# Tumor heterogeneity predicts metastatic potential in colorectal cancer

Je-Gun Joung[1†], Bo Young Oh[2†], Hye Kyung Hong[3], Hisham Al-Khalidi[4], Faisal Al-Alem[5], Hae-Ock Lee[1], Joon Seol Bae[1], Jinho Kim[1], Hong-Ui Cha[1], Maram Alotaibi[4], Yong Beom Cho[3], Mazen Hassanain[5*], Woong-Yang Park[1,6,7*], Woo Yong Lee[3*]

**Author's affiliations:**

[1]Samsung Genome Institute, Samsung Medical Center, Seoul 06351, Korea

[2]Department of Surgery, College of Medicine, Ewha Woman University, Seoul 07985, Korea

[3]Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea

[4]Department of Pathology, College of Medicine, King Saud University, Riyadh 12372, Saudi Arabia

[5]Department of Surgery, College of Medicine, King Saud University, Riyadh 12372, Saudi Arabia

[6]Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul 06351, Korea

[7]Department of Molecular Cell Biology, Sungkyunkwan University School of Medicine, Suwon 16419, Korea

[†]These authors contributed equally to this work.

**Running title:** Tumor heterogeneity in colorectal cancer liver metastasis

**Keywords:** Colorectal cancer, Tumor heterogeneity, Whole exome sequencing, Clonality,

Liver metastasis

**Grant support:** The study was supported in part by the Korea Health Technology R&D Project (HI13C2096 to W.-Y.P.) through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Korea.

*__Corresponding authors:__

Woong-Yang Park, M.D., Ph.D., Samsung Genome Institute, Samsung Medical Center, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea. Phone: +82-2-3410-6128; fax: +82-2-2148-9819; e-mail: woongyang.park@samsung.com

Mazen Hassanain, M.D., Ph.D., Department of Surgery, College of Medicine, King Saud University, Riyadh 12372 Saudi Arabia. Phone: +966-50-514-1090, e-mail: mhassanain@ksu.edu.sa

Woo Yong Lee, M.D., Ph.D., Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Samsung Medical Center, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea. Phone: +82-2-3410-0261; e-mail: wooyong123@samsung.com

## ABSTRACT

**Purpose:** Tumors continuously evolve to maintain growth; secondary mutations facilitate this process, resulting in high tumor heterogeneity. In this study, we compared mutations in paired primary and metastatic colorectal cancer (CRC) tumor samples to determine whether tumor heterogeneity can predict tumor metastasis.

**Experimental design:** Somatic variations in 46 pairs of matched primary–liver metastatic tumors and 42 primary tumors without metastasis were analyzed by whole exome sequencing. Tumor clonality was estimated from single nucleotide and copy number variations. The correlation between clinical parameters of patients and clonal heterogeneity in liver metastasis was evaluated.

**Results:** Tumor heterogeneity across CRC samples was highly variable; however, a high degree of tumor heterogeneity was associated with a worse disease free survival. Highly heterogeneous primary CRC was correlated with a higher rate of liver metastasis. Recurrent somatic mutations in *APC*, *TP53*, and *KRAS* were frequently detected in highly heterogeneous CRC. The variant allele frequency of these mutations was high, while somatic mutations in other genes such as *PIK3CA* and *NOTCH1* were low. The number and distribution of primary CRC sub-clones were preserved in metastatic tumors.

**Conclusions:** Heterogeneity of primary CRC tumors can predict the potential for liver metastasis and thus, clinical outcome of patients.

## TRANSLATIONAL RELEVANCE

Understanding the genetic heterogeneity of tumors is important for predicting patient treatment responses and disease progression. In colorectal cancer (CRC), tumor heterogeneity may be associated with worse prognosis and response to therapy. In this study, we demonstrate that a high degree of tumor heterogeneity is correlated with poor clinical outcome in CRC patients. Highly heterogeneous primary CRC was more likely to lead to liver metastasis and was linked to vascular invasion and the occurrence of CRC-associated somatic mutations. Interestingly, the subclonal pattern of primary CRC was largely preserved in metastatic tumors. Our results suggest that CRC liver metastasis develops through the collective spread of multiple clonal sub-populations in parallel. Thus, tumor heterogeneity plays a critical role in the metastatic progression of CRC and may serve as a marker for predicting clinical outcome.

## INTRODUCTION

Colorectal cancer (CRC) is one of the most commonly diagnosed cancers worldwide and a leading cause of cancer-related death (1,2). Early detection by endoscopic examination and more effective chemotherapy have led to increased survival of CRC patients, although the mortality rate remains high in cases of metastatic disease and/or recurrence (2). Synchronous metastases are detected in 20%–25% of patients at the time of diagnosis; in addition, approximately 50% develop metastasis during the disease course (3). Metastatic progression is a multi-step process involving phenotypic changes in primary tumor cells as a result of genetic and/or epigenetic alterations that facilitate dissemination and tissue invasion of tumor cells (4,5). Many proto-oncogenes and tumor suppressor genes associated with CRC have been identified; however, little is known regarding the molecular events responsible for metastatic progression of the disease.

Tumor heterogeneity—the concept that a single tumor consists of many tumor cell sub-clones—has become an important topic in cancer genomics (6). It is hypothesized to play a critical role in the progression of many cancer types and is a major obstacle to precision cancer therapy. During this process, sub-clones continuously arise via genomic mutation. The presence of sub-clones has been shown to adversely affect outcome in chronic lymphocytic leukemia (7), head and neck cancer (8), and lung adenocarcinoma (9). The so-called Big Bang model of tumor progression posits that malignant potential is mostly determined early in tumor development and is non-selective (10). However, the full complement of factors that lead to tumor heterogeneity during CRC progression is unknown.

Next-generation sequencing technology has recently been used in clinical applications to identify mutations and copy number variations in oncogenes. Specifically, whole-exome sequencing (WES) has revealed informative mutations within exome coding regions. Based on WES data, sub-clones of tumor cell populations can be inferred by computational methods

such as PyClone (11), SciClone (12), and EXPANDS (13). Clonal heterogeneity analysis may be useful for predicting patient prognosis and response to therapy. Early clonal events are potential therapeutic targets as they represent the tumor cell population, while sub-clonal events may be associated with specific targets linked to sub-optimal outcomes (14).

