing *KRAS* mutations for CMS4 were found in 21/25 training sets and chr20q amplifications in CMS2 were reproduced in 22/25 training sets (Fig. 6a).

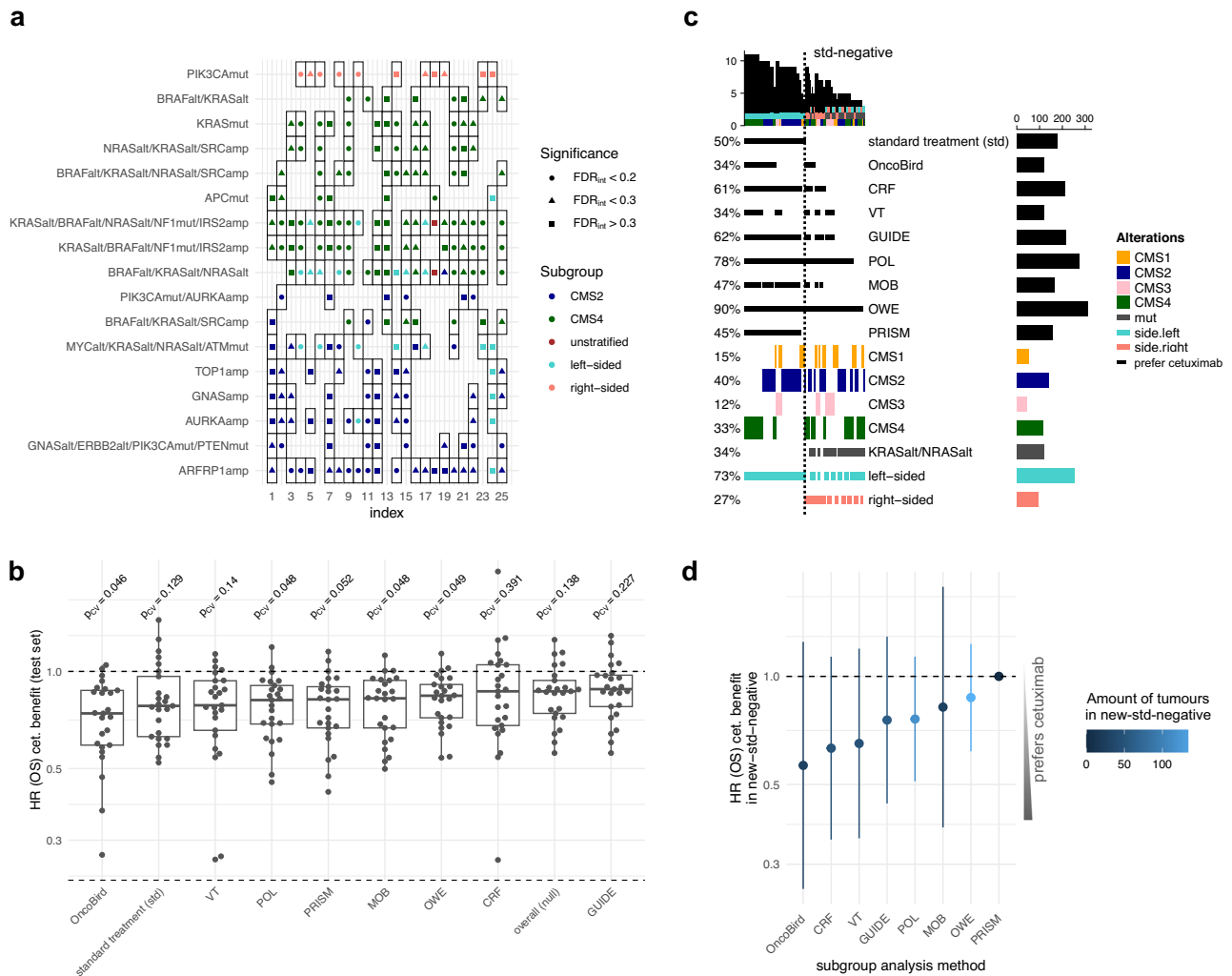
**Benchmarking of methods for subgroup analysis**

For benchmarking OncoBird, we compared it to alternative methods that can be used to investigate predictive biomarkers based on overall survival. Together with OncoBird, eight algorithms and implementations were used in order to identify subgroups with differential treatment effects, i.e., virtual twins (VT)<sup>9</sup>, model-based partitioning (MOB)<sup>8</sup>, an outcome-weighting method (OWE)<sup>11</sup>, causal random forests (CRF)<sup>12</sup>, policy learning (POL)<sup>44</sup>, GUIDE<sup>45</sup> and PRISM<sup>46</sup> (Supplementary Table 1; “Methods”; Fig. 6b).

For the evaluation, we first derived hazard ratios for cetuximab benefit based on OS in the subgroups according to the predicted biomarkers for all methods across five times 5-fold cross-validation (“Methods”). We also focused on the current treatment guidelines for mCRC, according to which patients should receive cetuximab if their tumours are *RAS* wild-type and left-sided (std; Fig. 6b)<sup>37</sup>. While the treatment benefit was not significant for the std-positive subgroup

(Fig. 6b, median HR = 0.78,  $p_{cv} = 0.129$ ), the methods that found the highest significant benefits were OncoBird (median HR = 0.74,  $p_{cv} = 0.046$ ), POL (median HR = 0.81,  $p_{cv} = 0.048$ ), MOB (median HR = 0.83,  $p_{cv} = 0.048$ ) and OWE (median HR = 0.84,  $p_{cv} = 0.049$ ) ordered by the magnitude of the hazard ratio (Fig. 6b).

Next, we leveraged the whole dataset to identify cetuximab sensitivity biomarkers with each method and compared them to the treatment guidelines. On average, 73% of methods identified cetuximab benefit for a patient in the std-positive subgroup, whereas only 33% of methods detected further benefits in the std-negative subgroup (Fig. 6c). 7/8 (88%) methods found mutually exclusive alterations in *KRAS*, *NRAS* or *BRAF* as a predictive biomarker, from which one, two and four methods proposed this marker in conjunction with tumour sidedness, CMS and across all patients, respectively (Supplementary Table 1). Only 2/8 (25%) methods highlighted *TOP1* amplifications as a potential biomarker (Supplementary Table 1). This highlights that current subgroup analysis methods mostly recover standard clinical practice, whilst sparsely identifying complementary predictive subgroups, thus highlighting the unmet need for cancer biology-driven frameworks such as OncoBird.



**Fig. 6 | Stability analysis and benchmark with other methods.** **a** The ten most significant biomarkers across 25 models of five times repeated 5-fold cross-validation. **b** Boxplots of treatment effects in terms of hazard ratios for the predicted subgroups in the 25 test sets for the benchmarked methods, including standard treatment guidelines (std) and overall across all patients (null). The centre line depicts the median; the box represents the inter-quartile range (IQR) and the whiskers the interval 1.5 times the IQR. **c** Oncoprint showing identified subgroups

for the benchmarked methods, including std, CMS subtypes, tumour sidedness and mutations in *KRAS* and *NRAS*. **d** Forest plot showing hazard ratios including 95% confidence intervals (CI) and amount of patients in the subgroup for which standard treatment is not recommended and which was found by subgroup analysis methods (new-std-negative). A Source Data file is provided, which contains the source data for the figure panels and the sample sizes of the conducted statistical tests.



