Dissertation

Tumor Genome Stratification and Network Analyses based on Non-invasive Liquid Biopsies

submitted by

Qing ZHOU

for the Academic Degree of

Doctor of Philosophy (PhD)

at the

MEDICAL UNIVERSITY OF GRAZ

Diagnostic & Research Institute of Human Genetics

Under the Supervision of

Univ.-Prof. Dr. Michael R. Speicher

2020

DECLARATION

I hereby declare that this dissertation is my own original work and that I have fully acknowledged by name all of those individuals and organizations that have contributed to the research for this dissertation. Due acknowledgement has been made in the text to all other material used. Throughout this dissertation and in all related publications I followed the "Guidelines of the Medical University of Graz on Good Scientific Practice".

Qing ZHOU; April 2020

DISCLOSURES

Please note that parts of this thesis has already been accepted for publication.

Zhou, Q., Perakis, S.O., Ulz, P., Mohan, S., Riedl, J.M., Talakic, E., Lax, S., Tötsch, M., Hoefler, G., Bauernhofer, T., Pichler, M., Gerger, A., Geigl, J.B., Heitzer, E. & Speicher, M.R. Cell-free DNA analysis reveals POLR1D-mediated resistance to bevacizumab in colorectal cancer. Genome Med 12, 20 (2020).

The following researchers have actively contributed to this publication:

Samantha O. Perakis; Jochen B. Geigl

Institute of Human Genetics, Diagnostic and Research Center for Molecular Biomedicine, Medical University of Graz, Graz, Austria.

Peter Ulz

Institute of Human Genetics, Diagnostic and Research Center for Molecular Biomedicine, Medical University of Graz, Graz, Austria.

Present: Freenome, South San Francisco, CA, USA.

Sumitra Mohan

Institute of Human Genetics, Diagnostic and Research Center for Molecular Biomedicine, Medical University of Graz, Graz, Austria.

Present: Cancer Research UK-Manchester Institute, Manchester, UK.

Jakob M. Riedl; Thomas Bauernhofer; Martin Pichler; Armin Gerger

Department of Internal Medicine, Division of Oncology, Medical University of Graz, Graz, Austria.

Emina Talakic

Division of General Radiology, Medical University of Graz, Graz, Austria.

Sigurd Lax

Johannes Kepler University Linz, Linz, Austria.

Martin Tötsch

Institute of Pathology, General Hospital Hochsteiermark, Leoben, Austria.

Gerald Hoefler

Institute of Pathology, Diagnostic and Research Center for Molecular Biomedicine, Medical University of Graz, Graz, Austria.

Ellen Heitzer

Institute of Human Genetics, Diagnostic and Research Center for Molecular Biomedicine, Medical University of Graz, Graz, Austria.

BioTechMed-Graz, Graz, Austria.

Christian Doppler Laboratory for Liquid Biopsies for Early Detection of Cancer, Graz, Austria.

Michael R. Speicher

Institute of Human Genetics, Diagnostic and Research Center for Molecular Biomedicine, Medical University of Graz, Graz, Austria.

BioTechMed-Graz, Graz, Austria.

All the co-authors have agreed to the inclusion of their published data in the dissertation and permission to reproduce illustrations and figures from own or third-party publications has been granted.

There are only two ways to live your life:

as though nothing is a miracle,

or as though everything is a miracle.

(Albert Einstein)

ACKNOWLEDGEMENTS

I would like to express my most profound appreciation to **Prof. Speicher** for allowing me to conduct my Ph.D. in his research team and under his supervision. Thank you for all of your scientific advice, discussions, corrections, and also for allowing me go to different laboratories and attend various scientific conferences to get more experience.

I would also like to thank **Prof. Heitzer** for all of her great advice, support, and also for allowing me to work on another great project. Thanks to **Prof. Geigl** for all of his clinically relevant advice and also thanks to **Peter** for solving all of my questions regarding bioinformatics.

Special thanks to **Samantha**, it's a pleasure to work with you. Thanks so much for your kindness and for helping me throughout all difficult times. Many thanks to **Tina, Sabrina, Julie, Ricarda, and Christine** for their help in the lab and a great working atmosphere.

Thanks to **Prof. Bauernhofer, Prof. Gerger, Prof. Pichler, and Dr. Riedl** for providing all of the blood samples and clinical data. Also, we are grateful to **Prof. Pichler** for providing us with the HT29 and SW480 cell lines. Thanks to **Prof. Höfler, Prof. Lax, and Dr. Tötsch** for providing the tumor tissue samples. Thanks to **Dr. Alberto Bardelli** for providing us with the OXCO-2 cell line. Thanks to **Dr. Talakic** for all the support regarding radiology. The results shown in this thesis are in part based on data generated by the **TCGA Research Network:** https://www.cancer.gov/tcga.

This work was supported by the **Medical University of Graz** (PhD Program in Molecular Medicine (MolMed)), by **CANCER-ID**, a project funded by the Innovative Medicines Joint Undertaking (IMI JU; #115749-1), by the **BioTechMed-Graz** flagship project **"EPIAge"**, and by the **Austrian Federal Ministry for Digital and Economic Affairs** (Christian Doppler Research Fund for Liquid Biopsies for Early Detection of Cancer).

Thanks to all of **my family members and my friends** for encouraging and supporting me. Special thanks to my husband, **Dr. Yang**, for your support and patience throughout my entire education and for helping me always find the right way.

TABLE OF CONTENTS

ABBREVIATIONS IX
LIST OF FIGURES XII
LIST OF TABLES XIV
ZUSAMMENFASSUNG XV
ABSTRACT XVII
1 INTRODUCTION 1
1.1 Colorectal Cancer 1
1.1.1 Diagnosis and monitoring 1
1.1.2 CRC treatment 2
1.1.3 Anti-EGFR 3
1.1.4 Anti-ERBB2 5
1.1.5 Anti-angiogenesis 7
1.2 Cell-free DNA 10
1.2.1 Biology of Cell-free DNA 10
1.2.2 Circulating tumor DNA 12
1.2.3 ctDNA and Colorectal Cancer 14
2 AIMS 16
3 MATERIALS AND METHODS 17
3.1 Ethics 17
3.2 Sample collection 17
3.3 Cell-free DNA isolation 17
3.4 Genomic DNA isolation 18
3.5 Cell-free DNA sample pre-screening 18
3.6 Plasma-seq: Whole-genome sequencing of primary tumor and plasma samples 22
3.7 SiMSen-Seq: Simple, Multiplexed, PCR-based barcoding of DNA for Sensitive mutation detection using Sequencing 23
3.8 TCGA data collection and analysis 25
3.10 Gene expression prediction 26

3.11 Cell lines and cell culture 26
3.12 Generation of stable FLT3-overexpressing cell line 28
3.13 siRNA knockdown assays 29
3.14 RNA isolation, quantitative RT-PCR 30
3.15 Whole transcriptome RNA-seq 31
3.16 Colony formation and cell viability assay 32
4 RESULTS 33
4.1 ctDNA can be used to monitor and predict the response of treatment in mCRC patients 33
4.1.1 Emergence of *KRAS* and *ERBB2* focal amplifications in cfDNA coincides with acquired resistance to anti-EGFR treatment 33
4.1.2 Increasing of *CDK6* focal amplification in ctDNA reveals emergence of acquired resistance to lapatinib 35
4.1.3 Establishment of SiMSen-Seq approach 38
4.1.4 Application of SiMSen-Seq for monitoring in mCRC patients 40
4.2 ctDNA can be used to identify novel oncogene in mCRC 43
4.2.1 Recurrent focal events identified by Plasma-Seq 43
4.2.2 The 13q12.13-12.3 amplicon is associated with late-stage clinical features 46
4.2.3 Definition of the minimally 13q12 amplified region and involved genes 50
4.2.4 The oncogene *FLT3* is not associated with proliferation of CRC cells 52
4.2.5 Identification of *POLR1D* as a potential driver gene in 13q12.2 54
4.2.6 POLR1D affects expression of VEGFA and EREG 59
4.2.7 Nucleosome positioning mapping to infer POLR1D expression in plasma 64
4.2.8 The emergence of *POLR1D* amplification correlates with resistance to bevacizumab 66
5 DISCUSSION 72
5.1 ctDNA can be used for monitoring and predicting the response of treatment in mCRC patients 74
5.1.1 Serial monitoring of patients under targeted therapy 74
5.2 ctDNA can be used to identify novel resistance-related genes in mCRC 75
5.2.1 Large-scale data analysis revels the relationship between 13q12.2 SCNA and disease progress 75

5.2.2 The oncogene *FLT3* is not the diver gene of CRC.....75
5.2.3 Identification of *POLR1D* as a potential driver gene in 13q12.2.....76
5.2.4 POLR1D affects the expression of VEGFA and EREG, and is involved in the resistance to bevacizumab.....77
5.3 Limitations.....79
6 CONCLUSION.....80
REFERENCES.....81
SUPPLEMENTARY TABLES.....100

ABBREVIATIONS

°C	Degree Celsius
AML	Acute Myeloid Leukemia
CA19-9	Carbohydrate antigen
CCLE	Cancer Cell Line Encyclopedia
CDK4	Cyclin-dependent Kinase 4
CDK6	Cyclin-dependent Kinase 6
CEA	Carcinoma Embryonic Antigen
cfDNA	Cell-free DNA
Chr	Chromosome
CRC	Colorectal Cancer
CT	Computerized Tomography
ctDNA	Circulating Tumor DNA
Cyclin D	D-type Cyclins
ddPCR	Droplet Digital PCR
DEME	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide Triphosphate
dPCR	Digital PCR
dsDNA	Double-Stranded Deoxyribonucleic acid
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
ERBB2	Receptor tyrosine-protein kinase erbB-2
ERBB4	Receptor tyrosine-protein kinase erbB-4
FDA	Food and Drug Administration
FFPE	Formalin-fixed paraffin-embedded
FLT3	Fms-like tyrosine kinase 3

FOLFIR	FOLinate/leucovorin, Fluorouracil, and IRInotecan
FOLFOX	FOLinic acid/leucovorin, Fluorouracil, and Oxaliplatin
GDC	NCI Genomic Data Commons
GISTIC	Genomic Identification of Significant Targets in Cancer
IMDM	Iscove's Modified Dulbecco's Medium
mCRC	Metastatic Colorectal Cancer
min	Minute
ml	Milliliter
NDR	Nucleosome-depleted Region
ng	Nanogram
NGS	Next-generation Sequencing
OS	Overall Survival
PCR	Polymerase Rhain Reaction
PD	Progressive Disease
PET-CT	Positron Emission Tomography-Computed Tomography
PFS	Progress-free Survival
pg	Picogram
Plasma-Seq	Whole-genome Sequencing of Plasma DNA
PlGF	Placental Growth Factor
POLR1D	DNA-directed RNA polymerases I and III subunit RPAC2
RECIST	Response Evaluation Criteria in Solid Tumors
RNA	Ribonucleic acid
RR	Response Rate
rRNA	Ribosomal ribonucleic acid
RT-PCR	Real-time Polymerase Chain Reaction
SCNA	Somatic Copy Number Alteration
SD	Stable Disease
SiMSenSeq	Simple, Multiplexed, PCR-based barcoding of DNA for Sensitive mutation detection using Sequencing
SLE	Systemic Lupus Erythematosus

SVM	Support Vector Machine
TCGA	The Cancer Genome Atlas
TE	Tris-EDTA
TF	Tumor fraction
TFBS	Transcription Factor Binding Sites
TSS	Transcription Start Site
U	Unit
UMI	Unique Molecular Identifier
VEGF	Vascular Endothelial Growth Factor
VEGFA	Vascular Endothelial Growth Factor A
VEGFR	VEGF receptor
µg	Microgram
µl	Microliter
µM	Micromolar

LIST OF FIGURES

Figure 1: The EGFR pathway and mechanisms of resistance to anti-EGFR therapies 4
Figure 2: The ERBB2 pathway and mechanisms of resistance to anti-ERBB2 therapies 6
Figure 3: VEGF subtypes, corresponding VEGFR, and anti- angiogenesis drugs 7
Figure 4: Circulating tumor DNA could be released into the blood through cell necrosis, apoptosis and active secretion 11
Figure 5: Inter-tumor and intra-tumor heterogeneity in cancer 13
Figure 6: Read Coverage of transcription start sites in plasma DNA predict the gene expression 15
Figure 7: Whole genome profiling of OXCO-2, HT29, and SW480 cell lines 27
Figure 8: Emergence of *KRAS* and *ERBB2* amplification during anti-EGFR treatment in patients C17 and C129 34
Figure 9: Changes of the *CDK6* and *ERBB2* amplification under lapatinib treatment in patient C79 36
Figure 10: Summary of SiMSen-Seq sensitivity test results 39
Figure 11: Change of *TP53* deletion and mutation under regorafenib and lonsurf treatment in patient C128 41
Figure 12: Summary of focal SCNA events identified in our and TCGA cohort 44
Figure 13: The 13q12.13-12.3 amplicon is associated with late-stage clinical features 48
Figure 14: Definition of the minimally 13q12 amplified region 51
Figure 15: The oncogene *FLT3* is not associated with proliferation of CRC cells 53
Figure 16: Validation of the relationship between gene copy number and gene expression in all candidate genes in TCGA and CCLE dataset 55
Figure 17: Silencing of *POLR1D* inhibited cell viability in HT29 and SW480 cells 57
Figure 18: POLR1D affects expression of VEGFA and EREG 60
Figure 19: POLR1D affects expression of PPP1R15A, MOSPD2, GARS, FAM84B, and KIF21B 62
Figure 20: TSS-nucleosome occupancy patterns predict the expression of 13q12.2 amplicon located genes 65

Figure 21: Emergence of the 13q12 amplicon under bevacizumab treatment in patient C216 68

Figure 22: Alternating *POLR1D* and *ERBB2* amplifications in serial plasma analyses of patient C129 70

LIST OF TABLES

Table 1: Summary of LINE-1 specific and sample specific primers 21

Table 2: Summary of SiMSen-Seq Target specific primers 25

Table 3: Summary of all siRNA used and the knockdown efficiency in HT29 and SW480 cell lines 29

Table 4: Summary of all qPCR primers used and the corresponding PCR protocols 31

Table 5: Summary of clinical information of all cases harboring 13q12.2 focal amplification 47

Table 6: Summary of genes differently expressed after *POLR1D* knockdown in both HT29 and SW480 cells and which showed consistent results in the TCGA dataset 61

ZUSAMMENFASSUNG

Darmkrebs ist weltweit die dritthäufigste diagnostizierte bösartige Erkrankung. Die Mortalität von Darmkrebs ist in den letzten Jahrzehnten aufgrund verbesserter Diagnoseverfahren, die die Identifizierung von Patienten im Frühstadium der Erkrankung ermöglichen, zurückgegangen. Das Überleben von Patienten mit fortgeschrittenen Krebsstadien ist jedoch immer noch unbefriedigend. In jüngster Zeit haben gezielte Therapien, z.B. anti-EGFR- und anti-VEGF-Behandlungen, bei Patienten mit Darmkrebs im Spätstadium vielversprechende Ergebnisse gezeigt. Allerdings ist der Überlebensvorteil gezielter Therapien in der Regel auf wenige Monate begrenzt, und es kommt häufig zu erworbenen Resistenzen. Die sogenannte „liquid biopsy“, die sich auf die Analyse von Charakteristika eines Tumors aus einer Körperflüssigkeit wie Blut bezieht, ist zu einer vielversprechenden Methode für die Frühdiagnose und die Behandlung von Patientinnen und Patienten geworden. In dem in dieser Arbeit beschriebenen Projekt haben wir die Sequenzierung des gesamten Genoms der Plasma-DNA, die bei Personen mit Krebs zirkulierende Tumor-DNA (ctDNA) enthält, eingesetzt, um das Tumorgenom von Betroffenen mit Darmkrebs (CRC) zu bewerten. Ein Schwerpunkt lag auf den Fällen, die eine gezielte Behandlung erhielten, um die Verwendung von ctDNA für erworbene Resistenz und die Vorhersage des Patientenergebnisses zu validieren und auch um neue Resistenzmechanismen zu bestimmen.

Mit Hilfe von longitudinalen Plasmaanalysen untersuchten wir die Entwicklung von Tumorgenomen in einer Kohorte mit metastasiertem Darmkrebs (CRC). Interessanterweise entwickelten zwei Fälle unter Anti-EGFR-Behandlung eine fokale Amplifikation in resistenzbezogenen Genen (*KRAS* und *ERBB2*), als die Krankheit fortschritt. Darüber hinaus wurde bei einem Patienten unter Anti-EGFR2-Behandlung eine resistenzbezogene fokale Amplifikation in *CDK6* gefunden. Diese Ergebnisse legten nahe, dass die liquid biopsy ein vielversprechendes Verfahren zur Identifizierung neu auftretender fokaler Amplifikationen sein kann, die dann Hinweise für Mechanismen für erworbene Resistenzen gegen eine verabreichte Therapie sein können. Zusätzlich zu fokalen Amplifikationen mit etablierten sogenannten „driver Genen“ (Treibergenen) identifizierten wir auch wiederkehrende fokale Amplifikationen, die noch nicht im Detail untersucht worden waren, wie z.B. ein Amplikon auf dem Chromosom 13q12.2.

Durch die Analyse von CRC-Fällen aus der TCGA-Datenbank bestätigten wir, dass die Amplifikation von 13q12.2 mit fortgeschrittenen Tumorstadien assoziiert war. Wir definierten zunächst die minimal amplifizierte Region und testeten dann die Funktion aller enthaltenen Gene mit zwei in vitro-Zellmodellen. Wir fanden Hinweise darauf, dass das Gen *POLR1D* das potenzielle driver Gen innerhalb des Amplikons war, das die Zellproliferation durch die Induktion der Expression einiger Onkogene (z.B. *VEGFA* und *EREG*) beeinflusst. Die Hochregulierung von *VEGFA*, einem wesentlichen Regulator der Angiogenese, wurde für die Resistenz gegen die Behandlung mit Bevacizumab ursächlich verantwortlich gemacht. Interessanterweise beobachteten wir bei der seriellen Überwachung von zwei Patienten unter Bevacizumab-Behandlung die Entstehung der 13q12.2-Amplifikation zum Zeitpunkt der Resistenzentwicklung gegen diese Therapie. Somit deuteten alle Ergebnisse darauf hin, dass *POLR1D* das potenzielle driver Gen in 13q12.2 ist und eine Rolle bei der Bevacizumab-Resistenz spielt.

Zusammenfassend haben wir bestätigt, dass die liquid biopsy als nicht-invasive Methode zur Vorhersage der erworbenen Resistenz und des Therapieansprechens in der Klinik anwendbar ist.

ABSTRACT

Colorectal cancer is the third most commonly diagnosed malignancy worldwide. The mortality of colorectal cancer has been decreasing in recent decades due to improved diagnostic procedures, which allow the identification of patients with early-stage disease. However, the survival of patients with advanced diseases is still unsatisfactory. Recently, targeted therapies, e.g., anti-EGFR and anti-VEGF treatments, have shown promising results in late-stage colorectal cancer patients. However, the survival benefit of targeted therapies is usually limited to a few months, and acquired resistance occurs frequently. Liquid biopsy, which refers to the analysis of tumor content from a body fluid such as blood, has become a promising method for early diagnosis and the management of patients. In the project described in this thesis, we employed whole-genome sequencing of plasma DNA, which in patients with cancer contains circulating tumor DNA (ctDNA) to evaluate the tumor genome of patients with colorectal cancer (CRC). A focus was on those patients who received targeted treatment, to validate the usage of ctDNA for acquired resistance and patient outcome prediction, and also to determine novel resistance mechanisms.

Using longitudinal plasma analyses, we studied the evolution of tumor genomes in a metastatic colorectal cancer cohort. Interestingly, two patients under anti-EGFR treatment showed focal amplification in resistance-related genes (i.e., *KRAS* and *ERBB2*) when the disease progressed. Moreover, focal amplification in *CDK6* (resistance-related) was found in one patient under anti-ERBB2 treatment. These results suggested that liquid biopsy is a promising tool for identification of emerging focal amplifications, which may be associated with acquired resistance prediction. In addition to focal amplifications with established driver genes, we also identified recurrent focal amplifications, which had not yet been investigated in detail, such as an amplicon on chromosome 13q12.2. By analysis of CRC cases from the TCGA database, we confirmed that amplification of 13q12.2 was associated with more advanced stages. We first defined the minimally amplified region and then tested the function of all genes included with two *in vitro* cell models. We found evidence for *POLR1D* being the potential driver gene within the amplicon, which impacted cell proliferation by inducing the expression of some oncogenes (e.g., *VEGFA* and *EREG*). The upregulation of VEGFA, an essential regulator of angiogenesis, has been implicated in the

resistance to bevacizumab treatment. Interestingly, by serial monitoring of two patients under bevacizumab treatment, we observed the emergence of the 13q12.2 amplification when the respective tumors developed resistance to the therapy. All the results indicated that *POLR1D* is the potential driver gene in 13q12.2 and plays a role in bevacizumab resistance.

To this end, we confirmed that liquid biopsy as a non-invasive method applicable for acquired resistance and patient outcome prediction in the clinic. Liquid biopsy is also a useful tool in cancer research, which allows the identification of novel driver genes.

1 INTRODUCTION

1.1 Colorectal Cancer

Colorectal cancer (CRC) remains a large global health problem, representing the third most commonly diagnosed malignancy worldwide and one of the major causes of morbidity and mortality across populations (Siegel, Miller & Jemal, 2019). Although the incidence and mortality rate of CRC have been decreasing over the past decades, its burden is projected to increase by 60% by the year 2030, with an estimated 2.2 million new cases and 1.1 million deaths (Ferlay et al., 2013, Arnold et al., 2017). Early diagnosis and surgical removal are still the only curative approach for CRC patients. However, making patient diagnoses in early stages is still a challenge; approximately 30% of patients already present with distant metastases at the time of diagnosis (Saltz et al., 2004). Even when diagnosed at an early stage, 50% of patients will still develop metastases and progress to a later stage (DeSantis et al., 2014). Therapeutic options for metastatic colorectal cancer (mCRC) are limited and in very advanced stages, only palliative approaches may remain to improve the quality of life for the patient.

1.1.1 Diagnosis and monitoring

Carcinoma embryonic antigen (CEA), a high molecular weight glycoprotein, is a widely used tumor marker in CRC patients discovered by Gold and Freedman in 1965 (Tong et al., 2018). CEA is expressed by large intestine cells and functions as an intercellular adhesion molecule. It has been reported that serum CEA levels are informative of tumor stage initial diagnosis in CRC patients (Stojkovic Lalosevic et al., 2017). CEA serum levels are used as a prognostic biomarker to monitor and predict potential disease progression in CRC. It has been published that increasing CEA levels after surgery is a predictive marker for recurrence (Irvine, Scott & Clark, 2007). CEA levels also highly correlate with tumor burden; larger tumor size is typically associated with higher CEA serum concentrations. Thus, abnormal CEA levels are rarely found in patients with tumor sizes smaller than 3 cm (Kannagi et al., 2004). Although it has been reported that increased concentrations of CEA in early-stage cancer indicate more progressive disease, CEA is rarely used in the early diagnosis of CRC (Stojkovic Lalosevic et al., 2017). Due to the low levels of CEA

concentration, CEA changes in early-stage disease are hard to distinguish with many non-neoplastic conditions, e.g., inflammatory bowel disease, pancreatitis, chronic obstructive pulmonary disease, hepatitis, liver cirrhosis and hypothyroidism (Vukobrat-Bijedic et al., 2013).

Carbohydrate antigen (CA19-9), a high molecular weight glycoprotein, is a cancer antigen whose serum concentration may be elevated in metastatic colon cancer patients (Vukobrat-Bijedic et al., 2013). Although CA19-9 is not a specific marker for CRC, it has been reported that the assessment of CA19-9, together with CEA, can increase the diagnostic sensitivity (Stiksma, Grootendorst & van der Linden, P W, 2014). CEA and CA19-9 are both biomarkers for late-stage disease and can be used as a prognostic marker to predict the risk of metastasis (Stojkovic Lalosevic et al., 2017) or to monitor mCRC patient response (Gao et al., 2018).

In addition to serum biomarkers, imaging also plays an important role in CRC patient screening and monitoring. Computed tomography (CT) is well-known to be used in the management of CRC patients, e.g., pre-treatment staging, detection of distant metastasis, and assessing patient response (Tan, Iyer, 2010). Recent advances in CT technologies such as positron emission tomography-computed tomography (PET-CT) have increased the staging accuracy but do not provide additional survival benefit (Sobhani et al., 2018). According to the American Cancer Society (ACS) guidelines, a five-year CT follow-up is recommended for asymptomatic patients with average risk (Levin et al., 2003).

In summary, both serum biomarkers and CT scans are useful tools for patient management for both patients undergoing curative surgery or palliative treatment. The CT scan is one of the recommended tools for patient screening. Further technologies may increase the detection rate of patients in early stages, which in turn may improve the outcome of CRC patients.

1.1.2 CRC treatment

CRC treatment recommendations depend on histology, stage and the overall condition of the patient. Currently, surgery is still the most frequently recommended treatment for early-stage patients, but even for metastatic hepatic lesions, a potentially curative approach through

metastasectomies exists (Reissfelder et al., 2011). For advanced patients, chemotherapy is frequently used, such as FOLFOX (folinic acid + fluorouracil (5FU) + oxaliplatin) and FOLFIRI (folinic acid + fluorouracil (5FU) + irinotecan) (Gustavsson et al., 2015). Targeted therapy is also a commonly used approach. Currently, anti-EGFR and anti-VEGF treatments are commonly applied to suitable patients in combination with chemotherapy (Seeber, Gastl, 2016). Anti-ERBB2 is frequently used in breast cancer patients and has shown some potential in CRC patients who harbor ERBB2 amplification (Seeber, Gastl, 2016).

1.1.3 Anti-EGFR

Epidermal growth factor receptor (EGFR) is a member of the ErbB family of receptors, which also includes HER2/neu (ERBB2), HER3 (ErbB3), and HER4 (ErbB4) (Yarden, Sliwkowski, 2001). EGFR can be activated by members of the epidermal growth factor family (EGF family), which activates subset signaling cascades such as RAS-RAF-MEK-ERK and the PI3K-AKT-mTOR pathway (Hynes, Lane, 2005). Activation of such pathways subsequently stimulate key processes involved in oncogenesis and tumor progression, including tumor cell proliferation, angiogenesis, and migration (Figure 1) (Ciardiello, Tortora, 2001).

Commonly used EGFR inhibitors include monoclonal antibodies against EGFR (e.g., cetuximab and panitumumab) and small molecular drugs, which inhibit the EGFR tyrosine kinase domain (e.g., gefitinib and erlotinib) (Dutta, Maity, 2007). Cetuximab, in combination with FOLFIRI, has been reported to improve overall survival (OS), progression-free survival (PFS), and patient response rate (RR) in *KRAS* and *NRAS* wild type mCRC patients (Sotelo et al., 2014). Cetuximab, in combination with FOLFOX, had similar results except for the OS (Sotelo et al., 2014). Similar results have been published for panitumumab when used in combination with FOLFOX or FOLFIRI (Ketzer et al., 2018). Currently, cetuximab or panitumumab in combination with FOLFOX or FOLFIRI is the gold standard in first-line mCRC care (Ohhara et al., 2016).

The most common mechanism of resistance to anti-EGFR therapy is the *KRAS* mutation, which activates the RAS-RAF-MAF pathway without EGFR activation (Figure 1) (Leto, Trusolino, 2014). Similarly, *BRAF, PIK3CA*, and *PTEN* mutations all were reported to be involved in the resistance

to anti-EGFR treatment by activating the pathway (Figure 1) (Leto, Trusolino, 2014). *EGFR* mutations are also a well-known mechanism of resistance to monoclonal antibodies against EGFR, which can be overcome by applying small molecular drugs (Figure 1) (Montagut et al., 2012). Amplification of *ERBB2* or mutations in the other ErbB family members also were reported as potential resistance mechanisms that create hyperactive bypass pathways (Figure 1) (Chong, Janne, 2013).

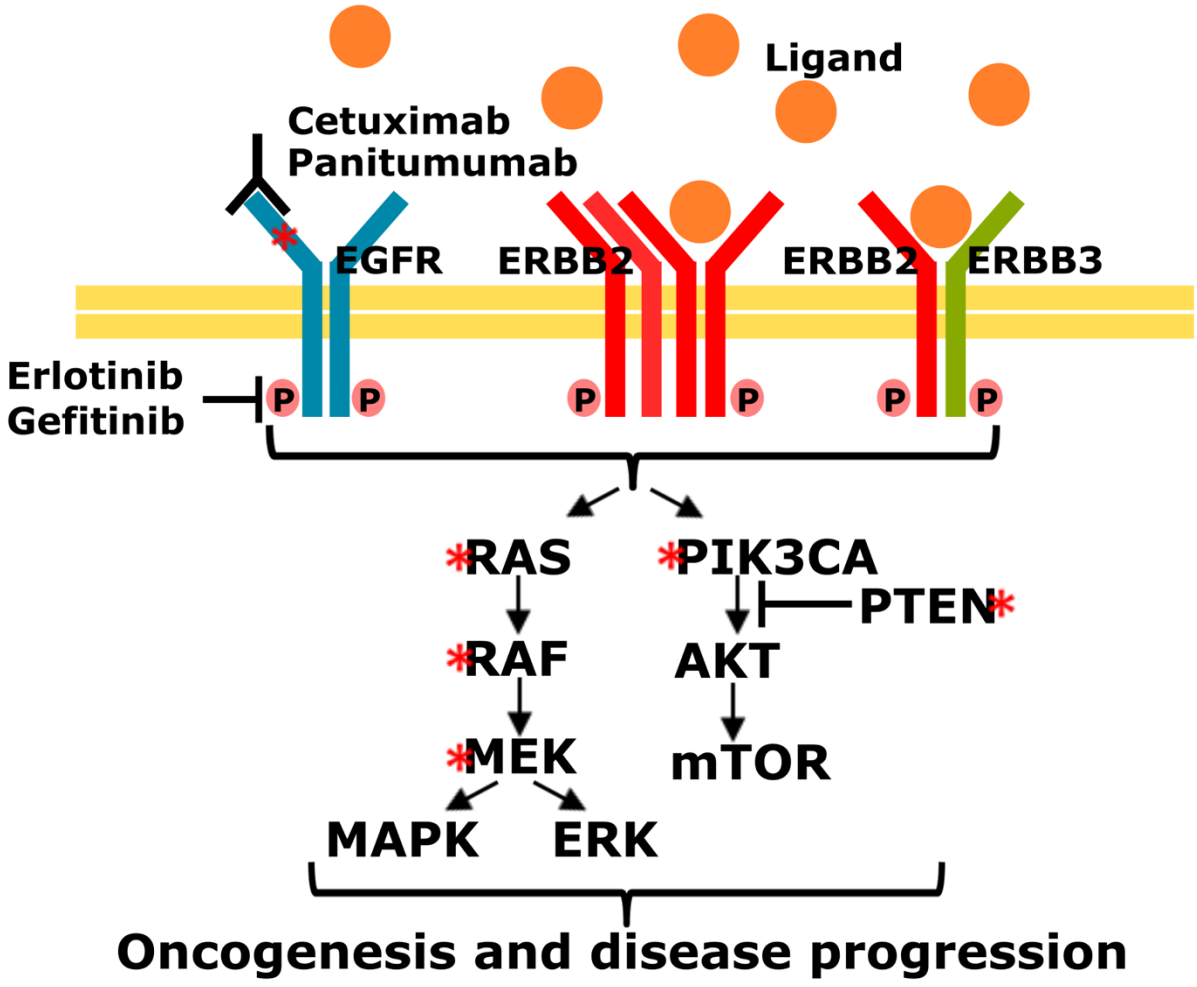


Figure 1: The EGFR pathway and mechanisms of resistance to anti-EGFR therapies (Red star represents for the potential mutation sites involved in resistance.)

1.1.4 Anti-ERBB2

As described above, *HER2/neu* (*ERBB2*) is also a member of the ErbB family of receptors. The activation mechanism and following signaling cascade are highly similar between *ERBB2* and *EGFR*, which include activation of the RAS-RAF-MEK-ERK pathway (Figure 2) (Feng et al., 2018). ERBB2-related pathways bypass the EGFR pathway, which has been identified as a potential mechanism of resistance for anti-EGFR therapy. Moreover, it has been reported that *ERBB2* amplification was more frequent in *KRAS* and *BRAF* wild type mCRC (Ross et al., 2018). Currently, anti-ERBB2 treatment has been widely used for HER2-positive breast cancer and the two most frequently used drugs are trastuzumab (a monoclonal antibody against ERBB2) and lapatinib (a small molecule drug inhibits the ERBB2 tyrosine kinase domain) (Jiang, N. et al., 2018). To date, evidence for a potentially beneficial role of anti-ERBB2 therapy in the treatment of ERBB2-positive mCRC patients in a small population has been reported (Greally, Kelly & Cercek, 2018). However, in order to be approved by the FDA, additional and larger clinical trials are still needed.

The most common resistance mechanisms of anti-ERBB2 therapy are mutations in the downstream pathway caused by *PTEN* deletions or *PIK3CA* mutations (Nishimura et al., 2013). Recently, mutations in *CDK4* (encodes cyclin-dependent kinase 4) and *CDK6* (encodes cyclin-dependent kinase 6) have been reported to be involved in acquired resistance to anti-ERBB2 therapy (Niu, Xu & Sun, 2019). By binding with D-type cyclins (Cyclin D), CDK4/6 becomes activated and can subsequently deactivate pRb, which promotes cell progression from the G1 to S phase of the cell cycle. Mutations in CDK4/6 lead to consistent activation of the Cyclin D – CDK4/6 complex, which then promotes cell cycle and rapid cell proliferation (Figure 2) (Portman et al., 2019). Recent publications have shown some potential benefits by treating anti-ERBB2 treatment-resistant tumors with CDK4/6 inhibitors (Niu, Xu & Sun, 2019).