To investigate whether tumor heterogeneity is associated with clinical outcome and is a driver for mutations in CRC, we performed WES of primary samples from CRC patients and their hepatic metastatic tumors. We identified numerous somatic mutations and sub-clones from individual tumors. Our results provide insight into the mechanism of metastasis in CRC and provide a basis for the development of novel therapeutic agents for disease treatment.

## MATERIALS AND METHODS

### Patients

This study was approved by the Institutional Review Boards of Samsung Medical Center (SMC) (IRB no. 2010-04-004) and King Saudi University (KSU) (IRB no. E-12-592). All procedures were conducted in accordance with the Declaration of Helsinki. Written, informed consent was obtained from all enrolled patients.

The study population included 88 CRC patients consisting of 42 stage I–III and 46 stage IV patients with synchronous liver metastasis. All had histologically confirmed primary colorectal adenocarcinoma and had undergone curative surgical resection for the primary tumor and metastatic lesions at the SMC and KSU from 2004 to 2014. Patients were excluded if they had recurrent disease, local excision, palliative surgery, or previous treatment prior to colonic surgery.

A total of 222 samples from 88 patients were obtained; the primary and metastatic tumor and blood (or normal colon) samples were obtained from metastatic patients, and primary tumor and blood or normal colon samples were obtained from non-metastatic patients.

### Whole exome sequencing

Genomic DNA was extracted from fresh frozen and formalin-fixed, paraffin-embedded (FFPE) tissue using the QIAamp DNA mini and a DNA FFPE Tissue kits (Qiagen, Hilden, Germany). The tissue was microdissected. Since paraffin fixation chemically modifies genomic (g)DNA and thereby reduces its quality, we applied the following quality control thresholds: i) purity: 260/280 > 1.8 and 260/230 > 1.8; ii) total amount > 250 ng; and iii) degradation (FFPE): $\Delta$Ct value < 2.0 or DNA median size > 0.35 kb. The highest quality gDNA in each sample was sheared with an S220 ultra-sonicator (Covaris, Woburn, MA, USA)

and used to construct a library with the SureSelect XT Human All Exon v5 and SureSelect XT reagent kit, HSQ (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's protocol. This kit is designed to enrich 335,756 exons of 21,058 genes, covering ~71 Mb of the human genome. Enriched exome libraries were multiplexed and sequenced on the HiSeq 2500 platform (Illumina, San Diego, CA, USA). Briefly, a paired-end DNA sequencing library was prepared through gDNA shearing, end-repair, A-tailing, paired-end adaptor ligation, and amplification. After hybridizing the library with bait sequences for 16 h, the captured library was purified and amplified with an indexing barcode tag and library quality and quantity were assessed. The exome library was sequenced using the 10-bp paired-end mode of the TruSeq Rapid PE Cluster and TruSeq Rapid SBS kits (Illumina).

**Exome sequence data analysis**

Sequencing reads were aligned to the University of California Santa Cruz hg19 reference genome (downloaded from http://genome.ucsc.edu) using the Burrows-Wheeler Aligner v. 0.6.2 (15) with default settings. PCR duplications were marked using Picard-tools-1.8 (http://picard.sourceforge.net/) and data cleanup was achieved using GATK-2.2.9 (16). Point mutations were identified with the MuTect tool (https://github.com/broadinstitute/mutect) in paired samples. Annovar was used to annotate variants. Signature analysis of mutational processes was carried out using the deconstructSigs tool (17). Hyper-mutated samples with more than 1000 mutations were excluded as they introduced bias. Copy number variations were detected using EXCAVATOR software (18).

**Clonality analysis**

Sub-clones were obtained using clustering cancer cell fractions (CCF) by PyClone, which de-convolves tumor sequences into sub-clones based on a hierarchical Bayesian clustering model (11). Input data were generated from somatic single-nucleotide variants (SNVs) detected by MuTect and corresponding copy number variations. SNVs were clustered according to similar CCF, after which each set of clustered mutations (a subclone) of cancer cells was identified. Step filtering of SNVs with large, credible intervals from the previous stage was applied to remove non-informative posterior distributions. Each tumor was classified into two types of clonality based on the number of sub-clones they possessed. We defined a tumor as an oligo-clone when there were one or two sub-clones; otherwise, the tumor was defined as a multi-clone. A clonal evolution model for metastatic CRCs was generated using ClonEvol R (https://github.com/hdng/clonevol).

**Measurement of heterogeneity**

Heterogeneity was measured by mutant-allele tumor heterogeneity (MATH) analysis, which was initially developed to measure intra-tumor heterogeneity in head and neck cancer samples (8). Mutation allele frequencies (MAFs) of mutated loci in each tumor were determined; the center and width of the distribution of MAFs among these loci were obtained, and the ratio of the width to the center of the distribution was calculated as median absolute deviation/median MAF.

**Assessment of clinical outcomes**

Patients were categorized according to clonality and groups were compared with respect to clinicopathologic features. To investigate the clinical relevance of clonality, we analyzed disease-free survival (DFS) according to clonality. Recurrence was established by biopsy following colonoscopy or based on imaging findings typical of cancer recurrence, and DFS

was defined as the period from surgery to recurrence. The primary endpoint of this study was the degree of clonality of each tumor, and the secondary endpoint was survival outcome based on clonality.

**Statistical analysis**

Statistical analyses were performed using R v.3.1.2 software (https://www.r-project.org/). Categorical variables were compared using the $\chi^2$ test. Survival rates were analyzed with the Kaplan–Meier method and log-rank test. Multivariate analysis was carried out using a Cox proportional hazard model to identify prognostic factors. P values were derived from two-tailed tests and those less than 0.05 were considered statistically significant.

