

OPEN

# Latency and interval therapy affect the evolution in metastatic colorectal cancer

Hamid Nikbakht<sup>1,2</sup>, Selin Jessa<sup>1,2</sup>, Mahadeo A. Sukhai<sup>3</sup>, Madeleine Arseneault<sup>1,2</sup>, Tong Zhang<sup>3</sup>, Louis Letourneau<sup>2</sup>, Mariam Thomas<sup>3</sup>, Mathieu Bourgey<sup>2</sup>, Michael H. A. Roehrl<sup>4,5</sup>, Robert Eveleigh<sup>2</sup>, Eric X. Chen<sup>3</sup>, Monika Krzyzanowska<sup>3</sup>, Malcolm J. Moore<sup>3</sup>, Amanda Giesler<sup>3</sup>, Celeste Yu<sup>3</sup>, Philippe L. Bedard<sup>3</sup>, Suzanne Kamel-Reid<sup>3</sup>, Jacek Majewski<sup>1,2</sup>, Lillian L. Siu<sup>3</sup>, Yasser Riazalhosseini<sup>1,2\*</sup> & Donna M. Graham<sup>3</sup>

While comparison of primary tumor and metastases has highlighted genomic heterogeneity in colorectal cancer (CRC), previous studies have focused on a single metastatic site or limited genomic testing. Combining data from whole exome and ultra-deep targeted sequencing, we explored possible evolutionary trajectories beyond the status of these mutations, particularly among patient-matched metastatic tumors. Our findings confirm the persistence of known clinically-relevant mutations (e.g., those of RAS family of oncogenes) in CRC primary and metastases, yet reveal that latency and interval systemic therapy affect the course of evolutionary events within metastatic lesions. Specifically, our analysis of patient-matched primary and multiple metastatic lesions, developed over time, showed a similar genetic composition for liver metastatic tumors, which were 21-months apart. This genetic makeup was different from those identified in lung metastases developed before manifestation of the second liver metastasis. These results underscore the role of latency in the evolutionary path of metastatic CRC and may have implications for future treatment options.

Results of somatic mutational analysis increasingly guide treatment decisions in oncology. In colorectal cancer (CRC), activating mutations in *KRAS* (codons 12, 13, 61, 117 and 146), *NRAS* (codons 12, 13, 61, 117 and 146), and *BRAF* (V600E) confer resistance to anti-epidermal growth factor receptor (EGFR) therapy including Cetuximab and Panitumumab<sup>1–4</sup>. This has resulted in extended *RAS* testing to more comprehensively determine treatment sensitivity in CRC<sup>5,6</sup>. Additionally, microsatellite instability (MSI) caused by deficiency of mismatch repair (MMR) proteins, has predictive value for response to adjuvant chemotherapy and immune checkpoint inhibitors in CRC<sup>7,8</sup>. Comparison of primary and metastatic tumors within individual patients underscores the heterogeneous nature of tumors and the varying responses of sub-clones to different treatment strategies<sup>9</sup>. Despite increasing use of molecular testing into clinical practice, this is typically performed using tumor tissue from a single anatomic location, at a single time point, obtained from archived diagnostic specimens or a freshly procured biopsy. Genomic evaluations of tumor tissue specimens from different sites within tumor or over time are rarely performed, due to several factors including risks associated with invasive procedures, cost, sample accessibility, and DNA quality. Current analyses of primary and metastatic CRC indicate low rates of variation between tumor sites for the clinically-relevant mutations that may impact upon treatment decisions<sup>10,11</sup>. However, these comparisons have surveyed the presence or absence of somatic mutations without specifically assessing their allelic frequency, which can reflect on genomic evolution of metastatic CRC. Likewise, our knowledge about the extent to which systemic therapy can influence this evolutionary path is limited. Here, we explore the evolution of metastatic CRC by analyzing the dynamic changes in frequency of somatic mutations identified through whole exome sequencing (WES) of patient-matched primary and metastatic tumors. Investigating the genetic architecture of different metastatic lesions from the same patient in the presence or absence of systemic therapy provides insight into both inherent and therapy-driven evolutionary paths. We further study the contribution of specific

<sup>1</sup>Department of Human Genetics, McGill University, Montreal, Québec, Canada. <sup>2</sup>McGill University and Génome Québec Innovation Centre, Montreal, Québec, Canada. <sup>3</sup>Princess Margaret Cancer Centre, Toronto, Ontario, Canada. <sup>4</sup>UHN Program in BioSpecimen Sciences, Toronto General Hospital, Toronto, Ontario, Canada. <sup>5</sup>Department of Pathology, Toronto General Hospital, Toronto, Ontario, Canada. \*email: [Yasser.riazalhosseini@mcgill.ca](mailto:Yasser.riazalhosseini@mcgill.ca)

mutations in six clinically-relevant genes (*APC*, *TP53*, *KRAS*, *NRAS*, *PIK3CA* and *BRAF*) to the mutational burden of CRC by measuring variant allele frequencies using targeted ultra-deep sequencing.

## Methods

**Patients and samples.** The study cohort comprised 350 patients with advanced malignancies enrolled in the initial phase of the Integrated Molecular Profiling in Advanced Cancer Trial (IMPACT) at the Princess Margaret Cancer Centre (NCT 01505400)<sup>12</sup>. All patient information and tissue were collected under the University of Toronto/University Health Network Research Ethics review board-approved protocol (REB#12-5453-TE) with written informed consent obtained from all participants. We reviewed medical records of these patients and selected those with multiple tumor samples ( $n = 62$ ) and with CRC as the primary diagnosis ( $n = 27$ ). From these, 15 patients had tissue available from multiple tumor samples with 11 having more than one tumor sample with adequate tissue for evaluation (26 samples in total) and were included in the current study (Supplementary Fig. 1). The cohort of 11 patients included eight patients with available tissue from the primary tumor and  $\geq 1$  metastatic site, two patients with pairs of metastases only, and one patient with an anastomotic recurrence five years after initial resection of the primary tumor (Supplementary Table 1). One sample from each patient had previously been genotyped using a customized 23-gene panel via the Sequenom MassARRAY<sup>12</sup>. All procedures and experiments were conducted in accordance with University of Toronto/University Health Network guidelines.

**Tissue preparation and targeted capture.** Representative slides or tissue blocks from tumor specimens were collected. A hematoxylin and eosin-stained section from each tumor was reviewed by a specialized gastrointestinal pathologist. Tumors were macrodissected to maximize tumor content where necessary, and the percentage of viable tumor cells was estimated by the pathologist. DNA was isolated from the 36 available formalin-fixed paraffin embedded (FFPE) tissue specimens using QIAamp DNA micro kit (QIAGEN Inc, Valencia, CA), following instructions provided by the manufacturer. Loci harboring clinically-relevant mutations of *TP53* (codons 175, 213, 245, 248, 273 and 306), *APC* (codons 1378 and 1450), *KRAS* (codons 12, 13, 22, 61, 117 and 146), *NRAS* (codons 12, 13 and 61), *PIK3CA* (codons 539, 542, 545 and 1047), and *BRAF* (codon 600) were captured by multiplex amplification on Fluidigm access arrays (South San Francisco, CA) using 50 ng of DNA. PCR primers specific to each locus are provided in Supplementary Table 2. Hotspots were selected following literature review and use of publicly available datasets, with input from our molecular diagnostics laboratory<sup>3,4,13-15</sup>. Of note, *NRAS* codon 117 and 146 mutations were not included in this analysis due to their relative rarity<sup>4</sup>.