Ideally, subgroup analysis should reveal subgroups with high treatment effects for refining treatment strategies and recover subgroups in the standard treatment strategy. Therefore, we evaluated the newly proposed subgroup for which standard treatment is not recommended (new-std-negative) for each method. We derived the hazard ratios for cetuximab benefit based on OS for all methods in the new-std-negative subgroups (Fig. 6d). Lower hazard ratios in new-std-negative patients indicate the discovery of off-label subgroups for which cetuximab is currently not recommended (Fig. 6d). Accordingly, OncoBird showed the numerically lowest hazard ratio  $HR = 0.57$  ( $p = 0.16$ ,  $N = 29$ ) for the new-std-negative subgroup compared to all other methods (Supplementary Table 1; Fig. 6d).

In summary, many of the computational methods reproduced the clinically established biomarkers, whilst OncoBird empowers advanced biomarker identification by thoroughly integrating biological priors in the form of tumour subtypes. The simplicity of statistical models leveraged in OncoBird further increases interpretability and transparency.

## Discussion

We demonstrated that OncoBird has the capabilities to characterise the molecular and biomarker landscape of RCTs. Here exemplified, we captured the established clinical biomarkers of FIRE-3, and proposed five predictive biomarker hypotheses ( $FDR_{int} < 0.2$ ). The biomarkers were based on either individual cancer genes or mutually exclusive patterns and exploited these genetic events in the context of well-characterised cancer subtypes. In addition, OncoBird thoroughly corrects for multiple hypothesis testing and includes resampling-based adjustments of treatment effects. In essence, OncoBird systematically investigated the molecular landscape of the FIRE-3 clinical trial, suggested biomarkers based on genetic alterations, performed a data-driven subgroup analysis, and finally, presented the results in an interpretable and intuitive way.

The statistical power of detecting biomarkers depends on the amount of screened genes and subtypes, sample sizes and magnitude of treatment effects. For example, subtype-specific analyses reduce patient subgroup sizes, thus limiting the power for detecting interactions. In order to gain statistical power to detect genetic biomarkers with low mutational frequency, OncoBird exploits mutually exclusive modules ("Methods"). Despite the use of resampling-based treatment effect estimation in the found subgroups, hypotheses generated by exploratory tools such as OncoBird ought to be replicated in independent clinical trials. Nonetheless, OncoBird identified promising patient subpopulations within the FIRE-3 and ADJUVANT clinical trials with supported biological interpretation, which indicated refined predictive benefits in cancer subtypes.

A limitation of data-driven subgroup analysis is that these may produce spurious results if not biologically interpretable<sup>47</sup>. To mitigate this risk, we used established tumour subtypes with distinct tumour biology in mCRC, i.e., here, the consensus molecular subtypes (CMS)<sup>29</sup> and primary tumour sidedness<sup>21</sup>. Furthermore, the grouping of functionally similar mutually exclusive somatic mutations in the cancer gene sequencing panel reinforced the identification of biological signals.

Somatic mutations may drive tumour subtypes, therefore, we systematically investigated mutational patterns within CMS1-4 and tumour sidedness. We found the majority of *BRAF* mutations in CMS1 and observed a co-occurrence between CMS2, left-sided tumours and amplifications in chr20q. In particular, CMS2 is characterised by a *MYC* signalling activation<sup>29</sup>, which may be co-regulated by activation of the co-amplified *AURKA*<sup>48</sup>. While we predominantly identified CMS-specific biomarkers, our results suggest that both primary tumour side and CMS subtypes play a major role in the landscape of predictive biomarkers. This highlights the need for OncoBird, an integrated biomarker discovery framework, which integrates the molecular landscape of RCTs with its biomarkers.

Several genes were co-amplified in chr20q, i.e., *ARFRP1*, *TOPI*, and *SRC*, thus determining the drivers among these biomarker candidates is challenging. Among the prominent chr20q amplifications, *TOPI* was previously proposed as a biomarker for irinotecan efficacy in metastatic colorectal cancer<sup>49,50</sup>, which is part of the chemotherapeutic backbone of the FIRE-3 trial. Literature suggests that *TOPI* abundance is essential for irinotecan-induced DNA double-strand breaks during DNA replication<sup>51</sup>. Additionally, *TOPI* was identified to regulate EGFR through an endogenous interaction with the transcription factor c-Jun<sup>52</sup>, which supports the hypothesis that *TOPI* amplifications may be the actionable biomarker. *SRC* has been reported to play a role in cancer progression<sup>53,54</sup>, whereas for *ARFRP1*, no functional evidence has been presented yet.

The resulting co-amplifications between these cancer genes complicate the determination of the genetic driver in chr20q. To understand the causality of cancer aetiologies, further efforts require additional treatment regimes. Alternative clinical trials for metastatic colorectal cancer often involve different chemotherapy backbones, i.e., fluorouracil, leucovorin, and oxaliplatin (FOLFOX) or fluorouracil, leucovorin, and irinotecan (FOLFIRI)<sup>30</sup>. The use of other therapy backbones may unravel the role of *ARFRP1*, *TOPI* and *SRC* amplifications regarding better efficacy for patients treated with cetuximab. However, discrepancies may arise due to the synergism and antagonism of the different chemotherapy backbones and targeted treatments<sup>55</sup>.

The prognostic potential of *APC* wild-type tumours for bevacizumab has been previously reported<sup>56</sup>, whereas OncoBird did not yield enough evidence to support this. Indeed, a confounding factor is the enrichment of *BRAF* mutations in the *APC* wild-type tumours ( $p = 1.4 \times 10^{-10}$ , Fisher's exact test). This is, 48% of *APC* wild-type tumours were *BRAF* mutated in the bevacizumab treatment arm, whereas in the cetuximab treatment arm, only 29% were *BRAF* mutated ( $p = 0.13$ , Fisher's exact test). Nevertheless, independently a correlation between *VEGFA* expression and the mutational status of *APC* has been previously observed in primary colorectal tumour samples<sup>57</sup>, suggesting that within *APC* mutated tumours, anti-VEGF treatment may indeed be beneficial.

Furthermore, *RAS/BRAF* mutations are known to harbour prognostic value in terms of overall survival<sup>38,43</sup>. Furthermore, we observed that *KRAS* mutations showed highly CMS-specific responses. In particular, treatment response differed for tumours classified as CMS4 by *KRAS* status, showing better response for cetuximab in *KRAS* wild-type and for bevacizumab in *KRAS* mutated tumours, respectively. CMS4 has been reported to be associated with VEGF pathway activation and is thus associated with angiogenesis<sup>29</sup>. Thus, patients with tumours resistant towards anti-EGFR treatment may benefit from VEGF inhibition. Further exclusion of *BRAF* mutations did not elevate the predictive potential of *KRAS* mutations in CMS4. However, the statistical power is limited by the fact that only six tumours harboured the prognostically unfavourable *BRAF* V600E mutation in CMS4<sup>20</sup>.