## RESULTS

### Identification of somatic mutations in CRCs

A total of 88 patients with CRC, including 46 with and 42 without liver metastasis (mCRC and nmCRC, respectively) were evaluated. We performed WES using 46 matched pairs (primary colon tumor, liver metastatic tumor, and blood [or normal colon] samples) from the mCRC group and 42 matched pairs (primary colon tumor and blood [or normal colon] samples) from the nmCRC group (**Supplementary Table 1**). Patient characteristics according to clonality are shown in Table 1; there were no differences between groups for most variables.

WES of the 222 samples yielded 33.7 billion reads. The average target depth for the analysis was 165.6× (s.d. 50.7) and whole-exome coverage was at least 99.4% (**Supplementary Table 1**). The depth of tumors and normal samples was 173.9× (s.d. 49.9) and 152.8× (s.d. 49.6), respectively.

We identified numerous somatic mutations including missense, nonsense, and splicing mutations in primary colon tumors obtained from mCRC and nmCRC groups (**Fig. 1**). These included several somatic mutations that frequently occur in CRCs, with the most common and well-described (19) identified at particularly high frequencies, including *adenomatous polyposis coli* (*APC*) (54.6%), *KRAS* (42.0%), and *TP53* (55.7%), consistent with previous reports (20). We also investigated whether there was a concordance of mutations between primary colon tumors and matched liver metastatic tumors in the mCRC group (**Fig. 1**). Most mutations were detected in both primary tumors and metastatic lesions, with high rates of concordance in *APC* (77.8%, 21/27 concordant mutations in primary tumors), *KRAS* (88.2%, 15/17), and *TP53* (87.5%, 21/24); this is consistent with previous studies of primary metastasis in CRC (21). Other CRC-associated genes such as *phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha* (*PIK3CA*), *mothers against decapentaplegic*

*homolog 4* (*SMAD4*), *NRAS*, *catenin β-1* (*CTNNB1*), *AKT1*, *NOTCH1*, and *anaplastic lymphoma kinase* (*ALK*)were less frequently mutated.

**Correlation between clonality and clinical features**

We investigated whether there was an association between clonality and primary tumor genomics or related clinical features. Clonality was determined based on sub-clones in each tumor. The number of sub-clones was positively correlated with tumor heterogeneity, as determined by MATH analysis (**Fig. 2A**). This indicates that tumors with a larger number of sub-clones exhibit a higher degree of heterogeneity.

We next performed a survival analysis to evaluate the clinical relevance of clonality, and found a significant association between DFS and number of sub-clones (**Fig. 2B**). CRC patients with multi-clone tumors showed worse clinical outcome relative to those with oligo-clone tumors ($P = 0.04$). Patients with a high tumor heterogeneity score showed a similar result ($P = 0.009$) (**Supplementary Fig. 1**). Tumor recurrence rate also differed between oligo- and multi-clone groups ($P < 0.012$; $\chi^2$ test) (**Fig. 2C**). Multivariate analyses were carried out using a Cox proportional hazard model to identify prognostic factors for DFS; these included variables that were significant in the univariate analysis. The results showed that the presence of multiple clones was an independent prognostic factor for reduced DFS (Table 2).

Mutations in the three major CRC-associated genes—i.e., *APC*, *KRAS*, and *TP53*—were more common, however the P values are all > 0.05. (**Fig. 2D**). Identified mutations are shown in **Figure 2E**. Other clinical features including preoperative carcinoembryonic antigen level, cell type, tumor stage at diagnosis, vascular invasion, lymphatic invasion, and perineural invasion were also compared to the number of sub-clones (1, 2, 3, 4, and ≥5). Among the assessed criteria, progression to an advanced stage and presence of vascular

invasion were significantly associated with multi-clone tumors (P = 0.046 and 0.015, respectively; $\chi^2$ test) (**Table 1**).

**Characteristics of somatic mutations according to clonality**

We categorized somatic mutations according to their order of CCFs to determine the relative timing of mutation acquisition. Mutations were classified into four categories based on CCF distribution (G1, 1st; G2, 2nd; G3, 3rd; and G4, 4th quantile) (**Fig. 3A**). G1 included events that occurred earlier than those in other categories. Mutations in *TP53*, *APC*, *KRAS*, *SMAD4*, *NRAS*, *CTNNB1*, and *F-box and WD repeat domain-containing 7*(*FBXW7*)—which are considered as early events—belonged to the high-CCF group. Interestingly, these mutations occurred predominantly in multi-clonal samples. For example, 71% of *APC* mutations belonged to the G1 and G2 groups (**Fig. 3B**). This supports the hypothesis that colorectal tumorigenesis is initiated by a mutation in *APC*. In multi-clonal samples, 74.2% of *APC* mutations were detected in the G1 and G2 groups as compared to 57.1% in these groups in oligo-clonal samples. This suggests that the occurrence of mutations decreased substantially during expansion into multiple sub-clonal populations. *KRAS* and *TP53* showed similar patterns in the relative timing of mutation acquisition (**Supplementary Figure 2**). In contrast, mutations in *PIK3CA* and *NOTCH1* belonged to a low-CCF group, indicating that they were later events.

We then carried out a signature analysis of the mutation process. Overall mutation patterns were similar to those of Signature 1, which is observed in most cancer types (22). We compared the substitution spectra of oligo- and multi-clones (**Fig. 3C**) and found that there were no differences in the patterns of nucleotide changes between oligo- and multi-clonal samples. However, when the signatures were decomposed, an additional one (Signature 19) was observed in a small number of oligo-clonal samples (**Supplementary Figure 3**),

282 although its etiology remains unclear.