**Whole-exome sequencing and data analysis.** WES libraries were generated using the Nextera Rapid Capture Enrichment library preparation kit (Illumina) according to the manufacturer's recommendations and were sequenced on an HiSeq. 2500 paired-end 100 base pair reads. Quality controlled reads processed by trimmomatic<sup>16</sup> (v0.32) to remove adapters and low quality reads were aligned to the human genome build GRCh37 using bwa-mem<sup>17</sup> (version 7.6). Mapped reads were further refined using GATK<sup>18</sup> (v3.2.2) and Picard program suites<sup>19</sup> (v1.118) to improve mapped reads near indels (GATK indel realignment) and improve quality scores (GATK base recalibration) as well as mark duplicate reads with the same paired start locations (Picard mark duplicates). Somatic calls generated using Mutect (v1.1.6) for SNVs and scalpel (v0.5.2) for indels<sup>20</sup> were further processed with the addition of functional annotations using snpEff<sup>21</sup> (v3.6b) and genomic annotation using Gemini<sup>22</sup> (v0.11.1a). All non-silent variants were inspected manually for quality control using Integrative Genome Viewers (IGV)<sup>23</sup>, and were predicted for functionality using CADD<sup>24</sup>, and fitness consequence scores<sup>25</sup>.

**Mutational signature analysis.** Mutational signature analysis was performed individually for each lesion using somatic base substitutions identified in that lesion by Mutect. The R packages SomaticSignature and BSgenome.Hsapiens.1000genomes.hs37d5<sup>26</sup> were used to extract the somatic mutational profiles and to provide the frequency for each of the 96 possible single base substitutions and their context combinations. Individual mutational profiles were then compared against the known signatures described by Alexandrov *et al.* using the DeconstructSigs R package to infer the contribution of known reference signatures to the individual somatic mutation profiles<sup>27,28</sup>.

**Copy number aberration analysis.** To analyze copy number variations in our samples we used an in-house program that exploits deviation of B allele frequency from 50% as well as read-depth, normalized based on the average coverage using WES data. Copy number variation (CNV) events were defined as follows: deviation from 50% in B allele frequency and a significant increase in normalized coverage was considered as a sign of amplification, deviation from 50% in B allele frequency, however, if was accompanied by decrease in normalized coverage, would be considered as deletion. If there was no change in the normalized coverage while B allele frequency represents deviation from 50%, we called it a potential copy-neutral loss of heterozygosity (LOH).

**Inferring clonal evolutionary trajectories within patients.** We used allele frequency values corrected<sup>29</sup> for both the tumor content and CNV events as the raw input to infer sub-clonal composition and evolutionary trajectories within each patient. We corrected the original allele frequency values based on the copy number aberration events harboring these variations. In our correction, we use the naïve assumption that the copy number aberration always affects the frequency of the altered allele to the largest extent. For example, in case of duplication, we assume the duplication occurred on the chromosome harboring the altered allele. Thus, we divide the number of altered allele reads into half to correct for the duplication and calculate the allele frequency (Duplication:  $Ref' = Ref$ ;  $Alt' = Alt/2$ ). In the case of deletion, we assume the deletion happens on the chromosome with reference allele. Thus, to correct for this, we add the number of reference allele reads to the number of altered allele reads and count it as the final number of reference allele read count (Deletion:  $Ref' = Ref + Alt$ ;  $Alt' = Alt$ ). For a gene on chromosome X (in male patients), this is treated as a loss of heterozygosity case. In this

case we add the number of alternate reads to the twice number of reference reads and use it as new reference read counts (Gene on X in male patients:  $Ref' = Ref * 2 + Alt$ ;  $Alt' = Alt$ ). In cases of copy neutral loss of heterozygosity, we add half of alternate reads to the number of reference reads and use it as the new value for reference read counts (Copy Neutral LOH:  $Ref' = Ref + Alt/2$ ;  $Alt' = Alt/2$ ). We developed an in-house program to cluster recurrent and exclusive mutations in primary versus metastatic lesions using expectation maximization algorithm using EMClust<sup>30,31</sup> library in R<sup>32</sup> and used the resulting putative subpopulations, their relative frequencies, and information about therapy administered between resections to infer evolutionary trajectories between lesions. The funnel plots representing these proposed scenarios are drawn manually using Adobe Photoshop® (CS3 extended version) program. The clustering plots representing subpopulations inferred from somatic mutations are created using ggplot2<sup>33</sup> library.

**Accession codes.** Raw sequence data have been deposited in the European Genome-phenome. Archive (EGA), under the Accession Code EGAS00001003646.

## Results

**Patient samples and whole exome sequencing.** We analyzed primary CRC, matched blood, and metastatic tumors (where available) from 11 patients, from which, all had germline DNA from blood samples, seven from both primary and more than one metastatic tumors, one from only the primary tumor, one from a primary tumor and an anastomotic recurrence, and two from metastatic tumor pairs only. Clinical characteristics and patient treatment strategies are outlined in Supplementary Table 1. Targeted ultra-deep sequencing (average of 25000×) was applied to clinically-actionable mutations for validation and measuring mutant allele frequencies. WES (see Methods) was applied to DNA isolated from macrodissected formalin-fixed paraffin-embedded (FFPE) tumor specimens and blood samples to an average coverage depth of 63X and 95X for blood and tumor samples, respectively.

**Somatic mutations.** On average, we detected 116 and 156 somatic exonic mutations in primary and metastatic lesions, respectively (excluding the hypermutated primary tumor in REACT-009-A, which harbored 1285 somatic mutations) (Supplementary Table 1).

We investigated global mutational patterns by analyzing base-substitution profiles and mutational signatures in tumor samples<sup>27,34</sup> (Supplementary Table 3). As reported recently, mutational signatures can shed light on mutagenic processes that affect cancer genomes. We observed that the primary tumor in patient-009 (REACT-009-A) and metastatic samples REACT-010-B and E are enriched for signature 6 (Supplementary Fig. 2). This signature is characterized by Cytosine deamination in repeated sequences with NpCpG pattern, and is associated with the inactivation of DNA mismatch repair (MMR) genes in colorectal cancer. Deficiency of MMR proteins results in microsatellite instability (MSI) pattern<sup>27</sup> which is reported in a variety of cancers including CRC, and with prognostic and predictive implications for CRC<sup>8,34,35</sup>. Our WES results from patient-009's primary tumor confirmed a hypermutated genome in this tumor, with mutations in the MMR genes *MSH2* and *MSH5* (Supplementary Tables 1 and 4).