In summary, OncoBird reproduced clinically established biomarkers and derived five hypotheses of biomarkers with predictive roles for FOLFIRI plus either cetuximab or bevacizumab. Highlighted examples include chr20q amplifications in CMS2 and *KRAS* mutations in CMS4, which may optimise patient stratification for metastatic colorectal cancer. Leveraging OncoBird for molecular profiling in the FIRE-3 clinical trial offered an expanded perspective on the molecular and biomarker landscape of these patients.

In the future, we anticipate that the analysis of clinical trials will progressively demand molecular patient tumour data, including pre-defined subtypes, highlighting the urgent need for integrative analysis tools such as OncoBird. Notably, OncoBird was developed for RCT designs and is generalisable to any trial designs for which the intention-to-treat population was defined before the treatment randomisation, i.e., the treatment assignment is independent of patient characteristics. According to this, OncoBird is applicable to modern clinical trial

designs based on master protocols<sup>58</sup>, i.e., basket, umbrella, and platform trials if control arms are included. In an emerging landscape of predictive molecular biomarkers in cancer, OncoBird may untangle complex dependencies between somatic alterations and tumour subtypes in RCTs. Furthermore, OncoBird is generalisable to any cancer entity, thus ultimately paving the way for the next generation of precision oncology therapies.

## Methods

### Clinical data of the FIRE-3 clinical trial

FIRE-3 is an open-label, randomised phase III trial to compare first-line treatment in *KRAS* exon 2 wild-type metastatic colorectal cancer patients (mCRC) with either cetuximab or bevacizumab in combination with 5-fluorouracil, leucovorin and irinotecan (FOLFIRI). The protocol and rules of conduct were previously published<sup>23,59</sup> (NCT00433927). The trial was conducted in accordance with the declaration of Helsinki (1996). All translational analyses were approved by the local ethics committee (University of Munich, registry no. 186-15). All patients included in this analysis provided written informed consent. 24% and 34% of the patients had female sex in the FOLFIRI plus cetuximab and bevacizumab arm, respectively. The sex is reported according to the study protocol<sup>23,59</sup>, and gender cannot be distinguished retrospectively. The biological sex of patients (i.e., male or female) was assigned by the study doctor of the respective trial centre and reported to the clinical research organisation (CRO). The original intention-to-treat population consisted of 752 patients in total. Primary and secondary endpoints of the FIRE-3 trial, including the median overall survival (OS) and progression-free survival (PFS), were expressed as months and defined as stated in the respective articles<sup>23,59</sup>. The objective response rate (ORR) was evaluated by the RECIST 1.0 criteria<sup>23,59</sup>.

### Next-generation sequencing and genetic alterations in FIRE-3

Primary tumour tissues from 373 patients have been molecularly characterised by next-generation sequencing (NGS) with the FoundationOne® panel (Foundation Medicine, Inc., MA, USA; catalogue number not available), which identified somatic mutations and copy number alterations, i.e., deletions and amplifications, of 277 key cancer genes, microsatellite instability (MSI) and tumour mutational burden<sup>20</sup>. Somatic alterations were delivered in the form of binary matrices, that reflect the mutant or wild-type status of a given gene based on single nucleotide variants (SV), copy number amplifications (AMP) and deletions (DEL). MSI is an important prognostic predictor and enriched in CMSI<sup>60</sup>, which is observed in our study, with 8 of 10 MSI-H tumours being classified as CMSI. However, MSI-H tumours are less prevalent in metastatic disease (~5%)<sup>60</sup>. Furthermore, only six and four MSI-H tumours were treated with bevacizumab and cetuximab, respectively.

### Gene expression profiling in FIRE-3

The genetic characterisation is complemented with gene expression profiles from Xcel® microarrays (Almac Ltd, Belfast, UK; catalogue number: 902016) in a subset of 451 patients. The clinical data and the layers of molecular characterisation led to 163 and 186 patients, which are fully characterised in the cetuximab and bevacizumab treatment arms, respectively.

### Tumour subtypes in FIRE-3

A clinically established subtype for mCRC is its primary tumour sidedness. Left-sided tumours were located in the left hemicolon, e.g., splenic flexure to the rectum. In contrast, right-sided tumours were located in the right colon, e.g., caecum to the transverse colon. In addition, annotations for molecular subtypes of mCRC were obtained from transcriptome data that has been previously used to classify

patients into their closest consensus molecular subtype (CMS)<sup>21,29</sup> using the *cmsclassifier* package with the SSP predictor. Thereby, 24 of out 373 patient tumours were not allocated to any CMS because of missing transcriptomics data and were left out of the CMS-specific analysis. The CMS classification was used as a complementary alternative to the primary tumour side and is currently discussed in multiple clinical settings<sup>61</sup>.

### Oncology Biomarker Discovery workflow

The Oncology Biomarker Discovery (OncoBird) framework applies to RCTs for which patients received either treatment  $t \in \{0,1\}$  according to the treatment indicator  $T$ , had an associated outcome  $Y$  and can be classified into  $q$  subtypes  $\{s_1, \dots, s_q\}$  according to the subtype variable  $S$  (clinical data). Additionally, patient tumours are characterised by  $m$  candidate genetic biomarkers  $\mathbf{X} = X_1, \dots, X_m$  with the observed biomarkers for patients  $\mathbf{x} = x_1, \dots, x_m$  (genetic data). The genetic data can be used to group functionally similar genes that can be added to the set of candidate biomarkers. Furthermore, it is possible to add additional binary features to  $\mathbf{X}$  such as binarised copy number alterations with appropriate cutoffs or the MSI status of a tumour. Both genetic data (*MUT*) and clinical data (*CLIN*) are required inputs to the OncoBird workflow (Supplementary Data 1), which is described in the following sections. All implemented thresholds of OncoBird can be adjusted by the user, thus empowering more lenient or stringent analyses.

### Characterising the molecular landscape in clinical trials

OncoBird first examines genetic features  $\mathbf{X}$  in tumour subtypes  $\{s_1, \dots, s_q\}$  independent of the treatment and patient response (function `GET-MUTATIONS-IN-SUBTYPES` in Supplementary Data 1). For examining enrichment or depletion of each genetic feature in tumour subtypes, one-sided hypergeometric tests are performed using the ‘phyper’ R function. Consecutively, the resulting  $p$ -values are corrected for multiple hypothesis testing with the Benjamini–Hochberg (BH) method<sup>62</sup>. The FDR cutoff for this analysis step is denoted by  $FDR_{\text{mol}}$  and controlled at  $FDR_{\text{mol}} = 0.05$  as our default setting. Our method generalises to any binary tumour characterisation, e.g., the MSI status in FIRE-3. As a default setting, we test genetic features that were mutated in at least ten tumours ( $n = 10$ ).