283

284 **Changes in clonality from primary to metastatic tumors**

285 We investigated quantitative changes (i.e., differences in the number of sub-clones) in
286 clonality during the progression from primary tumor to liver metastatic lesion. There are four
287 possible transitions of oligo- vs. multi-clones during this process (CRC-OO, oligo → oligo;
288 CRC-OM, oligo → multi, CRC-MO, multi → oligo; and CRC-MM, multi → multi) (**Fig. 4A**
289 and **Supplementary Table 2**). The CRC-MM transition (71.7%, 33/46 patients with
290 metastatic CRC) was predominant, indicating that the number of sub-clones was highly
291 consistent between multi-clonal primary tumors and liver metastatic lesions. A combination
292 of *APC*, *KRAS*, and *TP53* mutations was associated with both multi-clonal primary and
293 metastatic samples. For example, mCRC-35 showed a CRC-MM transition type (**Fig. 4B**);
294 the patient harbored mutations in *KRAS*, *TP53*, *CTNNB1*, and *APC*, which may be required
295 for clonal expansion. Interestingly, only *APC* and *TP53* mutations and not *KRAS* mutations
296 were associated with CRC-OO transitions. Nonetheless, we speculate that *KRAS* mutations
297 have an important role in clonal expansion.

## DISCUSSION

In this study, a comprehensive set of somatic mutations and sub-clones of individual primary and hepatic metastatic tumors from 88 CRC patients was identified by WES. We observed that tumor heterogeneity was significantly associated with survival outcome in CRC. Primary tumors with high heterogeneity were more likely to spread to the liver from the primary tumor site, suggesting an association with vascular invasion. It is highly likely that tumor cells in sub-clones spread together rather than individually to metastatic lesions. Six tumors in the mCRC group were hypermutated, but these did not always display high heterogeneity. For example, a *MutS homolog 6* (MSH6) mutation was oligo-clonal in both primary and metastatic lesions (Fig. 4B).

Mutations frequently occurring in CRC such as those in *APC*, *TP53*, and *KRAS* were more common in CRC patients with high as compared to low tumor heterogeneity and were considered as early clonal events in high-CCF groups in multi-clonal samples. These mutations would likely accumulate before clonal expansion into multiple sub-clonal populations. On the other hand, mutations in *PIK3CA* and *NOTCH1* in low CCFs occurred later through continuous clonal expansion, which is consistent with the Big Bang model of tumor progression (10). Thus, a rapidly growing clone within small tumor may arise from several early key mutations, with subsequent mutations improving the fitness of the local clonal expansion of tumor cell sub-populations. Somatic mutations in *PIK3CA* are late events during tumorigenesis (23,24), and those in *NOTCH1* are secondary events during T cell acute lymphoblastic leukemia (25). The low allele frequencies in *NOTCH1* observed in the present study support the notion that these mutations occur later. However, the effects of *NOTCH1* mutations in CRC and other cancer types remain to be determined.

There was no relationship between mutational signatures and the number of sub-clones, indicating that changes in mutational composition do not determine the number of

sub-clones during tumor progression. Although most of our analyses were not affected by hyper-mutated samples, mutation spectrum could be highly affected by hypermutation. Therefore, hypermutators were excluded from the mutation spectrum analysis as the patterns could reflect hypermutations rather than tumor heterogeneity. Previous studies of lung cancer reported that the spectrum of mutations caused by transformations associated with *apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like* (APOBEC) *cytidine deaminase* differed between the trunk (early events) and non-trunk (late events) (26). Additional studies are needed to identify the changes in the mutation spectrum that occur during CRC progression.

This study has a limitation that stage was differently distributed between groups according to clonality. Multiclonal patients had more metastatic disease at diagnosis than oligo-oclonal patients (60.7% vs. 33.3%). This may affect our result that the risk of relapse was higher in multi-clonal patients than oligo-clonal patients. Thus, multivariate analysis using a Cox proportional hazard model was performed to overcome the confounding bias and we identified that multi-clone was an independent poor prognostic factor for survival. Despite this limitation, our study had the strength of integrating tumor heterogeneity into clinical applications. Understanding tumor heterogeneity is essential for developing personalized treatment for patients with cancers, including CRC. Tumor heterogeneity resulting from the continuous accumulation of mutations during tumor progression determines tumor characteristics, influencing the clinical outcome of patients. Tumor heterogeneity, defined as the genetic diversity that exists within individual tumors, is one of the most challenging current issues in the field of cancer biology.

Tumor heterogeneity resulting from continuous accumulation of mutations can aid in predicting response and resistance to drugs and dictate clinical outcome. In this study, there was a significant association between the rate of primary CRC recurrence and number of sub-

clones, suggesting that recurrence is also related to high tumor heterogeneity. Our findings provide a basis for the development of personalized treatments for patients with CRC and other cancers.

## ACKNOWLEDGMENTS

We thank Seok Hyung Kim and Kyoung-Mee Kim (Department of Pathology, Samsung Medical Center) for pathological analyses. Biospecimens from Samsung Medical Center were provided by SMC BioBank. All sequencing was performed by Samsung Genome Institute.

## REFERENCES

1. Sui X, Xu Y, Yang J, Fang Y, Lou H, Han W, et al. Use of metformin alone is not associated with survival outcomes of colorectal cancer cell but AMPK activator AICAR sensitizes anticancer effect of 5-fluorouracil through AMPK activation. PLoS One 2014;9(5):e97781.
2. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA Cancer J Clin 2014;64(1):9-29.
3. Oh BY, Hong HK, Lee WY, Cho YB. Animal models of colorectal cancer with liver metastasis. Cancer Lett 2017;387:114-20.
4. Wanebo HJ, LeGolvan M, Paty PB, Saha S, Zuber M, D'Angelica MI, et al. Meeting the biologic challenge of colorectal metastases. Clin Exp Metastasis 2012;29(7):821-39.
5. Vignot S, Lefebvre C, Frampton GM, Meurice G, Yelensky R, Palmer G, et al. Comparative analysis of primary tumour and matched metastases in colorectal cancer patients: Evaluation of concordance between genomic and transcriptional profiles. Eur J Cancer 2015.
6. Schmidt F, Efferth T. Tumor Heterogeneity, Single-Cell Sequencing, and Drug Resistance. Pharmaceuticals (Basel) 2016;9(2).
7. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell 2013;152(4):714-26.
8. Mroz EA, Tward AD, Hammon RJ, Ren Y, Rocco JW. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. PLoS Med 2015;12(2):e1001786.

9. Zhang J, Fujimoto J, Wedge DC, Song X, Seth S, Chow CW, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science 2014;346(6206):256-9.

10. Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, et al. A Big Bang model of human colorectal tumor growth. Nat Genet 2015;47(3):209-16.

11. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. Nat Methods 2014;11(4):396-8.

12. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. PLoS Comput Biol 2014;10(8):e1003665.

13. Andor N, Harness JV, Muller S, Mewes HW, Petritsch C. EXPANDS: expanding ploidy and allele frequency on nested subpopulations. Bioinformatics 2014;30(1):50-60.

14. Turajlic S, McGranahan N, Swanton C. Inferring mutational timing and reconstructing tumour evolutionary histories. Biochim Biophys Acta 2015;1855(2):264-75.

15. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754-60.

16. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20(9):1297-303.

17. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol 2016;17:31.

18. Magi A, Tattini L, Cifola I, D'Aurizio R, Benelli M, Mangano E, et al. EXCAVATOR:

detecting copy number variants from whole-exome sequencing data. Genome Biol 2013;14(10):R120.

19. Conlin A, Smith G, Carey FA, Wolf CR, Steele RJ. The prognostic significance of K-ras, p53, and APC mutations in colorectal carcinoma. Gut 2005;54(9):1283-6.

20. Network TCGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487(7407):330-7.

21. Brannon AR, Vakiani E, Sylvester BE, Scott SN, McDermott G, Shah RH, et al. Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. Genome Biol 2014;15(8):454.

22. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature 2013;500(7463):415-21.

23. Verlaat W, Snijders PJ, van Moorsel MI, Bleeker M, Rozendaal L, Sie D, et al. Somatic mutation in PIK3CA is a late event in cervical carcinogenesis. J Pathol Clin Res 2015;1(4):207-11.

24. Samuels Y, Velculescu VE. Oncogenic mutations of PIK3CA in human cancers. Cell Cycle 2004;3(10):1221-4.

25. Mansour MR, Duke V, Foroni L, Patel B, Allen CG, Ancliff PJ, et al. Notch-1 mutations are secondary events in some patients with T-cell acute lymphoblastic leukemia. Clin Cancer Res 2007;13(23):6964-9.

26. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science 2014;346(6206):251-6.

## TABLES

**Table 1.** Demographic and clinical landscape of 88 CRC patients

| Characteristic | Oligo-clone (n = 27) | Multi-clone (n = 61) | P-value |
|---|---|---|---|
| Age, median (years) | 54.8 | 57.1 | 0.454 |
| Gender, n (%) | | | 0.846 |
| Male | 14 (51.9%) | 33 (54.1%) | |
| Female | 13 (48.1%) | 28 (45.9%) | |
| Preoperative CEA level, n (%) | | | 0.125 |
| < 5 ng/ml | 9 (33.3%) | 28 (45.9%) | |
| ≥ 5 ng/ml | 9 (33.3%) | 22 (36.1%) | |
| Unknown | 9 (33.3%) | 11 (18.0%) | |
| Location of primary tumor, n (%) | | | 0.940 |
| Colon | 21 (77.8%) | 47 (77.0%) | |
| Rectum | 6 (22.2%) | 14 (23.0%) | |
| Cell type, n (%) | | | 0.893 |
| WD/MD | 20 (74.1%) | 46 (75.4%) | |
| PD/MUC/SRC | 7 (25.9%) | 15 (24.6%) | |
| Stage at diagnosis, n (%) | | | 0.046 |
| I | 1 (3.7%) | 2 (3.3%) | |
| II | 7 (25.9%) | 14 (23.0%) | |
| III | 10 (37.0%) | 8 (13.1%) | |
| IV | 9 (33.3%) | 37 (60.7%) | |
| Vascular invasion, n (%) | | | 0.015 |
| Yes | 2 (7.4%) | 20 (32.8%) | |
| No | 25 (92.6%) | 41 (67.2%) | |
| Lymphatic invasion, n (%) | | | 0.930 |
| Yes | 10 (37.0%) | 22 (36.1%) | |
| No | 17 (63.0%) | 39 (63.9%) | |
| Perineural invasion, n (%) | | | 0.351 |
| Yes | 7 (25.9%) | 22 (36.1%) | |
| No | 20 (74.1%) | 39 (63.9%) | |
| Adjuvant treatment, n (%) | | | 0.520 |
| Yes | 24 (88.9%) | 51 (83.6%) | |
| No | 3 (11.1%) | 10 (16.4%) | |

*CRC* colorectal cancer, *CEA* carcinoembryonic antigen, *WD* well differentiated, *MD* moderately differentiated, *PD* poorly differentiated, *MUC* mucinous carcinoma, *SRC* signet ring cell carcinoma

437 **Table 2.** Multivariate analysis for disease-free survival in 88 CRC patients

| Variables | Univariate | Multivariate | |
|---|---|---|---|
| | p-value | HR (95% CI) | p-value |
| Clonality | | | |
| multi *vs*. oligo | 0.041 | 3.083(1.035-9.184) | 0.043 |
| Age (years) | | | |
| ≥ 65 *vs*. < 65 | 0.585 | | |
| Gender | | | |
| female *vs*. male | 0.622 | | |
| CEA level (ng/ml) | | | |
| ≥ 5 *vs*. < 5 | < 0.001 | 2.343(0.858-6.393) | 0.097 |
| Location of tumor | | | |
| rectum *vs*. colon | 0.001 | | |
| Stage | < 0.001 | | 0.014 |
| III *vs*. I-II | 0.404 | 1.639(0.390-6.887) | 0.500 |
| IV *vs*. I-II | < 0.001 | 6.006(1.659-21.743) | 0.006 |
| Cell type | | | |
| PD/MUC/SRC *vs*. WD/MD | 0.303 | | |
| Vascular invasion | | | |
| yes *vs*. no | 0.031 | | |
| Lymphatic invasion | | | |
| yes *vs*. no | 0.023 | | |
| Perineural invasion | | | |
| yes *vs*. no | < 0.001 | 5.736(2.443-13.468) | < 0.001 |
| Adjuvant treatment | | | |
| yes *vs*. no | 0.105 | | |

438

439

440

441

442

443

444

445

## FIGURE LEGENDS

**Figure 1. Heat map of somatic (missense, nonsense, and splicing) mutations detected in CRC-associated genes.** Listed genes are sorted by mutation frequency in TCGA cohorts of TCGA colorectal adenocarcinoma.