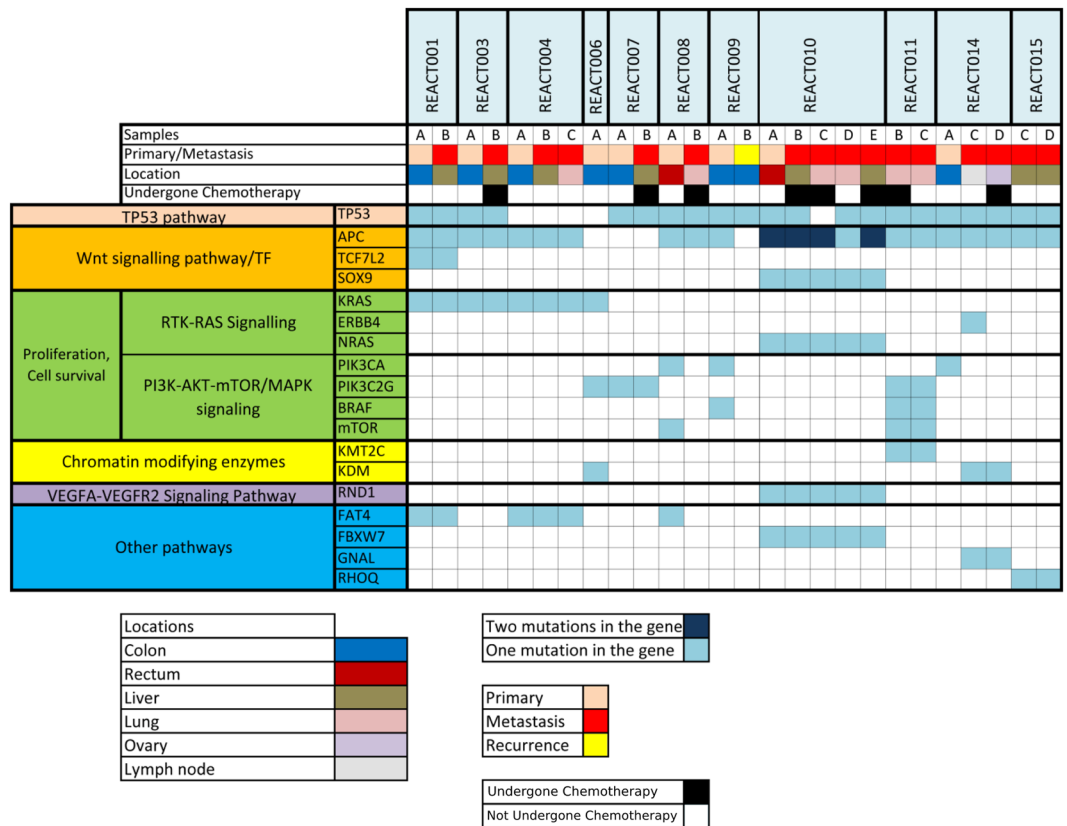
Next, we focused on non-silent somatic mutations in primary tumors, and identified 1116 such mutations affecting 1019 genes following manual *in silico* verification of mutations using Integrative Genomics Viewer (IGV)<sup>23</sup> (Supplementary Table 4). Forty-five of the affected genes were recurrently mutated in at least two primary tumor samples (Supplementary Table 4), among which, nine genes were affected in the primary tumor of at least 3 of the 9 patients whose primary tumors were analyzed by WES (33.3%); *APC*, *TP53*, *KRAS*, *SYNE1*, *FAT4*, *FBN2*, *PIK3CA*, *UHRF1* and *WDFY4*. Among these 9 patients, *APC* and *TP53* were the most frequently mutated genes affected by somatic mutations in 78% followed by *KRAS* which was mutated in 44% of primary tumors (Fig. 1), in line with previous studies reporting these genes as the most commonly mutated genes in CRC<sup>36,37</sup>.

All *APC*-mutated tumors were affected with at least one protein truncating mutation, which may lead to the activation of WNT pathway<sup>38</sup>. *TP53* was affected in nine patients in our study, with various mutations. Activating mutations of *KRAS* codon 12 were detected in four patients, and oncogenic mutations of *NRAS* (G13V) and *BRAF* (V600E) were each identified in one additional tumor, contributing to a prevalence of 54.4% for activating mutations of RAS pathway. Notably, these genes were not mutated concurrently in tumors highlighting a mutually-exclusive pattern for RAS pathway mutations (Fig. 1 and Supplementary Table 4). Three patients were affected with *PIK3CA* mutations, two of which (E545G and Q546K) have been previously reported in CRC<sup>39</sup>. A third *PIK3CA* mutation (H665D) with uncertain clinical significance has no previous record in COSMIC database, but was predicted as likely functional based on CADD and fitCons scores. Three additional tumors were affected with mutations in *PIK3C2G*, resulting in a frequency of 55.5% of primary samples affected with PI3K pathway mutations (Fig. 1 and Supplementary Table 4).

Among other recurrently mutated genes, *SYNE1* and *FAT4* have been reported with high prevalence of non-silent mutations in CRC patients<sup>36</sup>. Hypermethylated-*SYNE1* promoter has been found in 80% of patients with colitis-associated CRC<sup>40</sup>. Likewise, *UHRF1* has been shown to be involved in cellular proliferation and molecular pathogenesis of CRC in the right hemicolon<sup>41</sup>.

**Chromosomal copy number alterations.** Analysis of the chromosomal copy number alterations (CNAs) revealed highly recurrent amplifications of 13q in 75% (6/8) of primary tumors. This was followed by gains of 11q and 6q in 50% (4/8) and 37.5% (3/8) of primary tumors, respectively. Moreover, we detected 63% (5/8) cases of LOH and one case of deletion of 17p in a primary tumor (Supplementary Table 5).

Amongst seven patients with both primary and metastatic tumors, 3 patients showed the same CNA in both primary tumor and one or more metastatic lesion(s). These include recurrent amplifications of 6q and 13q (REACT-001 and REACT-010) as well as copy neutral loss of heterozygosity in 17p (REACT-004 and