### Identifying mutual exclusivity

For the identification of mutually exclusive modules, we used the Mutex algorithm<sup>39</sup> (function `GET-MUTATIONS-MODULES` in Supplementary Data 1). It leverages a signalling network<sup>63</sup> collecting interactions from Pathway Common<sup>64</sup>, SPIKE<sup>65</sup> and SignalLink<sup>66</sup> in order to scan for common downstream effects of combinations of somatic alterations  $\mathbf{X}$ . The default setting only uses somatic variants that were altered in at least ten tumours ( $n = 10$ ).

### Genetic and subtype-specific biomarkers

OncoBird tests single somatic alterations and previously derived mutually exclusive somatic alterations for differential prognosis in each treatment arm separately (function `GET-TREATMENT-SPECIFIC-BIOMARKERS` in Supplementary Data 1). The patient outcome  $Y(T = t, S = s_k)$  for the treatment arm  $T = t$  in subtype  $S = s_k$  with  $k = 1, \dots, q$  may be defined by survival data (OS or PFS) or a binary variable measuring the objective response rate (ORR). Depending on the type of outcome, this is modelled with either Cox proportional hazards regression models or logistic regression models expressed by their linear predictor function  $f(\mathbf{x}, t)$ . Using this classical approach for subgroup analysis, the treatment-specific regression models in subtypes take the form

$$f(\mathbf{x}, t) = \alpha_{0j} + \alpha_{1j}x_j + \sum_l C_l \quad (1)$$

Cox proportional hazards regression models for survival endpoints were implemented with the 'coxph' function from the *survival* R package or logistic regression models for binary response variables were implemented using the 'glm' function. We test each  $\mathbf{x} = x_1, \dots, x_m$  first across all tumours, and subsequently in tumour subtypes  $\{s_1, \dots, s_q\}$ , i.e., CMS or primary sidedness.  $\alpha_{ij}$  is the coefficient estimating the contribution of candidate biomarker  $j = 1, \dots, m$  for patient outcomes in the context of each treatment arm  $T = t$  in the subtype  $S = s_k$ . The predictors  $C_1, \dots, C_l$  include additional prognostic covariates and their coefficients.

The  $p$ -value  $p_{\alpha_{ij}}$  derived by a Wald test from the coefficient  $\alpha_{ij}$  is multiplicity-adjusted for each treatment arm  $t$  and across all biomarkers  $x_j$  with  $j = 1, \dots, m$  for either all patients or across subtypes  $s_k$  with  $k = 1, \dots, q$  and yields adjusted  $p$ -values  $\tilde{p}_{\alpha_{ij}}$  using the Benjamini–Hochberg (BH) method<sup>62</sup>. The default false discovery rates (FDR) are controlled at  $FDR_{\alpha} = 0.1$  for either treatment-specific component  $\alpha_{ij}$ .

The adjustable default setting of OncoBird is to only perform statistical tests if, for a given candidate biomarker  $x_j$  and tumour subtype  $s_k$ , at least  $n = 10$  samples were present in each mutant and wild-type population. Additionally, OncoBird only tested alterations for which its corresponding gene module had at least  $n$  tumours redistributed compared to the single gene alteration.

### Predictive components of biomarkers

For the subsequent comparison of treatment arms, OncoBird tests for significant statistical interactions between treatment arms and genetic alterations in tumour subtypes (function `GET-PREDICTIVE-BIOMARKERS` in Supplementary Data 1). For that, we modelled the outcome  $Y(S = s_k)$  in subtype  $S = s_k$  with  $k = 1, \dots, q$  using regression models with interactions between  $T$  and  $X_j$  which take the form

$$f(\mathbf{x}, t) = \beta_{0j} + \beta_{1j}x_j + \beta_{2j}x_jt + \sum_l C_l, \quad (2)$$

where the coefficients  $\beta_{1j}$  and  $\beta_{2j}$  estimate the prognostic and predictive component of biomarker  $x_j$  in subtype  $s_k$ , respectively. The  $p$ -value  $p_{\beta_{2j}}$  derived with a Wald test from the coefficient  $\beta_{2j}$  is multiplicity-adjusted across all  $m$  biomarkers for either all patients or across subtypes  $s_k$  with  $k = 1, \dots, q$  and yields BH adjusted  $p$ -values  $\tilde{p}_{\beta_{2j}}$ . The default FDR is controlled at  $FDR_{\beta} = 0.2$  for predictive components. The biomarker  $X_j$  in subtype  $s_k$  is a putatively predictive biomarker if  $\tilde{p}_{\alpha_{ij}} < FDR_{\alpha}$  for either  $t$  and  $\tilde{p}_{\beta_{2j}} < FDR_{\beta}$ .

Furthermore, OncoBird only performs statistical tests if for a given genetic alteration  $X_j$  and tumour subtype  $s_k$ , at least  $n = 10$  samples were present in each mutant and wild-type population for each treatment arm as default setting.

### Resampling for correction of conditional average treatment effects

Lastly, we estimate the conditional average treatment effect (CATE) for the found biomarkers (function `GET-PREDICTIVE-BIOMARKERS` in Supplementary Data 1). For each significant  $X_j$  in  $s_k$ , there is one CATE estimate in each found subpopulation with a positive (mutant) biomarker  $x_j = 1$  and negative (wild type) biomarker  $x_j = 0$ . In each population, we estimate the CATE by modelling the outcome  $Y$  by

$$f(\mathbf{x}, t) = \gamma_0 + \gamma_1t + \sum_l C_l, \quad (3)$$

where  $\gamma_1$  estimates the (biased) CATE in terms of either hazard ratios or odds ratios dependent on outcome type in the subgroup defined by biomarker  $x_j$  and subtype  $s_k$ . The population with the larger absolute estimate  $\gamma_1$  is used to estimate the subgroups  $A_{x_j, s_k}$ .

For each found subgroup  $A$ , we assess the significance to the associated CATE estimate  $\gamma_1$  and derive the  $p$ -value  $p_{\gamma_1}$  using a Wald test. Furthermore, we perform a multiplicity-adjustment of  $p_{\gamma_1}$  and derive honest estimates of the CATE.