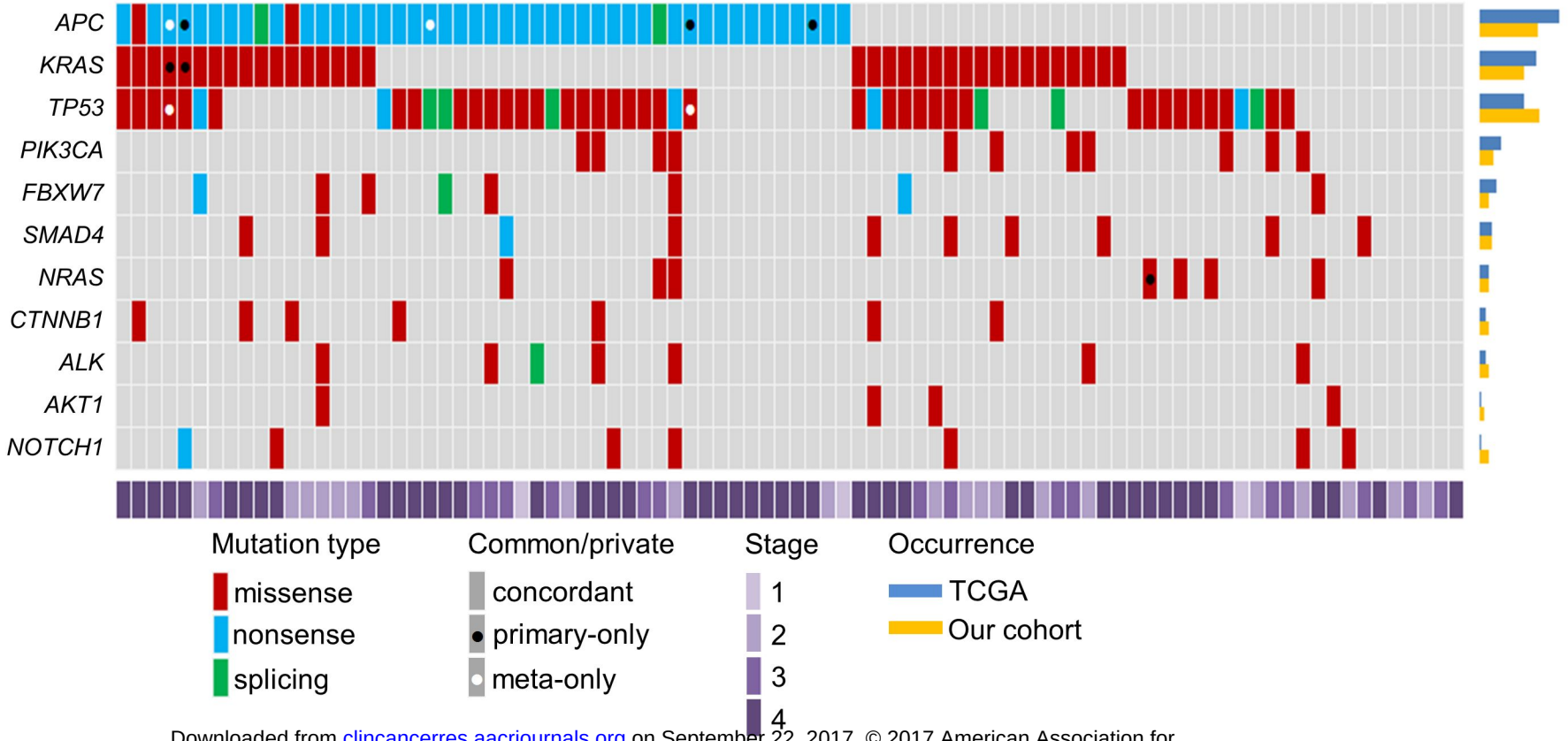
**Figure 2. Clinical significance of clonality in CRC. A.** Correlation between number of sub-clones and tumor heterogeneity. The former was inferred by clustering CCFs using PyClone, and the latter was determined by WES and MATH analysis of primary tumors (n = 88). **B.** Kaplan-Meier survival analysis of patient groups with oligo- and multi-clones. **C.** Patients with oligo- and multi-clone tumors were compared in terms of tumor recurrence rate. **D.** Comparison between proportions of oligo- and multi-clones in tumors with *APC*, *KRAS*, or *TP53* mutations. **E.** Mutation profiles of CRC-related genes (*APC*, *KRAS*, and *TP53*) and clinical features along with the number of sub-clones.

**Figure 3. Characteristics of mutations according to clonality. A.** Heat map of mutational frequency in four categories based on CCF distribution (G1, 1st; G2, 2nd; G3, 3rd; and G4, 4th quantile). G1 (or G4) corresponds to an early (or late) event. The size of each circle corresponds to the number of mutations. **B.** Circle diagram showing the number of *APC* mutations in each of the four categories (G1–G4). **C.** Spectrum of substitutions in oligo- and multi-clones. Each bar represents the proportion of each type of substitution.