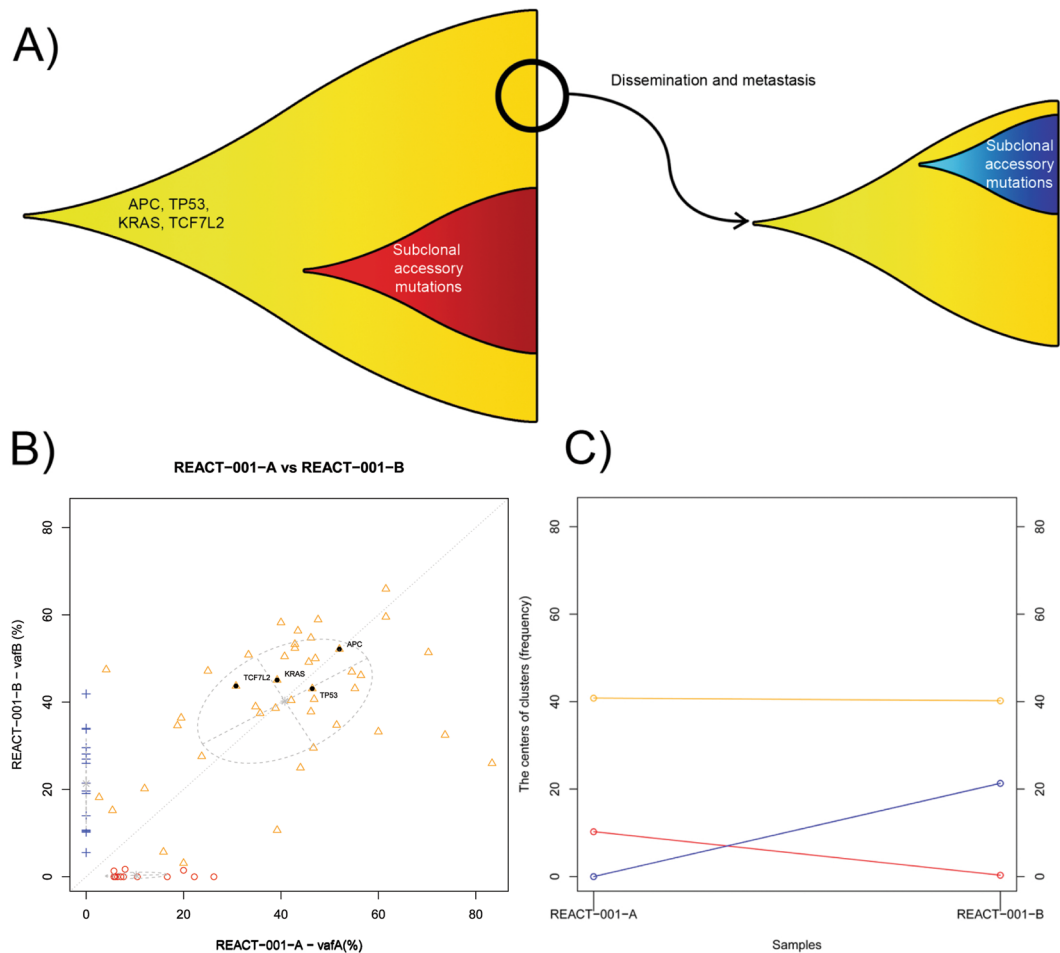


**Figure 1.** Mutational status of genes involved in CRC pathways across samples studied. Patient IDs are shown on top and individual samples procured from each patient are depicted beneath the corresponding patient ID. The type of sample (primary or metastasis), affected organ and information about the application of chemotherapy prior to the sampling are provided in colored legends.

REACT-010) (Supplementary Table 5). Patient REACT-011 shows a very similar CNA pattern between its two metastases (AMP at 3p, 5p, 6p, 6q, 11q, 12q, 13q, 19p and 21q.) This CNA similarity pattern between metastases was less pronounced in other patients. This patient also had similarity in non-silent somatic mutation patterns between the metastatic samples. There was a short time interval (6-months) between metastatic lesions for patient REACT-011 during which no treatment was administered. Our results show that patients who had received an interval treatment seem to show a greater decrease in the total number of complete arm aberrations, that is loss or gain of a complete chromosome arm, between their samples (REACT-007, REACT-008 and REACT-014) compared to patients who did not receive any treatment (REACT-011 and REACT-015) (Supplementary Table 5). Effects of therapy on the abundance of CNVs in post-treatment liver metastases have previously been reported<sup>42</sup>.

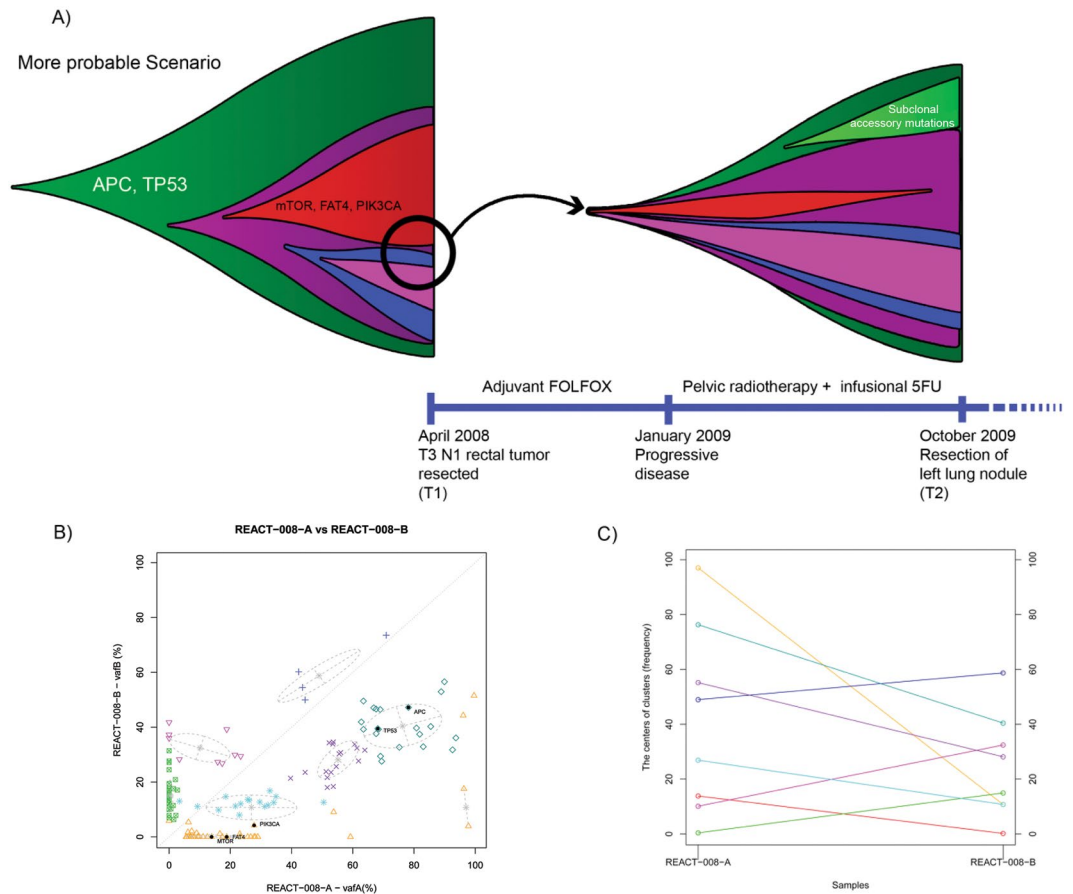
**Analysis of the clonal patterns.** We investigated the clonality of tumor samples and inferred evolutionary trajectories of metastatic tumors in patients with paired primary and metastatic lesions using allele frequency of somatic mutations (see Methods). Our analysis confirmed sub-clonal composition of both primary and metastatic tumors across all samples. Overall, whereas private mutations (i.e., mutations exclusive to primary or metastatic lesion) constituted a large proportion of somatic mutations in each specimen, mutations in known driver genes were consistently present in both primary and metastatic lesions, with few exceptions discussed below.