The  $p$ -values are adjusted for multiplicity using a permutation-based approach that takes into account the entire subgroup search strategy<sup>3</sup>. For that, we permuted the treatment labels  $U = 1000$  times to obtain null datasets without any differential treatment effects. Next, for each null dataset, we select significant subgroups  $A^{(u)}$  for the same thresholds and record the treatment effect  $p$ -value of the best subgroup  $p^{(u)}$  with  $u = 1, \dots, U$ . The adjusted  $p$ -values are then given by

$$\tilde{p}_{\gamma_1} = \frac{1}{U} \sum_{u=1}^U I_{\{p^{(u)} \leq p_{\gamma_1}\}}(p^{(u)}), \quad (4)$$

the fraction of  $p$ -values  $p^{(u)}$  that are smaller or equal than  $p_{\gamma_1}$  with the indicator function  $I$ . Furthermore, we derive an honest estimate of the treatment effect  $\gamma_1$ . Since subgroups  $A$  are derived from the same data as the treatment effect estimates, the estimates from the resubstitution  $\gamma_1(A_{x_j, s_k})$  will be biased. In order to derive a bias-corrected estimate  $\tilde{\gamma}_1$ , we use a previously proposed non-parametric bootstrap approach<sup>9</sup>. For that, we generated  $B = 500$  bootstrapped datasets. For each resampled dataset  $b = 1, \dots, B$  we estimate subgroups  $\hat{A}_{x_j, s_k}^{(b)}$ . The treatment effects can then be either estimated on the  $b$ -th resampled dataset  $\gamma_1^{(b)}(A^{(b)})$  or on the original dataset  $\gamma_1(A^{(b)})$ . The bias-corrected CATE estimate is then given by

$$\hat{\gamma}_1 = \frac{1}{B} \sum_{b=1}^B (\gamma_1(A) + \gamma_1(A^{(b)}) - \gamma_1^{(b)}(A^{(b)})). \quad (5)$$

The 95% confidence intervals are constructed by the 0.025 and 0.975 quantiles of the bootstrapped distribution.

### OncoBird parameterisation for FIRE-3

We used the function `GET-MUTATIONS-IN-SUBTYPES` to evaluate the primary tumour side and CMS as tumour subtypes with the default setting  $FDR_{mol} < 0.05$ . In total, we performed 156 and 312 statistical tests for the primary tumour sidedness and CMS, respectively. Using the `GET-MUTATIONS-MODULES` function with default settings, we analysed 42 genes which yielded 29 mutually exclusive modules. Mutations in *KRAS* or *NRAS* are the established clinical biomarkers for anti-EGFR treatment, thus we jointly modelled *KRAS* and *NRAS* as *RAS* mutations resulting in 10 additional modules.

The `GET-TREATMENT-SPECIFIC-BIOMARKERS` function was used with the number of metastatic sites and the information about a prior tumour resection as added covariates  $C_1, C_2$ . With the OncoBird default setting, we performed 816 statistical tests across all readouts  $Y$  (OS, PFS and ORR), the cetuximab and bevacizumab treatment arm and tumour subtypes, i.e., CMS1-4, left- and right-sided and across all tumours. FDR cutoffs are employed for each treatment arm separately and are denoted  $FDR_{cet}$  and  $FDR_{bev}$  for the analysis in the cetuximab and bevacizumab treatment arms, respectively. In total, we found 92 significant associations with the default setting  $FDR_{cet/bev} < 0.1$ . The criteria  $HR < 1$  and  $OR < 1$  corresponded to a better prognosis for the mutant tumours compared to the wild-type tumours and vice versa. To consistently report  $HR < 1$  and  $OR < 1$  as beneficial risk reduction, reciprocal values of HRs and ORs were used if wild-type tumours displayed a better prognosis. We represent  $p$ -values, hazard/odds ratios with the 95% confidence intervals (CI) in square brackets and the associated FDRs.



In FIRE-3, the `GET-PREDICTIVE-BIOMARKERS` function with default settings resulted in a total amount of 396 statistical tests across the readouts  $Y$  (OS, PFS and ORR) and the tumour subtypes  $S_k$ . FDR cutoffs for the interaction tests across both treatment arms are denoted by  $FDR_{int}$ . We explored 57 associations with  $FDR_{int} < 0.6$  and  $FDR_{cet/bev} < 0.1$  (Supplementary Data 3) and further focused on a subset of five biomarkers with default setting  $FDR_{int} < 0.2$  for OS, i.e., two gene modules and three single genes (Supplementary Data 4). For the cross-validation analysis, a more lenient  $FDR_{int} < 0.3$  was employed, which deviated from default setting to account for reduced sample sizes in the training and testing splits. HRs and ORs  $>1$  and  $<1$  corresponded to benefit with cetuximab and bevacizumab, respectively. To report the benefits of cetuximab treatment, the reciprocal values of HRs and ORs were used in the manuscript in order to display treatment benefits consistently with  $HR < 1$  and  $OR < 1$ . We reported  $p$ -values and hazard/odds ratios with the 95% CIs for the treatment comparison and the  $p$ -values and associated FDRs for the interaction tests.

### OncoBird parameterisation for ADJUVANT

The ADJUVANT clinical trial in *EGFR* mutant non-small cell lung cancer (NSCLC) aimed to assess the efficacy of gefitinib versus chemotherapy with vinorelbine and cisplatin (NCT01405079)<sup>34</sup>. The trial was previously approved by the research ethics boards of Guangdong Provincial People's Hospital and all other participating hospitals<sup>35</sup>. Of note, 58% and 59% of patients had female sex in the gefitinib and chemotherapy arm, respectively. The sex was reported according to the study protocol<sup>34</sup>, and gender cannot be distinguished retrospectively. We used the *EGFR* subtype, i.e., exon 19 deletion or exon 21 Leu858Arg, and the smoking history as putative tumour subtypes and clinical endpoints were disease-free survival (DFS) and overall survival (OS). We analysed 22 somatic alterations in 171 patients, from which 76 patients were treated with chemotherapy alone, and 95 were treated with gefitinib. For the subsequent analysis, we used the OncoBird default settings. The obtained results (Supplementary Data 2; Supplementary Figs. 3–6) and an associated extensive report can be reproduced in a runnable demo on Code Ocean (<https://codeocean.com/capsule/9911222/tree/v1>).

### Benchmarking of alternative methods with FIRE-3

For benchmarking the biomarker identification, we compared OncoBird to seven competing subgroup analysis algorithms leveraging the overall survival of FIRE-3 (Supplementary Table 1)<sup>8,9,11,12,44–46</sup>. We formed predictors by concatenating clinical annotations, including information about tumour resection, number of metastatic sites, age, gender, MSI and lung metastatic status. We added single genetic alterations and mutually exclusive modules observed across at least ten patients and in both investigated tumour subtypes, thus mirroring the OncoBird default settings. Furthermore, we investigated interactions between genetic alterations and tumour primary sidedness or CMS as predictors. Subgroups for the method evaluation were formed as the union of the subgroups showing cetuximab benefit according to the identified biomarkers (Supplementary Table 1).

All benchmarked models were 5-fold cross-validated with five repetitions. A univariate Cox proportional hazards model assessed performances leveraging the treatment effect based on OS in the subgroups with predicted benefits according to the found biomarkers. This included the treatment effect across the whole test set and in the subgroup defined by the current treatment guidelines, i.e., left-sided and *RAS* wild-type tumours<sup>37</sup>. The significance of the treatment effect in the subgroups of the test set was assessed using a modified  $t$ -test for resampled performance metrics<sup>67</sup>, denoted by  $p_{cv}$ .