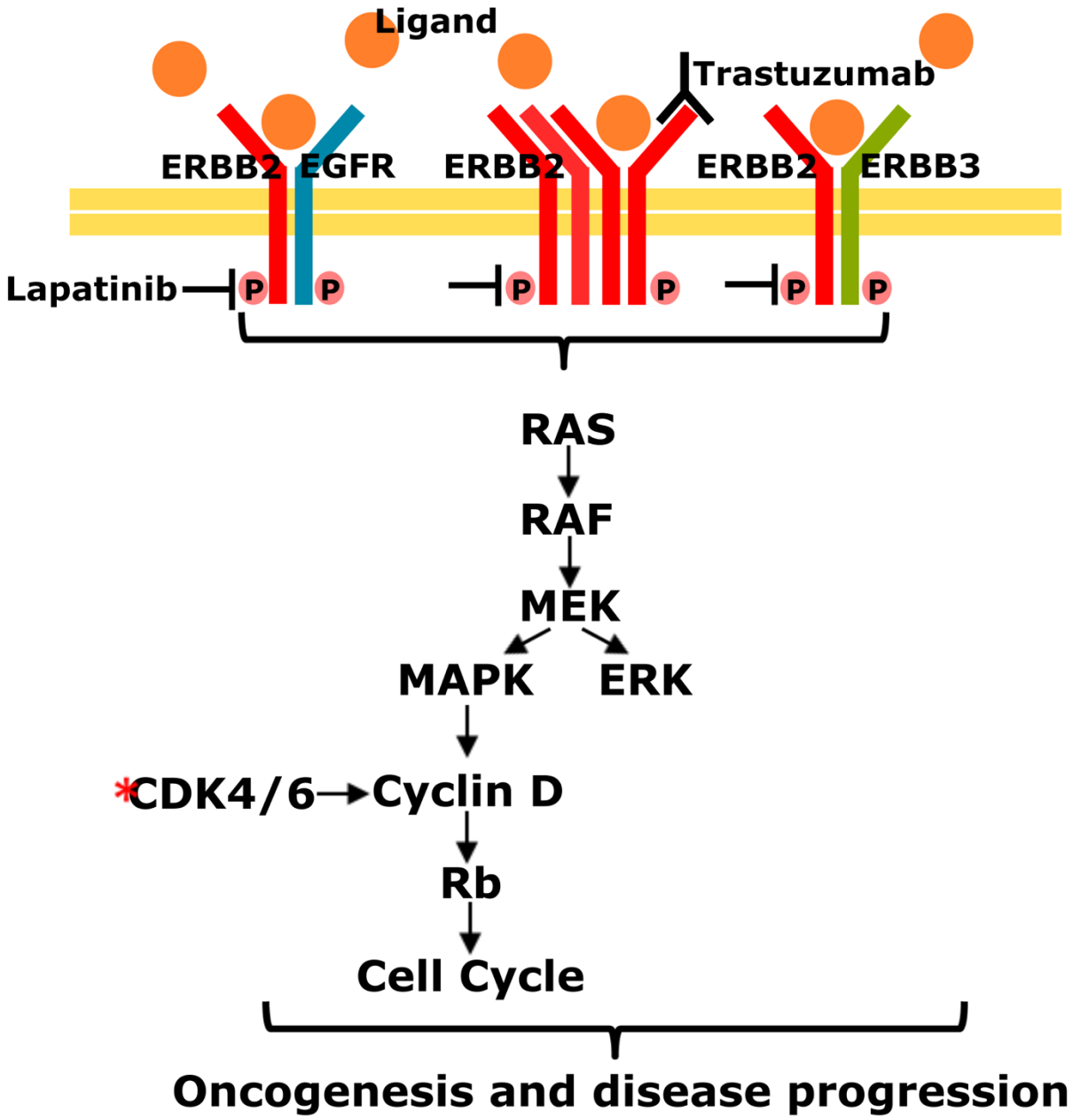


Figure 2: The ERBB2 pathway and mechanisms of resistance to anti-ERBB2 therapies (Red star represents for the potential mutation sites involved in resistance.)

1.1.5 Anti-angiogenesis

Vascular endothelial growth factor (VEGF) is a group of growth factors including VEGF-A, VEGF-B, VEGF-C, VEGF-D and placental growth factor (PlGF) (Figure 3) (Holmes, Zachary, 2005). VEGFA-A is a dominant isoform that has shown the highest expression and biological activity. VEGF-A can bind and activate VEGF receptor-1 (VEGFR-1) and VEGFR-2, which in turn stimulates angiogenesis (Figure 3) (Saif, 2013). A pro-angiogenic effect of VEGF-A in multiple tissues, including different types of tumor tissues, e.g. CRC, lung cancer, and breast cancer, was recently shown (Maj, Papiernik & Wietrzyk, 2016). VEGF-A also hase been reported to stimulate endothelial cell mitogenesis and cell migration (Herzog et al., 2011), which may play a role in tumor invasion and metastasis.

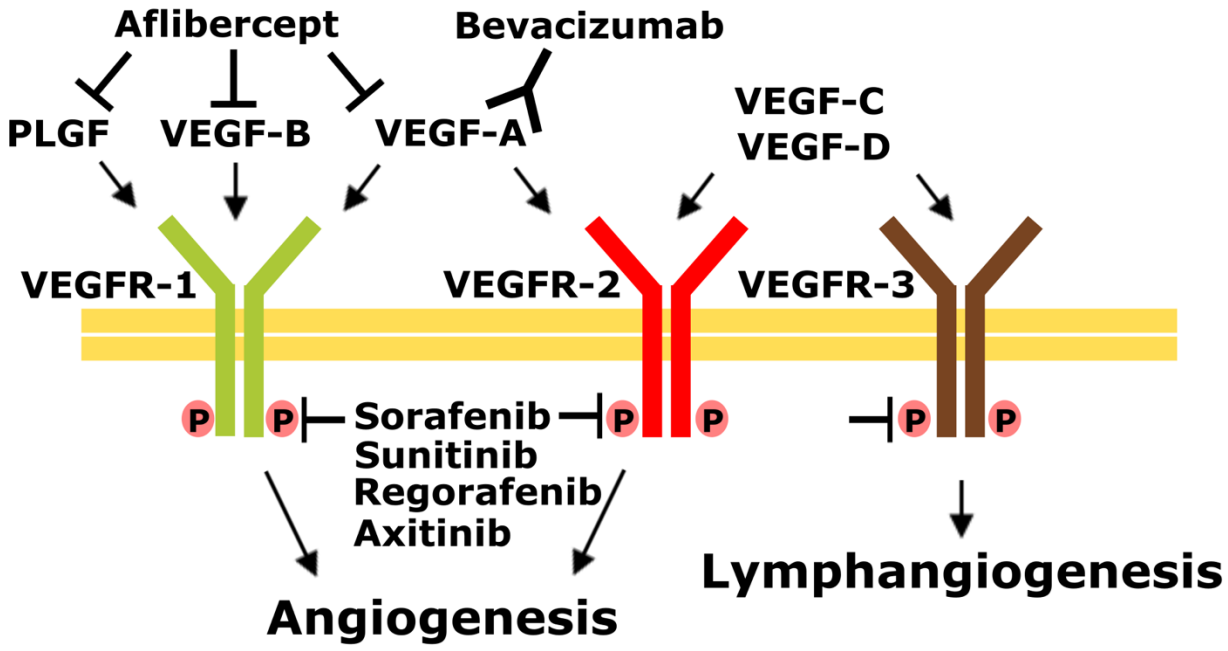


Figure 3: VEGF subtypes, corresponding VEGFR, and anti- angiogenesis drugs

Recent publications have shown a negative correlation between VEGF-A expression and patient survival in tumors overexpressing VEGF-A (Hegde et al., 2013). Hence, VEGF-A is a potential target for the treatment of cancer. Bevacizumab, a monoclonal antibody that can inhibit all isoforms of VEGF-A, is the first anti-angiogenesis drug approved by FDA for use in mCRC (Kramer, Lipp, 2007). Numerous studies have demonstrated improvements in patient survival in mCRC patients treated with bevacizumab in combination with chemotherapy in the first-line setting. For instance, patients receiving bevacizumab in combination with chemotherapy demonstrated better survival than the placebo cohort (median OS: 20.3 vs. 15.6 months, $P < 0.01$; median PFS: 10.6 vs. 6.2, $P < 0.01$) (Hurwitz et al., 2005). Moreover, patients treated with bevacizumab in combination with chemotherapy experienced longer survival compared to patients treated with chemotherapy alone (median OS: 18.7 vs. 16.1 months, $P < 0.01$; median PFS: 8.8 vs. 6.4, $P < 0.01$) (Hurwitz et al., 2013). Even after progression after first-line bevacizumab treatment, patients still benefit from continued bevacizumab therapy (Grothey et al., 2008). Hence, bevacizumab is currently widely used as a first- or second-line treatment for mCRC patients in combination with chemotherapy. However, the survival benefit of bevacizumab in mCRC patients is limited to a few months and acquired resistance mechanisms are largely unknown. Recently, alternatively spliced variants of VEGF-A were found and some of them demonstrated antiangiogenic activation (Hilmi, Guyot & Pages, 2012). In this case, a more selective antibody might be needed to prevent the inhibition of those VEGF-A variants. Moreover, bevacizumab may reduce tumor perfusion and oxygenation, which may lead to difficulty of drug distribution (Van der Veldt, A A et al., 2012). Bevacizumab treatment was recently reported to induce autocrine VEGF signaling, which results in upregulation of VEGF-A expression (Fan et al., 2011, Mesange et al., 2014). Higher VEGF-A levels subsequently limited the efficacy of bevacizumab.

In contrast to VEGF-A, VEGF-B only interacts with VEGFR-1 and plays a less important role in tumor angiogenesis, which may only involve the maintenance of newly formed blood vessels (Olofsson et al., 1998). VEGF-C and VEGF-D both interact with VEGFR-2 and VEGFR-3 and impact both angiogenesis and lymphangiogenesis (Saif, 2013). PlGF also interacts with VEGFR-1 and may be involved in pathological angiogenesis (Saif, 2013). Recently, drugs targeting the VEGFR tyrosine kinase domain, e.g., regorafenib, sorafenib, and sunitinib, also were applied to

mCRC patients and showed some promising results. One recent study indicates that regorafenib-treated patients demonstrated better survival compared with the placebo group (median OS: 6.4 vs. 5.0 months, $P < 0.01$; median PFS: 1.9 vs. 1.7, $P < 0.01$) (Grothey et al., 2013). Similar results were found in another recent study (regorafenib vs placebo; median OS: 8.8 vs. 6.3 months, $P < 0.01$; median PFS: 3.2 vs. 1.7, $P < 0.01$) (Li et al., 2015).

In summary, anti-angiogenesis therapy, especially bevacizumab, is widely used in the clinic in both first- and second-line settings. Patients have shown significant survival benefits, but this was limited to only several months. The resistance mechanism of anti-angiogenesis therapy is largely unknown and thus a predictive biomarker for bevacizumab efficacy is lacking.

1.2 Cell-free DNA

Cell-free DNA (cfDNA) is DNA released from cells to the blood or other body fluids, e.g., urine and cerebrospinal fluid (Heitzer et al., 2019). cfDNA was first described in 1948 by Mandel and Metais, who found the presence of cfDNA in blood from both healthy and sick humans (MANDEL, METAIS, 1948). Circulating tumor DNA (ctDNA) is a portion of cfDNA released from tumor cells. Stroun and colleagues suggested in 1989 that cfDNA from cancer patients may contain characteristics of tumor DNA (Stroun et al., 1989). This suggestion was indeed confirmed when the presence of ctDNA was verified in 1994. Two research groups reported that the presence of *NRAS* and *KRAS* mutation DNA fragments in cfDNA from acute myeloid leukemia and pancreatic cancer, respectively (Sorenson et al., 1994, Vasioukhin et al., 1994). In 2008, Ellinger et al. described the difference in cfDNA size distribution between bladder cancer and benign prostate hyperplasia, i.e. smaller (124bp) and larger (271bp) fragments were present in the cancer patients (Ellinger et al., 2008). With this evidence, researchers realized that cfDNA comprises tumor and non-tumor DNA in the blood. With the decreasing cost of next-generation sequencing, it is possible to specifically analyze ctDNA to monitor patient response and to evaluate disease progression. For instance, in 2009, one study reported that analysis of cfDNA can be used to detect the appearance of micro-metastasis lesions in cancer patients (Schwarzenbach et al., 2009). Moreover, it has also been reported that analysis of ctDNA might reveal cancer-related genetic and epigenetic alterations that are related to cancer progression (Schwarzenbach, Hoon & Pantel, 2011) .

1.2.1 Biology of Cell-free DNA

It is still not clear how cfDNA is released into the blood and how it is removed. One possible explanation is that the release of cfDNA happens during cellular turnover, e.g. cell apoptosis and necrosis (Figure 4) (Hashad et al., 2012, Delgado et al., 2013, Park et al., 2012, Schwarzenbach, 2013, Schwarzenbach, 2012, Schwarzenbach, Hoon & Pantel, 2011). Necrotic cells or apoptotic cells are engulfed by macrophages, causing the release of DNA fragments into the blood during the activation or death of macrophages. Another explanation is that cells actively release DNA into the blood (Figure 4) (Stroun et al., 2001). It has been reported from multiple studies that tumor cells can release DNA into the circulation through exosomes (Kahlert et al., 2014, Thakur et al.,

2014). The study of circulating exosomal DNA in cancer patients is another promising area of liquid biopsy. As different methods of release lead to different digestion processes of DNA, studying cfDNA size distribution is a way to investigate the source of cfDNA in more detail. It has been reported in multiple studies that the average fragment size of cfDNA is around 160bp to 180bp (Heitzer, Auer et al., 2013, Chan, K. C. et al., 2004, Mouliere et al., 2011), which corresponds to DNA wrapped around a nucleosome (around 142bp) plus one linker (around 20bp). This fragment size is closer to the DNA size released from apoptotic cells. In a previous study, in addition to those around 160-180bp, we also reported the detection of cfDNA with an average size of around 320-360bp (Heitzer et al., 2013). However, these DNA fragments are more likely from di-nucleosomes rather than necrosis, which leads to DNA fragments size longer than 10,000bp. In summary, much of the evidence indicates that apoptosis may be the main source of cfDNA.

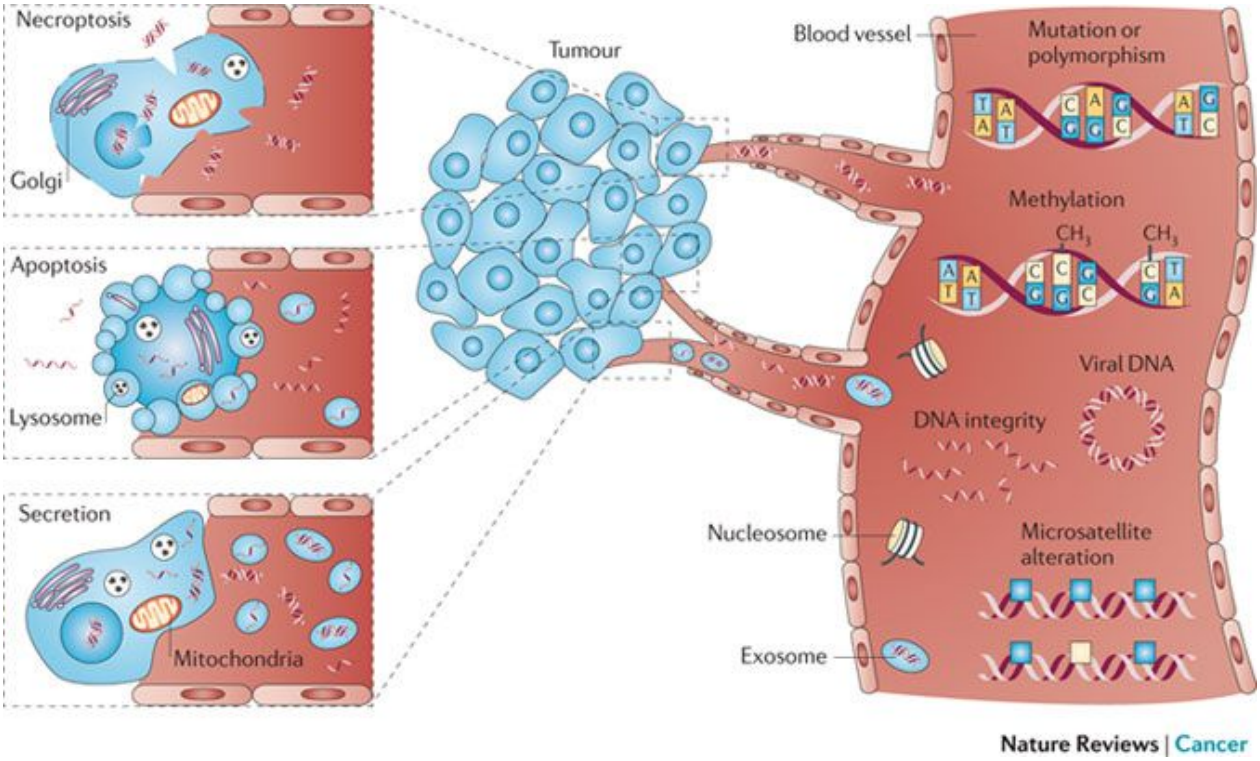


Figure 4: Circulating tumor DNA could be released into the blood through cell necrosis, apoptosis and active secretion (Schwarzenbach, Hoon & Pantel, 2011) (Permission for reuse of this figure has been granted from the publisher (Springer Nature))

The presence of cfDNA in the blood not only depends on release but is also highly related to degradation and clearance. One early study reported that cfDNA is cleared by the liver and kidney, which leads to a half-life of 16 minutes to several hours (Lo et al., 1999). Another study described that cfDNA has two clearance phases, i.e., a rapid phase (half-life around 1 hour) followed by a second phase (half-life around 23 hours) (Yu, S. C. et al., 2013). Different forms of cfDNA have different survival times, e.g. single-stranded DNA is digested more quickly than double-stranded DNA, and DNA protected by nucleosomes is more difficult to digest than naked DNA. To date, how cfDNA is cleared and which enzymes are involved in this process is still unknown. One recent study reported that knockout of DNase1-like 3 in mice leads to an increase in larger fragments of cfDNA and a decrease in small fragments, suggesting that DNase1-like 3 may be one of the enzymes involved in cfDNA degradation (Serpas et al., 2019). Moreover, the presence of anti-DNA antibodies in systemic lupus erythematosus (SLE) may lead to an enrichment of small cfDNA fragments (Chan, R. W. et al., 2014), which suggests the involvement of immune response in the clearance of cfDNA.

1.2.2 Circulating tumor DNA

As described before, cfDNA corresponds to DNA released from both non-tumor and tumor cells (i.e. ctDNA) in cancer patients. In other words, ctDNA is a fraction of cfDNA which originated from tumor cells. As a result, ctDNA contains some different biologic features compared to cfDNA released from healthy cells. Firstly, several studies conducting cfDNA size profiling have shown that ctDNA from cancer cells is shorter than cfDNA released from non-cancerous cells (Jiang, P. et al., 2015, Underhill et al., 2016, Jiang, P., Lo, 2016). Secondly, as ctDNA can be released from the primary tumor and metastatic lesions, ctDNA usually contains almost all genetic and epigenetic alterations from a tumor that can be used to monitor the disease (Diaz, Bardelli, 2014). Currently, tissue biopsies are still considered the gold standard for cancer diagnosis and monitoring. However, limitations of tissue biopsies include the invasiveness requiring surgery or needle biopsies. Tissue biopsies are only applicable if the location of the tumor is accessible by needle biopsy and moreover, tissue biopsies are usually only taken from one location, which may not represent the whole picture of the disease. It is increasingly recognized that primary tumor and metastatic lesions from different locations contain inter-tumor heterogeneity (Figure 5) (Burrell, Swanton, 2014). Even within the

same lesion, tumors may still contain a high degree of genetic and molecular heterogeneity, which is referred to as intra-tumor heterogeneity (Figure 5) (Burrell, Swanton, 2014). ctDNA, as mentioned before, contains genetic information from almost all lesions (Murtaza et al., 2015) and can be collected in a non-invasive way, providing a potential way to overcome the limitations of tissue biopsy.

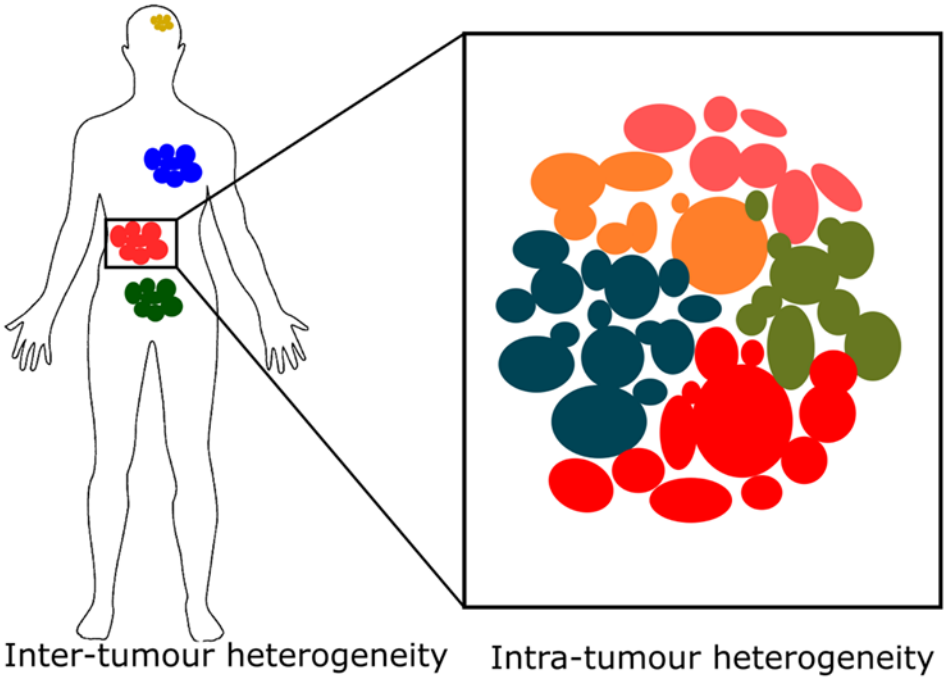


Figure 5: Inter-tumor and intra-tumor heterogeneity in cancer

Lastly, due to the short turnover time, ctDNA released into the blood provides real-time information about the disease. Recent studies have indicated that the amount of ctDNA in circulation correlates with tumor burden and therapeutic response (Wu et al., 2002, Forshew et al., 2012, Diaz et al., 2012). Moreover, some studies have found that the appearance of resistance mutations in ctDNA might be a predictive biomarker for patient response, especially for targeted therapy (Ulz, Belic et al., 2016, Mohan et al., 2014, Chabon et al., 2016).

1.2.3 ctDNA and Colorectal Cancer

ctDNA has shown promising clinical utility in the management of CRC, ranging from applications in early detection (Lofton-Day et al., 2008, Phallen et al., 2017, Bettegowda et al., 2014, Mead et al., 2011, Bedin et al., 2017, Cohen et al., 2018, Diehl et al., 2005), detection of relapse (Scholer et al., 2017, Tie et al., 2016, Reinert et al., 2019), identification of prognostic markers (El Messaoudi et al., 2016), molecular characterization of metastatic disease (Onstenk et al., 2016), as well as tracking response to therapy (Misale et al., 2012, Barault et al., 2018, Tie et al., 2015, Diehl et al., 2008, Diaz et al., 2012). For a detailed review, see Marcuello et al. (2019) (Marcuello et al., 2019). The analysis of ctDNA for specific gene mutations or somatic copy number alterations (SCNAs), e.g. *KRAS*, *BRAF*, *PIK3CA* and *TP53*, has shown promising results for early detection (Phallen et al., 2017, Mead et al., 2011) and response prediction (Misale et al., 2012, Tie et al., 2015, Diaz et al., 2012, Mohan et al., 2014) in CRC patients. Monitoring disease courses using specific mutations and SCNAs in ctDNA is one of the leading applications in liquid biopsy, which strengthens the role of ctDNA in the clinical management and understanding of disease progression in CRC. Recently, methylation changes in ctDNA have shown some promising results in CRC management. For instance, one study reported that analysis of methylation of *TMEFF2*, *NGFR* and *SEPT9* could be used for CRC screening (Lofton-Day et al., 2008). Moreover, another study indicated that monitoring hypermethylated regions within *BCAT1/IKZF1* allowed the prediction of CRC recurrence (Symonds et al., 2020).

ctDNA not only provides genetic or epigenetic information of the tumor, but also provides some information about gene expression. As described above, DNA protected by nucleosome is more difficult to be digested than naked DNA. During gene expression, nucleosomes need to be removed from the transcript start sites (TSSs), resulting in a nucleosome-depleted region (Figure 6) (Murtaza, Caldas, 2016). Due to the lack of nucleosomes in these areas, the respective cfDNA regions are digested more quickly than other plasma DNA segments and therefore sequencing coverages are lower in these regions (Figure 6) (Murtaza, Caldas, 2016). In summary, decreasing coverages in TSSs predict the higher expression of these genes. A recent study confirmed this finding by analysis of cfDNA from patients with breast cancer (Murtaza, Caldas, 2016, Ulz, Thallinger et al., 2016). Moreover, our recent study reported that analysis of the coverages at transcription factor binding

sites (TFBSs) allowed detection of early-stage CRC patients and tumor subclassification in prostate cancer (Ulz et al., 2019).

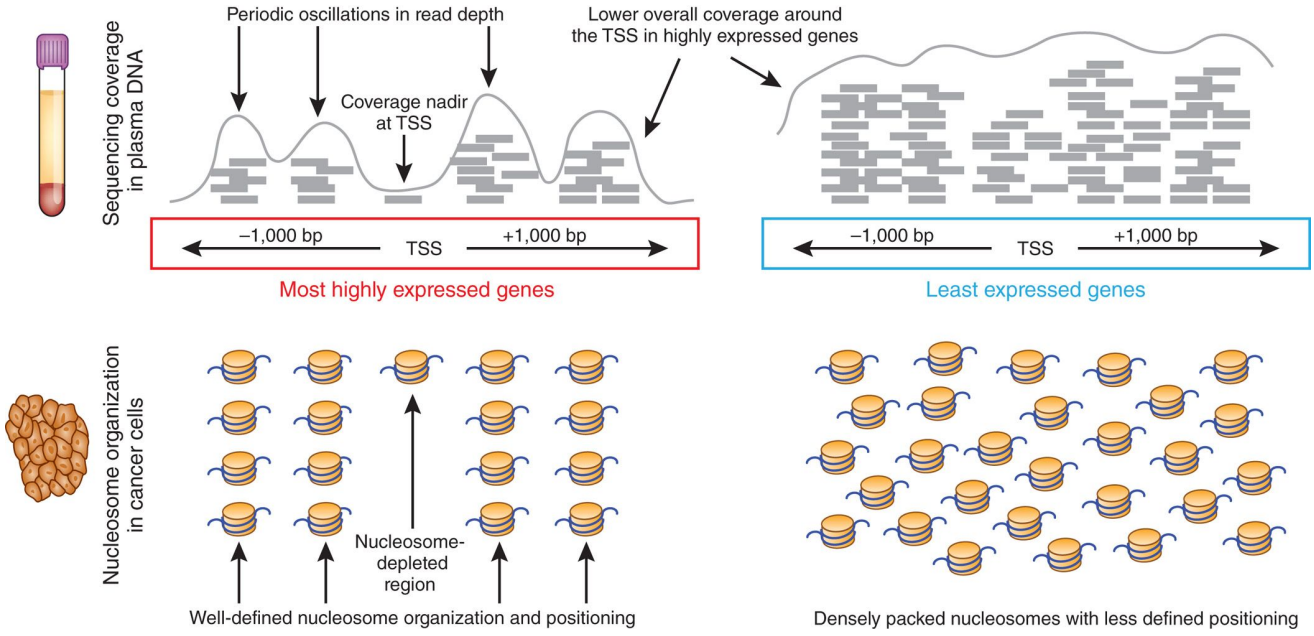


Figure 6: Read coverage at transcription start sites in plasma DNA predict gene expression (Murtaza, Caldas, 2016) (Permission for reuse of this figure has been granted from the publisher (Springer Nature))

2 AIMS

In this study, we focused our research on the analysis of ctDNA using whole-genome sequencing in mCRC patients. Firstly, we validated the application of ctDNA whole-genome sequencing in mCRC patient management, i.e. monitoring and predicting response to treatment in mCRC patients using ctDNA. Secondly, we validated the application of ctDNA whole-genome sequencing for monitoring disease progression, i.e. identification of novel driver genes in mCRC. To this end, we applied ctDNA whole-genome sequencing (Plasma-Seq) to 150 mCRC patients.

Specific aims of this project were as follows:

- Establishment of SiMSen-Seq (Simple, Multiplexed, PCR-based barcoding of DNA for Sensitive mutation detection using Sequencing)
- Serial monitoring of patients under targeted therapy to validate the predictive value of ctDNA for acquired resistance
- Serial monitoring of patients to validate the predictive value of ctDNA for disease progression
- Establishing landscapes of mCRC SCNAs to identify frequently altered genes and novel recurrently amplified genomic regions with no known driver gene, e.g. chr 13q12.2
- To investigate the consequences of chr 13q12.2 amplification and the potential driver gene within this region

3 MATERIALS AND METHODS

3.1 Ethics

The study was approved by the Ethics Committee of the Medical University of Graz (approval number 21-229 ex 09/10), conducted according to the Declaration of Helsinki, and written informed consent was obtained from all patients (Zhou et al., 2020).

3.2 Sample collection

All patients had metastatic CRC and were being treated at the Department of Internal Medicine, Division of Oncology, at the Medical University of Graz (Zhou et al., 2020). Serum CEA and CA19-9 values and CT scans were obtained as part of routine clinical care. Patients were evaluated to have stable disease or progressive disease in accordance with the response evaluation criteria in solid tumors (RECIST) (Schwartz et al., 2016). For each patient, 16 ml of whole blood were collected in 2 PAXgene Blood ccfDNA tubes (Qiagen, Hilden, Germany) at each time point. Samples were collected at multiple time points, including before treatment, during treatment (every 1-2 months) and at the end of treatment (after patient developed progressive disease). For some interesting patients, pre-treatment tumor specimens were available as a result of surgical or bioptic procedures.

3.3 Cell-free DNA isolation

Whole blood was processed as follows: blood was transferred into 15ml tubes and centrifuged at 200×g for 10 min with 0 brake and acceleration powers. Subsequently, blood was centrifuged for an additional 10 min at 1600×g and the supernatant was collected and transferred into a new 15ml tube. After an additional 1600×g 10 min centrifugation, plasma was collected and aliquoted into 2ml Eppendorf tubes and stored at -80℃. Cell-free DNA was isolated using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. In brief, 2ml of plasma were mixed with 200µl proteinase K and 1.6ml buffer ACL (containing 1µg carrier RNA). After incubation at 60℃ for 30 min, 3.6ml Buffer ACB were added into the mixture and incubated on ice for 5 minutes. The mixture was then transferred into QIAamp

Mini columns with a tube extender, which allowed the mixture to pass the QIAamp Mini membrane by vacuum force. After washing the membrane with Buffer ACW1, Buffer ACW2 and 96-100% ethanol, the QIAamp Mini column was centrifuged at 20,000×g for 3 minutes. The column was then incubated at 56℃ for 10 minutes in order to dry the membrane completely. Cell-free DNA was eluted with 90µl nuclease-free water (Qiagen, Hilden, Germany). The DNA concentration was measured using the Qubit™ dsDNA HS Assay kit (Thermo Fisher Scientific, Vienna, Austria) according to the manufacturer's instructions.

3.4 Genomic DNA isolation

Genomic DNA was isolated from FFPE tissue using the GeneRead™ DNA FFPE kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions with some modifications. Briefly, the section was collected in a 1.5ml Eppendorf tube and mixed with 1ml xylene (Sigma-Aldrich/Merck KGaA, Darmstadt, Germany) and after 10 minutes of incubation, the section was washed three times with 100% ethanol. Ethanol was removed and the section was dried at room temperature for 10 minutes. Subsequently, 55µl RNase-free water, 25µl Buffer FTB and 20µl proteinase K were added to the section and incubated at 56℃ overnight. After incubation at 90℃ for 1h, 115µl RNase-free water and 35µl UNG were added to the mixture, followed by incubation at 50℃ for 1 h. After incubation with 2µl RNase A at room temperature for 2 min, 250µl Buffer AL and 250µl ethanol (96-100%) were added to the mixture, which was then transferred into a QIAamp MinElute column. The column was centrifugated at 20,000×g for 1 min in order to let the lysate pass the membrane. The membrane was then washed with 500µl Buffer AW1, 500µl Buffer AW2 and 96-100% ethanol, individually. After drying the membrane via high-speed centrifugation, DNA was eluted with 30µl nuclease-free water (Qiagen, Hilden, Germany). The DNA concentration was measured using the Qubit™ dsDNA BR Assay kit (Thermo Fisher Scientific, Vienna, Austria) according to the manufacturer's instructions.

3.5 Cell-free DNA sample pre-screening

As the focus of this study was on copy number profiling of ctDNA, which requires tumor fractions of 5% and higher, we pre-screened our cell-free DNA using our previously published modified

Fast Aneuploidy Screening Test-Sequencing System (mFAST-SeqS) (Belic et al., 2015). In brief, Phusion Hot Start II DNA Polymerase (2U/µL; Thermo Fisher Scientific, Vienna, Austria) was used for the first polymerase chain reaction (PCR) as follows: 5-10µl cell-free DNA corresponding to 0.1-5ng of total DNA was mixed with 5× Phusion HF Buffer, 2U of Phusion Hot Start II Polymerase, 0.25µM of targeted-specific LINE-1 primers (Table 1) and 200µM dNTPs. PCR was conducted using the following conditions: 98℃ 2min - 98℃ 10s, 57℃ 120s, 72℃ 120s for 5 cycles. The PCR product was then purified using AMPureXP beads (Beckman Coulter, Brea, CA, USA). Briefly, the PCR product was mixed with 70µl beads and incubated at room temperature for 10 min. After 5 min magnet incubation, the supernatant was removed and the bead pellet was washed twice with 70% ethanol. After drying for 10 min, DNA was eluted with 12µl 1xTE buffer (Sigma-Aldrich/Merck KGaA, Darmstadt, Germany). 10µl product was then used in the second PCR: the product was mixed with 5× Phusion HF Buffer, 2U of Phusion Hot Start II Polymerase, 0.25µM of sample-specific index primers (Table 1) and 200µM dNTPs. PCR was conducted using the following conditions: 98℃ 2min - 98℃ 10s, 57℃ 120s, 72℃ 120s for 18 cycles. The PCR product was purified using AMPureXP beads as before. The purified product was checked on an Agilent Bioanalyzer with the 7500 DNA kit (Agilent, Santa Clara, USA) for quality and quantity. The final library was sequenced on an Illumina MiSeq or NextSeq 550 instrument (Illumina, San Diego, CA, USA) for the generation of 150bp single reads or 76bp paired-end reads with ~ 1 million reads per sample. The following analysis was done using an in-house script (Belic et al., 2015).