**Patients who received no interval treatment.** Analysis of primary tumor and metastases in patients who had received no interval therapy revealed consistency for presence and relative frequency of identified driver mutations. For example, in patient REACT-001, where no interval chemotherapy was administered, the relative frequency of main driver mutations (*APC* E1288\*, *TP53* L125LH, and *KRAS* G12A) did not change between the primary and metastatic tumors, suggesting that the cluster of cells disseminated from the primary tumor to form the metastatic tumor harbored the entire set of main driver mutations (i.e. the dissemination happens after the occurrence of all above mentioned alterations) (Fig. 2A). In both primary and metastatic tumor samples, some sub-clonal (passenger) mutations<sup>29</sup> emerged secondary to the main driver mutations. Similar evolutionary trajectories are presented in Supplementary Fig. 3 for patient REACT-004 who had not received interval therapy. The lack of multi-region sequencing in each of these samples leaves sampling bias as a potential alternative explanation.



**Figure 2.** (A) Proposed scenario for the evolution and clonality within the primary and metastatic lesions in patient REACT-001. B and C represent the clustering of somatic mutations in these two tumor samples, and the changes in allele frequency of the centers of these clusters. The center of a cluster inferred from somatic mutation allele frequencies corresponds to the average cellular proportion of the subpopulation characterized by mutations in that cluster. The allele frequency values were inferred from WES data, and corrected for CNA events and for tumor content, which was estimated from the histological examinations. (B) Clustering of the mutations in primary versus metastatic lesion into groups which characterize clonal subpopulations in the lesions. (C) Changes in the population proportion in different clusters. The yellow cluster maintains a ~40% cell population proportion while the red cluster of mutations is lost in the metastatic lesion and the new cluster (sub-clone) of mutations has arisen in the metastatic lesion (blue). Red subpopulation (exclusive to the primary tumor), harbor mutations accumulated in a subpopulation different from which the dissemination of metastatic cells happens or potentially emerged after the dissemination. The blue cluster (exclusive to the metastatic lesion), however, represent cells with new mutations emerging following dissemination from the primary tumor.

**Patients who received interval treatment.** A different pattern is evident following the analysis of cases where interval systemic treatment has been administered. Patient REACT-008 was diagnosed with a primary tumor in the rectum, and had treatment with adjuvant folinic acid, fluorouracil, and oxaliplatin (FOLFOX) between resection of the primary and a lung metastasis. In this case, the primary tumor had been completely resected with no residual disease detected at the time of chemotherapy administration. *TP53* (R114H) and *APC* (E1379\*) mutations exist in a sub-clone present in relatively high frequency in both primary and metastasis. However, relative abundance of a fraction of cells harboring *PIK3CA* (H665D) and *mTOR* (V1795M) mutations, which are predicted as likely functional mutations, declined substantially in the lung metastasis compared to the primary tumor (Fig. 3). This could be due to either cytotoxicity on the metastatic tumor or the absence of these mutations in the cluster of cells that disseminated from the primary tumor to form the lung metastasis. Interestingly, the most frequent cell population shows a relatively stable frequency in the metastatic lesion, which may be explained by the presence of the *TP53* mutation, possibly conferring resistance to chemotherapy in these cells. The possible elimination of the sub-clone harboring mutations in *mTOR*, *PIK3CA* and a *FAT4* due to adjuvant chemotherapy could have happened either in the primary tumor or in the metastasis. Given the chronology



**Figure 3.** (A) Proposed scenario for the evolution and clonality within the primary and metastatic lesion in patient REACT-008. (B) Clustering of the mutations in primary versus metastatic lesions. (C) Changes in the population proportion in different clusters.

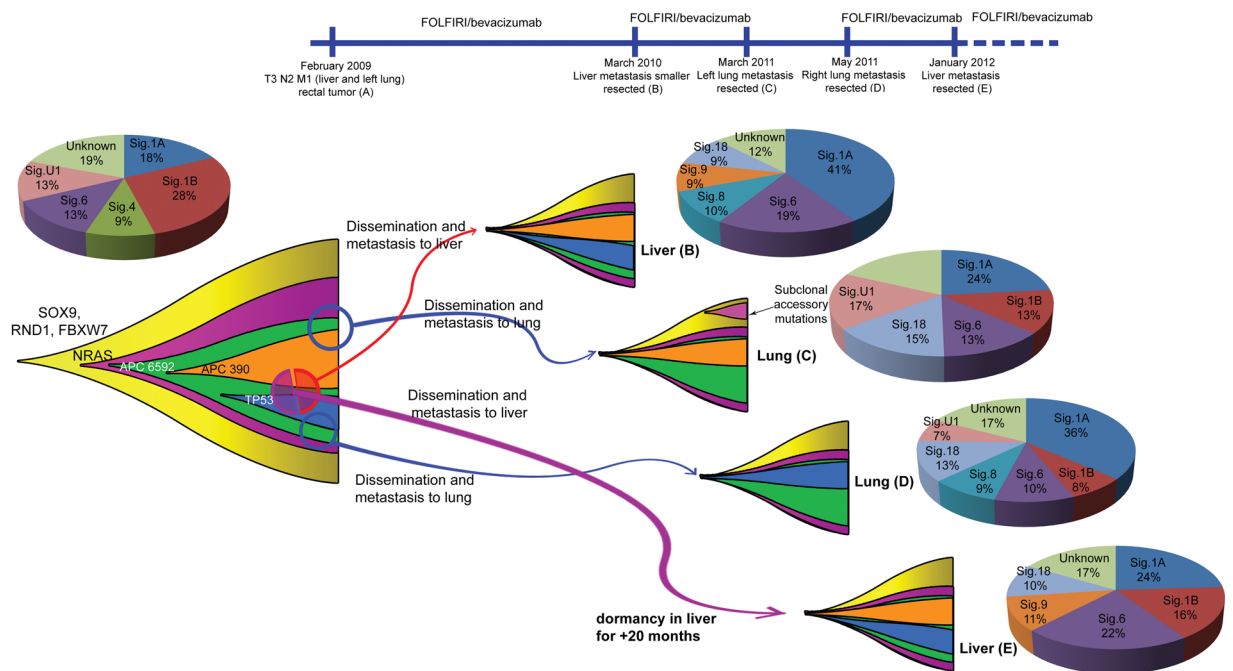
of systemic therapy in this case, the latter scenario seems more likely than the former. This scenario is represented in the funnel plot (Fig. 3).

Patient REACT-010 is interesting due to the number of metastatic samples and the timing of systemic therapy administered. This patient has a primary tumor in the rectum, and two pairs of metastasis, one pair in the liver (21-months apart) and one pair in the lungs (left and right; one-month apart) (Fig. 4). The patient received folinic acid, fluorouracil, and irinotecan (FOLFIRI)/bevacizumab treatment between the primary and the first metastasis (liver REACT-010-B), and between the first and second metastasis (left lung REACT-010-C). These two metastases were present at the time of resection of the primary tumor. Following treatment with FOLFIRI/bevacizumab the liver metastasis reduced, but the lung lesion remained stable. This patient was also treated with FOLFIRI/bevacizumab between the third (lung REACT-010-D) and fourth (liver REACT-010-E) metastases, but received no interval treatment between resection of the lung lesions (i.e. REACT-010-C and D).