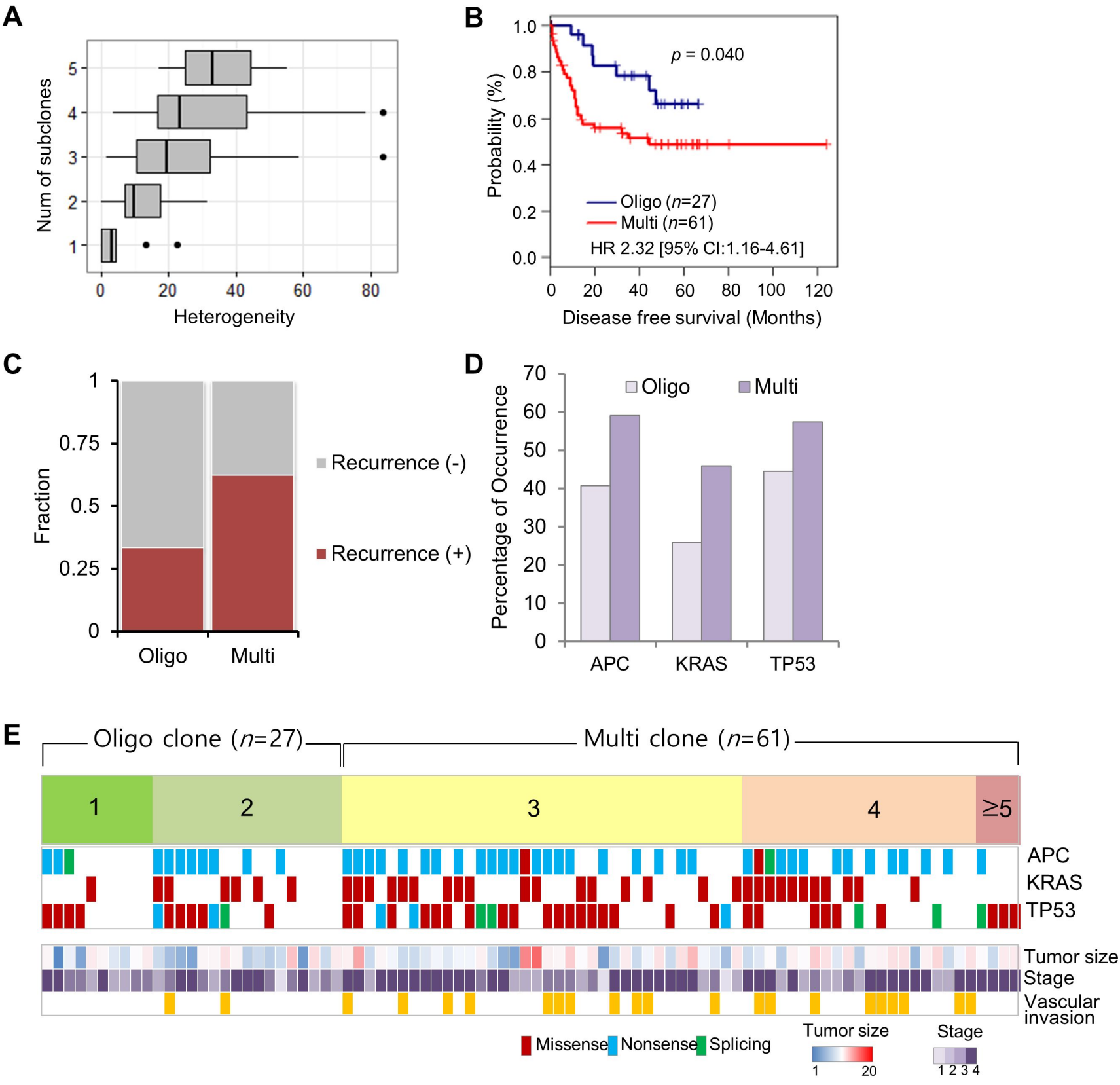
**Figure 4. Changes in clonality from primary to metastatic tumor. A.** Four possible cases of clonality transition for oligo- vs. multi-clones and primary vs. metastatic tumors (CRC-OO, oligo → oligo; CRC-OM, oligo → multi; CRC-MO, multi → oligo; and CRC-MM, multi → multi). Each triangle indicates one of *APC*, *KRAS*, and *TP53* mutations, and ∅ indicates that

none of them are found. Small arrow and triangle represented which mutations appeared between primary and metastatic tumor. (#) indicates the number of patients with the corresponding pattern. Large arrows indicate metastasis and their thickness are represented according to the number of patients with a transition type. **B.** Inferred clonal evolution in metastatic CRC. Each area indicates a sub-population containing a set of mutations. Cancer-associated genes are shown. A tree depicts the evolution of tumors. P# and M# indicate a sub-clone in a primary and metastatic tumors, respectively.