LINE-1 specific	
PRIMER	SEQUENCES
LINE1_F	TCTTTCCCTACACGACGCTCTTCCGATCTACACAGGGAGGGGAACAT
LINE1_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCCATGGTGGTTTGCT
Sample-specific	
PRIMER	SEQUENCES

Forward	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Rev_Index1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index9	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index10	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index11	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index12	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rev_Index13	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Rev_Index14	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT
Rev_Index15	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT
Rev_Index16	CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT
Rev_Index18	CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT
Rev_Index19	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT
Rev_Index20	CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT
Rev_Index21	CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT
Rev_Index22	CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT
Rev_Index23	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCT
Rev_Index25	CAAGCAGAAGACGGCATACGAGATATCAGTGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT
Rev_Index27	CAAGCAGAAGACGGCATACGAGATAGGAATGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT

Table 1: Summary of LINE-1-specific and sample-specific primers

3.6 Plasma-seq: Whole-genome sequencing of primary tumor and plasma samples

Whole-genome sequencing libraries were prepared and sequenced for plasma and tumor samples when available by methods published previously (Heitzer, Ulz et al., 2013, Zhou et al., 2020, Mohan et al., 2014, Heitzer et al., 2013). Shotgun libraries were prepared using the TruSeq DNA LT Sample Preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions for both primary tumor samples and for cell lines. In brief, 100ng genomic DNA was fragmented to 350bp using the Covaris S220 (Covaris Ltd. Brighton, UK). After clean-up with SPB, DNA was diluted to 60µl and mixed with 40µl ERP3. The mixture was incubated at 30℃ for 30 min for end repair. A double size selection was then applied to the mixture using 0.56× SPB and 0.12× SPB in order to remove large and small DNA fragments, respectively. In order to add one adenine nucleotide to the 3' ends of the blunt fragments, 12.5µl ATL2 was added to 17.5µl of size-selected DNA and the mixture was incubated in the following conditions: 37℃ 30 min - 70℃ 5 min – 4℃ 5 min. Subsequently, 2.5µl LIG, 2.5µl RSB and 2.5µl Illumina index adapter were added to the mixture and incubated at 30℃ for 10 min to ligate Illumina index adapters to the end of the fragments. After inhibiting the ligation with 5µl STL and purification repeated twice with 1× SPB, the product was processed with a PCR step to enrich DNA fragments as follows: reagents setting: 5µl PPC + 20µl EPM + 25µl product from the previous step. PCR setting: 95℃ 3 min - 98℃ 20s, 60℃ 15s, 72℃ 30s for 8 cycles - 72℃ 5 min - 4℃ ∞. The mixture was then purified with 1× SPB and the final library was eluted in 30µl RSB.

Shotgun libraries of cell-free DNA samples were also prepared using the TruSeq DNA LT Sample Preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions but with several modifications: 5–10ng of input DNA were used and the fragmentation step was omitted, the double size selection step was replaced with a single 1× SPB purification, since cell-free DNA is enriched for fragments in the range of 160 to 340bp, and 25 PCR cycles were used for the selective amplification step of library fragments.

150bp single reads or 76bp paired-end reads were generated for both genomic DNA and cell-free DNA libraries an were sequenced on the Illumina MiSeq or NextSeq 550 instrument (Illumina, San Diego, CA, USA). The average reads obtained for each sample was 5-10 million, which

represents a 0.1-0.2× coverage of the whole genome. Somatic copy number aberration analysis was conducted using an in-house script as published previously (Heitzer et al., 2013). Tumor fraction from cell-free DNA and tumor tissue was estimated with the ichorCNA algorithm, a probabilistic model for the simultaneous prediction of large-scale copy number alterations and estimation of tumor fraction (Adalsteinsson et al., 2017).

3.7 SiMSen-Seq: Simple, Multiplexed, PCR-based barcoding of DNA for Sensitive mutation detection using Sequencing

Barcoding of DNA and library construction were performed as published previously (Stahlberg et al., 2016, Stahlberg et al., 2017). Barcoding of DNA was conducted using Phusion™ Hot Start II High-Fidelity DNA Polymerase as follows: 2µl 5× Phusion HF Buffer, 0.05µl Phusion Polymerase (2U/µl), 0.2µl 10mM dNTP mix (Sigma-Aldrich/Merck KGaA, Darmstadt, Germany), 1µl 5M L-Carnitine (Sigma-Aldrich/Merck KGaA, Darmstadt, Germany), 0.4µl pre-designed target specific primer mix (1µM; Table 2;Integrated DNA Technologies, Inc., Iowa, USA) and 6.35µl cell-free DNA template (10-20ng). After incubation using the following conditions: 98°C 30s - 98°C 10s, 62°C 6 min for 3 cycles, 20µl protease (30ng/µl, Streptomyces griseus; Sigma-Aldrich/Merck KGaA, Darmstadt, Germany) was added to the mixture and incubated at 65°C for 15 min to digest the present polymerase, followed by an incubation at 95°C for 15 min to inactivate the protease. 10µl of the mixture from the previous step was mixed with 1.6µl sample specific primer (10µM; Table 1), 8.4µl nuclease-free water and 20µl Q5 Hot Start High-Fidelity 2× Master Mix (New England Biolabs GmbH, Frankfurt, Germany) to construct the final library. Afterwards, an incubation was carried out: 98°C 3 min - 98°C 10s, 80°C 1s, 72°C 30s, 76°C 30s for 28 cycles with ramping at 0.2°C/s - 4°C ∞, the final product was purified by 1× AMPure XP beads and eluted with 10µl TE buffer. The final library was sequenced on an Illumina NextSeq instrument (Illumina, San Diego, CA, USA) for the generation of 150bp single reads. An analysis tool kit (Debarcer) was used for the analysis of SiMSen-seq (Stahlberg et al., 2016, Stahlberg et al., 2017).

PRIMER	GENE	TARGET	SEQUENCES
KRAS_1-F	KRAS	Chr12:25398233-25398296	GGACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNATGGGAAAGAGTGTCCTTTACCTCTATTGTTGGATCATATTCGTCCA
KRAS_1-R			GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTGCTGAAAATGACTGAATATAAACTTGTG
TP53_1-F	TP53	Chr17:7577064-7577125	GGACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNATGGGAAAGAGTGTCCGTGGTGAGGCTCCCCTTT
TP53_1-R			GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACTGGGACGGAACAGCTTTG
TP53_7-F		Chr17:7577519-7577574	GGACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNATGGGAAAGAGTGTCCCCTGGAGTCTTCCAGTGTGATG
TP53_7-R			GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGACTGTACCACCATCCACTACAAC
TP53_993-F		Chr17:7576812-7576876	GGACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNATGGGAAAGAGTGTCCAACGGCATTTTGAGTGTTAGAC
TP53_993-R			GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGCCAAAGAAGAAACCACTG

PIK3CA_234-F	PIK3CA	Chr3:178952065-178952089	GGACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNATGGGAAAGAGTGTCCCTGAGCAAGAGGCTTTGGAGTATTTCATG
PIK3CA_234-R			GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCAATCCATTTTTGTTGTCCAGCCAC

Table 2: Summary of SiMSen-Seq Target specific primers

3.8 TCGA data collection and analysis

TCGA data analyzed in this work originated from TCGA-COADREAD projects (TCGA, 2016, Zhou et al., 2020). Clinical records and gene expression data were downloaded from Broad Institute GDAC Firehose (http://gdac.broadinstitute.org/). Absolute copy number results were downloaded from the NCI Genomic Data Commons (GDC; https://gdc.cancer.gov/about-data/publications/pancanatlas) (Grossman et al., 2016, Zhou et al., 2020). Patient samples were categorized as "balanced", "gain" or "amplification" according to 13q12.2 copy number (balanced: 1 < copy number <= 3, gain: 3 < copy number <= 6, amplification: copy number > 6) (Zhou et al., 2020). Statistical analyses were performed in R. Focal event calling was performed with an in-house script as published previously (Heitzer et al., 2013).

3.9 Digital PCR copy number assay

SCNAs of *POLR1D* and *ERBB2* were analyzed using digital PCR (dPCR), which was performed on the QuantStudio™ 3D platform (Life Technologies, Carlsbad, CA, USA). 10ng cell-free DNA was diluted to 6.6µl and mixed with 9µl Master Mix v2, 1.2µl predesigned TaqMan assays for copy number detection (POLR1D (Hs02926936_cn) or ERBB2 (Hs00450668_cn); Life Technologies,

Carlsbad, CA, USA) and 1.2µl reference assay (TERT; 4403315; Life Technologies, Carlsbad, CA, USA). 14.5µl of this mixed reaction was loaded onto the QuantStudio™ 3D Digital PCR Chip v2, and PCR was performed as follows: 96℃ 10 min - 60℃ 2 min, 98℃ 30s for 39 cycles - 60℃ 2 min - 10℃ ∞. After PCR amplification, the chip was imaged using the QuantStudio™ 3D Digital PCR instrument. An online analysis tool was then applied to the raw data with the confidence level set to 95% and the desired precision value set to 10%.

3.10 Gene expression prediction

Gene expression prediction analysis was performed as published previously (Ulz et al., 2016, Zhou et al., 2020). In brief, whole genome libraries prepared as described before were sequenced to obtain 200 million reads (5× coverage of the whole genome). After mapping all the reads to the reference genome (hg19), coverage values around transcription start sites (TSSs) were extracted from aligned BAM files. Read depths of these sites were normalized by the mean value of the combined regions: TSS – 3,000 to TSS – 1,000 and TSS + 1,000 and TSS + 3,000. The coverage from −1,000bp to +1,000bp (2K-TSS coverage) and from −150bp to +50bp (NDR coverage) was used for gene expression prediction with support vector machines (SVMs). For each sample, the SVM was trained based on a k-fold cross-validation with known expressed and unexpressed transcription start sites and the remaining genes were predicted from the trained model. Only genes that which are consistently predicted across folds (the same prediction in >95% of folds) were considered to be valid predictions.

3.11 Cell lines and cell culture

Human CRC cell lines (OXCO-2, SW480, and HT29) were selected based on their copy number variation profiles. OXCO-2 was provided by Dr. Alberto Bardelli, Laboratory Molecular Oncology at the Candiolo Cancer Institute IRCCS - Candiolo (Torino). SW480 and HT29 cell lines were provided by Prof. Martin Pichler, Department of Internal Medicine, Division of Oncology, Medical University of Graz. Whole-genome profiling was performed for all cell lines. Whole chromosome 13 gain was detected in SW480 cells, whereas 13q12.2 focal amplification was identified in HT29 cells. OXCO-2 cells showed balanced profiling on chromosome 13 (Figure 7).

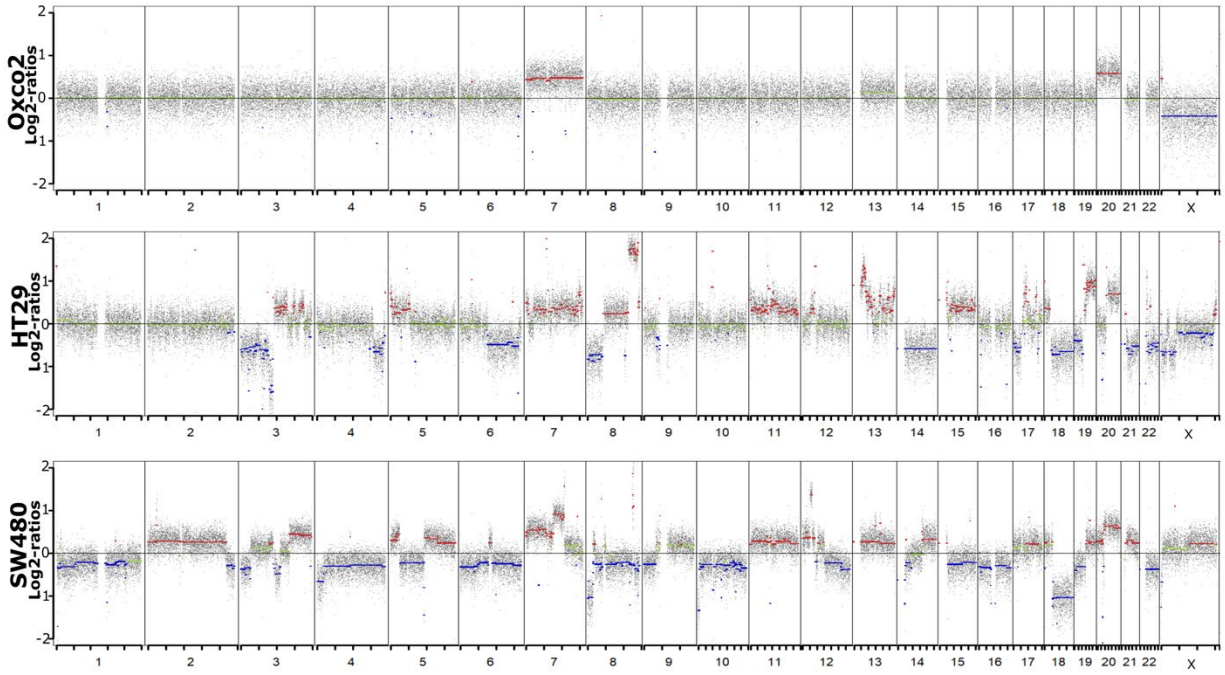


Figure 7: Whole genome profiling of OXCO-2, HT29 and SW480 cell lines [Figure and legend originally published in Genome Medicine (Zhou et al., 2020)]

All the medium and supplements were obtained from Life Technologies Gibco Ltd. OXCO-2 cells were maintained in Iscove's Modified Dulbecco's Medium (IMDM). HT29 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) and SW480 cells were maintained in RPMI 1640 medium. All media were supplemented with 5% fetal bovine serum (Gibco, Thermo Fisher Scientific, Vienna, Austria) and 1% penicillin/streptomycin (Gibco, Thermo Fisher Scientific, Vienna, Austria). The CRC cell lines were authenticated at the Cell Bank of the Core Facility of the Medical University of Graz, Austria, by performing an STR profiling analysis (Kit: Promega, PowerPlex 16HS System; Cat.No. DC2101, last date of testing: July 2019).

Defrosting cells

One vial of the frozen cell line was taken out from the liquid nitrogen tank and placed into a 37℃ water bath. When the cell solution was totally thawed, cells were transferred into 15ml falcon tubes containing 10ml 37℃ culture medium. Cells were then centrifuged at 1500rpm for 5 min; the cell pellet was re-suspended in 10ml fresh medium and plated into a T75 flask.

Freezing cells

Freezing medium was prepared by adding 20% FBS and 5% dimethyl sulfoxide (DMSO; Sigma-Aldrich/Merck KGaA, Darmstadt, Germany) in the growth medium. Cells were trypsinized, counted, and centrifuged at 1500rpm for 10min at 4℃. Cells were resuspended in the freezing medium such that 1 million cells were frozen per cryovial. The cryovials were then placed into a freezing container and placed at -80℃ and placed in liquid nitrogen the following day.

3.12 Generation of stable FLT3-overexpressing cell line

The OXCO-2 cell line, which does not harbor amplification nor expression of FLT3, was used to generate the stable FLT3-overexpressing cell line. In brief, OXCO-2 cells were seeded in a 6-well plate in IMDM media and grown to approximately 80% confluence. The FLT3 (NM_004119) GFP-tagged human cDNA ORF clone was ordered from OriGene and was transfected into cells with the FuGene HD (Promega, Mannheim, Germany) transfection reagent according to the manufacturer's recommendations. After 24 hours, cells were transferred into selective medium and supplemented with 1mg/ml Geneticin (Gibco, Thermo Fisher Scientific, Vienna, Austria). The medium was changed every 3 days until the formation of colonies. Colonies were collected via trypsin-soaked cloning discs (Sigma-Aldrich/Merck KGaA, Darmstadt, Germany), and once detached, each colony was transferred separately into a 96-well plate containing IMDM media and geneticin. Cells were grown in 96-well plates until they reached sufficient confluence and then transferred into a bigger plate to further select for clones stably expressing FLT3 (OXCO2-FLT3). The expression of FLT3 was tested by reverse transcription polymerase chain reaction (RT-PCR) as described below.

3.13 siRNA knockdown assays

siRNA	ID	Gene	Knockdown efficiency in HT29	Knockdown efficiency in SW480
siCDX2-2	s2876	CDX2	Successful	Successful
siCDX2-3	s2878	CDX2	Successful	Successful
siLNX2-1	s48196	LNX2	Successful	Successful
siLNX2-2	s48195	LNX2	Successful	Successful
siPAN3-1	s48721	PAN3	Insufficient knockdown	Successful
siPAN3-2	s48723	PAN3	Insufficient knockdown	Successful
siPAN3-3	s48722	PAN3	Insufficient knockdown	Successful
siPDX1-1	s7488	PDX1	Successful	Successful
siPDX1-2	s223944	PDX1	Insufficient knockdown	Successful
siPOLR1D-2	23515	POLR1D	Successful	Successful
siPOLR1D-3	119496	POLR1D	Successful	Successful
siPOLR1D-4	39457	POLR1D	Successful	Successful

Table 3: Summary of all siRNA used and the knockdown efficiency in HT29 and SW480 cell lines [Table and legend originally published in Genome Medicine (Zhou et al., 2020)]

Predesigned siRNAs (Table 3) targeting CDX2, LNX2, PAN3, PDX1, and POLR1D were purchased from Life Technologies Ambion Ltd. For a scrambled control, the AllStars Negative Control siRNA (Ambion, Life Technologies) was used. Reverse transfection was carried out with Lipofectamine™ RNAiMAX transfection reagent (Invitrogen, Thermo Fisher Scientific, Vienna, Austria) according to the suggestion of the supplier. In brief, for each well of 24-well plate, 50-100nM scrambled or targeted siRNA was mixed with 25µl of Opti-MEM media (Gibco, Thermo Fisher Scientific, Vienna, Austria). In another tube, 1µl of siRNA transfection reagent was mixed with 24µl of Opti-MEM media. After a 5 min room temperature incubation, these two mixtures were mixed in a 1:1 ratio and continually incubated at room temperature for 20 min. After incubation, 50µl of the mixture was applied to the middle of an empty well in a 24-well plate and

40,000 cells (suspended in 450µl media) were administered to the top of the mixture. Transfected cells were incubated for 72 hours before performing proliferation assays or harvesting for expression analysis.

3.14 RNA isolation, quantitative RT-PCR

RNA was isolated via the TRIzol (Invitrogen, Thermo Fisher Scientific, Vienna, Austria) method. Briefly, cells were washed with PBS and 500µl TRIzol was added to one well of a 24-well plate. After mixing well, the lysate was transferred into a 1.5ml tube and incubated at room temperature for 10 min. After adding 100µl chloroform (Sigma-Aldrich/Merck KGaA, Darmstadt, Germany) into the lysate and incubation for 3 min, the mixture was centrifuged at 13,000×g for 15 min at 4℃. The supernatant was moved to a new 1.5ml tube and 25µl isopropanol (Sigma-Aldrich/Merck KGaA, Darmstadt, Germany) was added. After 10 min room temperature incubation, the mixture was centrifuged at 13,000×g for 15 min at 4℃, and the liquid was removed. The RNA pellet was washed with ice-cold 75% ethanol twice and dried for 15 min and subsequently the pellet was resuspended in 20µl RNase-free water. The RNA concentration was quantified with the Nanodrop 1000 (Thermo Fisher Scientific, Vienna, Austria).

Gene	**Forward Primer Sequence (5′ → ‘3)**	**Reverse Primer Sequence (5′ → ‘3)**	**PCR protocol**
FLT3	TTTCACAGGACTTGGACAGAGATTT	GAGTCCGGGTGTATCTGAACTTCT	95°C/50°C/72°C
POLR1D (V1)	CCACCTGAGGATCCAGAAAC	CCTCGTGCAATACAAATGTCA	95°C/60°C
POLR1D (V2)	AAGAACTGCTTAAGGAGGCAA	TCTTCGCTGGTTCCTTATCG	95°C/60°C
PAN3	TTGGTGCCCTCAACATCTCT	TTGATCCCATCGGAACTAGC	95°C/60°C
CDX2	GAACCTGTGCGAGTGGATG	TCCTCCGGATGGTGATGTAG	95°C/60°C

PDX1	CCTTTCCCATGGATGAAG TC	TTCAACATGACAGCCAG CTC	95°C/60°C
LNX2	ATGCAACGTTGTGATCTG GA	CCAAACAGTCTGCTTCT GGA	95°C/60°C
GAPDH	CCAAAATCAAGTGGGGCG ATG	AAAGGTGGAGGAGTGGG TGTCG	95°C/60°C

Table 4: Summary of all qPCR primers used and the corresponding PCR protocols [Table and legend originally published in Genome Medicine (Zhou et al., 2020)]

Reverse transcription was done with the QuantiTect Reverse Transcription Kit (Qiagen, Hilden, Germany) according to the manufacturer's suggestions. 1µg total RNA was mixed with 2µl gDNA Wipeout Buffer and the mixture was scaled up to 14µl. After incubation at 42℃ for 2 min, 1µl Quantiscript Reverse Transcriptase, 4µl 5× Quantiscript RT Buffer and 1µl RT Primer Mix were added to the mixture and incubated at 42℃ for 15 min followed by a 95℃ 3 min incubation to inactivate the reverse transcriptase. Reversed cDNA was then used in the quantitative RT-PCR, which was performed on the ABI 7500 system to detect the expression of target genes. Biozym Blue S´Green qPCR Kit (Biozym, Hessisch Oldendorf, Germany) and pre-designed RT-PCR primers (Table 4; Microsynth AG, Switzerland) were used in quantitative RT-PCR according to the manufacturer's suggestions.

3.15 Whole transcriptome RNA-seq

Captured coding transcriptome RNA-seq libraries were prepared with the TruSeq RNA Exome kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions with 100ng total RNA input. Briefly, 100ng total RNA was diluted to 8.5µl and mixed with 8.5µl EPH. After incubation at 94℃ for 8 min, first strand cDNA synthesis was conducted by adding 8µl FSA (contains 10% SuperScript II) to the mixture and incubated as follows: 25℃ 10 min - 42℃ 15 min - 70℃ 15 min - 4℃ 5 min. To process the second strand cDNA synthesis, 5µl RSB and 20µl SMM were added to the mixture and incubated at 16℃ for 1h. After clean up with SPB, cDNA was used to prepare the shotgun library similar as described before (i.e., using TruSeq DNA LT Sample

preparation Kit for genomic DNA) but without end repair step, as the cDNA generated before already had blunt ends. Libraries were quantified using the Agilent High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA, USA) and 200ng of each DNA library were used for the transcriptome enrichment. 45µl of the DNA library pool (up to 4 libraries were pooled in one reaction) was mixed with 50µl CT3 and 5µl CEX and incubated as follows: 95℃ 10 min - 94℃ decrease to 58℃ in 18 cycles (2℃ per cycle) of 1 min incubation - 58℃ 90 min, in order to hybridize the pre-designed probes to the cDNA library. By mixing with 250µl SMB and incubating at room temperature for 25 min, the hybridized probes were captured by SMB and after washing twice with EWS, the library was eluted with 23µl EE1 (contain 5% HP3). The hybridization and capture steps were repeated once more and the transcriptome-enriched library was purified using 1.8× AMPure XP beads and amplified by mixing with 5µl PPC and 20µl EPM and incubated as follows: 98℃ 30s - 98℃ 10s, 60℃ 30s, 72℃ 30s for 10 cycles - 72℃ 5 min - 10℃ ∞. The final library was purified with 1.8× AMPsure XP beads and sequenced on an Illumina NextSeq instrument (Illumina, San Diego, CA, USA) for the generation of 75 bp paired-end reads. RNA-seq data was then analyzed with a pseudo-alignment based approach (Kallisto) (Bray et al., 2016), followed by differential gene expression analysis performed by DESeq2 Bioconductor package in R (Love, Huber & Anders, 2014).

3.16 Colony formation and cell viability assay

The colony formation assay was conducted as follows: cells were seeded in a 24-well plate for 72 hours and fixed in 100% methanol for 20 min. After staining with 0.5% crystal violet for 5 min, images of each well were acquired and colonies in each well were counted using ImageJ software.

Cell viability was performed with CellTiter-Glo Luminescent Cell Viability Assay (Promega, Mannheim, Germany) according to the manufacturer's instructions. Briefly, cells were seeded in 96-well plates for 72 hours and one volume of CellTiter-Glo reagent was added to each well. After mixing for 2 min and incubation at room temperature for 10 min, the luminescence was recorded on the LUMIstar OPTIMA Microplate Luminometer (BMG LABTECH GmbH, Ortenberg, Germany).

4 RESULTS

4.1 ctDNA can be used to monitor and predict the response of treatment in mCRC patients

4.1.1 Emergence of *KRAS* and *ERBB2* focal amplifications in cfDNA coincides with acquired resistance to anti-EGFR treatment

Previously, we had confirmed that focal amplification of *KRAS* can be used as a predictive biomarker for CRC patients treated with anti-EGFR treatment (Mohan et al., 2014). As shown in Figure 8A, the first plasma sample of patient C17 did not harbor *KRAS* amplification. However, after 5 months of anti-EGFR treatment, a focal amplification was detected in the genomic region where *KRAS* is located, which coincided with acquired resistance to the treatment as evidenced by a new liver metastasis. *KRAS* mutations are the most frequent acquired resistance mechanism to anti-EGFR therapy in CRC (Zhao et al., 2017). *KRAS* amplification, although not observed as frequently as *KRAS* mutations, also has been shown to be involved in the resistance to anti-EGFR treatment (Misale et al., 2014, Valtorta et al., 2013, Mohan et al., 2014). Another potential mechanism of acquired resistance to anti-EGFR therapy is *ERBB2* amplification (Bertotti et al., 2011). We also identified one patient (C129) where the first blood sample did not demonstrate *ERBB2* amplification, whereas after nine months of anti-EGFR treatment, when the tumor had acquired resistance to the treatment with enlarged liver metastasis lesions, an *ERBB2* focal amplification was detected in the ctDNA (Figure 8B).

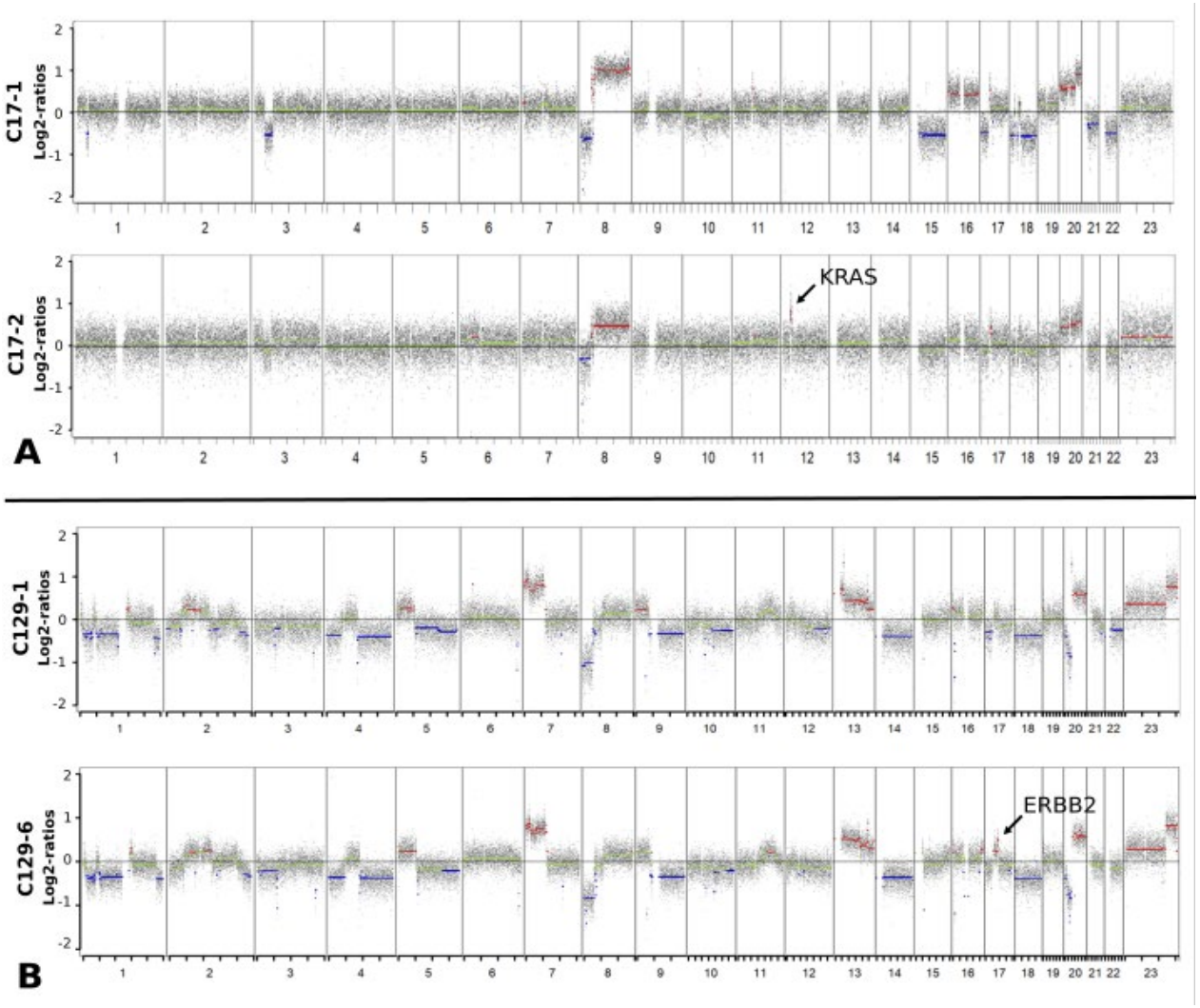


Figure 8: Emergence of *KRAS* and *ERBB2* amplification during anti-EGFR treatment in patients C17 and C129

A: Genome-wide log2-ratio plots of plasma samples from C17 obtained before anti-EGFR treatment (upper) and after 5 months of anti-EGFR treatment (bottom). The arrow represents the region that harbors the *KRAS* amplification. Copy number gains are shown in red, balanced regions in green and copy number losses in blue.

B: Genome-wide log2-ratio plots of plasma samples from C129 obtained before anti-EGFR treatment (upper) and after 9 months of anti-EGFR treatment (bottom). The arrow represents the region that harbors the *ERBB2* amplification. Copy number gains are shown in red, balanced regions in green and copy number losses in blue.

4.1.2 Increasing of *CDK6* focal amplification in ctDNA reveals emergence of acquired resistance to lapatinib

It has been previously published that defects in cell cycle regulation via increased levels of Cyclin D-CDK4/6 complex is a well-known mechanism of acquired resistance to lapatinib (Niu, Xu & Sun, 2019, Jiang et al., 2018). Patient C79 was under lapatinib treatment due to high-level focal amplification of *ERBB2* found in the first blood sample (Figure 9A). After a short increase of the *ERBB2* log2 ratio in the second sample, the log2 ratio of *ERBB2* decreased in the next two samples (Figure 9A & B). However, a consistent increase of the *CDK6* log2-ratio was found in all the samples (Figure 9A & B) and the patient showed resistance to the treatment after 77 days, with an increased size of liver metastasis lesions and appearance of pleural effusion in the right lung (Figure 9C).

These results suggested that ctDNA can be used to monitor the response of targeted treatment in mCRC patients by monitoring SCNAs harboring known oncogenes, which are related to the acquired resistance of the drug.

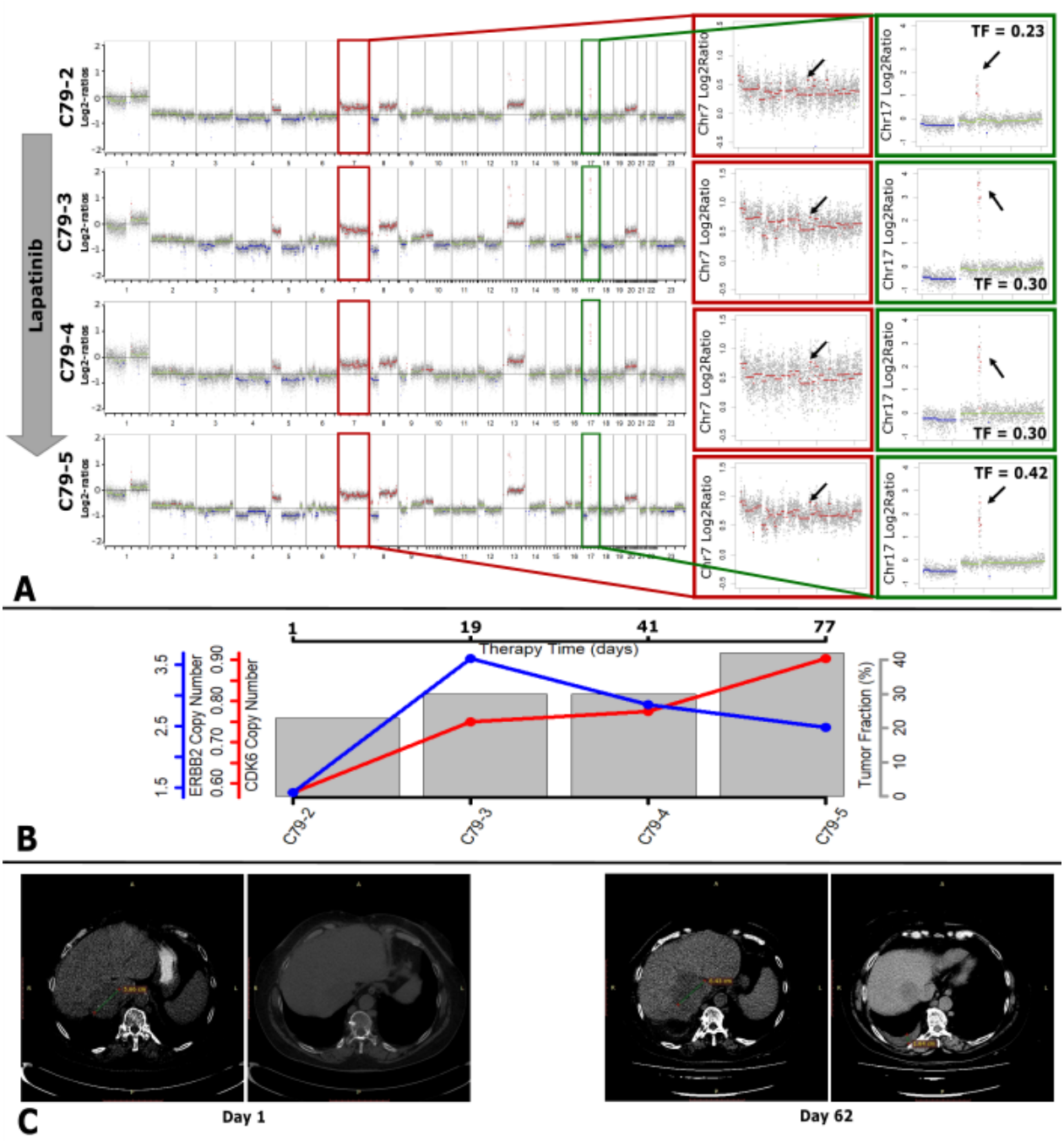


Figure 9: Changes in the *CDK6* and *ERBB2* amplification under lapatinib treatment in patient C79

A: Genome-wide log2-ratio plots of plasma samples from C79 obtained before treatment with lapatinib (1st), 19 days (2nd), 41 days (3rd), and 138 days (4th) after lapatinib treatment. The insets

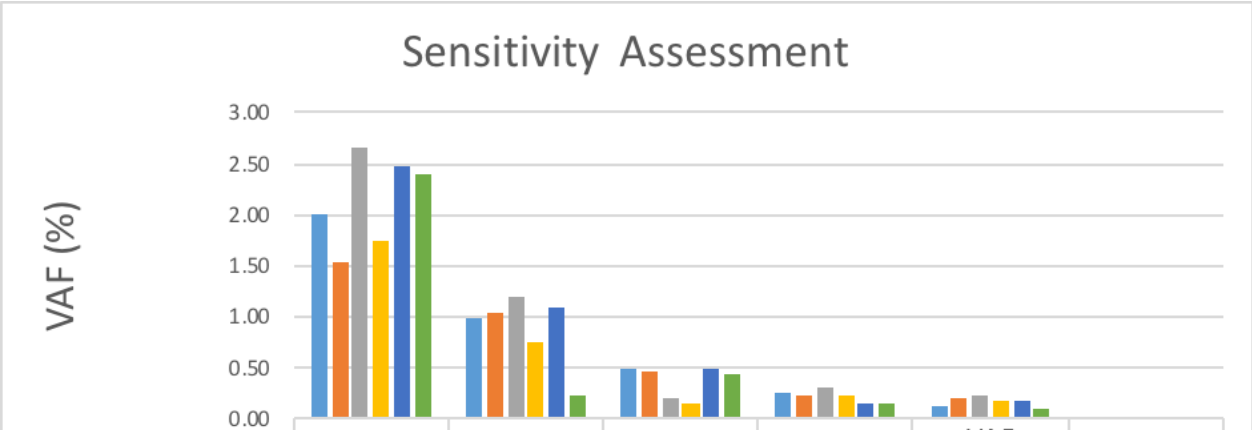
illustrate the respective tumor fraction (TF) for each analysis and enlarged log2-ratio plots of chromosome 7 and 17, arrows represent the region that harbors the *CDK6* (chromosome 7) and *ERBB2* (chromosome 17) amplification. Copy number gains are shown in red, balanced regions in green and copy number losses in blue.

