

Research Article

Non-Invasive Colorectal Cancer Mutation Profiling by Targeted Sequencing of Tumour-Derived DNA in the Tongue Coating of Cancer Patients

**Sze Chuen Cesar Wong^{1*}*, Xiao Meng Pei^{1*}, Emily Kai Yee Lam² *,
Lawrence Wing Chi Chan³, Hin Fung Tsang³**

¹Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region

²Department of Pathology, Hong Kong Children's Hospital, Hospital Authority, Hong Kong Special Administrative Region

³Department of Health Technology and Informatics, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region

***** Co-first Author

***Corresponding author:** Sze Chuen Cesar Wong, Department of Applied Biology and chemical Technology, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region

Citation: Cesar Wong SC, Pei XM, Yee Lam EK, Chi Chan LW, Tsang HF (2023) Non-Invasive Colorectal Cancer Mutation Profiling by Targeted Sequencing of Tumour-Derived DNA in the Tongue Coating of Cancer Patients. Int J Geriatr Gerontol 6: 146. DOI: 10.29011/2577-0748.100046

Received Date: 27 January, 2023; **Accepted Date:** 01 February, 2023; **Published Date:** 07 February, 2023

Abstract

Colorectal cancer is the third most common cancers worldwide. Early detection and treatment will improve the survival and prognosis. Despite the significant progress in non-invasive detection strategies, there are still a lot of limitations hindering them to become a routine practice. Therefore, it is necessary to explore other ways of non-invasive detection of colorectal cancer. Tongue coating is an attractive specimen which is completely non-invasive and safe for colorectal cancer detection as tongue is connected to the large intestine directly. In this study, we hypothesized that tongue coating may contain DNA arising from the primary tumor of the large intestine. To test this hypothesis, targeted sequencing which profiled somatic mutation hotspots in 48 cancer-associated genes was performed using tongue coating samples, the primary tumor and adjacent normal colorectal epithelial tissues from the same patient. We aimed to find out if there are tumor-associated mutations in the tongue coating samples. Results showed that 2 to 6 tumor-associated mutations, located in 8 genes of KIT, CDH1, CTNNB1, GNAS, RET, SMAD4, HNF1A and RB1, were found. Furthermore, 80-85% of the sequence variations were not detected in the corresponding buccal swab samples. Finally, 60-100% of the tumor-derived mutations in tongue coating disappeared after surgical operation. These interesting results are the first to demonstrate the presence of tumor-derived DNA mutations in the tongue coating of colorectal cancer patients. Moreover, these findings have shown the feasibility of using tongue coating samples for colorectal cancer detection and monitoring.

Keywords: Non-Invasive, Colorectal Cancer, Targeted Sequencing, Tumor-Derived DNA, Tongue Coating

Introduction

Colorectal cancer (CRC) is one of the most common cancers worldwide and it is the most common cancer and the second leading cause of cancer deaths in Hong Kong [1]. The disease is highly curable if detected at an early stage. However, early CRC is mostly symptomless [2]. A variety of screening tests have therefore been investigated for early detection of CRC [3,4]. Among them, faecal occult blood test has been the most extensively investigated, but it has low detection sensitivity on each round of screening [3,4]. Colonoscopy and sigmoidoscopy are the gold standards for examination of the colon and rectum. However, the cost, the need of full bowel preparation and sedation, and the small but definite risk of perforation make them less suitable for a widespread population screening [3,4]. Hence, there is a need to develop new non-invasive diagnostic methods for the detection and monitoring of CRC.

Carcinoembryonic antigen (CEA) is a widely used serum marker for CRC, but is unreliable in detecting postoperative recurrence [5]. Other biomarkers utilizing thymidylate synthase [6], vascular endothelial growth factor [7], loss of heterozygosity at 18q [8] and microsatellite instability [9] may either be prognostic or predictive of treatment response. However, they could not provide additional clinical values to the classical method using histopathologic features by pathologists. Imaging modalities such as positron emission tomography scan and magnetic resonance colonoscopy are useful in the prognosis of long-term survival of CRC patients, but these methods are too expensive for routine postoperative surveillance.