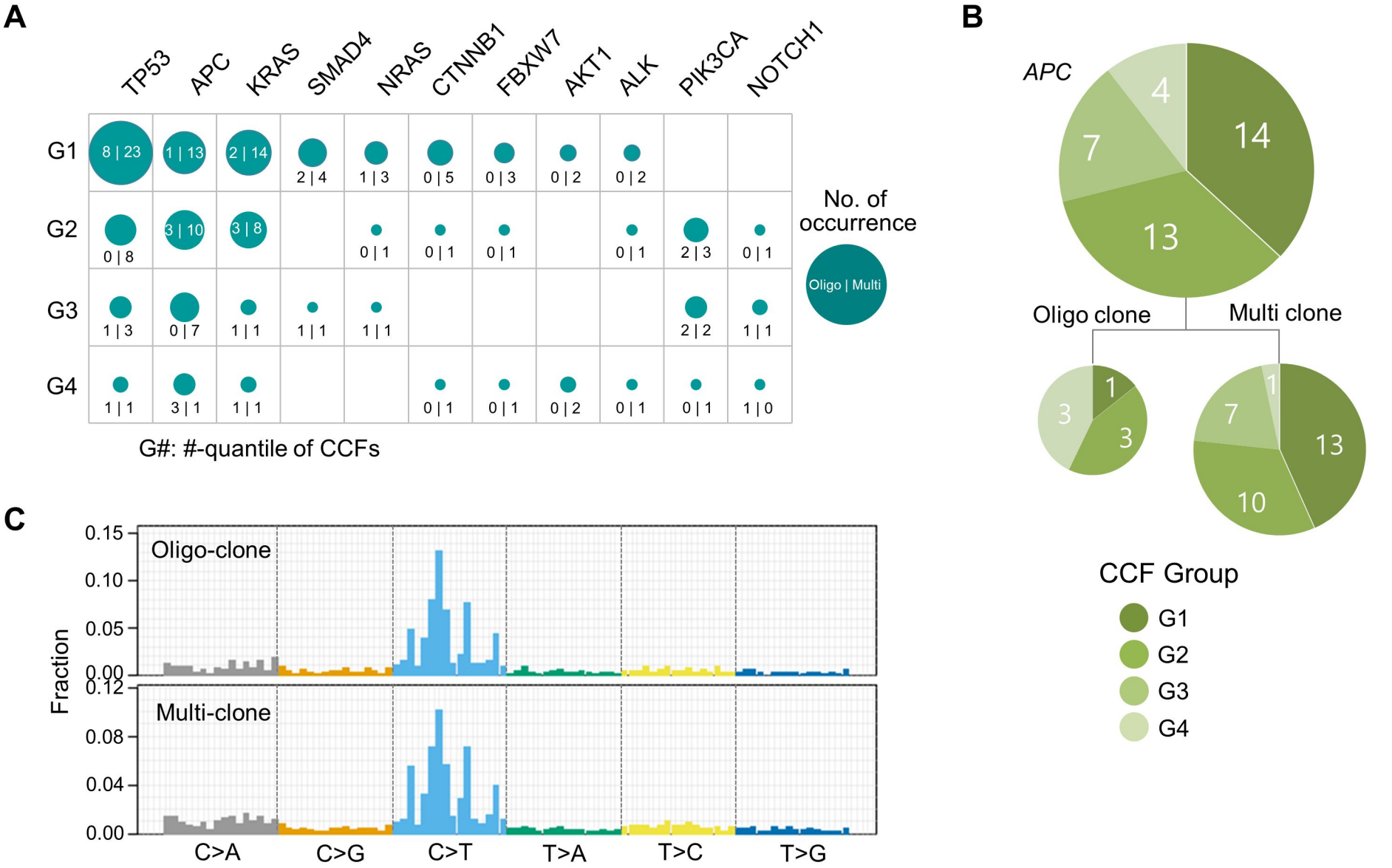
# Figure 1



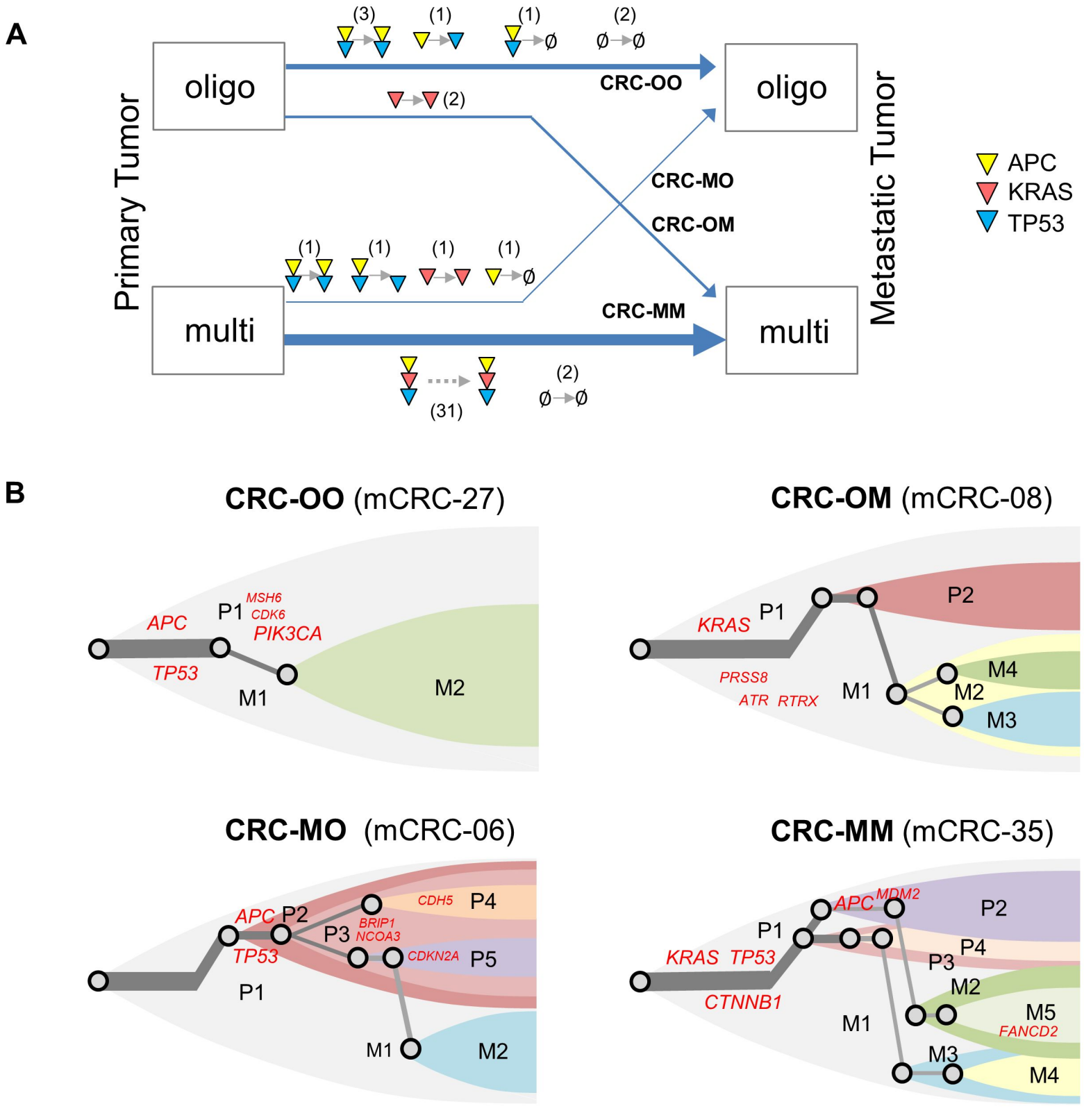
# Figure 2



# Figure 3



# Figure 4



AACR American Association for Cancer Research

Clinical Cancer Research

# Tumor heterogeneity predicts metastatic potential in colorectal cancer

Je-Gun Joung, Bo Young Oh, Hye Kyung Hong, et al.

*Clin Cancer Res* Published OnlineFirst September 22, 2017.