B: Plot illustrating all time points of blood collection and relative gene copy number changes. Red line: *CDK6* log2-ratio changes identified by Plasma-seq. Blue line: *ERBB2* log2-ratio changes identified by plasma-Seq. Gray bar: Tumor content identified in every sample using ichorCNA.

C: CT images obtained on day 1 and day 62 after lapatinib treatment. Compared to the first image, all liver metastasis lesions had become larger, and pleural effusion appeared in the right lung. ctDNA can be used as a prognostic biomarker in mCRC patients.

4.1.3 Establishment of SiMSen-Seq approach

Mutation detection in ctDNA may be hampered by certain technical issues, such as low abundancy of tumor DNA in plasma, PCR or sequencing errors. For a high sensitivity detection of mutations in ctDNA, we applied a UMI-based approach (i.e., SiMSen-Seq). We first tested the sensitivity of this approach with the Seraseq ctDNA V2 reference materials (SERACARE LIFE SCIENCES, Milford, MA, USA), which contain hotspot mutations in *KRAS*, *PIK3CA* and *TP53* with different allele frequencies. As shown in Figure 10, mutations were not detected in the wild type reference but were detected in the other reference with an allele frequency close to the digital droplet PCR (ddPCR) results reported by the provider, except for *TP53* p.C242fs*5 in the variant allele frequency (VAF) 1% group and *PIK3CA* p.H1047R and TP53 p.R273H in the VAF 0.5% group, where the detected allele frequencies according to our method were lower than the reported. As all mutations reported in the VAF 0.1% group were detected, we confirmed that SiMSen-Seq could be used to detect mutations with an allele frequency higher than 0.1%. As there were still some discrepancies between the SiMSen-Seq and ddPCR results, the SiMSen-Seq results still contain variants around ∓ 0.8%. In this case, we applied a threshold at a VAF 1% in order to ensure that all of the mutations could be detected.



	VAF 2%	VAF 1%	VAF 0.5%	VAF 0.25%	VAF 0.125%	WT
Expected	2.00	1.00	0.50	0.25	0.13	0.00
KRAS p.G12D	1.53	1.03	0.47	0.24	0.21	0.00
PIK3CA p.H1047R	2.67	1.19	0.21	0.30	0.24	0.00
TP53 p.R273H	1.75	0.76	0.16	0.24	0.19	0.00
TP53 p.R248Q	2.48	1.08	0.50	0.16	0.17	0.00
TP53 p.C242fs*5	2.39	0.24	0.44	0.16	0.10	0.00

Figure 10: Summary of SiMSen-Seq sensitivity test results

Seraseq ctDNA V2 reference materials contain *KRAS* p.G12D, *PIK3CA* p.H1047R, *TP53* p.R273H, *TP53* p.R248Q and *TP53* p.C242fs*5 mutations at different allele frequencies, as shown in the expected row and the light blue bar. Allele frequencies detected by SiMSen-Seq of different mutations are shown in the following rows and different colored bars, respectively.

4.1.4 Application of SiMSen-Seq for monitoring in mCRC patients

In our mCRC cohort, we found one interesting patient (C128), in which the *TP53* c.993+1 G>T mutation was detected at a VAF of 10% in the first blood draw and amplification of *TP53* was not detected (Figure 11 A & B). The patient was treated with regorafenib for 2 months and once he developed resistance to the treatment (i.e. increasing of serum CEA and CA19-9 levels and new metastasis lesions were found in the liver) (Figure 11 A & B), the allele frequency of the *TP53* mutation increased to 30% and a *TP53* focal deletion was detected with a log2-ratio of -0.6 (Figure 11 A & B). CT scan images were not available due to data transfer limitations between hospitals. The patient was then treated with Lonsurf and stable disease was reported for the next blood draw. Although the serum CEA and CA19-9 levels increased significantly at this time point, there were no significant changes detected in the CT scans, which is consistent with the *TP53* mutation results, as no significant change was identified (Figure 11 A & B). However, the *TP53* focal deletion still showed a significant decrease (Figure 11 A & B), which was potentially due to tumor heterogeneity. After treatment for an additional 2.5 months, the patient showed progressive disease, with a dramatic increase of serum CEA and CA19-9 levels (Figure 11 A & B) and enlarged metastasis lesions found on the liver, which is consistent with the ctDNA changes, i.e. significant increase in the allele frequency of the *TP53* mutation and decrease in the *TP53* log2 ratio (Figure 11 A & B). The plasma tumor content correlated well with the change of the TP53 mutation and the log2-ratio of the deletion, which means that these 2 changes may be primarily due to the higher amount of tumor content in the blood. As we have described previously, high tumor content typically correlates with higher tumor burden and monitoring the changes in the TP53 mutation allows us to monitor tumor burden changes in the patient.

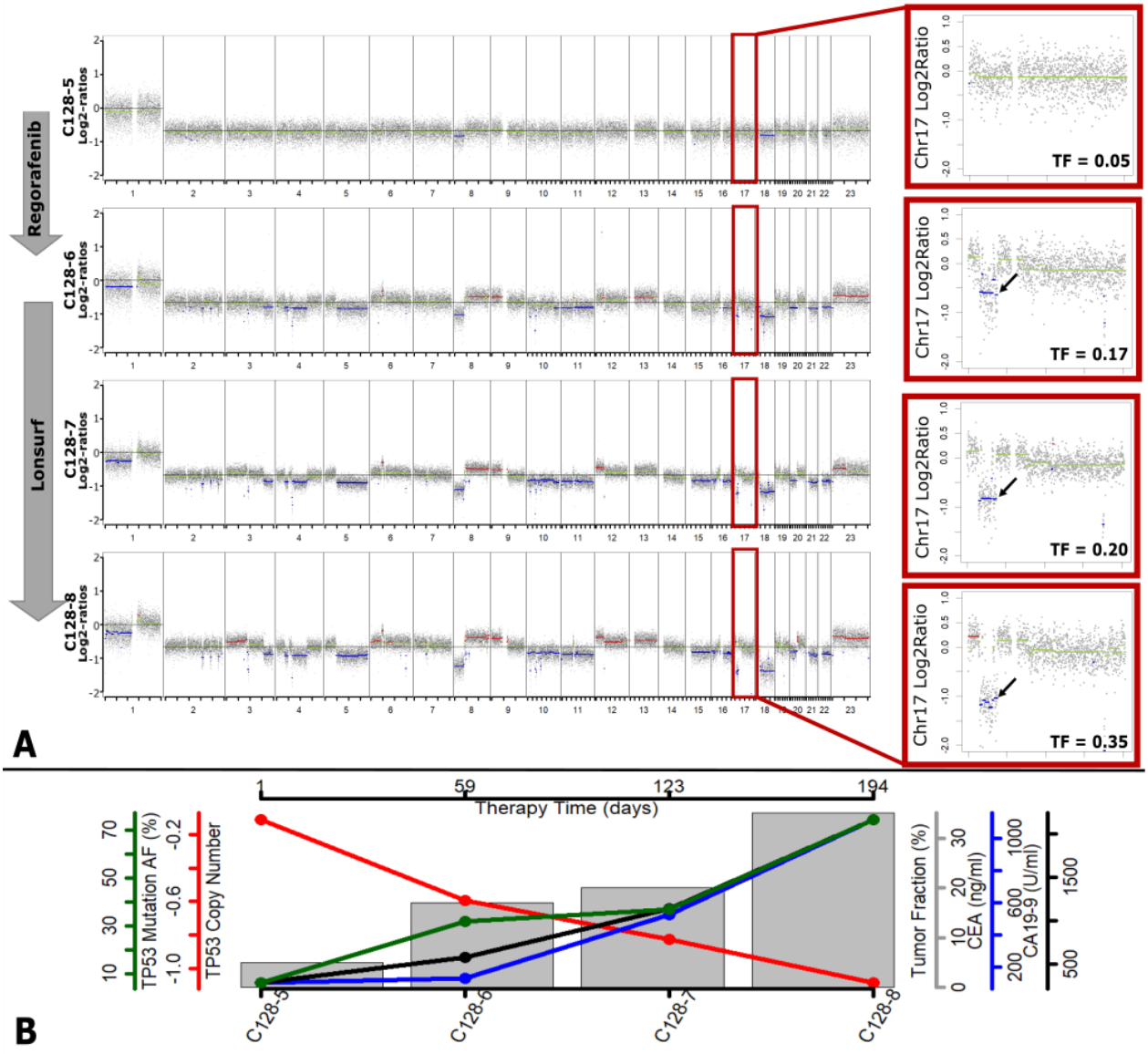


Figure 11: Changes in *TP53* deletion and mutation under regorafenib and Lonsurf treatment in patient C128

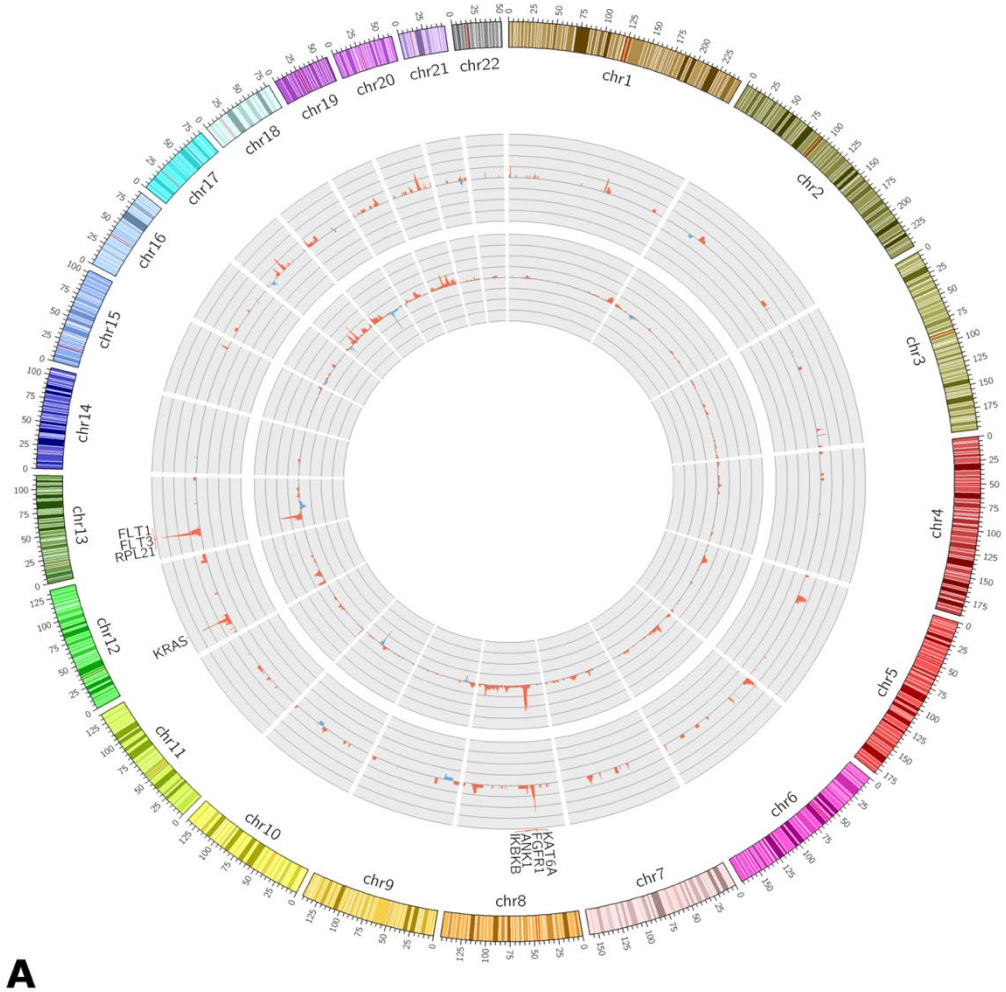
A: Genome-wide log2-ratio plots of plasma samples from C128 obtained before (1st), after 59 days (2nd) regorafenib treatment, after 64 days (3rd) and 135 days (4th) of Lonsurf treatment. The insets illustrate the respective tumor fraction (TF) for each analysis and enlarged log2-ratio plots of chromosome 17, arrows represent the region that harbors the TP53 deletion. Copy number gains are shown in red, balanced regions in green and copy number losses in blue.

B: Plot illustrating all time points of blood collection and relative biomarker changes. Red line: *TP53* log2-ratio changes identified by plasma-Seq. Green line: *TP53* mutation allele frequency. Blue line: Serum CEA levels. Black line: Serum CA19-9 levels. Gray bar: Tumor content identified in every sample using ichorCNA.

4.2 ctDNA can be used to identify novel oncogenes in mCRC

4.2.1 Recurrent focal events identified by plasma-Seq

Using plasma-Seq and our pre-defined criteria for focal SCNAs, we identified several recurrent focal events in a set of 150 mCRC patients (Figure 12A, Supp Table 1). Focal amplification of 12p12.1, 13q12.13-q12.3, and 8p11.23-p11.22 were found to be present in more than 5% of patients in our cohort (Figure 12B, Supp Table 1). To confirm our findings, we used publicly available SCNA data from 619 CRC patients included in the TCGA database (Figure 12A, Supp Table 2). Interestingly, we found that the 12p12.1 focal amplification, where *KRAS* is located, was present in around 6.7% of patients in our cohort, but in only 1.6% of patients of the TCGA cohort (P=0.002, Chi-squared; Figure 12B). This increase may be due to a higher representation of anti-EGFR therapy administered to our cohort, as prolonged anti-EGFR therapy can lead to the acquirement of a focal *KRAS* amplification. A focal amplification in 8p11.23-p11.22 was detected in both cohorts at a similar percentage, which reveals the frequent amplification of *FGFR1* in CRC (Figure 12B). Interestingly, recurrent focal amplification of 13q12.13-q12.3 was also identified in both cohorts, but the frequency in our cohort was much higher than that of the TCGA cohort (8.7% vs. 4.5%; P=0.043, Chi-squared; Figure 12B).



A

Cytoband	Type	Frequency (Our cohort;%)	Frequency (TCGA;%)	P-value	Potential driver genes	Genes
12p12.1	Gain	6.7	1.6	0.002	KRAS	CASC1,DD157417,KRAS,LYRM5
13q12.13-q12.3	Gain	8.7	4.5	0.043	RPL21, FLT3, FLT1	USP12,GSX1,GTF3A,LNX2,MTIF3,POLR1D, RASL11A,RPL21,SNORA27,SNORD102, PDX1,ATP5EP2,CDX2,PDX1,PRHOXNB, FLT3,PAN3,PAN3-AS1,FLT1,HV303147
8p11.23-p11.22	Gain	5.4	5.1	0.839	FGFR1	NSD3,LETM2,FGFR1,C8orf86,RNF5P1

B

Figure 12: Summary of focal SCNA events identified in our cohort and TCGA cohort [Figure and legend originally published in Genome Medicine (Zhou et al., 2020)]

A: Circos plot showing the recurrent focal events present in our mCRC cohort in comparison to the TCGA cohort. The three regions, including 12p12.1, 13q12.13-q12.3, and 8p11.23-p11.22,

were present in more than 5% of our cohort. Overlap between the two cohorts was seen on chr8 and 13, whereas our cohort had a higher percentage of 12p12.1 events when compared to the TCGA database, most likely the result of a higher frequency of patients treated with anti-EGFR therapy in our cohort.

B: Recurrent focal events from our patient cohort with a frequency higher than 5%. Potential driver genes were identified according to a machine learning–based method for driver gene prediction (Tokheim et al., 2016). The difference in these 3 recurrent focal events between our cohort and the TCGA cohort was analyzed using the Chi-squared test.

A recurrent focal amplification at 8p11.21 was detected at a higher frequency in the TCGA cohort than our cohort (6.1% vs. 1.3%; Supp Table 1 & 2). Of all 3 genes located in this region, *IKBKE* was reported to contribute to inflammation-related, angiogenic and metastatic processes in CRC (Agarwal et al., 2005). However, the reason as to why there was a higher representation in the TCGA dataset still needs further clarification. Some other known amplifications containing known driver genes, e.g., *ERBB2*, *EGFR* and *MYC*, and the deletion of tumor suppressor genes, e.g., *CDKN2A* and *MAP2K4*, were identified at a similar percentage in both cohorts (Supp Table 1 & 2). Another interesting finding was that the deletion of *PTEN* was detected to be lower than 1% in our cohort but present at ~3% in the TCGA dataset, which may be due to lower tumor content in our plasma cohort. However, a higher *PIK3CA* amplification was identified in our cohort (~2%) but was detected at lower than 1% in the TCGA dataset (Supp Table 1 & 2). As it is well-known that *PIK3CA* and *PTEN* both contribute to the PI3K/AKT/mTOR pathway and that increased function of *PIK3CA* and decreased function of *PTEN* have the same effect, this might potentially serve as an explanation for the different recurrent SCNAs of *PTEN* and *PIK3CA* between our cohort and the TCGA dataset.

The 13q12 amplification was presented at a high frequency in our cohort, but no known candidate driver gene was located in it. We therefore focused on a more detailed characterization of this region.

4.2.2 The 13q12.13-12.3 amplicon is associated with late-stage clinical features

As illustrated above, a 13q12.13-q12.3 focal amplification was frequently present in both cohorts but was higher in our patient group, which may be because we only recruited mCRC patients, whereas 55% of TCGA data represent localized stages (I and II). This suggests that the 13q12.13-12.3 amplicon may be more related to late-stage events of CRC. To test this, we compared the clinical features of patients with 13q12.13-q12.3 alterations and patients who did not harbor this SCNA in the TCGA dataset. As shown in Figure 13A, patients harboring 13q12.13-q12.3 SCNAs showed more late-stage disease (Stage III and Stage IV) ((P=1.23E-05; Chi-square), distant metastases (P=1.62E-04; Chi-square) and lymph node metastases (P=3.87E-05; Chi-square). One interesting finding was that presence of this SCNA was enriched in patients with primary tumors located in the rectum when compared with patients with primary tumors located in the colon (P=0.02013, Chi-squared). Hence, the TCGA analysis suggested that the 13q12.13-12.3 amplicon may be associated with late-stage clinical features.

To test this, we applied CNV-seq to the corresponding FFPE tumor tissue samples in 9 of 14 affected patients. As shown in Table 5, a focal amplification was already present in the primary tumor tissue in 4 patients (C123, C109, C178, and C74), whereas in 5 patients (C240, C118, C79, C206, and C166), this focal amplification was acquired at a later time point (Table 5, Figure 13B) (Zhou et al., 2020). In 3 of 5 patients with an undetected amplification in the primary tumor tissue (C240, C118, C79), the 13q12.13-q12.3 focal amplification appeared in one of the plasma DNA analyses after metastases were detected or the patient developed progressive disease (Table 5, Figure 13B) (Zhou et al., 2020). Furthermore, in one additional patient (C216), from whom an FFPE sample was not accessible, the amplification was not present in the first plasma sample but rather acquired after the patient exhibited progressive disease (Figure 13C) (Zhou et al., 2020).

Patient ID	Disease Stage (TC)	Disease Stage (PC)	Location	cM (TC)	cM (PC)	TBC (months)	Detection*
C74	IV B	IV B	Rectum	M1	M1	22.8	Yes
C79	NA	NA	NA	NA	NA	NA	No
C95	II A	IV B	Left flexure	M0	M1	43.4	NA
C109	III B	IV B	Rectum	M0	M1	101.7	Yes
C110	IV A	IV B	Colon ascending	M1	M1	0.7	NA
C118	IV B	IV B	Colon sigmoid	M0	M1	22.9	No
C112	III A	IV B	Rectum	NA	NA	27.1	NA
C123	III B	IV B	Colon sigmoid	M0	M1	76.5	Yes
C129	I	IV B	Colon sigmoid	M0	M1	121.3	NA
C166	IV A	IV A	Colon transverse	M1	M1	0.7	No
C178	III C	IV B	Cecum	M0	M1	5.4	Yes
C206	III B	III B	Rectum	M0	M0	7.6	No
C240	III B	IV B	Colon sigmoid	M0	M1	79.9	No
C216	NA	IV B	NA	NA	NA	NA	NA

Table 5: Summary of clinical information of all cases harboring 13q12.2 focal amplification

[Table and legend originally published in Genome Medicine (Zhou et al., 2020)]

*13q12.2 focal amplification detected in primary tumor tissue or not.

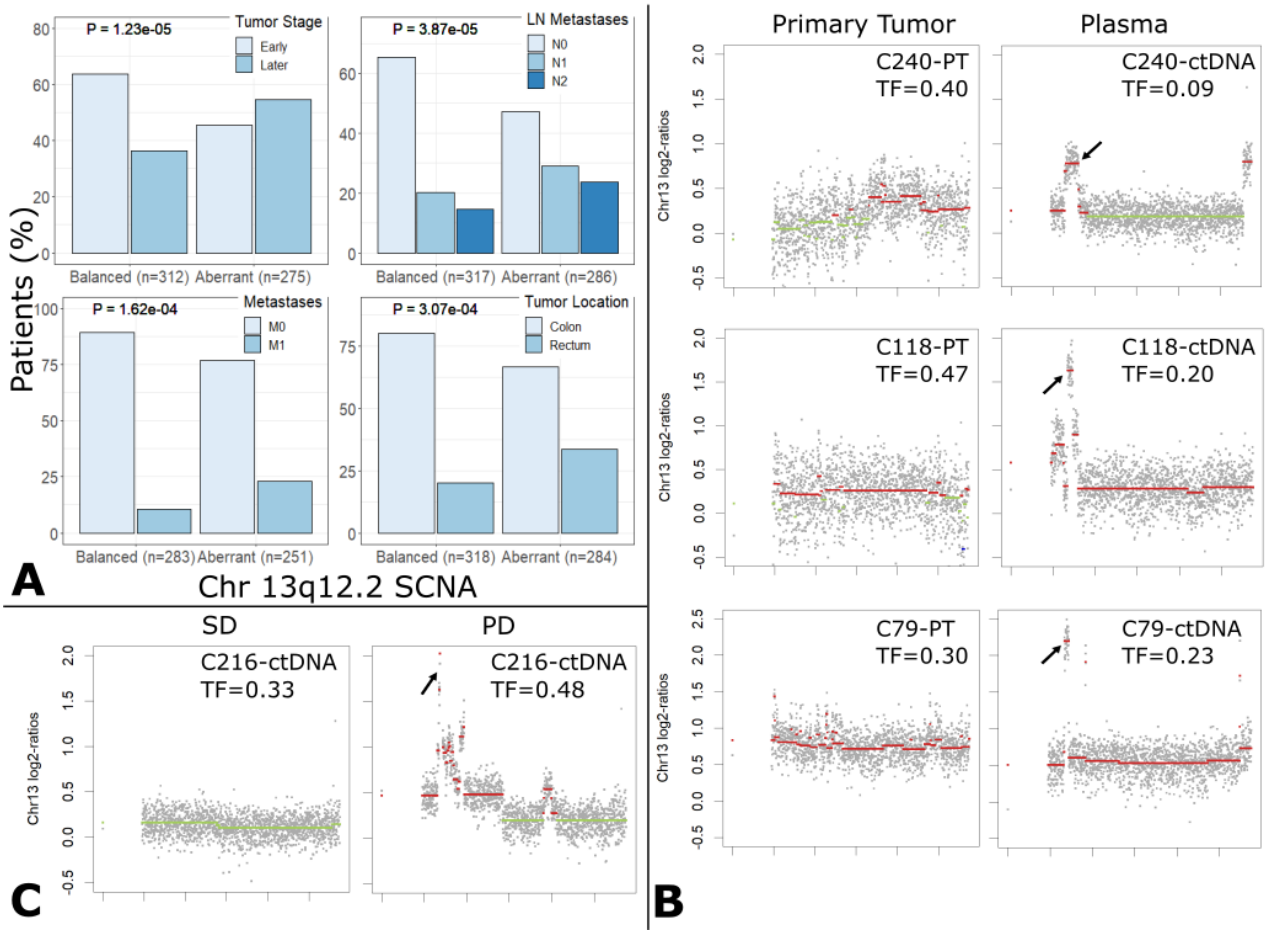


Figure 13: The 13q12.13-12.3 amplicon is associated with late-stage clinical features [Figure and legend originally published in Genome Medicine (Zhou et al., 2020)]

A: The TCGA cohort was separated into balanced and aberrant (including gain and amplification cases) groups, and statistically significant differences were detected in 4 clinical features, i.e., tumor stage, distant metastasis, lymph node metastasis and tumor location, as shown in the bar charts. P-values were calculated using the Chi-squared test.

B: Plots illustrating the log2-ratio changes on chromosome 13 for the primary tumor (PT) and one plasma sample (ctDNA). In patients C240, C118 and C79, the 13q12.2 focal amplification was not identified in the primary tumor but found in the plasma when patients developed progressive disease. Copy number gains are shown in red and balanced regions in green. Tumor fraction (TF) of every sample was estimated with ichorCNA (Adalsteinsson et al., 2017).

C: Plots illustrating the log2-ratio changes on chromosome 13 from two plasma samples. In C216, the 13q12.2 amplification was not detected in the first plasma sample, but later when the disease progressed. Copy number gains are shown in red and balanced regions in green. Tumor fraction (TF) of every sample was calculated using ichorCNA (Adalsteinsson et al., 2017) . (SD = stable disease; PD = progressive disease)

4.2.3 Definition of the minimally 13q12 amplified region and involved genes

The aforementioned data suggested that the 13q12.2 amplicon is associated with late-stage and progressive disease. But the driver gene in this genomic range was unknown. To investigate this, we first determined the minimal overlapping range of all the focal events in our patient cohort. By calculating the median log2-ratio of chr13 against all 14 patients with 13q12.2 focal amplification, an overlapping peak was selected by excluding the remaining areas of chromosome 13 (Figure 14). We then mapped all focal events of these 14 patients to this peak range and counted the frequency of each pre-defined 50kb bins covered (Figure 14). Using the bin that demonstrated the lowest frequency, we calculated a p-value in order to identify the statistically significant minimal overlapping range. We identified a broad peak ($p < 0.05$, Fisher's exact; chr13:27,708,804-28,667,235) and a focal peak ($p < 0.01$, Fisher's exact; chr13:28,441,650-28,667,235) in our patient cohort. Subsequently, we applied the same method to the TCGA cohort (broad peak: chr13:28,197,436-28,650,763; focal peak: chr13:28,382,214-28,604,579; Figure 14). Peaks identified in TCGA dataset were similar with the GISTIC analysis output (broad peak: chr13:28,192,985-28,773,237; focal peak: chr13: 28,391,954-28,558,679; Figure 14) (Mermel et al., 2011). To the end, seven genes (*POLR1D*, *GSX1*, *PDX1*, *ATP5EP2*, *CDX2*, *PRHOXNB*, and *FLT3*) were completely located within the broad peak for all three analyses (Figure 14).

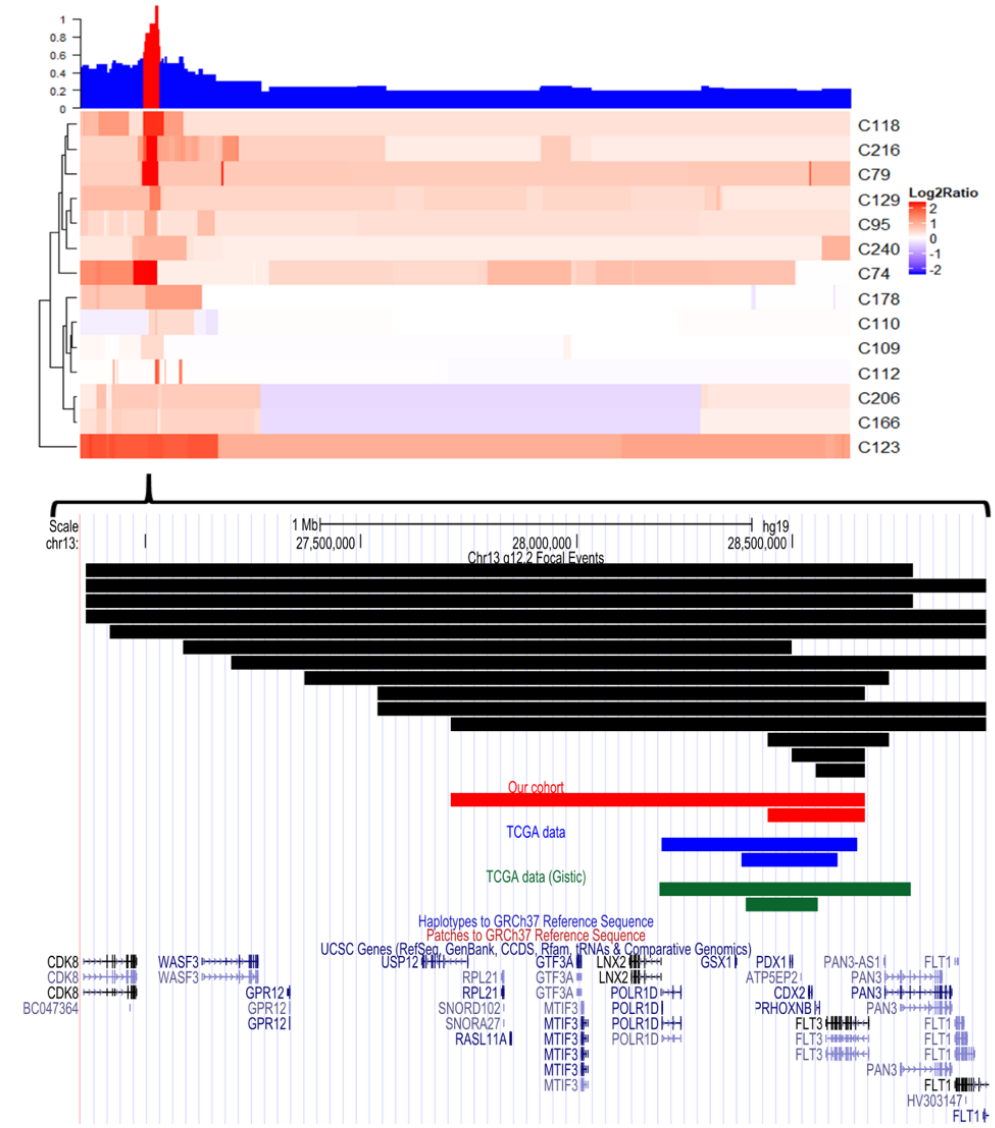


Figure 14: Definition of the minimally 13q12 amplified region [Figure and legend originally published in Genome Medicine (Zhou et al., 2020)]

Plots demonstrating the process of defining the minimally 13q12 amplified region (top) and involved genes (bottom). Top: The heatmap indicates the log2-ratios of chromosome 13 and the bar plot above illustrates the median log2-ratios for all 14 patients. Red bars: The selected peak. Bottom: Black bars: all focal events from our cohort. Red bars: Broad and focal peak of our dataset using an in-house method. Blue bars: Broad and focal peak of TCGA dataset using an in-house method. Green bars: Broad and focal peak of our dataset using the GISTIC method (Mermel et al., 2011).

4.2.4 The oncogene *FLT3* is not associated with proliferation of CRC cells

Seven genes (*POLR1D*, *GSX1*, *PDX1*, *ATP5EP2*, *CDX2*, *PRHOXNB*, and *FLT3*) were completely located in the minimally amplified region. *FLT3* is a well-known driver gene in hematological malignancies, which can be specifically targeted by sorafenib (Zauli et al., 2012). *FLT3* was considered to be a reasonable driver candidate. Therefore, we first checked whether *FLT3* plays a role in CRC oncogenesis. We checked for an association of gene expression and copy number of *FLT3* in the TCGA or the CCLE CRC cell line dataset. Consistent with previous reports (Braxton et al., 2016), we did not observe a correlation of mRNA expression and copy number of *FLT3* (Figure 15A & B), suggesting that *FLT3* is not the driving event of the 13p12.2 amplification. To confirm this observation, we generated a CRC cell line stably expressing FLT3, as almost no CRC cell line showed a high expression of FLT3. As expected, overexpression of FLT3 in OXCO-2 cells did not lead to a significant change in cell proliferation (Figure 15C). As a summary, these results suggested that *FLT3* may not function as a driver gene in 13q12.2.