For comparing computational methods and their predicted biomarkers, the models were fitted on the whole dataset. The parameterisation of these methods was followed according to the suggested default settings unless in conflict with the above outlined

use case. For example, for tree-based methods, the features contained in the resulting tree were used as biomarkers with tree depths = 2, with a minimum subgroup size of  $n = 10$ . For the implementation of the virtual twins method (VT)<sup>9</sup>, we used the R package *randomForestSRC* with default parameters and averaged predictions over 10 times repeated 10-fold cross-validation. Subsequently, a regression tree was fitted to the original data. In order to perform model-based recursive partitioning<sup>8</sup>, we used the R package *model4you*<sup>68</sup> using an exponential model with default conditional inference tree control parameters. The PRISM method<sup>46</sup> was implemented in the R package *StratifiedMedicine*, for which we used Cox proportional hazards regression. We used the implementation of causal survival forests<sup>69</sup> (CRF) in the R package *grf*<sup>60</sup> for estimating conditional treatment effects. The propensity scores were set as constant and the target estimand was set to restricted mean survival time (RMST) with horizon = 100. After model fitting, variable importance scores were extracted, and biomarkers were selected according to predictors with significant linear projections of the conditional average treatment effects ( $p < 0.05$ ). Next, we employed policy learning (POL)<sup>44</sup> to find optimal treatment regimens using the R package *policytree*<sup>71</sup>. We used the 50 most important predictors according to the CRF causal survival forest model variable importance scores and their treatment effect estimates to produce a decision tree.

The remaining methods were not based on trees. For the outcome weighted method (OWE)<sup>11</sup>, implemented in the R package *personalized*<sup>72</sup>, we used a constant propensity score, lasso loss and 10-fold cross-validation. The GUIDE method<sup>45</sup> was available as a binary executable under <https://pages.cs.wisc.edu/~loh/guide.html>. We used Cox proportional hazards regression with interactions tests and mean-based trees with pruning. For the SIDES method (R package *SIDES*)<sup>7</sup>, we used `level_control=0` and `alpha=0.05`.

### Statistics and reproducibility

The investigators were not blinded to the randomised treatment allocation during the data collection and outcome assessment. Since the conducted subgroup analysis is retrospective, the sample sizes were not predetermined. No data were excluded from the analysis. Details of the conducted statistical tests are provided in the figure captions, Supplementary Data 2–4 and Source Data. The results of the statistical analysis of the ADJUVANT clinical trial are reproducible from a demo run on Code Ocean (<https://codeocean.com/capsule/9911222/tree/v1>).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The clinical data summary from the FIRE-3 clinical trial analysed in this study has been deposited in the Pharmnet.bund online platform of the German Federal Ministry of Health ([https://portal.dimdi.de/data/ctr/O-0329\\_01-2-1-B80630-20190731152224.pdf](https://portal.dimdi.de/data/ctr/O-0329_01-2-1-B80630-20190731152224.pdf)) and was published before<sup>19</sup>. The clinical and molecular data is available under restricted access due to data privacy laws. The raw and processed data can be obtained through the corresponding author at [volker.heinemann@med.uni-muenchen.de](mailto:volker.heinemann@med.uni-muenchen.de). The data from the results of OncoBird v0.1.0 executed on the FIRE-3 trial are available in Supplementary Data 3 and Source Data. The processed data from the ADJUVANT clinical trial is available on Zenodo<sup>33,35</sup>. The data from the results of OncoBird v0.1.0 executed on the ADJUVANT trial are available in Supplementary Data 2, Source Data and on Code Ocean (<https://codeocean.com/capsule/9911222/tree/v1>). Source data are provided with this paper.

### Code availability

Oncology Biomarker Discovery (OncoBird) is publicly available at <https://github.com/MendenLab/OncoBird>. The repository contains an



R package as well as a Shiny application with a graphical user interface in a local docker container (Supplementary Fig. 1). Additionally, a demo run of OncoBird v0.1.0 used for analysis is available on Code Ocean (<https://codeocean.com/capsule/9911222/tree/v1>).