Metastases in the liver share the same mutational profile for the driver mutations along with some private mutations in each (Fig. 4 and Supplementary Fig. 4). This suggests the possibility of dissemination of the founding clusters from the same region of the primary tumor sharing their ancestral sub-clonal structure. In contrast, for two lung metastatic tumors, which are only one-month apart, each harbor an exclusive driver mutation which is absent in the other. These are *APC* Q1349\* (depicted by orange color in Fig. 4), which is present only in REACT-010-C, and a start loss mutation in *TP53* (depicted by dark blue in Fig. 4), which is present only in REACT-010-D. The most plausible scenario here is that the two metastases have originated from two clusters of cells disseminating from different regions of the primary tumor with different sub-clonal structures. The clinical course may explain why the same *APC* mutation is present in the liver resected samples and first lung lesion (REACT-010-C).

The similar profiles of driver mutations in the liver lesions (B and E) suggest that although these metastases manifested at different times, their founding cell-clusters disseminated around the same time. This pattern can be explained by dormancy of the disseminated cells that formed the second liver metastasis, or simply by differential growth rates of seeded cells, which formed the metastases. Our targeted deep-sequencing results confirm the reconstructed clonal structure (Supplementary Fig. 4). Figure 4 represents the possible scenario for the evolutionary trajectories within primary and metastatic lesions for patient REACT-010. The clustering schemes within these samples are represented in Supplementary Fig. 4.

The clustering and changes in the sub-clone population proportions for all the other patients with primary tumor are represented in Supplementary Fig. 3.



**Figure 4.** Proposed evolution and clonality and changes in the mutation signature within primary tumor and four metastatic lesions in patient REACT-010.

**Patients with no primary tumor available for analysis.** Patient REACT-011 had two metastatic tumors in lung (4-months apart), with no available tissue from the primary tumor. Treatment with chemoradiation (5FU) + adjuvant FOLFOX was administered before development and resection of the first metastasis, however, no treatment was administered between metastases. Mutations in driver genes (nonsense mutations in both *APC* and *TP53*) are consistently present in both lesions. The tumors did not have mutations in any RAS pathway gene, but they share a missense mutation in the epigenetic modifier gene *KMT2C*, which shows high allele frequency in both lesions. Presence of the truncating *TP53* mutation may explain chemotherapy resistance. The shift in the allele frequencies of all the mutations towards higher values in the lung metastasis REACT-011-C suggests that relatively higher allele frequency of the main driver mutations in this tumor could be due to differences in tumor content between the two metastatic samples (40% vs 60%) (Supplementary Fig. 5A–C).

## Discussion

Our study is the first that applies a combination of whole-exome and ultra-deep targeted sequencing to the primary colorectal cancers as well as matched series of multiple metastatic lesions from individual patients. While our findings support previous studies highlighting the persistence of known clinically-relevant mutations between primary and metastatic CRC lesions<sup>10,11</sup>, we showed that, when present in the primary tumor, clinically-relevant *KRAS* and *NRAS* mutations are transmitted to the metastatic lesions of different organs and likely represent ubiquitous driver mutations<sup>43</sup>. Systemic therapies may apply selective pressure to the clonal composition of tumors, thereby contributing to resistance based upon changes in genomic aberrations. In EGFR-mutated non-small cell lung cancer (NSCLC), biopsies upon disease progression following targeted treatment have shown emergence of resistant clones in over 50% of cases, and altered histological subtype in 14%<sup>44</sup>. Treatment with temozolomide may influence the genomic tumor profile in patients with recurrent glioma, facilitating evolution to high-grade disease<sup>45</sup>. In CRC, analysis of cell lines and biopsy tissue from patients with tumors previously wild-type for *KRAS* mutations has shown that cetuximab treatment results in the emergence of *KRAS*-mutated sub-clones, as detected by analyzing circulating tumor DNA (ctDNA)<sup>11</sup>. However, our results did not reveal such emergence of RAS mutations in metastatic lesions, possibly due to small sample size. In NSCLC, higher *EGFR*-mutation allele frequency has been shown to correlate with response to EGFR-tyrosine kinase inhibitors<sup>46,47</sup>. In the event of development of appropriate targeted therapies for *KRAS*-mutated CRC, our results may have similar relevance. Notably, however, in the lung metastasis of patient REACT-008 we observed that cells affected by *mTOR* or *PIK3CA* mutations were eliminated following chemotherapy. Likewise, our mutational signature analysis indicated that interval chemotherapy can alter exonic mutational profiles of metastatic CRC tumors by promoting the enrichment or depletion of certain mutational signatures. These findings extend our knowledge about the effect of systemic therapy on the genomic architecture of metastatic tumors, beyond the status of *KRAS* and *NRAS* mutations, with potential translational implications. Furthermore, we also investigated mutational profiles of primary and an anastomotic recurrence tumor from a patient (patient 009) in our study. Interestingly, these tumors exhibited different mutational landscapes as WES revealed a hypermutated genome only in the primary and not in the recurrent tumor. Given that anastomotic recurrences are rare events as compared to metastasis in colorectal cancer patients, we found one recent study that compared mutations between primary CRC and anastomotic

recurrences. By examining mutations of candidate genes, it has been suggested that anastomotic recurrences are clonally related to primary tumors<sup>48</sup>. However, this study included only microsatellite stable tumors, whereas tumor 9A in our study is a hypermutated tumor with mutations in *MSH2* and *MSH5* genes, suggesting that this tumor is affected by microsatellite instability. Further investigation may be required to understand the evolution of anastomotic recurrences where microsatellite stability differs.

Development of malignancy through clonal expansion may occur due to somatic mutations in key genes, for example in *KRAS* and *NRAS* in CRC, leading to dysregulation of the normal homeostatic mechanisms. Mutations conferring growth advantage are drivers of the carcinogenic and metastatic process and provide clinically-relevant targets for treatment<sup>49</sup>. Across tumor sites, genomes of disseminated cancer cells may have similarities at onset of metastatic disease. However, substantial changes in the genetic composition occur spatially and over time. It is therefore plausible that the expansion of aggressive driver clone(s), and emergence of relevant sub-clones, correlate with development of incurable disease and potential therapeutic resistance<sup>50–53</sup>. We were able to study these patterns in the mutational profile of patient REACT-10 for whom in addition to the primary tumor, four distinct metastatic specimens were available for mutational analysis. Our analysis revealed a similar mutational signature profile for the two liver metastases that was different from those observed in lung metastases. Furthermore, driver gene mutations were constantly present with high frequency in both liver lesions, a pattern which was not the same in lung metastatic tumors. Given that the liver metastases are at the two ends of the disease course in this patient, our results suggest that while two liver metastases were most likely originated from a common cell cluster in the primary tumor, the disseminated cells that formed the second liver metastasis experienced a dormant phase or a much slower growth rate before manifesting. The original seeding clusters may have resided in the same region within the primary tumor or may originate from different regions. The latter can be explained by ‘early cell mixing’ phenomenon, which has recently been shown for CRC<sup>54</sup>. These findings highlight the complex nature of evolution of metastatic colorectal cancers, and the importance of dormancy in this context. The mechanisms that underlie tumor dormancy and reactivation in metastatic CRC are not well known; however, a recent study suggests that non-genetic factors including chemotherapy play key roles here<sup>55</sup>.