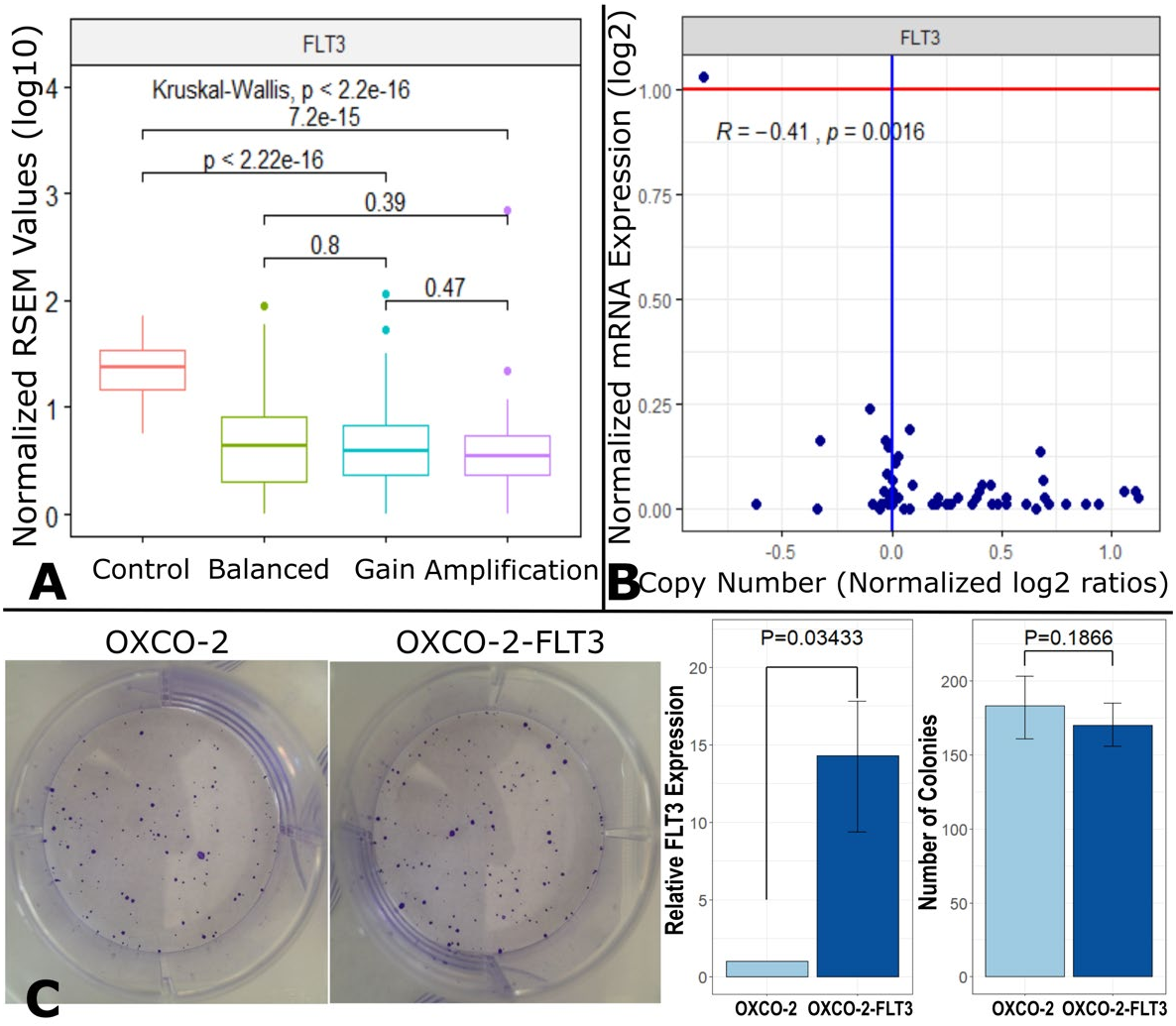


Figure 15: The oncogene *FLT3* is not associated with proliferation of CRC cells [Figure and legend originally published in Genome Medicine (Zhou et al., 2020)]

A: Box plot indicating no significant correlation between *FLT3* gene copy number and *FLT3* mRNA expression (log10(normalized RSEM value + 1)) in the TCGA cohort (Control/matched normal tissue: n=51; Balanced: n=196; Gain: n=129; Amplification: n=46).

B: The scatter plot indicates no correlation in *FLT3* copy number and *FLT3* mRNA expression (log2 (TPM + 1)) in 58 CRC cell lines (R=-0.41, P=0.0016; Pearson). The red line represents the noise threshold. (TPM = 1)

C: OXCO-2-FLT3 cell line showed significant overexpression of FLT3 (P=0.03433, t-test) but no significant changes were identified in cell proliferation (colony formation assay; P=0.1866, t-test).

4.2.5 Identification of *POLR1D* as a potential driver gene in 13q12.2

The abovementioned results suggested that *FLT3* may not function as a driver gene in CRC. To identify the potential driver gene in 13q12.2, we investigated the other 5 candidate genes (pseudogene *ATP5EP2* was excluded), which were fully located in the overlapping broad peak regions as well as the first immediate upstream and first immediate downstream genes, *LNX2* and *PAN3*, respectively. After testing the association between copy number and mRNA expression in the TCGA dataset and the CCLE dataset (Figure 16), we identified 5 genes that demonstrated a positive correlation: *LNX2*, *POLR1D*, *CDX2*, *PDX1* and *PAN3* (Figure 16A & B).

To further characterize a potential involvement of these genes *in vitro*, we induced a transient siRNA knockdown of these five genes in the two CRC cell lines HT29 and SW480, where 13q12.2 is overrepresented either due to a focal amplification (HT29) or gain of the entire chromosome 13 (SW480) (Figure 7). To this end, only the silencing of *POLR1D* showed a significant reduction in cell viability in both cell lines (Figure 17A & B). For validation, we silenced *POLR1D* in HT29 and SW480 with 3 different siRNAs and created a 72-hour time curve by monitoring the cell viability every day. As shown in Figure 17C, after 72h, a significant reduction of cell viability was found in both cell lines compared with the control group.

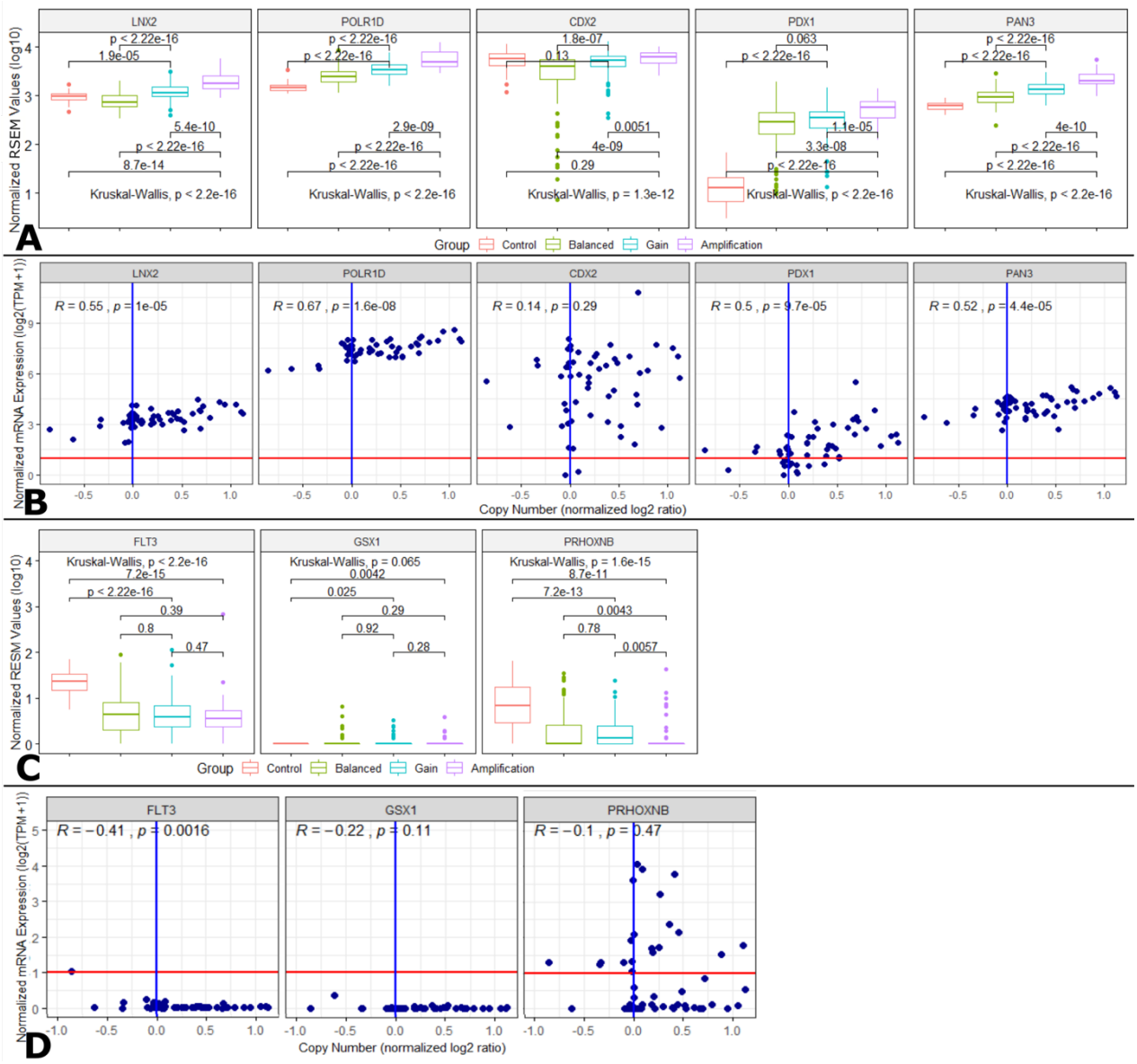


Figure 16: Validation of the relationship between gene copy number and gene expression in all candidate genes in TCGA and CCLE dataset [Figure and legend originally published in Genome Medicine (Zhou et al., 2020)]

A: Box plots showing significant positive correlation between gene copy number and mRNA expression (log10(normalized RSEM value + 1)) in 5 genes (i.e., *LNX2*, *POLR1D*, *CDX2*, *PDX1* and *PAN3*) in the TCGA cohort. (Control/matched normal tissue: n=51; Balanced: n=196; Gain: n=129; Amplification: n=46)

B: Scatter plots showing a positive correlation in gene copy number and mRNA expression (log2 (TPM + 1)) in 5 genes (i.e. *LNX2*, *POLR1D*, *CDX2*, *PDX1* and *PAN3*) in 58 CRC cell lines. R values and P values were calculated with the Pearson correlation test. The red line represents the noise threshold (TPM = 1).

C: No significant correlation between gene copy number and mRNA expression (log10(normalized RSEM value + 1)) was found in *FLT3*, *GSX1*, and *PRHOXNB* genes in the TCGA cohort. (Control/matched normal tissue: n=51; Balanced: n=196; Gain: n=129; Amplification: n=46)

D: No significant correlation between gene copy number and mRNA expression (log2 (TPM + 1)) in *FLT3*, *GSX1*, and *PRHOXNB* genes in 58 CRC cell lines. R values and P values were calculated with the Pearson correlation test. The red line represents the noise threshold. (TPM = 1)

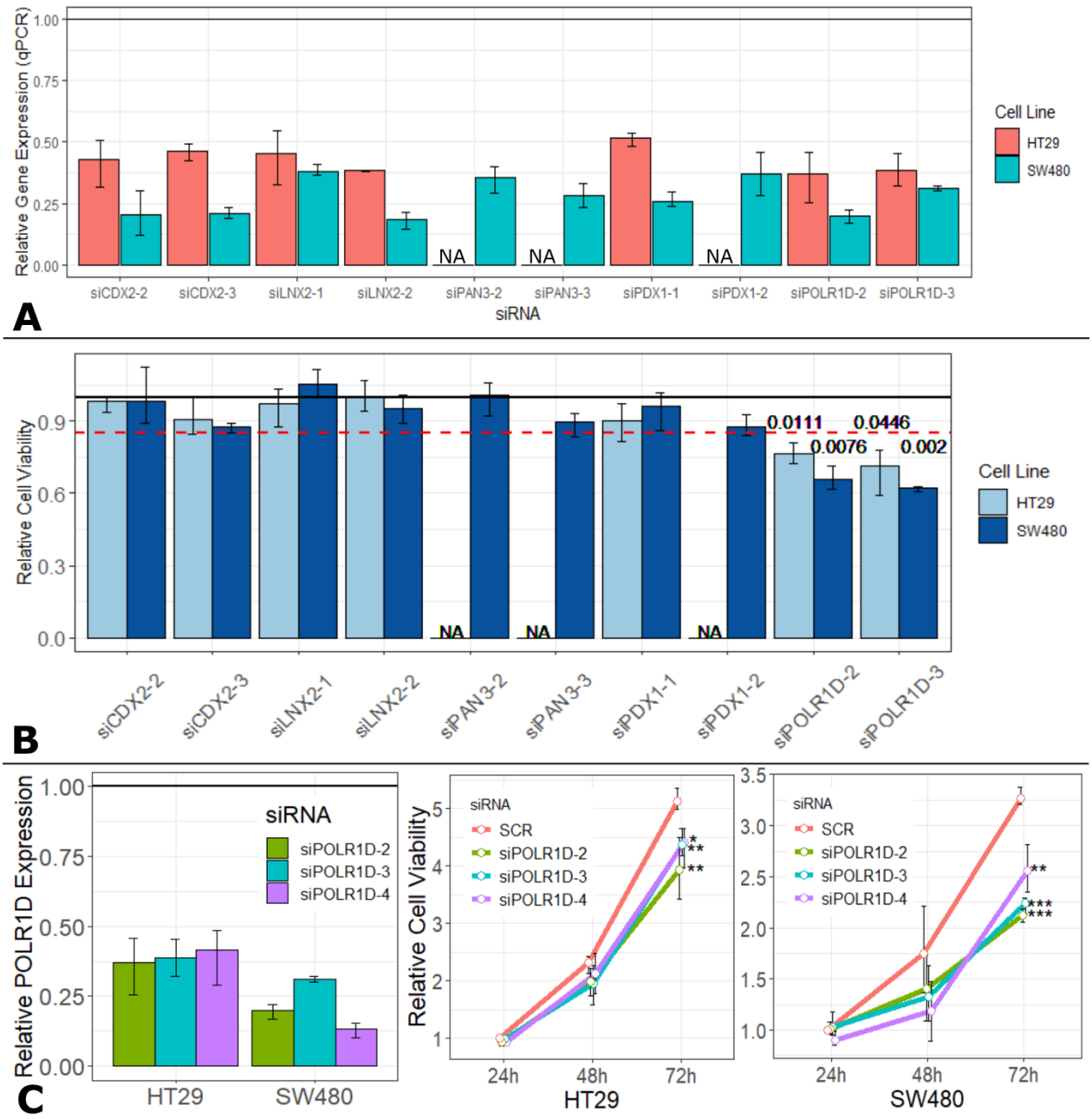


Figure 17: Silencing of *POLR1D* inhibited cell viability in HT29 and SW480 cells [Figure and legend originally published in Genome Medicine (Zhou et al., 2020)]

A: Bar plot showing the relative gene expression of respective genes after silencing. RT-PCR showing that silencing provided sufficient knockdown.

B: Bar plot showing the relative cell viability changes after silencing of different genes. Only silencing of *POLR1D* showed significant reduction of cell viability in both cell lines.
C: Left: Silencing of *POLR1D* with 3 different siRNA constructs. RT-PCR showing that silencing provided sufficient knockdown of POLR1D expression in both cell lines. Right: Cell viability time curve illustrating significant reduction of cell viability after knockdown of POLR1D expression in HT29 and SW480 cells (*: $P < 0.1$; **: $P<0.05$; ***: $P<0.01$, t-test).

These results indicated that *POLR1D* may function as a driver gene in CRC by affecting cell proliferation. POLR1D is a subunit of both RNA polymerases I and III. RNA polymerase I is involved in the production of 18S, 5.8S, and 28S rRNAs, while RNA polymerase III synthesizes small RNAs (Martinez-Calvillo et al., 2007). Despite a recent report describing the frequent overexpression of POLR1D in CRC (Wang, M. et al., 2018), the relationship between POLR1D and CRC has not been investigated in the literature (Zhou et al., 2020).

4.2.6 POLR1D affects expression of VEGFA and EREG

As described before, *POLR1D* may be a potential driver gene in 13q12.2 in CRC. We sought to elucidate the underlying oncogenic mechanisms of *POLR1D* using RNA-seq analysis. We applied mRNA-seq to our *POLR1D*-silenced and control cells and found 45 genes (POLR1D showed ~2-fold reduction in expression in the silenced cells; Supp Table 3) which were differentially expressed in both the HT29-POLR1D-silenced and SW480-POLR1D-silenced cell lines compared to the controls. Moreover, for 8 of these 45 genes (i.e., *PPP1R15A*, *MOSPD2*, *FAM84B*, *GARS*, *POLR1D*, *KIF21B*, *VEGFA*, and *EREG*) gene expression changes could also be identified in the TCGA mRNAseq data sets (Figure 18 & 19, Table 6). Hence, these 8 genes showed consistent reductions in 2 silenced cell lines and demonstrated upregulation in expression in patients harboring 13q12 SCNA compared to patients with a balanced 13q12 region.

Except *POLR1D*, all 7 genes (i.e. *PPP1R15A*, *MOSPD2*, *FAM84B*, *GARS*, *KIF21B*, *VEGFA*, *EREG*) have demonstrated involvement in the progression of various cancers. For instance, *FAM84B*, which encodes family with sequence similarity 84, member B protein, was reported to be involved in the progression of prostate cancer and esophageal squamous cell carcinoma (Cheng et al., 2016, Wong et al., 2017). *GARS* encodes glycyl-tRNA synthetase and has been shown to be involved in neddylation, a post-translational modification that controls cell cycle and proliferation and thus may play a role in cancer progression (Mo et al., 2016). *KIF21B* encodes a member of the kinesin superfamily and was described to be significantly associated with poor prognosis of prostate cancer patients (Arai et al., 2019). Mutation of *PPP1R15A*, which encodes protein phosphatase 1 regulatory subunit 15A, has been shown to be a valuable biomarker for mCRC patients sensitive to bevacizumab regimens (Roh et al., 2016). *MOSPD2* encodes motile sperm domain-containing protein 2 and has recently been reported to promote the metastasis of breast cancer (Salem et al., 2019). (Zhou et al., 2020)

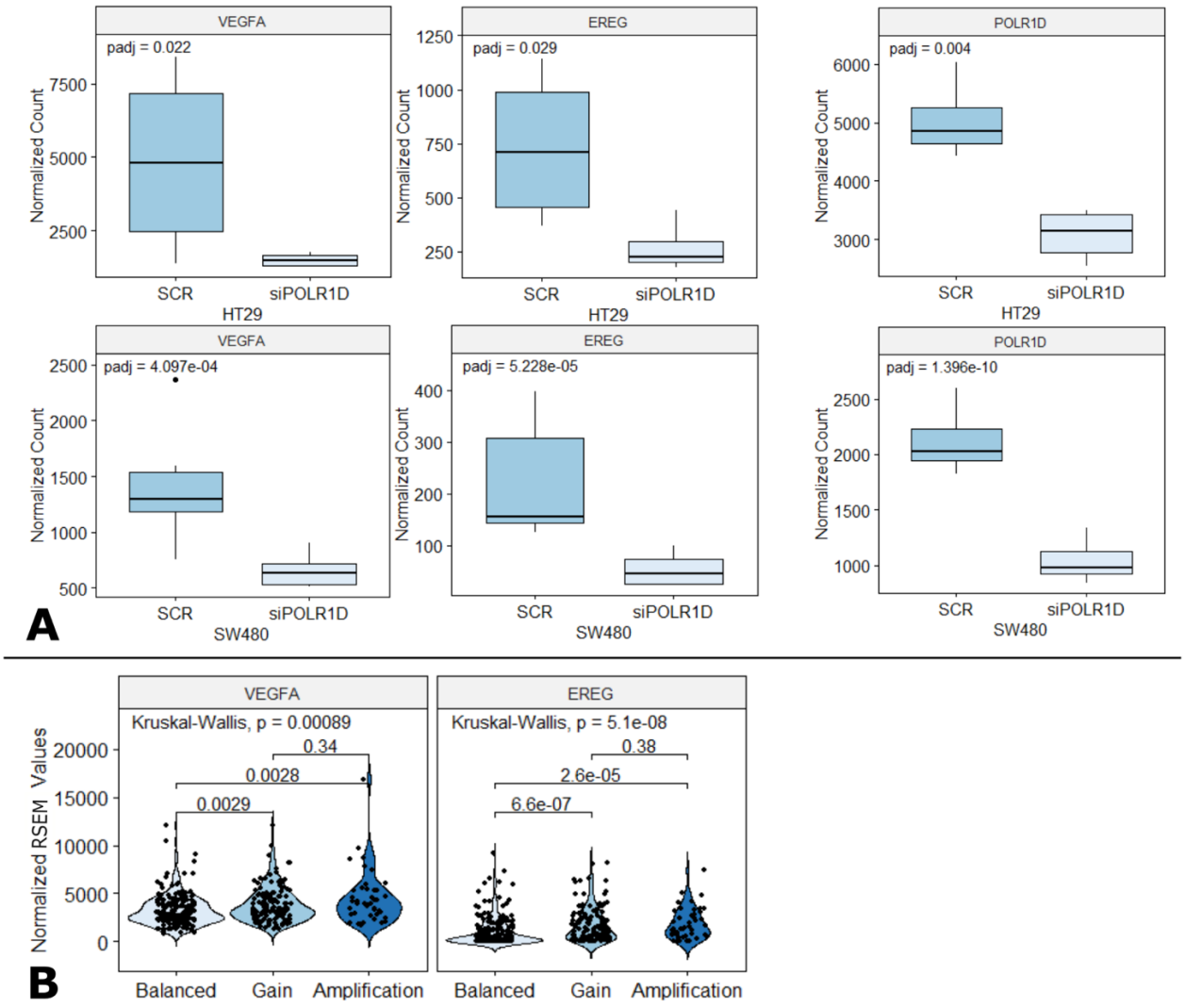


Figure 18: POLR1D affects expression of VEGFA and EREG [Figure and legend originally published in Genome Medicine (Zhou et al., 2020)]

A: Box plot illustrating the different expression of VEGFA, EREG, and POLR1D between scrambled siRNA control (SCR) and POLR1D silenced SW480 (SCR: n=6; siPOLR1D2: n=3; siPOLR1D3: n=3) and HT29 (SCR: n=4; siPOLR1D2: n=2; siPOLR1D3: n=2) cell lines. Gene expression was compared by normalized DESeq2 read counts. VEGFA and EREG expression were suppressed after POLR1D silencing. Adjusted p-values were calculated by DESeq2, an R package. **B:** Violin plots showing the different expression in VEGFA and EREG expression in 13q12.2 balanced (n=196), gain (n=129), and amplification (n=46) groups. Gene expression was

represented by normalized RSEM value. VEGFA and EREG expression were significantly increased in gain and amplification group compared with the balanced group.

Gene	HT29 Log2FC	HT29 Padj*	SW480 Log2FC	SW480 Padj*	TCGA P value+	TCGA Balanced Median	TCGA Aberrant Median
POLR1D	-0.71	0.004	-1.03	1.4E-10	5.0E-26	2514	3794
EREG	-1.45	0.029	-2.09	5.2E-05	9.5E-09	373	1302
VEGFA	-1.69	0.022	-1.11	4.1E-04	2.8E-04	2939	3403
GARS	-0.86	0.023	-0.96	6.1E-08	9.2E-03	3417	3774
FAM84B	-0.66	0.010	-0.80	4.6E-03	4.6E-02	1522	1709
PPP1R15A	-2.40	0.048	-2.09	1.2E-10	1.6E-02	1668	1846
KIF21B	-0.85	0.025	-1.36	1.2E-06	4.4E-02	645	710
MOSPD2	-0.98	0.025	-0.91	6.2E-04	2.3E-04	238	287

Table 6: Summary of genes differently expressed after *POLR1D* knockdown in both HT29 and SW480 cells and which showed consistent results in the TCGA dataset [Table and legend originally published in Genome Medicine (Zhou et al., 2020)]

Eight genes showed consistent expression change in HT29 and SW480 cell lines after POLR1D knockdown, and TCGA dataset, i.e., between 13q12.2 balanced and aberrant groups. Log2 fold change and P-value are listed in the table. (Log2FC: log2 fold change of gene expression (Log2 (siPOLR1D/SCR)). *: adjusted P-value calculated by DESeq2 package. +: P-value calculated using the Wilcoxon test.)

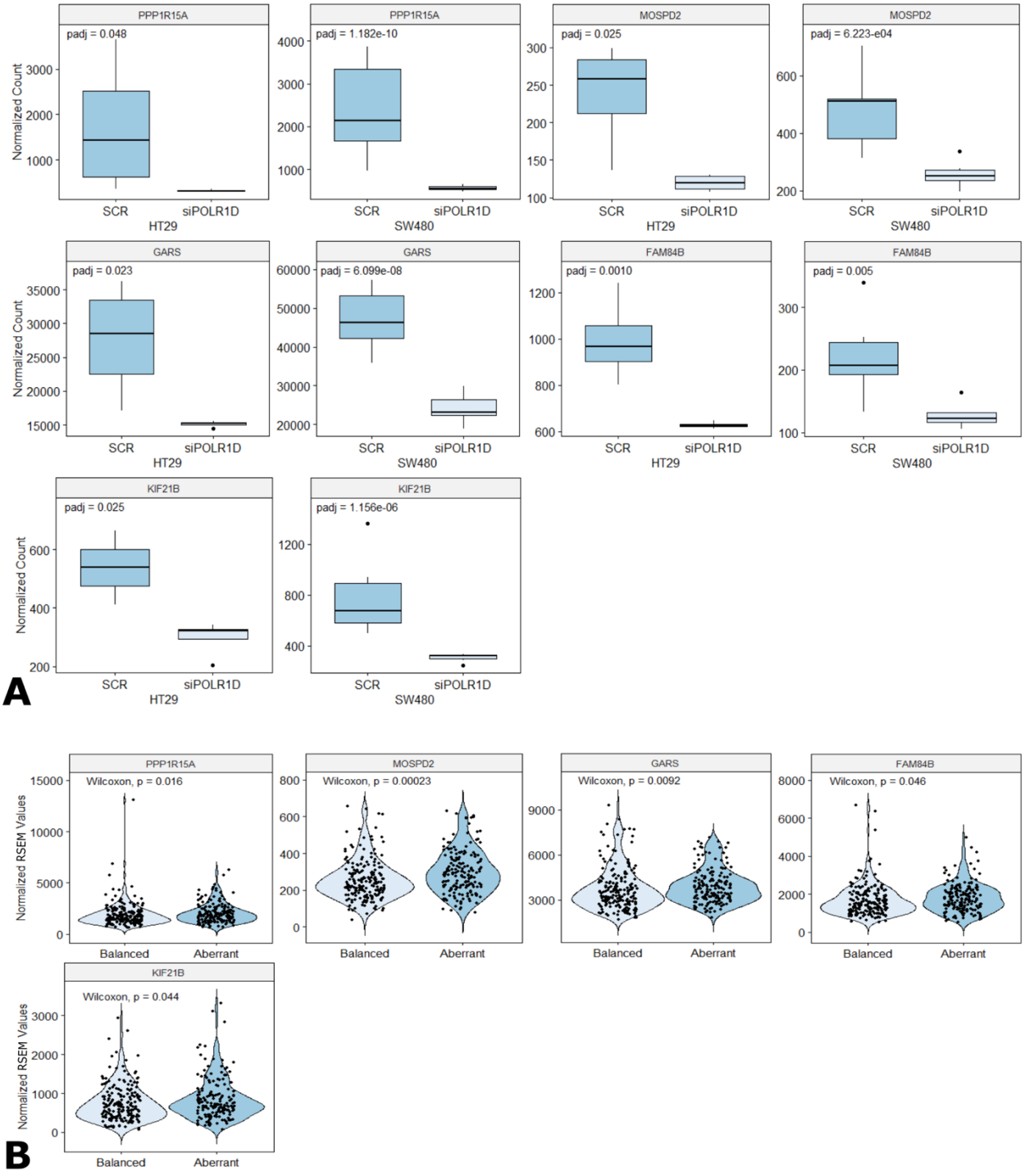


Figure 19: POLR1D affects expression of PPP1R15A, MOSPD2, GARS, FAM84B, and KIF21B [Figure and legend originally published in Genome Medicine (Zhou et al., 2020)]

A: Box plot illustrating the different expression of PPP1R15A, MOSPD2, GARS, FAM84B, and KIF21B between scrambled siRNA control (SCR) and POLR1D silenced SW480 (SCR: n=6; siPOLR1D2: n=3; siPOLR1D3: n=3) and HT29 (SCR: n=4; siPOLR1D2: n=2; siPOLR1D3: n=2) cell lines. Gene expression was represented by normalized DESeq2 read count. PPP1R15A, MOSPD2, GARS, FAM84B, and KIF21B expression were suppressed after POLR1D silencing. Adjusted p-values were calculated by DESeq2, an R package.
B: Violin plots showing the different expression in PPP1R15A, MOSPD2, GARS, FAM84B, and KIF21B expression in 13q12.2 balanced (n=196) and aberrant (n=175) groups. Gene expression was represented by normalized RSEM value. PPP1R15A, MOSPD2, GARS, FAM84B, and KIF21B expression were significantly increased in aberrant group compared with the balanced group.

Of particular relevance appeared to be *VEGFA* and *EREG*. *VEGFA* encodes vascular endothelial growth factor A, which is an important regulator of angiogenesis and plays a role in the development, progression and metastasis of CRC (George et al., 2001). Recent publications have shown that bevacizumab, an anti-VEGF monoclonal antibody approved for the treatment of advanced CRC by the FDA, induces autocrine VEGF signaling, which increases the expression of VEGF in the tumor cells and may contribute to metastasis and progressive disease (Hammond, Swaika & Mody, 2016, Mesange et al., 2014). Another interesting gene is *EREG*, encoding epiregulin, a member of the epidermal growth factor (EGF) family, which can bind to and activate EGFR and the structurally related erb-b2 receptor tyrosine kinase 4 (ERBB4) (Shirakata et al., 2000). Higher EREG expression is considered as a sign of higher activation of the EGFR pathway, which, in turn, means better response to anti-EGFR treatment (Pentheroudakis et al., 2013) but more resistant to drugs that target the other ERBB family members, e.g., ERBB2 (Auf et al., 2013, Takahashi et al., 2016). Epiregulin has been reported to be involved in the progression of various cancers (Takahashi et al., 2016, Auf et al., 2013).

4.2.7 Nucleosome positioning mapping to infer POLR1D expression in plasma

Although the TCGA data demonstrated a nice correlation between *POLR1D* SCNA and gene expression, the observation of a novel amplification in ctDNA does not allow the conclusion that the genes located within the amplified region are actually expressed in our patients. As described before, cfDNA consists predominantly of nucleosome-protected DNA shed into the bloodstream by cells undergoing apoptosis (Lo et al., 2010, Diehl et al., 2005). Distinct nucleosome footprints of TSSs can be used to identify expressed genes (Valouev et al., 2011). Our recent study reported that TSS-nucleosome occupancy patterns in cfDNA could be used to predict which cancer driver genes in regions with somatic copy number gains are expressed with high accuracy (Ulz et al., 2016). To test the expression status of genes in 13q12.2, we selected two plasma samples with focal amplification on 13q12.2 from two different patients (C129-8 and C216-12) to conduct our previously described TSS profiling analyses (Ulz et al., 2016).

These analyses predicted that all 5 genes (*POLR1D*, *LNX2*, *CDX2*, *PDX1*, and *PAN3*) with a positive correlation between SCNAs and mRNA expression in TCGA dataset, were expressed in both plasma samples (Figure 20). In contrast, *FLT3* was predicted as unexpressed in C216-12 and inconclusive in C129-8 (Figure 20). Hence, we confirmed in our cohort that tumors with focal 13q12.2 amplification indeed express POLR1D but not FLT3. This result also confirmed our previous finding that *FLT3* is not the cancer driver gene within this amplicon in colorectal cancer.

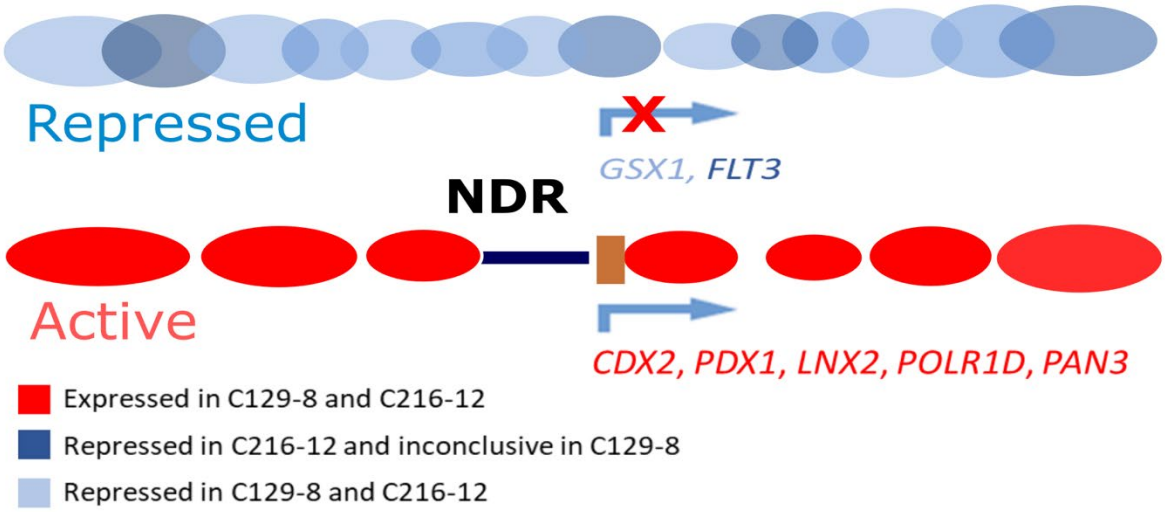


Figure 20: TSS-nucleosome occupancy patterns predict the expression of 13q12.2 amplicon located genes [Figure and legend originally published in Genome Medicine (Zhou et al., 2020)] Schematic indicating how TSS-nucleosome occupancy patterns differ between repressed and active genes. Expressed genes have a nucleosome-depleted region (NDR, dark blue line), and the nucleosome organization is well-defined. In contrast, the nucleosome organization of silenced genes is not well-defined, and often no NDR can be identified. With our previously published method (Ulz et al., 2016), we determined the expression status of genes within the 13q12.2 amplicon. *POLR1D*, *LNX2*, *CDX2*, *PDX1*, and *PAN3* are predicted to be expressed (red), whereas *FLT3* is predicted to be unexpressed in C216-12 and inconclusive in C129-8 (dark blue). In addition to the genes discussed in the text, we added the gene *GSX1* (light blue) as an example of a repressed gene (Part of the figure adapted from (Valouev et al., 2011)).

4.2.8 The emergence of *POLR1D* amplification correlates with resistance to bevacizumab

Serial monitoring of tumor genomes by plasma DNA analyses may reveal focal amplifications, which may harbor oncogenes as a mechanism of resistance to administered therapies in CRC. Due to the finding that patients with 13q12.2 amplification showed higher expression of VEGFA and that high VEGFA expression has been reported to be involved in the resistance to bevacizumab, tumors with a 13q12.2 amplification might be more resistant to bevacizumab. Therefore, we serially monitored plasma from two patients, i.e., C216 and C129, who both were treated with bevacizumab.

Patient C216 did not harbor 13q12.2 focal amplification in the first blood draw (C216-7; Figure 21A & B). After 7 months of bevacizumab treatment, we detected the emergence of the 13q12.2 focal amplification in sample C216-11, which correlated with the development of progressive disease and resistance to bevacizumab, i.e. increase in serum CEA and CA19-9 tumor markers (Figure 21A & B) and increase in the lesion size of liver metastases (Figure 21C). The treatment with bevacizumab was continued for another 2 months and a further increase of the log2-ratio of the 13q12.2 focal amplification was detected in the last sample (C216-12) as well as continued progression of the disease, as evidenced by increasing blood CEA and CA19-9 levels (Figure 21A & B) and the lesion size of liver metastases (Figure 21C). Digital PCR was used on all serial samples to confirm the changes of *POLR1D* copy number under bevacizumab treatment (Figure 21B).

In patient C129, a 13q12.2 amplification was identified in the first plasma sample (Figure 22A & B). After treatment with anti-EGFR for 9 months, the patient showed acquired resistance with escalated serum CEA values (Figure 22A & B), an enlarged lung metastasis lesion and appearance of pleural effusion in the right lung (Figure 22C). The progression of the disease was consistent with the changes in ctDNA, i.e. a novel focal amplification on 17q12, including *ERBB2*, which represents an established mechanism of resistance to anti-EGFR therapy (Figure 22A & B) (Bertotti et al., 2011). Moreover, the focal amplification in 13q12.2 detected in the first sample disappeared (Figure 22A & B). After switching to anti-VEGF treatment for 5 months, the amplification on 13q12.2 appeared again, whereas the amplification on 17q12 vanished, which was

accompanied by increasing CEA levels (Figure 22A & B), an enlarged lung metastasis lesion and increasing pleural effusion in the right lung (Figure 22C). Again, digital PCR was applied to all samples to validate copy number amplification detected by plasma-Seq in order to confirm the apparent clonal switch between *POLR1D* and *ERBB2* (Figure 22B).