## References

- Ting, N., Cappelleri, J. C., Ho, S. & Chen, D.-G. (eds) *Design and Analysis of Subgroups with Biopharmaceutical Applications* (Springer, 2020).
- European Medicines Agency. *Guideline on the Investigation of Subgroups in Confirmatory Clinical Trials*. Draft. European Medicines Agency/Committee for Medicinal Products for Human Use. EMA/CHMP/539146/2013 (EMA, 2014).
- Lipkovich, I., Dmitrienko, A. & D'Agostino Sr, B. R. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Stat. Med.* **36**, 136–196 (2017).
- Zhang, Z., Seibold, H., Vettore, M. V., Song, W.-J. & François, V. Subgroup identification in clinical trials: an overview of available methods and their implementations with R. *Ann. Transl. Med.* **6**, 122 (2018).
- Loh, W., Cao, L. & Zhou, P. Subgroup identification for precision medicine: a comparative review of 13 methods. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **9**, e1326 (2019).
- Lipkovich, I., Dmitrienko, A., Denne, J. & Enas, G. Subgroup identification based on differential effect search—a recursive partitioning method for establishing response to treatment in patient subpopulations. *Stat. Med.* **30**, 2601–2621 (2011).
- Lipkovich, I. & Dmitrienko, A. Strategies for identifying predictive biomarkers and subgroups with enhanced treatment effect in clinical trials using SIDES. *J. Biopharm. Stat.* **24**, 130–153 (2014).
- Seibold, H., Zeileis, A. & Hothorn, T. Model-based recursive partitioning for subgroup analyses. *Int. J. Biostat.* **12**, 45–63 (2016).
- Foster, J. C., Taylor, J. M. G. & Ruberg, S. J. Subgroup identification from randomized clinical trial data. *Stat. Med.* **30**, 2867–2880 (2011).
- Xu, Y. et al. Regularized outcome weighted subgroup identification for differential treatment effects. *Biometrics* **71**, 645–653 (2015).
- Chen, S., Tian, L., Cai, T. & Yu, M. A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics* **73**, 1199–1209 (2017).
- Wager, S. & Athey, S. Estimation and inference of heterogeneous treatment effects using random forests. *J. Am. Stat. Assoc.* **113**, 1228–1242 (2018).
- Künzel, S. R., Sekhon, J. S., Bickel, P. J. & Yu, B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl Acad. Sci. USA* **116**, 4156–4165 (2019).
- Cremolini, C. et al. First-line chemotherapy for mCRC—a review and evidence-based algorithm. *Nat. Rev. Clin. Oncol.* **12**, 607–619 (2015).
- Thomas, R. K. et al. High-throughput oncogene mutation profiling in human cancer. *Nat. Genet.* **39**, 347–351 (2007).
- Kawazoe, A. et al. A retrospective observational study of clinicopathological features of KRAS, NRAS, BRAF and PIK3CA mutations in Japanese patients with metastatic colorectal cancer. *BMC Cancer* **15**, 258 (2015).
- Van Cutsem, E. et al. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N. Engl. J. Med.* **360**, 1408–1417 (2009).
- Saltz, L. B. et al. Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study. *J. Clin. Oncol.* **26**, 2013–2019 (2008).
- Heinemann, V. et al. FOLFIRI plus cetuximab or bevacizumab for advanced colorectal cancer: final survival and per-protocol analysis of FIRE-3, a randomised clinical trial. *Br. J. Cancer* **124**, 587–594 (2021).
- Stahler, A. et al. Single-nucleotide variants, tumour mutational burden and microsatellite instability in patients with metastatic colorectal cancer: next-generation sequencing results of the FIRE-3 trial. *Eur. J. Cancer* **137**, 250–259 (2020).
- Stintzing, S. et al. Consensus molecular subgroups (CMS) of colorectal cancer (CRC) and first-line efficacy of FOLFIRI plus cetuximab or bevacizumab in the FIRE3 (AIO KRK-0306) trial. *J. Clin. Orthod.* **35**, 3510–3510 (2017).
- Laurent-Puig, P. et al. MiR-31-3p is a predictive biomarker of cetuximab response in FIRE3 clinical trial. *Ann. Oncol.* **27**, vi151 (2016).
- Heinemann, V. et al. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer (FIRE-3): a randomised, open-label, phase 3 trial. *Lancet Oncol.* **15**, 1065–1075 (2014).
- Duarte, S. et al. Right vs left-sided RAS wild-type metastatic colorectal cancer treated with EGFR inhibitors: prognostic differences. *Ann. Oncol.* **30**, iv53 (2019).
- Stintzing, S. et al. Randomized study to investigate FOLFOXIRI plus either bevacizumab or cetuximab as first-line treatment of BRAF V600E-mutant mCRC: the phase-II FIRE-4.5 study (AIO KRK-0116). *J. Clin. Orthod.* **39**, 3502–3502 (2021).
- Peeters, M. et al. Massively parallel tumor multigene sequencing to evaluate response to panitumumab in a randomized phase III study of metastatic colorectal cancer. *Clin. Cancer Res.* **19**, 1902–1912 (2013).
- Seymour, M. T. et al. Panitumumab and irinotecan versus irinotecan alone for patients with KRAS wild-type, fluorouracil-resistant advanced colorectal cancer (PICCOLO): a prospectively stratified randomised trial. *Lancet Oncol.* **14**, 749–759 (2013).
- Dienstmann, R., Salazar, R. & Tabernero, J. Overcoming resistance to anti-EGFR therapy in colorectal cancer. *Am. Soc. Clin. Oncol. Educ. Book* **35**, e149–e156 (2015).
- Guinney, J. et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
- Lenz, H.-J. et al. Impact of consensus molecular subtype on survival in patients with metastatic colorectal cancer: results from CALGB/SWOG 80405 (Alliance). *J. Clin. Oncol.* **37**, 1876–1885 (2019).
- Mooi, J. K. et al. The prognostic impact of consensus molecular subtypes (CMS) and its predictive effects for bevacizumab benefit in metastatic colorectal cancer: molecular analysis of the AGITG MAX clinical trial. *Ann. Oncol.* **29**, 2240–2246 (2018).
- Sveen, A., Kopetz, S. & Lothe, R. A. Biomarker-guided therapy for colorectal cancer: strength in complexity. *Nat. Rev. Clin. Oncol.* **17**, 11–32 (2020).
- cancer-oncogenomics. *cancer-oncogenomics/minerva-adjuvant-nslc: adjuvant minerva study v1.0.0*. Zenodo <https://doi.org/10.5281/zenodo.5242512> (2021).
- Zhong, W.-Z. et al. Gefitinib versus vinorelbine plus cisplatin as adjuvant treatment for stage II-IIIa (N1-N2) EGFR-mutant NSCLC (ADJUVANT/CTONG1104): a randomised, open-label, phase 3 study. *Lancet Oncol.* **19**, 139–148 (2018).
- Liu, S.-Y. et al. Genomic signatures define three subtypes of EGFR-mutant stage II-III non-small-cell lung cancer with distinct adjuvant therapy outcomes. *Nat. Commun.* **12**, 6450 (2021).
- Holch, J. W., Ricard, I., Stintzing, S., Modest, D. P. & Heinemann, V. The relevance of primary tumour location in patients with metastatic colorectal cancer: a meta-analysis of first-line clinical trials. *Eur. J. Cancer* **70**, 87–98 (2017).
- Chiorean, E. G. et al. Treatment of patients with late-stage colorectal cancer: ASCO Resource-Stratified Guideline. *JCO Glob. Oncol.* **6**, 414–438 (2020).
- Ptashkin, R. N. et al. Chromosome 20q amplification defines a subtype of microsatellite stable, left-sided colon cancers with wild-