Although presence of *KRAS* mutations using standardized clinical techniques may predict for resistance to anti-EGFR therapy, responses in *KRAS* wild-type tumors are not universal (single-agent response rates of 10–41%)<sup>2,56,57</sup>. Reasons for this are incompletely understood. Use of alternative technologies may detect low-level, or less frequent, *KRAS* mutations of uncertain clinical significance, undetectable by standard direct-sequencing and may also detect regional intratumor heterogeneity possibly directing treatment resistance<sup>58–60</sup>. Use of ultra-deep sequencing in our patient cohort ensured high sensitivity for *KRAS* and additional driver mutation analysis and permitted accurate quantification of mutant alleles in each specimen, necessary to study dynamics of mutational patterns throughout disease course. However, sampling was limited to a single region within each tumor specimen, precluding the evaluation of geographic heterogeneity within each lesion.

Data from our patient cohort highlights tumors lacking known clinically-relevant *KRAS* mutations with a complex, heterogeneous mutational landscape characterized by the presence of non-RAS mutations, as well as other *KRAS* mutations of uncertain clinical significance. We observed in tumors of three out of seven patients wild-type for *KRAS* codon 12 and 13, that *APC* was affected by protein-truncating mutations, an early event in CRC tumorigenesis<sup>61</sup>. Additional research including larger number of samples is warranted to investigate the mutational landscape of tumors that lack *KRAS* codon 12 and 13 mutations.

Sub-classification of CRC has been challenging and, despite the identification of a number of prognostic subgroups<sup>62,63</sup>, understanding of key driver mutations remains critical for optimal treatment of metastatic CRC. As there is greater understanding of the molecular landscape of CRC and as non-invasive methods of molecular testing become more frequently utilised<sup>64,65</sup> targeted ultra-deep sequencing of driver mutations provides a powerful tool to monitor tumor behavior in clinic, and to identify recurrent lesions, molecularly distinct from primary tumors. This technique demonstrates high sensitivity for mutation detection and its use in this study has confirmed the persistence of clinically-relevant mutations between primary tumor and metastatic lesions. Additionally, this work has highlighted the impact of latency and systemic treatment pressure effects in the evolution of metastatic CRC lesions, resulting in modulation of mutation frequency for driver mutations. By enhancing our knowledge about the behavior of driver mutations, these findings may have translational value for future clinical management of patients.

Received: 19 October 2017; Accepted: 27 December 2019;

Published online: 17 January 2020

## References

1. Karapetis, C. S. *et al.* *K-ras* Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer. *N. Engl. J. Med.* **359**, 1757–1765 (2008).
2. Amado, R. G. *et al.* Wild-Type *KRAS* Is Required for Panitumumab Efficacy in Patients With Metastatic Colorectal Cancer. *J. Clin. Oncol.* **26**, 1626–1634 (2008).
3. Di Nicolantonio, F. *et al.* Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J. Clin. Oncol.* **26**, 5705–12 (2008).
4. Douillard, J.-Y. *et al.* Panitumumab–FOLFOX4 Treatment and RAS Mutations in Colorectal Cancer. *N. Engl. J. Med.* **369**, 1023–1034 (2013).
5. Bokemeyer, C. *et al.* FOLFOX4 plus cetuximab treatment and RAS mutations in colorectal cancer. *Eur. J. Cancer* **51**, 1243–1252 (2015).
6. Van Cutsem, E. *et al.* Fluorouracil, leucovorin, and irinotecan plus cetuximab treatment and RAS mutations in colorectal cancer. *J. Clin. Oncol.* **33**, 692–700 (2015).
7. Gavin, P. G., Paik, S., Yothers, G. & Pogue-Geile, K. L. Colon Cancer Mutation: Prognosis/Prediction–Response. *Clin. Cancer Res.* **19** (2013).
8. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).



9. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
10. Artale, S. *et al.* Mutations of KRAS and BRAF in primary and matched metastatic sites of colorectal cancer. *J. Clin. Oncol.* **26**, 4217–9 (2008).
11. Misale, S. *et al.* Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532 (2012).
12. Stockley, T. L. *et al.* Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial. *Genome Med.* **8**, 109 (2016).
13. Brana, I. & Siu, L. L. Clinical development of phosphatidylinositol 3-kinase inhibitors for cancer treatment. *BMC Med.* **10**, 161 (2012).
14. Wellcome Trust Sanger Institute. COSMIC: Catalogue of Somatic Mutations in Cancer v80. 13-Feb-2017 Available at, <http://cancer.sanger.ac.uk/cosmic>. (Accessed: 1st March 2017).
15. World Health Agency. International Agency for Research on Cancer (IARC). IARC TP53 Database. April 2016 Available at, <http://p53.iarc.fr/>. (Accessed: 1st March 2017).
16. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
17. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (2013).
18. Van der Auwera, G. A. *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. in *Current Protocols in Bioinformatics* **43**, 11.10.1–11.10.33 (John Wiley & Sons, Inc., 2013).
19. BroadInstitute. Picard Tools - GitHub. Release 2.20.8 (2019). Available at, <https://github.com/broadinstitute/picard/>. (Accessed: 30th September 2019).
20. Narzisi, G. *et al.* Accurate de novo and transmitted indel detection in exome-capture data using microassembly. *Nat. Methods* **11**, 1033–1036 (2014).
21. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)*. **6**, 80–92 (2012).
22. Paila, U., Chapman, B. A., Kirchner, R. & Quinlan, A. R. GEMINI: Integrative Exploration of Genetic Variation and Genome Annotations. *PLoS Comput. Biol.* **9**, e1003153 (2013).
23. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–6 (2011).
24. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
25. Gulko, B., Hubisz, M. J., Gronau, I. & Siepel, A. A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nat. Genet.* **47**, 276–83 (2015).
26. Gehring, J. S., Fischer, B., Lawrence, M. & Huber, W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. *Bioinformatics* **31**, 3673–5 (2015).
27. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
28. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).
29. Nikbakht, H. *et al.* Spatial and temporal homogeneity of driver mutations in diffuse intrinsic pontine glioma. *Nat. Commun.* **7**, 11185 (2016).
30. EM Algorithm for Model-Based Clustering of Finite Mixture Gaussian Distribution [R package EMCluster version 0.2-6].
31. Chen, W.-C., Maitra, R. & Melnykov, V. A Quick Guide for the EMCluster Package.
32. R Development Core Team. R: a language and environment for statistical computing | GBIF.ORG. R Foundation for Statistical Computing (2016). Available at, <http://www.r-project.org/>. (Accessed: 1st March 2017).
33. Wickham, H. *ggplot2*, <https://doi.org/10.1007/978-0-387-98141-3> (Springer New York, 2009).
34. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **15**, 585–598 (2014).
35. Tran, B. *et al.* Impact of BRAF mutation and microsatellite instability on the pattern of metastatic spread and prognosis in metastatic colorectal cancer. *Cancer* **117**, 4623–32 (2011).
36. Yu, J. *et al.* Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer. *Gut* **64** (2015).
37. van den Broek, E. *et al.* High Prevalence and Clinical Relevance of Genes Affected by Chromosomal Breaks in Colorectal Cancer. *PLoS One* **10**, e0138141 (2015).
38. Novellademunt, L., Antas, P. & Li, V. S. W. Targeting Wnt signaling in colorectal cancer. A Review in the Theme: Cell Signaling: Proteins, Pathways and Mechanisms. *Am. J. Physiol. - Cell Physiol.* **309** (2015).
39. Samuels, Y. *et al.* High Frequency of Mutations of the PIK3CA Gene in Human Cancers. *Science (80-)*. **304** (2004).
40. Papadia, C. *et al.* FOXE1 and SYNE1 Genes Hypermethylation Panel as Promising Biomarker in Colitis-associated Colorectal Neoplasia. *Inflamm. Bowel Dis.* **20**, 271–277 (2014).
41. Kofunato, Y. *et al.* UHRF1 expression is upregulated and associated with cellular proliferation in colorectal cancer. *Oncol. Rep.* **28**, 1997–2002 (2012).
42. Kawamata, F. *et al.* Copy number profiles of paired primary and metastatic colorectal cancers. *Oncotarget* **9**, 3394–3405 (2018).
43. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).
44. Sequist, L. V. *et al.* Genotypic and Histological Evolution of Lung Cancers Acquiring Resistance to EGFR Inhibitors. *Sci. Transl. Med.* **3**, 75ra26–75ra26 (2011).
45. Johnson, B. E. *et al.* Mutational Analysis Reveals the Origin and Therapy-Driven Evolution of Recurrent Glioma. *Science (80-)*. **343** (2014).
46. Martin, P. *et al.* Clinical impact of mutation fraction in epidermal growth factor receptor mutation positive NSCLC patients. *Br. J. Cancer* **114**, 616–622 (2016).
47. Ono, A. *et al.* Mutant allele frequency predicts the efficacy of EGFR-TKIs in lung adenocarcinoma harboring the L858R mutation. *Ann. Oncol.* **25**, 1948–1953 (2014).
48. Vakiani, E., Janakiraman, M., Shen, R. & Sinha, R. Comparative genomic analysis of primary versus metastatic colorectal carcinomas. *J. Clin.* (2012).
49. Stratton, M. R. Exploring the Genomes of Cancer Cells: Progress and Promise. *Science (80-)*. **331** (2011).
50. Klein, C. A. Selection and adaptation during metastatic cancer progression. *Nature* **501**, 365–372 (2013).
51. Turner, N. C. & Reis-Filho, J. S. Genetic heterogeneity and cancer drug resistance. *Lancet. Oncol.* **13**, e178–85 (2012).
52. Vermaat, J. S. *et al.* Primary Colorectal Cancers and Their Subsequent Hepatic Metastases Are Genetically Different: Implications for Selection of Patients for Targeted Treatment. *Clin. Cancer Res.* **18** (2012).
53. Mathis, R. A., Sokol, E. S. & Gupta, P. B. Cancer cells exhibit clonal diversity in phenotypic plasticity. *Open Biol.* **7** (2017).
54. Ryser, M. D., Min, B.-H., Siegmund, K. D. & Shibata, D. Spatial mutation patterns as markers of early colorectal tumor cell mobility. *Proc. Natl. Acad. Sci.* **115**, 5774–5779 (2018).
55. Kreso, A. *et al.* Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* **339**, 543–8 (2013).

56. Jonker, D. J. *et al.* Cetuximab for the Treatment of Colorectal Cancer. *N. Engl. J. Med.* **357**, 2040–2048 (2007).
57. De Roock, W. *et al.* KRAS wild-type state predicts survival and is associated to early radiological response in metastatic colorectal cancer treated with cetuximab. *Ann. Oncol.* **19**, 508–515 (2007).
58. Perez, K. *et al.* Heterogeneity of colorectal cancer (CRC) in reference to KRAS proto-oncogene utilizing WAVE technology. *Exp. Mol. Pathol.* **95**, (74–82 (2013)).
59. Tougeron, D. *et al.* Effect of low-frequency KRAS mutations on the response to anti-EGFR therapy in metastatic colorectal cancer. *Ann. Oncol.* **24**, 1267–1273 (2013).
60. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (80-.)*. **344** (2014).
61. Fearhead, N. S., Britton, M. P. & Bodmer, W. F. The ABC of APC. *Hum. Mol. Genet.* **10**, 721–33 (2001).
62. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
63. Dunne, P. D. *et al.* Challenging the Cancer Molecular Stratification Dogma: Intratumoral Heterogeneity Undermines Consensus Molecular Subtypes and Potential Diagnostic Value in Colorectal Cancer. *Clin. Cancer Res.* **22** (2016).
64. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345 (2013).
65. Crowley, E., Di Nicolantonio, F., Loupakis, F. & Bardelli, A. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat. Rev. Clin. Oncol.* **10**, 472–484 (2013).

## Acknowledgements

We appreciate the support from the Canadian Centre for Computational Genomics for bioinformatics analysis. This work was generously funded by Cancer Care Ontario-Applied Cancer Research Units Grant, and Princess Margaret Cancer Foundation.

## Author contributions

H.N. participated in the design of the study, reconstructed the evolutionary structures within samples with help from S.J., analyzed the copy number variations, provided statistical analysis, contributed to the discussion, and drafted the manuscript. D.G. conceived of the study, participated in its design, coordination and analysis and drafted the manuscript. M.A. carried out the molecular genetic studies and assisted with statistical analysis. M.S. was involved in sample acquisition and carried out molecular genetic studies. L.L., M.B. and R.E. performed the bioinformatics analysis of sequencing data. M.K. carried out molecular genetic studies and assisted with statistical analysis. T.Z. was involved in sample acquisition and carried out molecular genetic studies. M.T. was involved in sample acquisition and carried out molecular genetic studies. M.R. was involved in sample acquisition and preparation. E.C. assisted with patient identification and critically revised the manuscript. M.M. assisted with patient identification and critically revised the manuscript. A.G. performed sample acquisition for analysis and was involved in study coordination. C.Y. assisted with patient identification, study coordination, and performed sample acquisition for analysis. P.L.B. participated in the design and coordination of the study and assisted with patient identification. S.K. was involved in sample acquisition and molecular genetic studies. JM contributed to the discussions and revision of the manuscript. L.L.S. conceived of the study, and participated in its design and coordination and helped to draft the manuscript. Y.R. participated in the design of the study, carried out the molecular genetic studies, performed the statistical analysis and helped to draft the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-57476-y>.

**Correspondence** and requests for materials should be addressed to Y.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020