In summary, our results indicated that *POLR1D* may function as a potential driver gene in the 13q12.2 amplification and may be involved in tumor progression by increasing the expression of VEGFA and EREG. Due to the higher expression of VEGFA, amplification of 13q12.2 may be involved in the acquired resistance to bevacizumab. Moreover, ctDNA can be used to identify novel oncogenes in mCRC.

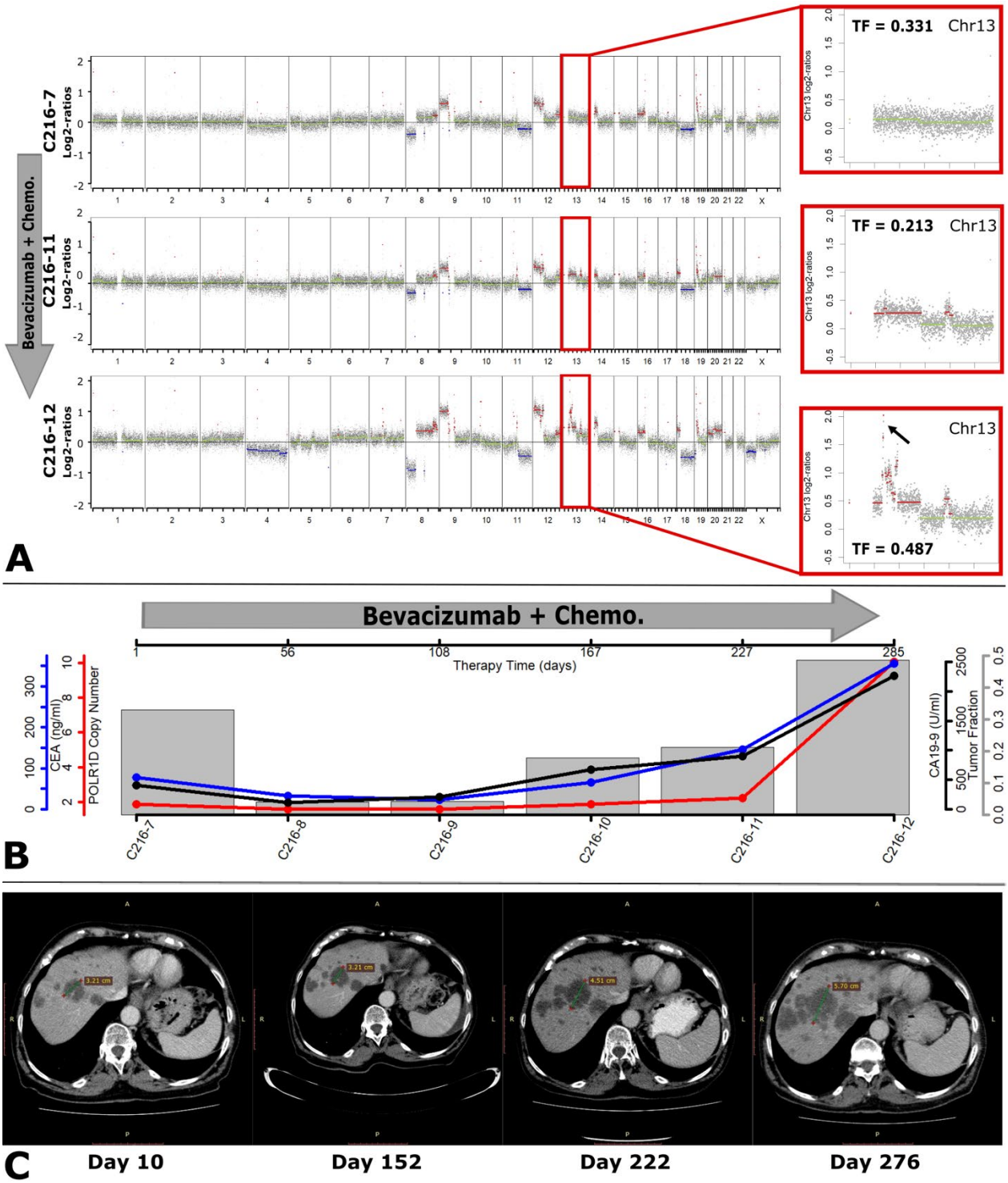


Figure 21: Emergence of the 13q12 amplicon under bevacizumab treatment in patient C216

[Figure and legend originally published in Genome Medicine (Zhou et al., 2020)]

A: Genome-wide log2-ratio plots of plasma samples from C216 obtained before bevacizumab treatment (1st), after 227 days (2nd), and after 285 days of bevacizumab treatment (3rd). The insets illustrate the respective tumor fraction (TF) for each analysis and enlarged log2-ratio plots of chromosome 13; the arrow represents the region that harbors the 13q12.2 amplification, which can also be detected in the second sample. Copy number gains are shown in red, balanced regions in green, and copy number losses in blue.

B: Plot illustrating all time points of blood collection and relative marker changes. Red line: *POLR1D* copy number changes identified by digital PCR. Blue line: Serum CEA level changes. Black line: Serum CA19-9 level changes. Gray bar: Tumor content identified in every sample using ichorCNA.

C: CT images obtained in 4 different time points, i.e. day 10, day 152, day 222 and day 276 after bevacizumab treatment. Compared to the first image, no significant changes were identified on day 152. On day 222, the pre-present liver metastasis lesions enlarged with occurrence of new micrometastasis lesions. On day 276, all liver metastasis lesions had become larger.

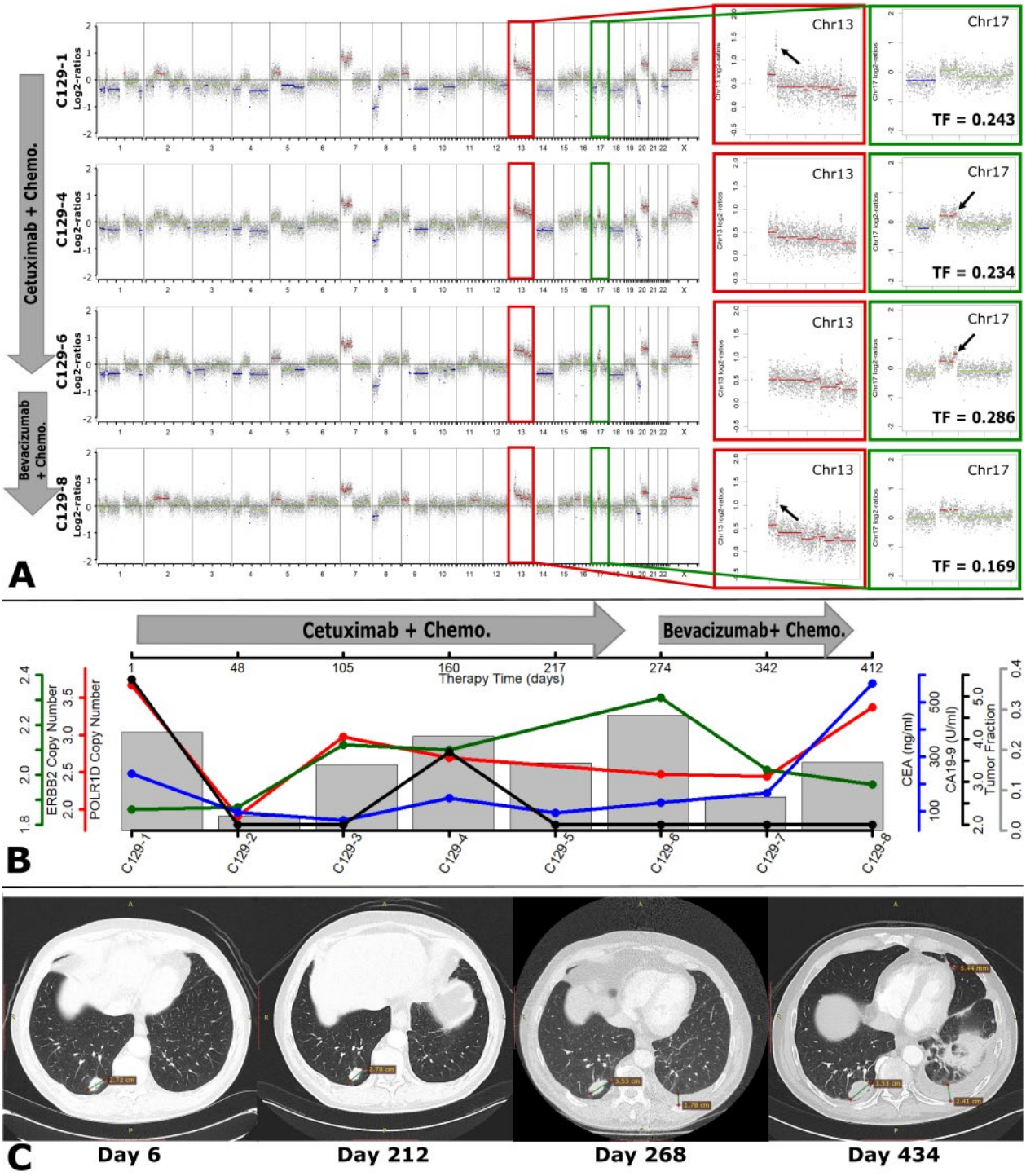


Figure 22: Alternating *POLR1D* and *ERBB2* amplifications in serial plasma analyses of patient C129 [Figure and legend originally published in Genome Medicine (Zhou et al., 2020)]

A: Genome-wide log2-ratio plots of plasma samples from C129 obtained before treatment with cetuximab (1st), 160 days (2nd) after cetuximab treatment, before treated with bevacizumab (3rd) and 138 days (4th) after bevacizumab treatment. The insets illustrate the respective tumor fraction (TF) for each analysis and enlarged log2-ratio plots of chromosome 13 and 17; the arrow represents the region that harbors the 13q12.2 and *ERBB2* amplification. Copy number gains are shown in red, balanced regions in green, and copy number losses in blue.

B: Plot illustrating all time points of blood collection and relative marker changes. Red line: *POLR1D* copy number changes identified by digital PCR. Green line: *ERBB2* copy number changes identified by digital PCR. Blue line: Serum CEA level changes. Black line: Serum CA19-9 level changes. Gray bar: Tumor content identified in every sample using ichorCNA.

C: Two CT images obtained on day 6 and day 212 of cetuximab treatment, and another two CT images obtained before bevacizumab treatment (Day 268) and 160 days after bevacizumab treatment (Day 434). No significant changes were identified on day 212. On day 268, the pre-present lung metastasis lesion became larger in the right lung, and pleural effusion appeared in the left lung. On day 434, this pre-present lesion became larger, and new metastasis lesions appeared with increasing of pleural effusion.

5 DISCUSSION

Colorectal cancer represents a large global health problem, being the third most commonly diagnosed malignancy worldwide and representing one of the major causes of morbidity and mortality across populations (Siegel, Miller & Jemal, 2019). In 2018 worldwide, 1.80 million new cases of CRC were diagnosed, and 862,000 patients died from CRC (Ahmed, 2020). However, efficient treatment options are available for CRC; over 70% of patients have shown a 5-year survival if patients were diagnosed at an early stage (stage I and II) (Ahmed, 2020). In contrast, for late-stage patients, often the only remaining treatment options are palliative to improve quality of life and to prolong overall survival. Chemotherapy is still the gold standard of mCRC treatment. Due to tumor heterogeneity, treating patients based only on the disease type and the stage is likely to be ineffective and may be associated with adverse drug reactions and side effects (Yau, 2019). In this case, the concept of precision medicine, where treatment decisions are mostly based on characteristics of the tumor genome, appear very promising and may be applied to more mCRC patients in the future.

Targeted therapy is one major concept of precision medicine, which pursues the aim to specifically kill tumor cells by targeting the particular biological characteristics of cancer cells with minimal toxicity to normal cells. Currently, anti-VEGF (e.g. bevacizumab) and anti-EGFR (e.g. cetuximab and panitumumab) monoclonal antibodies are the two most widely used types of drugs with promising improvements in quality and length of patients' survival (Hurwitz et al., 2005, Hurwitz et al., 2013, Ketzer et al., 2018, Sotelo et al., 2014, Ohhara et al., 2016). However, there are still limitations. Firstly, predictive biomarkers are needed in order to stratify patients into appropriate treatment groups, but, currently, suitable predictive biomarkers for most targeted treatments are lacking. Even with the currently known biomarkers, the limited access to tumor tissues in most late-stage CRC patients limits the usage. Secondly, acquired resistance happen in almost all patients treated with targeted drugs. In these cases, identification of biomarkers predicting the possible appearance of acquired resistance and tracking of those markers over time is very necessary and can help to switch treatment in time. Although a considerable number of patients with mCRC will experience progression and eventually exhaust standard therapies approved for

CRC, many of these patients remain candidates for further treatment strategies if they demonstrate an adequate performance score and lack significant comorbidities (Zhou et al., 2020). Lastly, there are patients where no actionable target can be found. Hence, identification of as many possible driver genes, which may be targetable in the future, may contribute to improving the specificity of targeted therapy.

Liquid biopsies, i.e., the analysis of tumor components (e.g. ctDNA and CTC) in bodily fluids such as blood (Zhou et al., 2020), may contribute to an improved application of targeted therapies. Liquid biopsy, as a non-invasive method, is an approach that can be applied to all mCRC, especially those patients with inaccessible tumors. Moreover, liquid biopsies also allow continuous monitoring of the patient, e.g. the evolution of the tumor genome (Bardelli, Pantel, 2017, Wan et al., 2017, Heitzer et al., 2019, Siravegna et al., 2017). Also, due to the short half-life of ctDNA, liquid biopsies were considered to provide more real-time information compared with other diagnostic methods. Another application of liquid biopsies is the identification of novel driver genes or novel relationships between known driver genes and resistance to present drugs (Nong et al., 2018, Thompson et al., 2016, Yamamoto et al., 2019, Said et al., 2020).

In this study, we focused on the analysis of ctDNA using whole-genome sequencing in mCRC patients to validate the application of liquid biopsy in mCRC patient management, i.e. monitoring and predicting the response of treatment in mCRC patients, and in researching disease progression, i.e. identification of novel driver genes in mCRC. To this end, we highlighted the importance of ctDNA application in mCRC in both the clinic and research areas.

5.1 ctDNA can be used for monitoring and predicting the response of treatment in mCRC patients

We first focused on the analysis of ctDNA using whole-genome sequencing in mCRC patients to validate the application of liquid biopsy in mCRC patient management. Our group has published before that serial monitoring of ctDNA allows the prediction of acquired resistance in mCRC and prostate cancer patients (Mohan et al., 2014, Ulz et al., 2016). Also, in some recent publications, ctDNA has been reported to be a prognostic biomarker for various cancers (Chen et al., 2020, Nicolini, Ferrari & Duffy, 2018, Provencio et al., 2017, Belic et al., 2018). To further validate these two usages, we applied serial monitoring to over 100 mCRC patients and interesting results were detected in some patients.

5.1.1 Serial monitoring of patients under targeted therapy

In 2 mCRC patients undergoing anti-EGFR therapy, the emergence of focal amplifications in resistance-related genes (i.e., *KRAS* and *ERBB2*) were detected when patients developed progressive disease. Moreover, in another patient under anti-ERBB2 treatment, *CDK6* focal amplification was identified. As described before, amplification of these genes has been shown to be directly related to anti-EGFR and anti-ERBB2 resistance (Misale et al., 2014, Valtorta et al., 2013, Mohan et al., 2014, Bertotti et al., 2011). The emergence of focal amplification in resistance-related genes correlated with progressive disease and hence clinical outcome and, therefore, can be used to predict the development of acquired resistance. However, this usage still has some limitations, such as the often low tumor content of cfDNA. Furthermore, a reliable resistance-related gene database is missing. More work is still needed to guarantee the accuracy of the results and widen the application.

5.2 ctDNA can be used to identify novel resistance-related genes in mCRC

Identification of novel driver genes and novel resistance mechanisms are both critical research areas of liquid biopsy, which may not only allow a better understanding of disease progression, but may also play an important role in precision medicine. In this project, we focused on mCRC SCNAs. The landscape of 150 mCRC SCNAs from ctDNA indicated some recurrent focal events with known driver genes, e.g. 12p12.1 and 8p11.23-p11.22 focal amplifications. In addition to these events, 13q12.2 was detected frequently without a known CRC driver gene. To explore this further, we applied *in vitro* and large-scale data analyses.

5.2.1 Large-scale data analysis reveals the relationship between 13q12.2 SCNA and disease progression

We started with verifying whether this recurrent SCNA is indeed related to CRC disease progression. In our dataset of 9 mCRC patients who had accessible tumor tissue and plasma samples, we identified the 13q12.2 focal amplification as a potential late event in CRC progression. In 5 of these 9 cases, this focal event was not present in the primary tumor, but acquired at a later time point, which was detected in the ctDNA. Moreover, in 3 of these 5 patients, focal amplification appeared in one of the ctDNA analyses after the patient developed progressive disease. Due to the limited patient number in our cohort, we validated this finding in an available public dataset (i.e. TCGA dataset). As expected, late-stage (stage III and IV), distant metastases and lymph node metastases cases were significantly enriched in CRC patients with 13q12.2 SCNA, compared with those who have balanced a 13q12.2 region.

These results indicated that 13q12.2 is not only a recurrent amplification event, but also indeed plays a potential role in CRC disease progression.

5.2.2 The oncogene *FLT3* is not a driver gene in CRC

One way to confirm the role of 13q12.2 SCNA in CRC disease progression is to identify the driver gene located in this amplicon. After minimum overlap analysis, we found that seven genes (i.e. *POLR1D*, *GSX1*, *PDX1*, *ATP5EP2*, *CDX2*, *PRHOXNB* and *FLT3*) were completely located within

the overlapping region. *FLT3* encodes fms like tyrosine kinase 3, which is involved in the proliferation and differentiation pathways in hematopoietic stem cells (Grafone et al., 2012). Also, *FLT3* is a known oncogene for hematological malignancies (Daver et al., 2019, Grafone et al., 2012). Although there is no evidence indicating that *FLT3* may also play a role in the tumorigenesis of CRC, one publication reported response to sorafenib (an anti-FLT3 compound) in one particular patient with 13q12.2 amplification (Moreira, Peixoto & de Sousa Cruz, M R, 2015). In this case, *FLT3* was considered to be the most reasonable candidate driver gene located in the 13q12.2 amplicon. However, by analyzing the TCGA data and CCLE cell line data, we found that *FLT3* amplification is not associated with evaluated FLT3 expression. Moreover, the TSS profiling from 2 in-house patients (i.e. C129 and C216) with 13q12.2 amplification predicted that *FLT3* was not expressed in these 2 cases. To further exclude *FLT3* from the driver gene list in CRC, we established a stable FLT3-overexpressing CRC cell line and found no significant proliferation advantage.

We further confirmed that *FLT3* is not an oncogene in CRC by appropriate *in vitro* experiments. Our findings are consistent with a previous study, which showed that *FLT3* amplification does not seem to be an actionable target FLT3 inhibitors (Lim et al., 2017). Actually, the tumor described in the aforementioned case report with the response to sorafenib, also harbored a *CDK8* amplification, which is an actionable target for sorafenib (Moreira, Peixoto & de Sousa Cruz, M R, 2015).

5.2.3 Identification of *POLR1D* as a potential driver gene in 13q12.2

To identify the novel driver gene, we further validated the function of other genes located in the overlapping range (i.e. *POLR1D*, *GSX1*, *PDX1*, *CDX2*, and *PRHOXNB*) together with the first upstream (*LNX2*) and downstream (*PAN3*) located genes. A positive correlation between copy number and gene expression was identified in 5 of them in the TCGA and CCLE data sets (i.e. *POLR1D*, *LNX2*, *CDX2*, *PDX1*, and *PAN3*). This finding is consistent with the TSS nucleosome positioning of our 13q12.2-amplified patients, which predicted that *POLR1D*, *LNX2*, *CDX2*, *PDX1* and *PAN3* are indeed expressed. Furthermore, we validated the function of these 5 genes in 2 CRC

cell lines. Interestingly, only silencing of *POLR1D* showed significant cell viability inhibition in both cell lines. Thus, we suggested *POLR1D* is the potential driver gene in this amplicon.

Two previous studies also attempted to identify the potential driver gene in 13q12.2 with varied results (Camps et al., 2013, Salari et al., 2012). One study reported both *LNX2* and *POLR1D* as potential driver genes, while another study suggested *CDX2*. In the first study, the SW480 (whole chr13 gain) and DLD1 (chr13 balanced) cell lines were used, which thus excluded cell lines harboring amplification of 13q12 (Camps et al., 2013). In the second study, cell lines harboring 13q12.2 amplification were used, but resulted in demonstration of high expression of CDX2 (Salari et al., 2012). However, according to our TCGA and CCLE analysis, the *CDX2* copy number is significant but poorly correlated with gene expression. Moreover, there have been multiple publications suggesting that *CDX2* functions as a tumor suppressor in CRC (Graule et al., 2018, Sandberg et al., 2019, Yu, J. et al., 2019). In this case, we think *CDX2* may act as an oncogene in CRC patients in cases with a very high expression of CDX2, but this is not necessarily applicable to all cases of CRC with 13q12.2 aberration (Zhou et al., 2020).

To this end, we identified *POLR1D* as the potential driver gene in 13q12.2 amplicon.

5.2.4 POLR1D affects the expression of VEGFA and EREG and is involved in resistance to bevacizumab

To further explore the involvement of *POLR1D* in the disease progression of CRC, we applied whole transcriptome sequencing to POLR1D-knockdown and control cells. As described before, we identified 45 genes that showed consistent changes in 2 different CRC cell lines. Alteration of 8 of them (i.e. *POLR1D*, *VEGFA*, *EREG*, *PPP1R15A*, *MOSPD2*, *FAM84B*, *GARS* and *KIF21B*) could also be detected in TCGA RNA-seq data.

Except for *POLR1D*, all other 7 genes have been reported to be involved in the progression of various cancers. *VEGFA* encodes VEGF-A, which is an essential regulator of angiogenesis and plays a role in the development, progression and metastasis of CRC (Wang, D. et al., 2017, Herzog et al., 2011, Maj, Papiernik & Wietrzyk, 2016, Holmes, Zachary, 2005). *EREG* encodes epiregulin,

which is overexpressed in CRC patients with higher activation of the EGFR pathway (Pentheroudakis et al., 2013), and plays a role in the resistance to anti-ERBB2 treatment (Auf et al., 2013, Takahashi et al., 2016). *FAM84B* encodes family with sequence similarity 84 member B protein, which has been reported to be involved in the progression of prostate cancer and esophageal squamous cell carcinoma (Cheng et al., 2016, Wong et al., 2017). *GARS* encodes glycyl-tRNA synthetase and is involved in the cell cycle regulation and thus play a role in cancer progression (Mo et al., 2016). The expression of KIF21B was reported to be significantly associated with poor prognosis of prostate cancer (Arai et al., 2019), and the mutation of *PPP1R15A* may be a biomarker predicting the response of mCRC patients treated with bevacizumab (Roh et al., 2016). *MOSPD2* encodes motile sperm domain-containing protein 2, which has been shown involved in the metastasis of breast cancer (Salem et al., 2019). Hence, expression of VEGFA, EREG, PPP1R15A, MOSPD2, FAM84B, GARS and KIF21B may contribute to the progression of CRC through POLR1D.

As described before, bevacizumab treatment induces autocrine VEGF signaling, which ultimately leads to the overexpression of VEGF-A and resistance to the drug (Fan et al., 2011, Mesange et al., 2014) . Our results indicated that CRC patients with 13q12.2 amplification have higher expression of POLR1D, which then induced the expression of VEGF-A. This result suggested a potential involvement of *POLR1D* amplification in bevacizumab resistance in CRC. In fact, serial plasma DNA analyses revealed that in two of our patients, the 13q12.2 amplification evolved under treatment and was in both cases linked to progressive disease (Zhou et al., 2020).

In summary, these results suggest *POLR1D* acts as a potential oncogene in the 13q12.2 amplicon. *POLR1D* amplification may be a novel resistance mechanism against bevacizumab. A larger cohort will be needed in order to confirm these findings.

5.3 Limitations

The main limitation of this study is the low tumor content of CRC patients, which restricts the usage of cfDNA. To conduct reliable copy number analyses, a tumor content higher than 5-10% in plasma is needed. Moreover, to monitor the ctDNA SCNA changes, tumor fractions should be similar in serial plasma samples. To avoid false-negative results and test if the tumor content is indeed similar in serial samples, we established the tumor fraction with the ichorCNA algorithm. Samples with tumor fractions lower than 5% were excluded, and the contribution of tumor content in SCNA changes in serial samples was considered.

Another limitation is that only one SCNA event was fully investigated, although other somatic alterations may modulate the therapeutic response to bevacizumab as well. For instance, the deletion of 18q11.2-q12.1 was recently reported to be a predictive marker of mCRC patient survival under bevacizumab treatment (van Dijk et al., 2018), which co-occurred in 10 of our 14 13q12.2 amplified patients. This suggests that a variety of somatic alterations may govern response to bevacizumab treatment such that further investigations are warranted.

6 CONCLUSION

Our results suggest that monitoring somatic focal events may allow the prediction of acquired resistance to targeted treatment and disease progression in mCRC patients. ctDNA, therefore, can be used as a non-invasive predictive and prognostic biomarker to manage mCRC patients under different treatments to improve the quality and length of life. Moreover, to those patients whose tumor tissues are inaccessible, ctDNA is an alternative way to access the tumor genome and to help the clinical management of patients.

Monitoring somatic focal events also can be used to identify novel driver genes in mCRC, which has meaningful implications for the identification of novel driver genes associated with late-stage cancers. As we reported in this study, we found a frequently amplified genomic region (i.e. 13q12.2), which previously had not been well-studied. Our investigations of this amplicon revealed that it may be related to tumor stage and metastasis. *POLR1D*, a subunit of RNA polymerases I and III, was then established as the most likely driver gene in this frequently amplified region. RNA-seq and big data analysis indicated that *POLR1D* impacts tumorigenesis of CRC, possibly by affecting the expression of *VEGFA*, *EREG*, *PPP1R15A*, *MOSPD2*, *FAM84B*, *GARS* and *KIF21B*. As involvement of *VEGFA* in the acquired resistance to bevacizumab is known, we decided to investigate the relationship between 13q12.2 amplification and bevacizumab resistance. According to the serial monitoring of 2 patients under bevacizumab treatment, we confirmed that *POLR1D* plays a potential role in the acquired resistance to bevacizumab and is a potential therapeutic target for mCRC.

Multiple publications employing liquid biopsies have reported promising clinical and basic research results, but limitations remain. One limitation of our study is that a high tumor content is needed to accurately identify SCNA changes in ctDNA. We are currently testing various new methods in our lab to improve our detection limits such that our methods may be even used in patients with very low ctDNA levels in plasma.

In summary, our results suggested that liquid biopsy as a non-invasive method can be used for both clinical (i.e. patient management) and research (i.e. novel driver gene identification) applications.

REFERENCES

Adalsteinsson, V.A., Ha, G., Freeman, S.S., Choudhury, A.D., Stover, D.G., Parsons, H.A., Gydush, G., Reed, S.C., Rotem, D., Rhoades, J., Loginov, D., Livitz, D., Rosebrock, D., Leshchiner, I., Kim, J., Stewart, C., Rosenberg, M., Francis, J.M., Zhang, C.Z., Cohen, O., Oh, C., Ding, H., Polak, P., Lloyd, M., Mahmud, S., Helvie, K., Merrill, M.S., Santiago, R.A., O'Connor, E.P., Jeong, S.H., Leeson, R., Barry, R.M., Kramkowski, J.F., Zhang, Z., Polacek, L., Lohr, J.G., Schleicher, M., Lipscomb, E., Saltzman, A., Oliver, N.M., Marini, L., Waks, A.G., Harshman, L.C., Tolaney, S.M., Van Allen, E.M., Winer, E.P., Lin, N.U., Nakabayashi, M., Taplin, M.E., Johannessen, C.M., Garraway, L.A., Golub, T.R., Boehm, J.S., Wagle, N., Getz, G., Love, J.C. & Meyerson, M. 2017, "Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors", *Nature communications,* vol. 8, no. 1, pp. 1324-017.

Agarwal, A., Das, K., Lerner, N., Sathe, S., Cicek, M., Casey, G. & Sizemore, N. 2005, "The AKT/I kappa B kinase pathway promotes angiogenic/metastatic gene expression in colorectal cancer by activating nuclear factor-kappa B and beta-catenin", *Oncogene,* vol. 24, no. 6, pp. 1021-1031.

Ahmed, M. 2020, "Colon Cancer: A Clinician's Perspective in 2019", *Gastroenterology research,* vol. 13, no. 1, pp. 1-10.

Arai, T., Kojima, S., Yamada, Y., Sugawara, S., Kato, M., Yamazaki, K., Naya, Y., Ichikawa, T. & Seki, N. 2019, "Pirin: a potential novel therapeutic target for castration-resistant prostate cancer regulated by miR-455-5p", *Molecular oncology,* vol. 13, no. 2, pp. 322-337.

Arnold, M., Sierra, M.S., Laversanne, M., Soerjomataram, I., Jemal, A. & Bray, F. 2017, "Global patterns and trends in colorectal cancer incidence and mortality", *Gut,* vol. 66, no. 4, pp. 683-691.

Auf, G., Jabouille, A., Delugin, M., Guerit, S., Pineau, R., North, S., Platonova, N., Maitre, M., Favereaux, A., Vajkoczy, P., Seno, M., Bikfalvi, A., Minchenko, D., Minchenko, O. & Moenner, M. 2013, "High epiregulin expression in human U87 glioma cells relies on IRE1alpha and promotes autocrine growth through EGF receptor", *BMC cancer,* vol. 13, pp. 597-2407.

Barault, L., Amatu, A., Siravegna, G., Ponzetti, A., Moran, S., Cassingena, A., Mussolin, B., Falcomata, C., Binder, A.M., Cristiano, C., Oddo, D., Guarrera, S., Cancelliere, C., Bustreo, S., Bencardino, K., Maden, S., Vanzati, A., Zavattari, P., Matullo, G., Truini, M., Grady, W.M., Racca, P., Michels, K.B., Siena, S., Esteller, M., Bardelli, A., Sartore-Bianchi, A. & Di Nicolantonio, F. 2018, "Discovery of methylated circulating DNA biomarkers for comprehensive non-invasive monitoring of treatment response in metastatic colorectal cancer", *Gut,* vol. 67, no. 11, pp. 1995-2005.

Bardelli, A. & Pantel, K. 2017, "Liquid Biopsies, What We Do Not Know (Yet)", *Cancer cell,* vol. 31, no. 2, pp. 172-179.

Bedin, C., Enzo, M.V., Del Bianco, P., Pucciarelli, S., Nitti, D. & Agostini, M. 2017, "Diagnostic and prognostic role of cell-free DNA testing for colorectal cancer patients", *International journal of cancer,* vol. 140, no. 8, pp. 1888-1898.

Belic, J., Graf, R., Bauernhofer, T., Cherkas, Y., Ulz, P., Waldispuehl-Geigl, J., Perakis, S., Gormley, M., Patel, J., Li, W., Geigl, J.B., Smirnov, D., Heitzer, E., Gross, M. & Speicher, M.R. 2018, "Genomic alterations in plasma DNA from patients with metastasized prostate cancer receiving abiraterone or enzalutamide", *International journal of cancer,* vol. 143, no. 5, pp. 1236-1248.

Belic, J., Koch, M., Ulz, P., Auer, M., Gerhalter, T., Mohan, S., Fischereder, K., Petru, E., Bauernhofer, T., Geigl, J.B., Speicher, M.R. & Heitzer, E. 2015, "Rapid Identification of Plasma DNA Samples with Increased ctDNA Levels by a Modified FAST-SeqS Approach", *Clinical chemistry,* vol. 61, no. 6, pp. 838-849.

Bertotti, A., Migliardi, G., Galimi, F., Sassi, F., Torti, D., Isella, C., Cora, D., Di Nicolantonio, F., Buscarino, M., Petti, C., Ribero, D., Russolillo, N., Muratore, A., Massucco, P., Pisacane, A., Molinaro, L., Valtorta, E., Sartore-Bianchi, A., Risio, M., Capussotti, L., Gambacorta, M., Siena, S., Medico, E., Sapino, A., Marsoni, S., Comoglio, P.M., Bardelli, A. & Trusolino, L. 2011, "A molecularly annotated platform of patient-derived xenografts ("xenopatients") identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer", *Cancer discovery,* vol. 1, no. 6, pp. 508-523.

Bettegowda, C., Sausen, M., Leary, R.J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B.R., Wang, H., Luber, B., Alani, R.M., Antonarakis, E.S., Azad, N.S., Bardelli, A., Brem, H., Cameron, J.L., Lee, C.C., Fecher, L.A., Gallia, G.L., Gibbs, P., Le, D., Giuntoli, R.L., Goggins, M., Hogarty, M.D., Holdhoff, M., Hong, S.M., Jiao, Y., Juhl, H.H., Kim, J.J., Siravegna, G., Laheru, D.A., Lauricella, C., Lim, M., Lipson, E.J., Marie, S.K., Netto, G.J., Oliner, K.S., Olivi, A., Olsson, L., Riggins, G.J., Sartore-Bianchi, A., Schmidt, K., Shih, l., Oba-Shinjo, S.M., Siena, S., Theodorescu, D., Tie, J., Harkins, T.T., Veronese, S., Wang, T.L., Weingart, J.D., Wolfgang, C.L., Wood, L.D., Xing, D., Hruban, R.H., Wu, J., Allen, P.J., Schmidt, C.M., Choti, M.A., Velculescu, V.E., Kinzler, K.W., Vogelstein, B., Papadopoulos, N. & Diaz, L.A. 2014, "Detection of circulating tumor DNA in early- and late-stage human malignancies", *Science translational medicine,* vol. 6, no. 224, pp. 224ra24.

Braxton, D.R., Zhang, R., Morrissette, J.D., Loaiza-Bonilla, A. & Furth, E.E. 2016, "Clinicopathogenomic analysis of mismatch repair proficient colorectal adenocarcinoma uncovers novel prognostic subgroups with differing patterns of genetic evolution", *International journal of cancer,* vol. 139, no. 7, pp. 1546-1556.

Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. 2016, "Near-optimal probabilistic RNA-seq quantification", *Nature biotechnology,* vol. 34, no. 5, pp. 525-527.

Burrell, R.A. & Swanton, C. 2014, "The evolution of the unstable cancer genome", *Current opinion in genetics & development,* vol. 24, pp. 61-67.

Camps, J., Pitt, J.J., Emons, G., Hummon, A.B., Case, C.M., Grade, M., Jones, T.L., Nguyen, Q.T., Ghadimi, B.M., Beissbarth, T., Difilippantonio, M.J., Caplen, N.J. & Ried, T. 2013, "Genetic amplification of the NOTCH modulator LNX2 upregulates the WNT/beta-catenin pathway in colorectal cancer", *Cancer research,* vol. 73, no. 6, pp. 2003-2013.