- type RAS/RAF and better overall survival. *Mol. Cancer Res.* **15**, 708–713 (2017).
39. Babur, Ö. et al. Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. *Genome Biol.* **16**, 45 (2015).
40. Hsu, H.-C. et al. Mutations of KRAS/NRAS/BRAF predict cetuximab resistance in metastatic colorectal cancer patients. *Oncotarget* **7**, 22257–22270 (2016).
41. Díaz-Rubio, E. et al. Role of Kras status in patients with metastatic colorectal cancer receiving first-line chemotherapy plus bevacizumab: a TTD group cooperative study. *PLoS ONE* **7**, e47345 (2012).
42. Modest, D. P. et al. Outcome according to KRAS-, NRAS- and BRAF-mutation as well as KRAS mutation variants: pooled analysis of five randomized trials in metastatic colorectal cancer by the AIO colorectal cancer study group. *Ann. Oncol.* **27**, 1746–1753 (2016).
43. Zhang, B., Yao, K., Zhou, E., Zhang, L. & Cheng, C. Chr20q amplification defines a distinct molecular subtype of microsatellite stable colorectal cancer. *Cancer Res.* **81**, 1977–1987 (2021).
44. Athey, S. & Wager, S. Policy learning with observational data. *Econometrica* **89**, 133–161 (2021).
45. Loh, W.-Y. & Zhou, P. The GUIDE approach to subgroup identification. *Design and Analysis of Subgroups with Biopharmaceutical Applications* (eds Ting, N. et al.) 147–165 (Springer, 2020).
46. Jemielita, T. O. & Mehrotra, D. V. PRISM: patient response identifiers for stratified medicine. Preprint at <https://arxiv.org/abs/1912.03337> (2019).
47. Dmitrienko, A., Muysers, C., Fritsch, A. & Lipkovich, I. General guidance on exploratory and confirmatory subgroup analysis in late-stage clinical trials. *J. Biopharm. Stat.* **26**, 71–98 (2016).
48. Takahashi, Y. et al. The AURKA/TPX2 axis drives colon tumorigenesis cooperatively with MYC. *Ann. Oncol.* **26**, 935–942 (2015).
49. Nygård, S. B. et al. DNA topoisomerase I gene copy number and mRNA expression assessed as predictive biomarkers for adjuvant irinotecan in stage II/III colon cancer. *Clin. Cancer Res.* **22**, 1621–1631 (2016).
50. Palshof, J. A. et al. Topoisomerase I copy number alterations as biomarker for irinotecan efficacy in metastatic colorectal cancer. *BMC Cancer* **17**, 48 (2017).
51. Xu, Y. & Her, C. Inhibition of topoisomerase (DNA) I (TOP1): DNA damage repair and anticancer therapy. *Biomolecules* **5**, 1652–1670 (2015).
52. Mialon, A. et al. DNA topoisomerase I is a cofactor for c-Jun in the regulation of epidermal growth factor receptor expression and cancer cell proliferation. *Mol. Cell. Biol.* **25**, 5040–5051 (2005).
53. Chen, J., Elfiky, A., Han, M., Chen, C. & Saif, M. W. The role of Src in colon cancer and its therapeutic implications. *Clin. Colorectal Cancer* **13**, 5–13 (2014).
54. Koh, H. M. et al. Aurora kinase A is a prognostic marker in colorectal adenocarcinoma. *J. Pathol. Transl. Med.* **51**, 32–39 (2017).
55. Aderka, D., Stintzing, S. & Heinemann, V. Explaining the unexplainable: discrepancies in results from the CALGB/SWOG 80405 and FIRE-3 studies. *Lancet Oncol.* **20**, e274–e283 (2019).
56. Wang, C., Ouyang, C., Sandhu, J. S., Kahn, M. & Fakih, M. Wild-type APC and prognosis in metastatic colorectal cancer. *J. Clin. Orthod.* **38**, 223–223 (2020).
57. Easwaran, V. et al. beta-Catenin regulates vascular endothelial growth factor expression in colon cancer. *Cancer Res.* **63**, 3145–3153 (2003).
58. Meyer, E. L. et al. The evolution of master protocol clinical trial designs: a systematic literature review. *Clin. Ther.* **42**, 1330–1360 (2020).
59. Stintzing, S. et al. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab for metastatic colorectal cancer (FIRE-3): a post-hoc analysis of tumour dynamics in the final RAS wild-type subgroup of this randomised open-label phase 3 trial. *Lancet Oncol.* **17**, 1426–1434 (2016).
60. Battaglin, F., Naseem, M., Lenz, H.-J. & Salem, M. E. Microsatellite instability in colorectal cancer: overview of its clinical significance and novel perspectives. *Clin. Adv. Hematol. Oncol.* **16**, 735–745 (2018).
61. Fontana, E., Eason, K., Cervantes, A., Salazar, R. & Sadanandam, A. Context matters—consensus molecular subtypes of colorectal cancer as biomarkers for clinical trials. *Ann. Oncol.* **30**, 520–527 (2019).
62. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Stat. Methodol.* **57**, 289–300 (1995).
63. Babur, Ö. et al. Pattern search in BioPAX models. *Bioinformatics* **30**, 139–140 (2014).
64. Cerami, E. G. et al. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **39**, D685–D690 (2011).
65. Paz, A. et al. SPIKE: a database of highly curated human signaling pathways. *Nucleic Acids Res.* **39**, D793–D799 (2011).
66. Fazekas, D., Koltai, M. & Túrei, D. SignalLink 2—a signaling pathway resource with multi-layered regulatory networks. *BMC Syst. Biol.* **7**, 7 (2013).
67. Bouckaert, R. R. & Frank, E. Evaluating the replicability of significance tests for comparing learning algorithms. *Advances in Knowledge Discovery and Data Mining*, 3–12 (Springer, 2004).
68. Seibold, H., Zeileis, A. & Hothorn, T. Model4you: an R package for personalised treatment effect estimation. *J. Open Res. Softw.* **7**, 17 (2019).
69. Cui, Y., Kosorok, M. R., Sverdrup, E., Wager, S. & Zhu, R. Estimating heterogeneous treatment effects with right-censored data via causal survival forests. *J. R. Stat. Soc. Series B Stat. Methodol.* **85**, 179–211 (2023).
70. Athey, S., Tibshirani, J. & Wager, S. Generalized random forests. *AOS* **47**, 1148–1178 (2019).
71. Sverdrup, E., Kanodia, A., Zhou, Z., Athey, S. & Wager, S. policytree: policy learning via doubly robust empirical welfare maximization over trees. *J. Open Source Softw.* **5**, 2232 (2020).
72. Huling, J. D. & Yu, M. Subgroup identification using the personalized package. *J. Stat. Softw.* **98**, 1–60 (2018).

## Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 950293, M.P.M.). The clinical study received industrial funding from Merck KGaA, Darmstadt, Germany and Pfizer GmbH, Germany. The transcriptome-based microarray for gene expression using Xcel<sup>®</sup> Array received funding from Almac Ltd, Belfast, UK. The FoundationOne<sup>®</sup> based sequencing analysis (MSI) received funding from Roche Pharma AG, Grenzach, Germany (grant numbers: n/a, V.H., S.S.).

## Author contributions

Conceptualisation, M.P.M. and V.H.; Data curation, A.S., S.S., D.P.M., U.V., T.D., M.M. and A.J.O.; Formal analysis, A.J.O.; Methodology, A.J.O., A.S. and M.P.M.; Supervision, V.H. and M.P.M.; Visualisation, A.J.O. and L.H.; Writing original draft, A.J.O., A.S., V.H. and M.P.M.; Writing, review and editing, all authors.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

A.S. served on advisory boards for BMS and Novocure, received honoraria for talks by Roche, Servier and Taiho Pharmaceuticals and

received reimbursement for travel by Roche, Merck KGaA, MSD Sharp & Dohme, Pfizer, Lilly Oncology, and Amgen. V.H., S.S. and D.P.M. received honoraria for talks, advisory boards and travel expenses by Merck KGaA, Amgen, Roche, Pfizer, BMS, MSD, AstraZeneca, Novartis, Terumo, Oncosil, Nordic, Seagen, GSK, Takeda, Servier, Pierre Fabre, Taiho, Lilly Oncology, Servier, Sanofi and Bayer Pharmaceuticals. M.P.M. is a former employee at AstraZeneca, academically collaborates with AstraZeneca, GSK and Roche, and receives funding from GSK and Roche. J.W.H. served on an advisory board for Roche, has received honoraria from Roche, and travel support from Novartis. M.M. received honoraria for advisory boards or talks by Amgen, BMS, Roche, Merck KGaA, MSD Sharp & Dohme, Lilly Oncology, Servier, Pierre Fabre, Taiho Sanofi and Bayer Pharmaceuticals and serves as officer for the European Organisation on Research and Treatment of Cancer (EORTC), and Arbeitsgemeinschaft internistische Onkologie (AIO). C.B.W. has received honoraria from Amgen, Bayer, Chugai, Celgene, GSK, MSD, Merck, Janssen, Ipsen, Roche, Servier, SIRTEx, Taiho; served on advisory boards for Bayer, BMS, Celgene, Servier, Shire/Baxalta, Rafael Pharmaceuticals, RedHill, Roche, has received travel support by Bayer, Celgene, RedHill, Roche, Servier, Taiho and research grants (institutional) by Roche. C.B.W. serves as an officer for the European Society of Medical Oncology (ESMO), Deutsche Krebshilfe (DKH) and AIO. The remaining authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-41011-4>.

**Correspondence** and requests for materials should be addressed to Volker Heinemann or Michael P. Menden.

**Peer review information** *Nature Communications* thanks Saskia Wilting and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023