Chabon, J.J., Simmons, A.D., Lovejoy, A.F., Esfahani, M.S., Newman, A.M., Haringsma, H.J., Kurtz, D.M., Stehr, H., Scherer, F., Karlovich, C.A., Harding, T.C., Durkin, K.A., Otterson, G.A., Purcell, W.T., Camidge, D.R., Goldman, J.W., Sequist, L.V., Piotrowska, Z., Wakelee, H.A., Neal, J.W., Alizadeh, A.A. & Diehn, M. 2016, "Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients", *Nature communications,* vol. 7, pp. 11815.

Chan, K.C., Zhang, J., Hui, A.B., Wong, N., Lau, T.K., Leung, T.N., Lo, K.W., Huang, D.W. & Lo, Y.M. 2004, "Size distributions of maternal and fetal DNA in maternal plasma", *Clinical chemistry,* vol. 50, no. 1, pp. 88-92.

Chan, R.W., Jiang, P., Peng, X., Tam, L.S., Liao, G.J., Li, E.K., Wong, P.C., Sun, H., Chan, K.C., Chiu, R.W. & Lo, Y.M. 2014, "Plasma DNA aberrations in systemic lupus erythematosus revealed by genomic and methylomic sequencing", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 111, no. 49, pp. E5302-11.

Chen, F.P., Huang, X.D., Lv, J.W., Wen, D.W., Zhou, G.Q., Lin, L., Kou, J., Wu, C.F., Chen, Y., Zheng, Z.Q., Li, Z.X., He, X.J. & Sun, Y. 2020, "Prognostic potential of liquid biopsy tracking in the posttreatment surveillance of patients with nonmetastatic nasopharyngeal carcinoma", *Cancer,* .

Cheng, C., Cui, H., Zhang, L., Jia, Z., Song, B., Wang, F., Li, Y., Liu, J., Kong, P., Shi, R., Bi, Y., Yang, B., Wang, J., Zhao, Z., Zhang, Y., Hu, X., Yang, J., He, C., Zhao, Z., Wang, J., Xi, Y., Xu, E., Li, G., Guo, S., Chen, Y., Yang, X., Chen, X., Liang, J., Guo, J., Cheng, X., Wang, C., Zhan, Q. & Cui, Y. 2016, "Genomic analyses reveal FAM84B and the NOTCH pathway are associated with the progression of esophageal squamous cell carcinoma", *GigaScience,* vol. 5, pp. 1-015.

Chong, C.R. & Janne, P.A. 2013, "The quest to overcome resistance to EGFR-targeted therapies in cancer", *Nature medicine,* vol. 19, no. 11, pp. 1389-1400.

Ciardiello, F. & Tortora, G. 2001, "A novel approach in the treatment of cancer: targeting the epidermal growth factor receptor", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 7, no. 10, pp. 2958-2970.

Cohen, J.D., Li, L., Wang, Y., Thoburn, C., Afsari, B., Danilova, L., Douville, C., Javed, A.A., Wong, F., Mattox, A., Hruban, R.H., Wolfgang, C.L., Goggins, M.G., Dal Molin, M., Wang, T.L., Roden, R., Klein, A.P., Ptak, J., Dobbyn, L., Schaefer, J., Silliman, N., Popoli, M., Vogelstein, J.T., Browne, J.D., Schoen, R.E., Brand, R.E., Tie, J., Gibbs, P., Wong, H.L., Mansfield, A.S., Jen, J., Hanash, S.M., Falconi, M., Allen, P.J., Zhou, S., Bettegowda, C., Diaz, L.A., Tomasetti, C., Kinzler, K.W., Vogelstein, B., Lennon, A.M. & Papadopoulos, N. 2018, "Detection and localization of surgically resectable cancers with a multi-analyte blood test", *Science (New York, N.Y.),* vol. 359, no. 6378, pp. 926-930.

Daver, N., Schlenk, R.F., Russell, N.H. & Levis, M.J. 2019, "Targeting FLT3 mutations in AML: review of current knowledge and evidence", *Leukemia,* vol. 33, no. 2, pp. 299-312.

Delgado, P.O., Alves, B.C., Gehrke Fde, S., Kuniyoshi, R.K., Wroclavski, M.L., Del Giglio, A. & Fonseca, F.L. 2013, "Characterization of cell-free circulating DNA in plasma in patients with prostate cancer", *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine,* vol. 34, no. 2, pp. 983-986.

DeSantis, C.E., Lin, C.C., Mariotto, A.B., Siegel, R.L., Stein, K.D., Kramer, J.L., Alteri, R., Robbins, A.S. & Jemal, A. 2014, "Cancer treatment and survivorship statistics, 2014", *CA: a cancer journal for clinicians,* vol. 64, no. 4, pp. 252-271.

Diaz, L.A. & Bardelli, A. 2014, "Liquid biopsies: genotyping circulating tumor DNA", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 32, no. 6, pp. 579-586.

Diaz, L.A., Williams, R.T., Wu, J., Kinde, I., Hecht, J.R., Berlin, J., Allen, B., Bozic, I., Reiter, J.G., Nowak, M.A., Kinzler, K.W., Oliner, K.S. & Vogelstein, B. 2012, "The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers", *Nature,* vol. 486, no. 7404, pp. 537-540.

Diehl, F., Li, M., Dressman, D., He, Y., Shen, D., Szabo, S., Diaz, L.A., Goodman, S.N., David, K.A., Juhl, H., Kinzler, K.W. & Vogelstein, B. 2005, "Detection and quantification of mutations in the plasma of patients with colorectal tumors", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, no. 45, pp. 16368-16373.

Diehl, F., Schmidt, K., Choti, M.A., Romans, K., Goodman, S., Li, M., Thornton, K., Agrawal, N., Sokoll, L., Szabo, S.A., Kinzler, K.W., Vogelstein, B. & Diaz, L.A. 2008, "Circulating mutant DNA to assess tumor dynamics", *Nature medicine,* vol. 14, no. 9, pp. 985-990.

Dutta, P.R. & Maity, A. 2007, "Cellular responses to EGFR inhibitors and their relevance to cancer therapy", *Cancer letters,* vol. 254, no. 2, pp. 165-177.

El Messaoudi, S., Mouliere, F., Du Manoir, S., Bascoul-Mollevi, C., Gillet, B., Nouaille, M., Fiess, C., Crapez, E., Bibeau, F., Theillet, C., Mazard, T., Pezet, D., Mathonnet, M., Ychou, M. & Thierry, A.R. 2016, "Circulating DNA as a Strong Multimarker Prognostic Tool for Metastatic Colorectal Cancer Patient Management Care", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 22, no. 12, pp. 3067-3077.

Ellinger, J., Bastian, P.J., Ellinger, N., Kahl, P., Perabo, F.G., Buttner, R., Muller, S.C. & Ruecker, A. 2008, "Apoptotic DNA fragments in serum of patients with muscle invasive bladder cancer: a prognostic entity", *Cancer letters,* vol. 264, no. 2, pp. 274-280.

Fan, F., Samuel, S., Gaur, P., Lu, J., Dallas, N.A., Xia, L., Bose, D., Ramachandran, V. & Ellis, L.M. 2011, "Chronic exposure of colorectal cancer cells to bevacizumab promotes compensatory pathways that mediate tumour cell migration", *British journal of cancer,* vol. 104, no. 8, pp. 1270-1277.

Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., Ji, X., Liu, W., Huang, B., Luo, W., Liu, B., Lei, Y., Du, S., Vuppalapati, A., Luu, H.H., Haydon, R.C., He, T.C. & Ren, G. 2018, "Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis", *Genes & diseases,* vol. 5, no. 2, pp. 77-106.

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J.W., Comber, H., Forman, D. & Bray, F. 2013, "Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012", *European journal of cancer (Oxford, England : 1990),* vol. 49, no. 6, pp. 1374-1403.

Forshew, T., Murtaza, M., Parkinson, C., Gale, D., Tsui, D.W., Kaper, F., Dawson, S.J., Piskorz, A.M., Jimenez-Linan, M., Bentley, D., Hadfield, J., May, A.P., Caldas, C., Brenton, J.D. & Rosenfeld, N. 2012, "Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA", *Science translational medicine,* vol. 4, no. 136, pp. 136ra68.

Gao, Y., Wang, J., Zhou, Y., Sheng, S., Qian, S.Y. & Huo, X. 2018, "Evaluation of Serum CEA, CA19-9, CA72-4, CA125 and Ferritin as Diagnostic Markers and Factors of Clinical Parameters for Colorectal Cancer", *Scientific reports,* vol. 8, no. 1, pp. 2732-018.

George, M.L., Tutton, M.G., Janssen, F., Arnaout, A., Abulafi, A.M., Eccles, S.A. & Swift, R.I. 2001, "VEGF-A, VEGF-C, and VEGF-D in colorectal cancer progression", *Neoplasia (New York, N.Y.),* vol. 3, no. 5, pp. 420-427.

Grafone, T., Palmisano, M., Nicci, C. & Storti, S. 2012, "An overview on the role of FLT3-tyrosine kinase receptor in acute myeloid leukemia: biology and treatment", *Oncology reviews,* vol. 6, no. 1, pp. e8.

Graule, J., Uth, K., Fischer, E., Centeno, I., Galvan, J.A., Eichmann, M., Rau, T.T., Langer, R., Dawson, H., Nitsche, U., Traeger, P., Berger, M.D., Schnuriger, B., Hadrich, M., Studer, P., Inderbitzin, D., Lugli, A., Tschan, M.P. & Zlobec, I. 2018, "CDX2 in colorectal cancer is an independent prognostic factor and regulated by promoter methylation and histone deacetylation in tumors of the serrated pathway", *Clinical epigenetics,* vol. 10, no. 1, pp. 120-018.

Greally, M., Kelly, C.M. & Cercek, A. 2018, "HER2: An emerging target in colorectal cancer", *Current problems in cancer,* vol. 42, no. 6, pp. 560-571.

Grossman, R.L., Heath, A.P., Ferretti, V., Varmus, H.E., Lowy, D.R., Kibbe, W.A. & Staudt, L.M. 2016, "Toward a Shared Vision for Cancer Genomic Data", *The New England journal of medicine,* vol. 375, no. 12, pp. 1109-1112.

Grothey, A., Sugrue, M.M., Purdie, D.M., Dong, W., Sargent, D., Hedrick, E. & Kozloff, M. 2008, "Bevacizumab beyond first progression is associated with prolonged overall survival in metastatic colorectal cancer: results from a large observational cohort study (BRiTE)", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 26, no. 33, pp. 5326-5334.

Grothey, A., Van Cutsem, E., Sobrero, A., Siena, S., Falcone, A., Ychou, M., Humblet, Y., Bouche, O., Mineur, L., Barone, C., Adenis, A., Tabernero, J., Yoshino, T., Lenz, H.J., Goldberg, R.M., Sargent, D.J., Cihon, F., Cupit, L., Wagner, A., Laurent, D. & CORRECT Study Group 2013, "Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial", *Lancet (London, England),* vol. 381, no. 9863, pp. 303-312.

Gustavsson, B., Carlsson, G., Machover, D., Petrelli, N., Roth, A., Schmoll, H.J., Tveit, K.M. & Gibson, F. 2015, "A review of the evolution of systemic chemotherapy in the management of colorectal cancer", *Clinical colorectal cancer,* vol. 14, no. 1, pp. 1-10.

Hammond, W.A., Swaika, A. & Mody, K. 2016, "Pharmacologic resistance in colorectal cancer: a review", *Therapeutic advances in medical oncology,* vol. 8, no. 1, pp. 57-84.

Hashad, D., Sorour, A., Ghazal, A. & Talaat, I. 2012, "Free circulating tumor DNA as a diagnostic marker for breast cancer", *Journal of clinical laboratory analysis,* vol. 26, no. 6, pp. 467-472.

Hegde, P.S., Jubb, A.M., Chen, D., Li, N.F., Meng, Y.G., Bernaards, C., Elliott, R., Scherer, S.J. & Chen, D.S. 2013, "Predictive impact of circulating vascular endothelial growth factor in

four phase III trials evaluating bevacizumab", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 19, no. 4, pp. 929-937.

Heitzer, E., Auer, M., Hoffmann, E.M., Pichler, M., Gasch, C., Ulz, P., Lax, S., Waldispuehl-Geigl, J., Mauermann, O., Mohan, S., Pristauz, G., Lackner, C., Hofler, G., Eisner, F., Petru, E., Sill, H., Samonigg, H., Pantel, K., Riethdorf, S., Bauernhofer, T., Geigl, J.B. & Speicher, M.R. 2013, "Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer", *International journal of cancer,* vol. 133, no. 2, pp. 346-356.

Heitzer, E., Haque, I.S., Roberts, C.E.S. & Speicher, M.R. 2019, "Current and future perspectives of liquid biopsies in genomics-driven oncology", *Nature reviews.Genetics,* vol. 20, no. 2, pp. 71-88.

Heitzer, E., Ulz, P., Belic, J., Gutschi, S., Quehenberger, F., Fischereder, K., Benezeder, T., Auer, M., Pischler, C., Mannweiler, S., Pichler, M., Eisner, F., Haeusler, M., Riethdorf, S., Pantel, K., Samonigg, H., Hoefler, G., Augustin, H., Geigl, J.B. & Speicher, M.R. 2013, "Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing", *Genome medicine,* vol. 5, no. 4, pp. 30.

Herzog, B., Pellet-Many, C., Britton, G., Hartzoulakis, B. & Zachary, I.C. 2011, "VEGF binding to NRP1 is essential for VEGF stimulation of endothelial cell migration, complex formation between NRP1 and VEGFR2, and signaling via FAK Tyr407 phosphorylation", *Molecular biology of the cell,* vol. 22, no. 15, pp. 2766-2776.

Hilmi, C., Guyot, M. & Pages, G. 2012, "VEGF spliced variants: possible role of anti-angiogenesis therapy", *Journal of nucleic acids,* vol. 2012, pp. 162692.

Holmes, D.I. & Zachary, I. 2005, "The vascular endothelial growth factor (VEGF) family: angiogenic factors in health and disease", *Genome biology,* vol. 6, no. 2, pp. 209-2005.

Hurwitz, H.I., Fehrenbacher, L., Hainsworth, J.D., Heim, W., Berlin, J., Holmgren, E., Hambleton, J., Novotny, W.F. & Kabbinavar, F. 2005, "Bevacizumab in combination with fluorouracil and leucovorin: an active regimen for first-line metastatic colorectal cancer", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 23, no. 15, pp. 3502-3508.

Hurwitz, H.I., Tebbutt, N.C., Kabbinavar, F., Giantonio, B.J., Guan, Z.Z., Mitchell, L., Waterkamp, D. & Tabernero, J. 2013, "Efficacy and safety of bevacizumab in metastatic colorectal cancer: pooled analysis from seven randomized controlled trials", *The oncologist,* vol. 18, no. 9, pp. 1004-1012.

Hynes, N.E. & Lane, H.A. 2005, "ERBB receptors and cancer: the complexity of targeted inhibitors", *Nature reviews.Cancer,* vol. 5, no. 5, pp. 341-354.

Irvine, T., Scott, M. & Clark, C.I. 2007, "A small rise in CEA is sensitive for recurrence after surgery for colorectal cancer", *Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland,* vol. 9, no. 6, pp. 527-531.

Jiang, N., Lin, J.J., Wang, J., Zhang, B.N., Li, A., Chen, Z.Y., Guo, S., Li, B.B., Duan, Y.Z., Yan, R.Y., Yan, H.F., Fu, X.Y., Zhou, J.L., Yang, H.M. & Cui, Y. 2018, "Novel treatment strategies for patients with HER2-positive breast cancer who do not benefit from current targeted therapy drugs", *Experimental and therapeutic medicine,* vol. 16, no. 3, pp. 2183-2192.

Jiang, P., Chan, C.W., Chan, K.C., Cheng, S.H., Wong, J., Wong, V.W., Wong, G.L., Chan, S.L., Mok, T.S., Chan, H.L., Lai, P.B., Chiu, R.W. & Lo, Y.M. 2015, "Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 112, no. 11, pp. E1317-25.

Jiang, P. & Lo, Y.M.D. 2016, "The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics", *Trends in genetics : TIG,* vol. 32, no. 6, pp. 360-371.

Kahlert, C., Melo, S.A., Protopopov, A., Tang, J., Seth, S., Koch, M., Zhang, J., Weitz, J., Chin, L., Futreal, A. & Kalluri, R. 2014, "Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer", *The Journal of biological chemistry,* vol. 289, no. 7, pp. 3869-3875.

Kannagi, R., Izawa, M., Koike, T., Miyazaki, K. & Kimura, N. 2004, "Carbohydrate-mediated cell adhesion in cancer metastasis and angiogenesis", *Cancer science,* vol. 95, no. 5, pp. 377-384.

Ketzer, S., Schimmel, K., Koopman, M. & Guchelaar, H.J. 2018, "Clinical Pharmacokinetics and Pharmacodynamics of the Epidermal Growth Factor Receptor Inhibitor Panitumumab in the Treatment of Colorectal Cancer", *Clinical pharmacokinetics,* vol. 57, no. 4, pp. 455-473.

Kramer, I. & Lipp, H.P. 2007, "Bevacizumab, a humanized anti-angiogenic monoclonal antibody for the treatment of colorectal cancer", *Journal of clinical pharmacy and therapeutics,* vol. 32, no. 1, pp. 1-14.

Leto, S.M. & Trusolino, L. 2014, "Primary and acquired resistance to EGFR-targeted therapies in colorectal cancer: impact on future treatment strategies", *Journal of Molecular Medicine (Berlin, Germany),* vol. 92, no. 7, pp. 709-722.

Levin, B., Brooks, D., Smith, R.A. & Stone, A. 2003, "Emerging technologies in screening for colorectal cancer: CT colonography, immunochemical fecal occult blood tests, and stool screening using molecular markers", *CA: a cancer journal for clinicians,* vol. 53, no. 1, pp. 44-55.

Li, J., Qin, S., Xu, R., Yau, T.C., Ma, B., Pan, H., Xu, J., Bai, Y., Chi, Y., Wang, L., Yeh, K.H., Bi, F., Cheng, Y., Le, A.T., Lin, J.K., Liu, T., Ma, D., Kappeler, C., Kalmus, J., Kim, T.W. &

CONCUR Investigators 2015, "Regorafenib plus best supportive care versus placebo plus best supportive care in Asian patients with previously treated metastatic colorectal cancer (CONCUR): a randomised, double-blind, placebo-controlled, phase 3 trial", *The Lancet.Oncology,* vol. 16, no. 6, pp. 619-629.

Lim, S.H., Kim, S.Y., Kim, K., Jang, H., Ahn, S., Kim, K.M., Kim, N.K., Park, W., Lee, S.J., Kim, S.T., Park, S.H., Park, J.O., Park, Y.S., Lee, S.H., Lim, H.Y., Park, K., Kang, W.K. & Lee, J. 2017, "The implication of FLT3 amplification for FLT targeted therapeutics in solid tumors", *Oncotarget,* vol. 8, no. 2, pp. 3237-3245.

Lo, Y.M., Chan, K.C., Sun, H., Chen, E.Z., Jiang, P., Lun, F.M., Zheng, Y.W., Leung, T.Y., Lau, T.K., Cantor, C.R. & Chiu, R.W. 2010, "Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus", *Science translational medicine,* vol. 2, no. 61, pp. 61ra91.

Lo, Y.M., Zhang, J., Leung, T.N., Lau, T.K., Chang, A.M. & Hjelm, N.M. 1999, "Rapid clearance of fetal DNA from maternal plasma", *American Journal of Human Genetics,* vol. 64, no. 1, pp. 218-224.

Lofton-Day, C., Model, F., Devos, T., Tetzner, R., Distler, J., Schuster, M., Song, X., Lesche, R., Liebenberg, V., Ebert, M., Molnar, B., Grutzmann, R., Pilarsky, C. & Sledziewski, A. 2008, "DNA methylation biomarkers for blood-based colorectal cancer screening", *Clinical chemistry,* vol. 54, no. 2, pp. 414-423.

Love, M.I., Huber, W. & Anders, S. 2014, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2", *Genome biology,* vol. 15, no. 12, pp. 550-014.

Maj, E., Papiernik, D. & Wietrzyk, J. 2016, "Antiangiogenic cancer treatment: The great discovery and greater complexity (Review)", *International journal of oncology,* vol. 49, no. 5, pp. 1773-1784.

MANDEL, P. & METAIS, P. 1948, "Les acides nucleiques du plasma sanguin chez l'homme [in French]", *Comptes rendus des seances de la Societe de biologie et de ses filiales,* vol. 142, no. 3-4, pp. 241-243.

Marcuello, M., Vymetalkova, V., Neves, R.P.L., Duran-Sanchon, S., Vedeld, H.M., Tham, E., van Dalum, G., Flugen, G., Garcia-Barberan, V., Fijneman, R.J., Castells, A., Vodicka, P., Lind, G.E., Stoecklein, N.H., Heitzer, E. & Gironella, M. 2019, "Circulating biomarkers for early detection and clinical management of colorectal cancer", *Molecular aspects of medicine,* vol. 69, pp. 107-122.

Martinez-Calvillo, S., Saxena, A., Green, A., Leland, A. & Myler, P.J. 2007, "Characterization of the RNA polymerase II and III complexes in Leishmania major", *International journal for parasitology,* vol. 37, no. 5, pp. 491-502.

Mead, R., Duku, M., Bhandari, P. & Cree, I.A. 2011, "Circulating tumour markers can define patients with normal colons, benign polyps, and cancers", *British journal of cancer,* vol. 105, no. 2, pp. 239-245.

Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R. & Getz, G. 2011, "GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers", *Genome biology,* vol. 12, no. 4, pp. R41-2011.

Mesange, P., Poindessous, V., Sabbah, M., Escargueil, A.E., de Gramont, A. & Larsen, A.K. 2014, "Intrinsic bevacizumab resistance is associated with prolonged activation of autocrine VEGF signaling and hypoxia tolerance in colorectal cancer cells and can be overcome by nintedanib, a small molecule angiokinase inhibitor", *Oncotarget,* vol. 5, no. 13, pp. 4709-4721.

Misale, S., Di Nicolantonio, F., Sartore-Bianchi, A., Siena, S. & Bardelli, A. 2014, "Resistance to anti-EGFR therapy in colorectal cancer: from heterogeneity to convergent evolution", *Cancer discovery,* vol. 4, no. 11, pp. 1269-1280.

Misale, S., Yaeger, R., Hobor, S., Scala, E., Janakiraman, M., Liska, D., Valtorta, E., Schiavo, R., Buscarino, M., Siravegna, G., Bencardino, K., Cercek, A., Chen, C.T., Veronese, S., Zanon, C., Sartore-Bianchi, A., Gambacorta, M., Gallicchio, M., Vakiani, E., Boscaro, V., Medico, E., Weiser, M., Siena, S., Di Nicolantonio, F., Solit, D. & Bardelli, A. 2012, "Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer", *Nature,* vol. 486, no. 7404, pp. 532-536.

Mo, Z., Zhang, Q., Liu, Z., Lauer, J., Shi, Y., Sun, L., Griffin, P.R. & Yang, X.L. 2016, "Neddylation requires glycyl-tRNA synthetase to protect activated E2", *Nature structural & molecular biology,* vol. 23, no. 8, pp. 730-737.

Mohan, S., Heitzer, E., Ulz, P., Lafer, I., Lax, S., Auer, M., Pichler, M., Gerger, A., Eisner, F., Hoefler, G., Bauernhofer, T., Geigl, J.B. & Speicher, M.R. 2014, "Changes in colorectal carcinoma genomes under anti-EGFR therapy identified by whole-genome plasma DNA sequencing", *PLoS genetics,* vol. 10, no. 3, pp. e1004271.

Montagut, C., Dalmases, A., Bellosillo, B., Crespo, M., Pairet, S., Iglesias, M., Salido, M., Gallen, M., Marsters, S., Tsai, S.P., Minoche, A., Seshagiri, S., Serrano, S., Himmelbauer, H., Bellmunt, J., Rovira, A., Settleman, J., Bosch, F. & Albanell, J. 2012, "Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer", *Nature medicine,* vol. 18, no. 2, pp. 221-223.

Moreira, R.B., Peixoto, R.D. & de Sousa Cruz, M R 2015, "Clinical Response to Sorafenib in a Patient with Metastatic Colorectal Cancer and FLT3 Amplification", *Case reports in oncology,* vol. 8, no. 1, pp. 83-87.

Mouliere, F., Robert, B., Arnau Peyrotte, E., Del Rio, M., Ychou, M., Molina, F., Gongora, C. & Thierry, A.R. 2011, "High fragmentation characterizes tumour-derived circulating DNA", *PloS one,* vol. 6, no. 9, pp. e23418.

Murtaza, M. & Caldas, C. 2016, "Nucleosome mapping in plasma DNA predicts cancer gene expression", *Nature genetics,* vol. 48, no. 10, pp. 1105-1106.

Murtaza, M., Dawson, S.J., Pogrebniak, K., Rueda, O.M., Provenzano, E., Grant, J., Chin, S.F., Tsui, D.W.Y., Marass, F., Gale, D., Ali, H.R., Shah, P., Contente-Cuomo, T., Farahani, H., Shumansky, K., Kingsbury, Z., Humphray, S., Bentley, D., Shah, S.P., Wallis, M., Rosenfeld, N. & Caldas, C. 2015, "Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer", *Nature communications,* vol. 6, pp. 8760.

Nicolini, A., Ferrari, P. & Duffy, M.J. 2018, "Prognostic and predictive biomarkers in breast cancer: Past, present and future", *Seminars in cancer biology,* vol. 52, no. Pt 1, pp. 56-73.

Nishimura, R., Arima, N., Toyoshima, S., Ohi, Y., Anan, K., Sagara, Y., Mitsuyama, S. & Tamura, K. 2013, "Evaluation of PTEN loss and PIK3CA mutations and their correlation with efficacy of trastuzumab treatment in HER2-positive metastatic breast cancer: A retrospective study (KBC-SG 1001)", *Molecular and clinical oncology,* vol. 1, no. 1, pp. 47-52.

Niu, Y., Xu, J. & Sun, T. 2019, "Cyclin-Dependent Kinases 4/6 Inhibitors in Breast Cancer: Current Status, Resistance, and Combination Strategies", *Journal of Cancer,* vol. 10, no. 22, pp. 5504-5517.

Nong, J., Gong, Y., Guan, Y., Yi, X., Yi, Y., Chang, L., Yang, L., Lv, J., Guo, Z., Jia, H., Chu, Y., Liu, T., Chen, M., Byers, L., Roarty, E., Lam, V.K., Papadimitrakopoulou, V.A., Wistuba, I., Heymach, J.V., Glisson, B., Liao, Z., Lee, J.J., Futreal, P.A., Zhang, S., Xia, X., Zhang, J. & Wang, J. 2018, "Circulating tumor DNA analysis depicts subclonal architecture and genomic evolution of small cell lung cancer", *Nature communications,* vol. 9, no. 1, pp. 3114-018.

Ohhara, Y., Fukuda, N., Takeuchi, S., Honma, R., Shimizu, Y., Kinoshita, I. & Dosaka-Akita, H. 2016, "Role of targeted therapy in metastatic colorectal cancer", *World journal of gastrointestinal oncology,* vol. 8, no. 9, pp. 642-655.

Olofsson, B., Korpelainen, E., Pepper, M.S., Mandriota, S.J., Aase, K., Kumar, V., Gunji, Y., Jeltsch, M.M., Shibuya, M., Alitalo, K. & Eriksson, U. 1998, "Vascular endothelial growth factor B (VEGF-B) binds to VEGF receptor-1 and regulates plasminogen activator activity in endothelial cells", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 95, no. 20, pp. 11709-11714.

Onstenk, W., Sieuwerts, A.M., Mostert, B., Lalmahomed, Z., Bolt-de Vries, J.B., van Galen, A., Smid, M., Kraan, J., Van, M., de Weerd, V., Ramirez-Moreno, R., Biermann, K., Verhoef, C., Grunhagen, D.J., IJzermans, J.N., Gratama, J.W., Martens, J.W., Foekens, J.A. & Sleijfer, S.

2016, "Molecular characteristics of circulating tumor cells resemble the liver metastasis more closely than the primary tumor in metastatic colorectal cancer", *Oncotarget,* vol. 7, no. 37, pp. 59058-59069.

Park, J.L., Kim, H.J., Choi, B.Y., Lee, H.C., Jang, H.R., Song, K.S., Noh, S.M., Kim, S.Y., Han, D.S. & Kim, Y.S. 2012, "Quantitative analysis of cell-free DNA in the plasma of gastric cancer patients", *Oncology letters,* vol. 3, no. 4, pp. 921-926.

Pentheroudakis, G., Kotoula, V., De Roock, W., Kouvatseas, G., Papakostas, P., Makatsoris, T., Papamichael, D., Xanthakis, I., Sgouros, J., Televantou, D., Kafiri, G., Tsamandas, A.C., Razis, E., Galani, E., Bafaloukos, D., Efstratiou, I., Bompolaki, I., Pectasides, D., Pavlidis, N., Tejpar, S. & Fountzilas, G. 2013, "Biomarkers of benefit from cetuximab-based therapy in metastatic colorectal cancer: interaction of EGFR ligand expression with RAS/RAF, PIK3CA genotypes", *BMC cancer,* vol. 13, pp. 49-2407.

Phallen, J., Sausen, M., Adleff, V., Leal, A., Hruban, C., White, J., Anagnostou, V., Fiksel, J., Cristiano, S., Papp, E., Speir, S., Reinert, T., Orntoft, M.W., Woodward, B.D., Murphy, D., Parpart-Li, S., Riley, D., Nesselbush, M., Sengamalay, N., Georgiadis, A., Li, Q.K., Madsen, M.R., Mortensen, F.V., Huiskens, J., Punt, C., van Grieken, N., Fijneman, R., Meijer, G., Husain, H., Scharpf, R.B., Diaz, L.A., Jones, S., Angiuoli, S., Orntoft, T., Nielsen, H.J., Andersen, C.L. & Velculescu, V.E. 2017, "Direct detection of early-stage cancers using circulating tumor DNA", *Science translational medicine,* vol. 9, no. 403, pp. 10.1126/scitranslmed.aan2415.

Portman, N., Alexandrou, S., Carson, E., Wang, S., Lim, E. & Caldon, C.E. 2019, "Overcoming CDK4/6 inhibitor resistance in ER-positive breast cancer", *Endocrine-related cancer,* vol. 26, no. 1, pp. R15-R30.

Provencio, M., Torrente, M., Calvo, V., Perez-Callejo, D., Gutierrez, L., Franco, F., Perez-Barrios, C., Barquin, M., Royuela, A., Garcia-Garcia, F., Bueno, C., Garcia-Grande, A., Camps, C., Massuti, B., Sotomayor, E. & Romero, A. 2017, "Prognostic value of quantitative ctDNA levels in non small cell lung cancer patients", *Oncotarget,* vol. 9, no. 1, pp. 488-494.

Reinert, T., Henriksen, T.V., Christensen, E., Sharma, S., Salari, R., Sethi, H., Knudsen, M., Nordentoft, I., Wu, H.T., Tin, A.S., Heilskov Rasmussen, M., Vang, S., Shchegrova, S., Frydendahl Boll Johansen, A., Srinivasan, R., Assaf, Z., Balcioglu, M., Olson, A., Dashner, S., Hafez, D., Navarro, S., Goel, S., Rabinowitz, M., Billings, P., Sigurjonsson, S., Dyrskjot, L., Swenerton, R., Aleshin, A., Laurberg, S., Husted Madsen, A., Kannerup, A.S., Stribolt, K., Palmelund Krag, S., Iversen, L.H., Gotschalck Sunesen, K., Lin, C.J., Zimmermann, B.G. & Lindbjerg Andersen, C. 2019, "Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer", *JAMA oncology,* .

Reissfelder, C., Timke, C., Schmitz-Winnenthal, H., Rahbari, N.N., Koch, M., Klug, F., Roeder, F., Edler, L., Debus, J., Buchler, M.W., Beckhove, P., Huber, P.E. & Weitz, J. 2011, "A

randomized controlled trial to investigate the influence of low dose radiotherapy on immune stimulatory effects in liver metastases of colorectal cancer", *BMC cancer,* vol. 11, pp. 419-2407.

Roh, S.A., Park, I.J., Yoon, Y.S., Kwon, Y.H., Chung, J.H., Kim, T.W., Cho, D.H., Lim, B.H., Kim, S.K., Kim, S.Y., Kim, Y.S. & Kim, J.C. 2016, "Feasibility of novel PPP1R15A and proposed ANXA11 single nucleotide polymorphisms as predictive markers for bevacizumab regimen in metastatic colorectal cancer", *Journal of cancer research and clinical oncology,* vol. 142, no. 8, pp. 1705-1714.

Ross, J.S., Fakih, M., Ali, S.M., Elvin, J.A., Schrock, A.B., Suh, J., Vergilio, J.A., Ramkissoon, S., Severson, E., Daniel, S., Fabrizio, D., Frampton, G., Sun, J., Miller, V.A., Stephens, P.J. & Gay, L.M. 2018, "Targeting HER2 in colorectal cancer: The landscape of amplification and short variant mutations in ERBB2 and ERBB3", *Cancer,* vol. 124, no. 7, pp. 1358-1373.

Said, R., Guibert, N., Oxnard, G.R. & Tsimberidou, A.M. 2020, "Circulating tumor DNA analysis in the era of precision oncology", *Oncotarget,* vol. 11, no. 2, pp. 188-211.

Saif, M.W. 2013, "Anti-VEGF agents in metastatic colorectal cancer (mCRC): are they all alike?", *Cancer management and research,* vol. 5, pp. 103-115.

Salari, K., Spulak, M.E., Cuff, J., Forster, A.D., Giacomini, C.P., Huang, S., Ko, M.E., Lin, A.Y., van de Rijn, M. & Pollack, J.R. 2012, "CDX2 is an amplified lineage-survival oncogene in colorectal cancer", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 109, no. 46, pp. E3196-205.

Salem, Y., Yacov, N., Propheta-Meiran, O., Breitbart, E. & Mendel, I. 2019, "Newly characterized motile sperm domain-containing protein 2 promotes human breast cancer metastasis", *International journal of cancer,* vol. 144, no. 1, pp. 125-135.

Saltz, L.B., Meropol, N.J., Loehrer PJ, S., Needle, M.N., Kopit, J. & Mayer, R.J. 2004, "Phase II trial of cetuximab in patients with refractory colorectal cancer that expresses the epidermal growth factor receptor", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 22, no. 7, pp. 1201-1208.

Sandberg, T.P., Sweere, I., van Pelt, G.W., Putter, H., Vermeulen, L., Kuppen, P.J., Tollenaar, R A E M & Mesker, W.E. 2019, "Prognostic value of low CDX2 expression in colorectal cancers with a high stromal content - a short report", *Cellular oncology (Dordrecht),* vol. 42, no. 3, pp. 397-403.

Scholer, L.V., Reinert, T., Orntoft, M.W., Kassentoft, C.G., Arnadottir, S.S., Vang, S., Nordentoft, I., Knudsen, M., Lamy, P., Andreasen, D., Mortensen, F.V., Knudsen, A.R., Stribolt, K., Sivesgaard, K., Mouritzen, P., Nielsen, H.J., Laurberg, S., Orntoft, T.F. & Andersen, C.L. 2017, "Clinical Implications of Monitoring Circulating Tumor DNA in Patients with

Colorectal Cancer", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 23, no. 18, pp. 5437-5445.

Schwartz, L.H., Litiere, S., de Vries, E., Ford, R., Gwyther, S., Mandrekar, S., Shankar, L., Bogaerts, J., Chen, A., Dancey, J., Hayes, W., Hodi, F.S., Hoekstra, O.S., Huang, E.P., Lin, N., Liu, Y., Therasse, P., Wolchok, J.D. & Seymour, L. 2016, "RECIST 1.1-Update and clarification: From the RECIST committee", *European journal of cancer (Oxford, England : 1990),* vol. 62, pp. 132-137.

Schwarzenbach, H. 2013, "Circulating nucleic acids as biomarkers in breast cancer", *Breast cancer research : BCR,* vol. 15, no. 5, pp. 211.

Schwarzenbach, H. 2012, "Circulating nucleic acids and protease activities in blood of tumor patients", *Expert opinion on biological therapy,* vol. 12 Suppl 1, pp. S163-9.

Schwarzenbach, H., Alix-Panabieres, C., Muller, I., Letang, N., Vendrell, J.P., Rebillard, X. & Pantel, K. 2009, "Cell-free tumor DNA in blood plasma as a marker for circulating tumor cells in prostate cancer", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 15, no. 3, pp. 1032-1038.

Schwarzenbach, H., Hoon, D.S. & Pantel, K. 2011, "Cell-free nucleic acids as biomarkers in cancer patients", *Nature reviews.Cancer,* vol. 11, no. 6, pp. 426-437.

Seeber, A. & Gastl, G. 2016, "Targeted Therapy of Colorectal Cancer", *Oncology research and treatment,* vol. 39, no. 12, pp. 796-802.

Serpas, L., Chan, R.W.Y., Jiang, P., Ni, M., Sun, K., Rashidfarrokhi, A., Soni, C., Sisirak, V., Lee, W.S., Cheng, S.H., Peng, W., Chan, K.C.A., Chiu, R.W.K., Reizis, B. & Lo, Y.M.D. 2019, "Dnase1l3 deletion causes aberrations in length and end-motif frequencies in plasma DNA", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 116, no. 2, pp. 641-649.

Shirakata, Y., Komurasaki, T., Toyoda, H., Hanakawa, Y., Yamasaki, K., Tokumaru, S., Sayama, K. & Hashimoto, K. 2000, "Epiregulin, a novel member of the epidermal growth factor family, is an autocrine growth factor in normal human keratinocytes", *The Journal of biological chemistry,* vol. 275, no. 8, pp. 5748-5753.

Siegel, R.L., Miller, K.D. & Jemal, A. 2019, "Cancer statistics, 2019", *CA: a cancer journal for clinicians,* vol. 69, no. 1, pp. 7-34.

Siravegna, G., Marsoni, S., Siena, S. & Bardelli, A. 2017, "Integrating liquid biopsies into the management of cancer", *Nature reviews.Clinical oncology,* vol. 14, no. 9, pp. 531-548.

Sobhani, I., Itti, E., Luciani, A., Baumgaertner, I., Layese, R., Andre, T., Ducreux, M., Gornet, J.M., Goujon, G., Aparicio, T., Taieb, J., Bachet, J.B., Hemery, F., Retbi, A., Mons, M., Flicoteaux, R., Rhein, B., Baron, S., Cherrak, I., Rufat, P., Le Corvoisier, P., de'Angelis, N., Natella, P.A., Maoulida, H., Tournigand, C., Durand Zaleski, I. & Bastuji-Garin, S. 2018, "Colorectal cancer (CRC) monitoring by 6-monthly 18FDG-PET/CT: an open-label multicentre randomised trial", *Annals of oncology : official journal of the European Society for Medical Oncology,* vol. 29, no. 4, pp. 931-937.

Sorenson, G.D., Pribish, D.M., Valone, F.H., Memoli, V.A., Bzik, D.J. & Yao, S.L. 1994, "Soluble normal and mutated DNA sequences from single-copy genes in human blood", *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology,* vol. 3, no. 1, pp. 67-71.

Sotelo, M.J., Garcia-Paredes, B., Aguado, C., Sastre, J. & Diaz-Rubio, E. 2014, "Role of cetuximab in first-line treatment of metastatic colorectal cancer", *World journal of gastroenterology,* vol. 20, no. 15, pp. 4208-4219.

Stahlberg, A., Krzyzanowski, P.M., Egyud, M., Filges, S., Stein, L. & Godfrey, T.E. 2017, "Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing", *Nature protocols,* vol. 12, no. 4, pp. 664-682.

Stahlberg, A., Krzyzanowski, P.M., Jackson, J.B., Egyud, M., Stein, L. & Godfrey, T.E. 2016, "Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing", *Nucleic acids research,* vol. 44, no. 11, pp. e105.

Stiksma, J., Grootendorst, D.C. & van der Linden, P W 2014, "CA 19-9 as a marker in addition to CEA to monitor colorectal cancer", *Clinical colorectal cancer,* vol. 13, no. 4, pp. 239-244.

Stojkovic Lalosevic, M., Stankovic, S., Stojkovic, M., Markovic, V., Dimitrijevic, I., Lalosevic, J., Petrovic, J., Brankovic, M., Pavlovic Markovic, A. & Krivokapic, Z. 2017, "Can preoperative CEA and CA19-9 serum concentrations suggest metastatic disease in colorectal cancer patients?", *Hellenic journal of nuclear medicine,* vol. 20, no. 1, pp. 41-45.

Stroun, M., Anker, P., Maurice, P., Lyautey, J., Lederrey, C. & Beljanski, M. 1989, "Neoplastic characteristics of the DNA found in the plasma of cancer patients", *Oncology,* vol. 46, no. 5, pp. 318-322.

Stroun, M., Lyautey, J., Lederrey, C., Olson-Sand, A. & Anker, P. 2001, "About the possible origin and mechanism of circulating DNA apoptosis and active DNA release", *Clinica chimica acta; international journal of clinical chemistry,* vol. 313, no. 1-2, pp. 139-142.

Symonds, E.L., Pedersen, S.K., Murray, D., Byrne, S.E., Roy, A., Karapetis, C., Hollington, P., Rabbitt, P., Jones, F.S., LaPointe, L., Segelov, E. & Young, G.P. 2020, "Circulating epigenetic biomarkers for detection of recurrent colorectal cancer", *Cancer,* .

Takahashi, N., Furuta, K., Taniguchi, H., Sasaki, Y., Shoji, H., Honma, Y., Iwasa, S., Okita, N., Takashima, A., Kato, K., Hamaguchi, T., Shimada, Y. & Yamada, Y. 2016, "Serum level of hepatocyte growth factor is a novel marker of predicting the outcome and resistance to the treatment with trastuzumab in HER2-positive patients with metastatic gastric cancer", *Oncotarget,* vol. 7, no. 4, pp. 4925-4938.

Tan, C.H. & Iyer, R. 2010, "Use of computed tomography in the management of colorectal cancer", *World journal of radiology,* vol. 2, no. 5, pp. 151-158.

TCGA 2016, *TCGA*.

Thakur, B.K., Zhang, H., Becker, A., Matei, I., Huang, Y., Costa-Silva, B., Zheng, Y., Hoshino, A., Brazier, H., Xiang, J., Williams, C., Rodriguez-Barrueco, R., Silva, J.M., Zhang, W., Hearn, S., Elemento, O., Paknejad, N., Manova-Todorova, K., Welte, K., Bromberg, J., Peinado, H. & Lyden, D. 2014, "Double-stranded DNA in exosomes: a novel biomarker in cancer detection", *Cell research,* vol. 24, no. 6, pp. 766-769.

Thompson, J.C., Yee, S.S., Troxel, A.B., Savitch, S.L., Fan, R., Balli, D., Lieberman, D.B., Morrissette, J.D., Evans, T.L., Bauml, J., Aggarwal, C., Kosteva, J.A., Alley, E., Ciunci, C., Cohen, R.B., Bagley, S., Stonehouse-Lee, S., Sherry, V.E., Gilbert, E., Langer, C., Vachani, A. & Carpenter, E.L. 2016, "Detection of Therapeutically Targetable Driver and Resistance Mutations in Lung Cancer Patients by Next-Generation Sequencing of Cell-Free Circulating Tumor DNA", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 22, no. 23, pp. 5772-5782.

Tie, J., Kinde, I., Wang, Y., Wong, H.L., Roebert, J., Christie, M., Tacey, M., Wong, R., Singh, M., Karapetis, C.S., Desai, J., Tran, B., Strausberg, R.L., Diaz, L.A., Papadopoulos, N., Kinzler, K.W., Vogelstein, B. & Gibbs, P. 2015, "Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer", *Annals of oncology : official journal of the European Society for Medical Oncology,* vol. 26, no. 8, pp. 1715-1722.

Tie, J., Wang, Y., Tomasetti, C., Li, L., Springer, S., Kinde, I., Silliman, N., Tacey, M., Wong, H.L., Christie, M., Kosmider, S., Skinner, I., Wong, R., Steel, M., Tran, B., Desai, J., Jones, I., Haydon, A., Hayes, T., Price, T.J., Strausberg, R.L., Diaz, L.A., Papadopoulos, N., Kinzler, K.W., Vogelstein, B. & Gibbs, P. 2016, "Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer", *Science translational medicine,* vol. 8, no. 346, pp. 346ra92.

Tokheim, C.J., Papadopoulos, N., Kinzler, K.W., Vogelstein, B. & Karchin, R. 2016, "Evaluating the evaluation of cancer driver genes", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 113, no. 50, pp. 14330-14335.

Tong, G., Xu, W., Zhang, G., Liu, J., Zheng, Z., Chen, Y., Niu, P. & Xu, X. 2018, "The role of tissue and serum carcinoembryonic antigen in stages I to III of colorectal cancer-A retrospective cohort study", *Cancer medicine,* vol. 7, no. 11, pp. 5327-5338.

Ulz, P., Belic, J., Graf, R., Auer, M., Lafer, I., Fischereder, K., Webersinke, G., Pummer, K., Augustin, H., Pichler, M., Hoefler, G., Bauernhofer, T., Geigl, J.B., Heitzer, E. & Speicher, M.R. 2016, "Whole-genome plasma sequencing reveals focal amplifications as a driving force in metastatic prostate cancer", *Nature communications,* vol. 7, pp. 12008.

Ulz, P., Perakis, S., Zhou, Q., Moser, T., Belic, J., Lazzeri, I., Wolfler, A., Zebisch, A., Gerger, A., Pristauz, G., Petru, E., White, B., Roberts, C.E.S., John, J.S., Schimek, M.G., Geigl, J.B., Bauernhofer, T., Sill, H., Bock, C., Heitzer, E. & Speicher, M.R. 2019, "Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection", *Nature communications,* vol. 10, no. 1, pp. 4666-019.

Ulz, P., Thallinger, G.G., Auer, M., Graf, R., Kashofer, K., Jahn, S.W., Abete, L., Pristauz, G., Petru, E., Geigl, J.B., Heitzer, E. & Speicher, M.R. 2016, "Inferring expressed genes by whole-genome sequencing of plasma DNA", *Nature genetics,* vol. 48, no. 10, pp. 1273-1278.

Underhill, H.R., Kitzman, J.O., Hellwig, S., Welker, N.C., Daza, R., Baker, D.N., Gligorich, K.M., Rostomily, R.C., Bronner, M.P. & Shendure, J. 2016, "Fragment Length of Circulating Tumor DNA", *PLoS genetics,* vol. 12, no. 7, pp. e1006162.

Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z. & Sidow, A. 2011, "Determinants of nucleosome organization in primary human cells", *Nature,* vol. 474, no. 7352, pp. 516-520.

Valtorta, E., Misale, S., Sartore-Bianchi, A., Nagtegaal, I.D., Paraf, F., Lauricella, C., Dimartino, V., Hobor, S., Jacobs, B., Ercolani, C., Lamba, S., Scala, E., Veronese, S., Laurent-Puig, P., Siena, S., Tejpar, S., Mottolese, M., Punt, C.J., Gambacorta, M., Bardelli, A. & Di Nicolantonio, F. 2013, "KRAS gene amplification in colorectal cancer and impact on response to EGFR-targeted therapy", *International journal of cancer,* vol. 133, no. 5, pp. 1259-1265.

Van der Veldt, A A, Lubberink, M., Bahce, I., Walraven, M., de Boer, M.P., Greuter, H.N., Hendrikse, N.H., Eriksson, J., Windhorst, A.D., Postmus, P.E., Verheul, H.M., Serne, E.H., Lammertsma, A.A. & Smit, E.F. 2012, "Rapid decrease in delivery of chemotherapy to tumors after anti-VEGF therapy: implications for scheduling of anti-angiogenic drugs", *Cancer cell,* vol. 21, no. 1, pp. 82-91.

van Dijk, E., Biesma, H.D., Cordes, M., Smeets, D., Neerincx, M., Das, S., Eijk, P.P., Murphy, V., Barat, A., Bacon, O., Prehn, J.H.M., Betge, J., Gaiser, T., Fender, B., Meijer, G.A., McNamara,

D.A., Klinger, R., Koopman, M., Ebert, M.P.A., Kay, E.W., Hennessey, B.T., Verheul, H.M.W., Gallagher, W.M., O'Connor, D.P., Punt, C.J.A., Loupakis, F., Lambrechts, D., Byrne, A.T., van Grieken, N C T & Ylstra, B. 2018, "Loss of Chromosome 18q11.2-q12.1 Is Predictive for Survival in Patients With Metastatic Colorectal Cancer Treated With Bevacizumab", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 36, no. 20, pp. 2052-2060.

Vasioukhin, V., Anker, P., Maurice, P., Lyautey, J., Lederrey, C. & Stroun, M. 1994, "Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia", *British journal of haematology,* vol. 86, no. 4, pp. 774-779.

Vukobrat-Bijedic, Z., Husic-Selimovic, A., Sofic, A., Bijedic, N., Bjelogrlic, I., Gogov, B. & Mehmedovic, A. 2013, "Cancer Antigens (CEA and CA 19-9) as Markers of Advanced Stage of Colorectal Carcinoma", *Medical archives (Sarajevo, Bosnia and Herzegovina),* vol. 67, no. 6, pp. 397-401.

Wan, J.C.M., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J.D., Caldas, C., Pacey, S., Baird, R. & Rosenfeld, N. 2017, "Liquid biopsies come of age: towards implementation of circulating tumour DNA", *Nature reviews.Cancer,* vol. 17, no. 4, pp. 223-238.

Wang, D., Sun, H., Wei, J., Cen, B. & DuBois, R.N. 2017, "CXCL1 Is Critical for Premetastatic Niche Formation and Metastasis in Colorectal Cancer", *Cancer research,* vol. 77, no. 13, pp. 3655-3665.

Wang, M., Niu, W., Hu, R., Wang, Y., Liu, Y., Liu, L., Zhong, J., Zhang, C., You, H., Zhang, J., Lu, L., Wei, L. & Xiao, W. 2018, "POLR1D promotes colorectal cancer progression and predicts poor prognosis of patients", *Molecular carcinogenesis,* .

Wong, N., Gu, Y., Kapoor, A., Lin, X., Ojo, D., Wei, F., Yan, J., de Melo, J., Major, P., Wood, G., Aziz, T., Cutz, J.C., Bonert, M., Patterson, A.J. & Tang, D. 2017, "Upregulation of FAM84B during prostate cancer progression", *Oncotarget,* vol. 8, no. 12, pp. 19218-19235.

Wu, T.L., Zhang, D., Chia, J.H., Tsao, K.-., Sun, C.F. & Wu, J.T. 2002, "Cell-free DNA: measurement in various carcinomas and establishment of normal reference range", *Clinica chimica acta; international journal of clinical chemistry,* vol. 321, no. 1-2, pp. 77-87.

Yamamoto, Y., Uemura, M., Fujita, M., Maejima, K., Koh, Y., Matsushita, M., Nakano, K., Hayashi, Y., Wang, C., Ishizuya, Y., Kinouchi, T., Hayashi, T., Matsuzaki, K., Jingushi, K., Kato, T., Kawashima, A., Ujike, T., Nagahara, A., Fujita, K., Imamura, R., Nakagawa, H. & Nonomura, N. 2019, "Clinical significance of the mutational landscape and fragmentation of circulating tumor DNA in renal cell carcinoma", *Cancer science,* vol. 110, no. 2, pp. 617-628.

Yarden, Y. & Sliwkowski, M.X. 2001, "Untangling the ErbB signalling network", *Nature reviews.Molecular cell biology,* vol. 2, no. 2, pp. 127-137.

Yau, T.O. 2019, "Precision treatment in colorectal cancer: Now and the future", *JGH open : an open access journal of gastroenterology and hepatology,* vol. 3, no. 5, pp. 361-369.

Yu, J., Liu, D., Sun, X., Yang, K., Yao, J., Cheng, C., Wang, C. & Zheng, J. 2019, "CDX2 inhibits the proliferation and tumor formation of colon cancer cells by suppressing Wnt/beta-catenin signaling via transactivation of GSK-3beta and Axin2 expression", *Cell death & disease,* vol. 10, no. 1, pp. 26-018.

Yu, S.C., Lee, S.W., Jiang, P., Leung, T.Y., Chan, K.C., Chiu, R.W. & Lo, Y.M. 2013, "High-resolution profiling of fetal DNA clearance from maternal plasma by massively parallel sequencing", *Clinical chemistry,* vol. 59, no. 8, pp. 1228-1237.

Zauli, G., Voltan, R., Tisato, V. & Secchiero, P. 2012, "State of the art of the therapeutic perspective of sorafenib against hematological malignancies", *Current medicinal chemistry,* vol. 19, no. 28, pp. 4875-4884.

Zhao, B., Wang, L., Qiu, H., Zhang, M., Sun, L., Peng, P., Yu, Q. & Yuan, X. 2017, "Mechanisms of resistance to anti-EGFR therapy in colorectal cancer", *Oncotarget,* vol. 8, no. 3, pp. 3980-4000.

Zhou, Q., Perakis, S.O., Ulz, P., Mohan, S., Riedl, J.M., Talakic, E., Lax, S., Totsch, M., Hoefler, G., Bauernhofer, T., Pichler, M., Gerger, A., Geigl, J.B., Heitzer, E. & Speicher, M.R. 2020, "Cell-free DNA analysis reveals POLR1D-mediated resistance to bevacizumab in colorectal cancer", *Genome medicine,* vol. 12, no. 1, pp. 20-020.

SUPPLEMENTARY TABLES

Supplementary Tables 1: Summary of all recurrent focal events with a frequency over 0.01 in our CRC cohort. Potential driver genes were identified as in Figure 12B [Table and legend originally published in Genome Medicine (Zhou et al., 2020)]

Chromosome	Start	End	CytoBand	Frequency	Type	Potential driver gene
chr1	150044901	150495806	1q21.1 - 1q21.3	0.02000	Gain	TARS2
chr1	150722211	151342717	1q21.3	0.03333	Gain	CTSK, SEMA6C, SETDB1, ZNF687
chr1	154458740	156896478	1q21.3-1q23.1	0.01333	Gain	ASH1L, CHRNB2, DAP3, RIT1, ZBTB7B, TDRD10, INSRR, NTRK1, NES
chr10	76781343	77515310	10q22.2	0.01333	Gain	ZNF503
chr12	16946077	18814875	12p12.3	0.01333	Gain	PIK3C2G
chr12	21872717	24534720	12p12.1	0.02667	Gain	ABCC9, C2CD5, KCNJ8, SOX5
chr12	25324271	25436943	12p12.1	0.06667	Gain	KRAS
chr12	26342358	26794292	12p12.1-p11.23	0.02000	Gain	ITPR2
chr12	120398658	124005022	12q24.23-q24-31	0.01333	Gain	ACADS, ANAPC5, C12orf43, CLIP1, DENR, GATC, HNF1A, KNTC1, MSI1, SETD1B
chr13	25677187	26749971	13q12.13	0.02667	Gain	ATP8A2
chr13	27708804	29118120	13q12.13-q12.3	0.08667	Gain	FLT3, RPL21, FLT1
chr13	29399846	31723136	13q12.3	0.02667	Gain	KATNAL1, MTUS2, TEX26
chr13	32291648	33193829	13q13.1	0.01333	Gain	BRCA2, RXFP2
chr15	90329711	91571230	15q26.1	0.02000	Gain	ANPEP, BLM, CRTC3, IDH2, IQGAP1, BLM, VPS33B
chr17	25558360	26860022	17q11.1-q11.2	0.02000	Gain	SARM1
chr17	36441327	36838077	17q12	0.02000	Gain	ARHGAP23
chr17	37458601	38022528	17q12	0.03333	Gain	ERBB2, MED1, CDK12, PPP1R1B, PNMT, IKZF3
chr17	38812164	38868521	17q21.2	0.01333	Gain	KRT222

chr17	39662987	39775716	17q21.2	0.01333	Gain	KRT15
chr17	40170346	40339418	17q21.2	0.01333	Gain	RAB5C, NKIRAS2
chr17	11651944	12274316	17p12	0.01333	Loss	MAP2K4
chr18	249470	988653	18p11.32	0.02000	Gain	COLEC12
chr19	38882717	39336407	19q13.2	0.02000	Gain	RYR1, ACTN4
chr19	39449097	41441718	19q13.2	0.02000	Gain	SUPT5H, LGALS13, ZNF780A, ZNF546
chr2	60888652	65519367	2p16.1-p14	0.01333	Gain	REL, USP34, XPO1
chr20	29807477	42005957	20q11.21-q13.11	0.01333	Gain	CCM2L, DEFB118, PTPRT
chr20	42517406	42743051	20q13.12	0.03333	Gain	TOX2
chr20	43532342	44717354	20q13.12	0.03333	Gain	STK4, SEMG2, MATN4, SLC12A5
chr20	52173316	52512073	20q13.2	0.01333	Gain	ZNF217
chr22	22281984	22565196	22q11.22	0.01333	Gain	TOP3B
chr3	169876604	170215035	3q26.2	0.01333	Gain	PRKCI, PHC3
chr3	178871529	179041343	3q26.32-q26.33	0.02000	Gain	PIK3CA
chr5	34685364	39659257	5p13.2-p13.1	0.02000	Gain	UGT3A1,SPEF2,PRLR,IL7R,CAPSL,BRIX1,AGXT2,C9, DAB2, EGFLAM, FYB, LIFR, UGT3A2, RANBP3L, OSMR, NIPBL
chr5	40397174	41757881	5p13.1	0.02000	Gain	PTGER4,MROH2B,CARD6,C7,C6,PLCXD3
chr6	221883	2716626	6p25.3-p25.2	0.02667	Gain	DUSP22, FOXQ1
chr6	3622361	6569112	6p25.2-p25.1	0.01333	Gain	FAM50B,F13A1
chr6	41684263	42643866	6p21.1	0.01333	Gain	USP49,CCND3
chr6	43095330	43377105	6p21.1	0.01333	Gain	ZNF318, SLC22A7
chr6	168332123	168557886	6q27	0.01333	Gain	MLLT4
chr7	55038885	55548664	7p11.2	0.01333	Gain	EGFR
chr7	72122259	76266287	7q11.22-q11.23	0.01333	Gain	LIMK1, MDH2, WBSCR16
chr7	100276775	100447258	7q22.1	0.01333	Gain	ZAN,EPO

chr7	114506009	114846100	7q31.1-q31.2	0.01333	Gain	MDFIC
chr7	116259840	116487551	7q31.2	0.02000	Gain	MET
chr7	116543893	118409416	7q31.2-q31.31	0.01333	Gain	WNT2, CTTNBP2, CFTR, ASZ1
chr8	37546484	37715573	8p11.23	0.01333	Gain	ZNF703
chr8	37828508	38110537	8p11.23	0.04000	Gain	EIF4EBP1, BAG4
chr8	38223403	39915647	8p11.23-p11.21	0.05333	Gain	FGFR1, ADAM2, ADAM18
chr8	41219037	41671065	8p11.21	0.01333	Gain	ANK1
chr8	48829123	48941964	8q11.21	0.01333	Gain	MCM4
chr8	128584220	128810067	8q24.21	0.02000	Gain	MYC
chr9	21556676	22802849	9p21.3	0.01333	Loss	CDKN2A
chr9	33948775	37113914	9p13.3-p13.2	0.01333	Gain	DNAJB5, FANCG, GLIPR2, KIAA1045, VCP, PAX5, PIGO

Supplementary Tables 2: Summary of all recurrent focal events with a frequency over 0.01 in TCGA cohort. Potential driver genes were identified as in Figure 12B [Table and legend originally published in Genome Medicine (Zhou et al., 2020)]

Chromosome	Start	End	CytoBand	Frequency	Type	Potential driver gene
chr1	240055405	247813706	1q43-q44	0.0134	Gain	NLRP3,OR2B11,OR2C3,OR2G2,OR2G3,FMN2,RGS7,EXO1,C1orf100,ZNF124,PLD5,GCSAML
chr10	87970222	90374638	10q23.2-q23.31	0.0357	Loss	PTEN,WAPAL,GLUD1,GRID1
chr12	889902	4699133	12p13.33-p13.32	0.0193	Gain	WNK1,LRTM2,FGF6,CACNA1C,DYRK4
chr12	24274378	24372426	12p12.1	0.0104	Gain	SOX5
chr12	24373828	24393754	12p12.1	0.0104	Gain	SOX5
chr12	24502128	28473666	12p12.1-p11.22	0.0163	Gain	KRAS,PTHLH,ITPR2,MRPS35,SOX5
chr13	19883672	21822880	13q12.11	0.0163	Gain	TPTE2,IFT88
chr13	23973888	30275107	13q12.12-q12.3	0.0446	Gain	FLT3,SPATA13,RPL21,FLT1,MTUS2,ATP12A,RNF17,ATP8A2
chr13	48656553	48880148	13q14.2	0.0104	Loss	RB1
chr13	48888473	49409343	13q14.2	0.0134	Loss	RB1
chr17	25270517	28450667	17q11.1-q11.2	0.0238	Gain	RAB34,TIAF1,SARM1,TAOK1,GIT1
chr17	36410559	58350644	17q12-q23.2	0.0386	Gain	ERBB2,SPOP,KANSL1,CDK12,BRCA1,KRT15,CA10,WFIKKN2,ABCC3,ARHGAP23,ARHGAP27,COL1A1,COPZ2,DGKE,FZD2,GFAP,HOXB1,HOXB3,IGF2BP1,IKZF3,KIF2B,KRT12,KRT13,KRT20,KRT222,KRTAP4-11,KRTAP4-5,LUC7L3,MED1,NKIRAS2,PNMT,PPP1R1B,RAB5C,RNF43,RPL27,SPAG9,STAT3,TMEM100,UBE2Z
chr17	61252252	61724695	17q23.3	0.0104	Gain	ACE
chr17	79042975	79513952	17q25.3	0.0104	Gain	ACTG1
chr17	11764238	12044287	17p12	0.0163	Loss	MAP2K4,DNAH9
chr18	329586	7831265	18p11.32-p11.23	0.0178	Gain	LRRC30,EPB41L3,COLEC12,LAMA1,PTPRM

chr18	7837860	10072276	18p11.23-p11.22	0.0134	Gain	TXNDC2,NDUFV2,ANKRD12,PTPRM
chr18	47957621	50052729	18q21.1-q21.2	0.0416	Loss	SMAD4,MAPK4,DCC
chr19	29489042	31284806	19q12	0.0149	Gain	VSTM2B,ZNF536
chr19	31290599	33013506	19q12-q13.11	0.0134	Gain	TSHZ3
chr20	29843973	31267648	20q11.21	0.0297	Gain	DEFB118,CCM2L,BCL2L1,ASXL1
chr20	41371926	44471340	20q12-q13.12	0.0253	Gain	SEMG2,TOX2,STK4,MATN4,PTPRT
chr20	44491614	44701824	20q13.12	0.0134	Gain	SLC12A5
chr5	37667440	38806643	5p13.2-p13.1	0.0104	Gain	EGFLAM,LIFR
chr5	38807123	38918715	5p13.1	0.0104	Gain	OSMR
chr5	38919619	39630569	5p13.1	0.0104	Gain	DAB2,OSMR,FYB,C9
chr6	41744948	44405268	6p21.1	0.0163	Gain	CCND3,POLR1C,TMEM151B,SLC22A7,ZNF318,HSP90AB1,USP49
chr7	54684099	55425140	7p11.2	0.0119	Gain	EGFR
chr8	35877471	42933372	8p12-p11.21	0.0565	Gain	ZNF703,FGFR1,EIF4EBP1,BAG4,ADAM2,ANK1,IKBKB,ADAM18,ANK1,KAT6A,KCNU1
chr8	101572814	102300630	8q22.2-q22.3	0.0149	Gain	YWHAZ
chr8	127039648	129858603	8q24.13-q24.21	0.0253	Gain	MYC
chr9	4953916	5087087	9p24.1	0.0104	Gain	JAK2,JAK2
chr9	21626573	22241415	9p21.3	0.0149	Loss	CDKN2A

Supplementary Tables 3: Summary of genes differently expressed after POLR1D knockdown in both HT29 and SW480 cells and their expression levels in TCGA dataset. 45 genes showed consistent expression change in HT29 and SW480 cell lines after POLR1D knockdown, log2 fold change and P-value are listed in the table. Additionally, expression levels of these genes were compared between 13q12.2 balanced and aberrant groups in the TCGA dataset. (Log2FC: log2 fold change of gene expression (Log2 (siPOLR1D/SCR)). *: adjusted P-value calculated by DESeq2 package. +: P-value calculated using Wilcoxon test.) [Table and legend originally published in Genome Medicine (Zhou et al., 2020)]

Gene	HT29 Log2FC	HT29 Padj*	SW480 Log2FC	SW480 Padj*	TCGA P value+	TCGA Balanced Median	TCGA Aberrant Median
POLR1D	-0.7128229	0.00449661	-1.0301794	1.40E-10	4.99054E-26	2513.89995	3793.959
EREG	-1.4452929	0.0293995	-2.0894708	5.23E-05	9.54296E-09	373.00665	1301.689
VEGFA	-1.6946917	0.02176345	-1.1063407	0.0004098	0.000279851	2938.6879	3403.325
GARS	-0.8629731	0.0233373	-0.9643462	6.10E-08	0.009231291	3417.0136	3774.436
FAM84B	-0.659154	0.00968512	-0.804392	0.00455962	0.04643275	1522.4487	1709.143
PPP1R15A	-2.4018937	0.04823467	-2.0948666	1.18E-10	0.01555854	1667.57915	1845.758
KIF21B	-0.8513724	0.02504416	-1.3588441	1.16E-06	0.04393476	644.7362	710.1464
MOSPD2	-0.9839675	0.02504416	-0.9077483	0.00062231	0.000228129	238.4314	287.4251
MUC12	2.188039	0.0233373	1.6257477	0.01996535	1.21902E-07	987.96985	2679.441
DPEP1	1.0391931	0.03239612	1.0525786	0.01479816	0.000155794	2265.18	3783.097
HSD11B2	1.2780104	0.00190125	0.9474168	0.00388939	0.006967245	1732.68055	2161.517
SULT2B1	0.6447917	0.01868489	1.0435684	0.01782547	0.007513332	365.1714	484.6336
DOCK11	2.0156599	0.02907911	1.0497497	0.0318923	0.00976697	348.3723	435.7683
PROX1	1.9051311	0.04074343	0.8948489	0.00071571	0.003398216	163.9231	205.283
MORN4	-3.2213428	0.0007858	-0.6895181	0.02557299	8.29056E-05	87.48645	72.8302
CTH	-1.9229918	0.01859367	-1.5466323	4.52E-06	0.001264022	122.837	103.0207
CHAC1	-2.3938038	5.73E-05	-1.3336525	0.01750432	0.004218698	168.54035	141.1333
SLFN5	-1.768586	0.00524378	-2.2540744	5.60E-05	0.001609381	293.0512	242.3706

SESN2	-1.9412111	0.00598858	-2.0298841	8.62E-05	0.001005439	454.92015	391.0093
SLC38A2	-1.1251007	0.00175032	-1.1995282	0.01106125	0.03182495	2945.4884	2855.136
SLC7A11	-1.6614718	0.00196317	-1.678202	0.00072969	0.000110188	488.9114	378.706
APOL6	-1.0914653	0.00060994	-1.5501895	0.00354982	1.02788E-06	1858.86865	1568.716
ANXA1	-0.6573977	0.03453267	-1.9886396	1.25E-08	0.001765714	1445.2876	1114.838
MTHFD2	-1.3845903	0.0123018	-0.9386858	0.00348556	2.39712E-08	2013.44365	1544.529
XBP1	-0.9617556	0.02530649	-0.9092207	7.07E-05	1.01505E-05	5308.0528	4634.824
DDIT4	-2.2934595	5.73E-05	-1.6754201	0.00909115	0.000306201	3003.3448	2195.472
TUBB	1.208217	0.02504902	1.7314544	0.02410687	0.05132316	16494.4767	15594.8
ULBP1	-2.3746428	0.00293501	-1.4778163	0.0048564	0.06732695	16.90435	13.9894
BEST1	-1.6286437	0.04111525	-1.581836	0.00190779	0.07428151	20.07455	24.1673
LONRF1	-2.3994092	0.00255374	-0.6202471	0.03162924	0.1227335	291.9928	317.0184
SP5	2.4616033	0.01598578	0.704191	0.00226095	0.1627161	189.3337	132.5616
HSPA13	-1.5490545	0.03706914	-1.0112029	1.75E-05	0.1939381	600.5065	568.3804
IL3RA	1.500118	0.0233373	1.4207262	0.01876354	0.2459109	111.17865	135.8867
ATF4	-0.7287908	0.01288611	-0.7945738	4.41E-05	0.4245156	6539.15705	6439.585
STRC	2.1308239	0.00293501	1.3832091	0.01774951	0.4389994	5.28165	5.5633
GDF15	-2.5897728	0.04074343	-1.6993912	0.0022295	0.5149138	2799.57955	2431.004
GPX8	-1.6006957	0.00011944	-1.1492875	0.00342902	0.5306847	307.09955	298.0921
SLC1A4	-2.2725932	0.00074873	-0.9699159	3.38E-06	0.5551172	1261.316	1267.928
INSIG2	-0.6693009	0.0142507	-0.8393069	0.00472022	0.5708263	491.41485	500.7547
GPT2	-0.9519746	0.02176345	-0.9747738	2.99E-06	0.6544763	1572.77765	1526.729
LGALS1	-0.799721	0.02530649	-0.744645	0.00746905	0.6621949	2338.41245	2124.263
AARS	-1.4168308	0.04716745	-0.8610822	0.00056494	0.7566797	3537.6686	3404.814
ATF3	-3.618937	0.00568068	-2.114485	0.00057144	0.7566797	773.1627	785.7408
CLIP4	-1.6842405	0.04890417	-1.3437026	0.00056158	0.7603699	46.61535	49.3107
DDIT3	-3.4201171	0.01604729	-1.9581415	4.04E-05	0.9748566	356.5413	369.2023