In Hong Kong as well as in other Western countries, 25% and 30% of CRC patients, present respectively with Tumor -Node-Metastasis (TNM) stages II and III [10], have a high risk of post-operative recurrence [11]. Historically, their 3-year disease-free survival was about 45-55%, and the 5-year overall survival was only 60% among patients who were treated with surgery alone [12]. The use of adjuvant 5-fluorouracil (5-FU)-based chemotherapy improves the disease-free survival by an absolute margin of around 16-18%, and overall survival of around 10-12% [13]. Oxaliplatin further adds to the benefit of 5-FU [14]. However, both 5-FU and oxaliplatin have acute and long term side effects, and not all CRC patients benefit from such treatment [14]. Hence, these adjuvant chemotherapies should be applied carefully. TNM classification is the most commonly used method in making therapeutic decision, but it is not reliable in identifying patients with “high-risk stage IIB” CRC who may need more aggressive adjuvant chemotherapy [6].

As the clinical behaviour of CRC is resulted from genetic, epigenetic and environmental interactions at multiple levels, a thorough understanding of the molecular basis of CRC is of utmost important to develop effective control measures. With the advent of high-throughput Next Generation Sequencing (NGS)

technologies, it is an appropriate time to discover more clinically relevant markers which will improve the patient outcome.

Targeted sequencing, which involves a selective enrichment or amplification of genomic regions of interest before NGS, has provided a cost-effective alternative for mutation profiling of a relatively large number of samples. For example, Han *et al* have sequenced the exons of 183 cancer-related genes in 60 colorectal adenocarcinomas and identified *APC*, *TP53*, and *KRAS* as the most commonly altered genes [15]. Moreover, Shao *et al* have sequenced the whole gene-body regions of 28 genes related to CRC in 30 cancer samples, and found that two single-nucleotide polymorphisms in the genes of *antigen presenting tapasin binding protein* and *transcription factor 3* were associated with patient survival [16]. Targeted NGS may therefore represent a valuable tool for a systematic screening and identification of novel tumor markers which could potentially improve the diagnostic accuracy as well as tumor classification of CRC [17].

Tongue coating is a potentially attractive kind of specimen for clinical diagnostic laboratory because its collection is simple and completely safe. However, only a few studies on the disease association of tongue coating have been reported thus far. In the patients with chorionic gastritis, microbial ribosomal RNA patterns and markers of glucose metabolites in tongue coating have been investigated [18]. A differential expression of keratin has also been observed in the tongue coating of patients with hepatitis B [19]. Recently, Jiang *et al* have correlated traditional Chinese medicine based tongue diagnosis with the tongue coating microbiomes as determined by NGS in chronic atrophic gastritis patients [20]. Apart from microbial nucleic acid analysis, there is a lack of molecular study on the non-oral-derived DNA in tongue coating. Being both the organs of the digestive system, the tongue is all the way linked to the large intestine. We therefore hypothesize that tongue coating may contain DNA arising from the primary tumor of the large intestine. To test this hypothesis, targeted NGS was performed for all of the samples using the TruSeq Amplicon Cancer Panel (Illumina, Inc, San Diego, USA) which profiled somatic mutation hotspots in 48 cancer-associated genes.

Results

Targeted sequencing allows a high-level sample multiplexing and deep-coverage profiling of the presumably low fractional amount of colorectal-derived DNA in tongue coating. According to the manufacturer's information, a single targeted sequencing run with 20 pooled samples would give an average sequencing depth of over 1000-fold per sample (Illumina). Hence, a 20-plex targeted NGS was performed in this study.

We obtained average sequencing coverages of 1325x, 1273x and 1873x for the tumor tissues, adjacent normal tissues and tongue coating samples, respectively. For each patient, we identified tumor - associated somatic mutations by sorting for sequence variations that were present in the tumor tissue but absent in the paired adjacent normal tissue (Figure. 1). After that, we determined if these tumor - associated mutations were found in the tongue coating of the respective patient (Figure. 1). As shown

in (Table 1), we identified 2 to 6 tumor - associated mutations in the tongue coating samples. Interesting, these mutations were located in 8 genes, i.e., KIT, CDH1, CTNNB1, GNAS, RET, SMAD4, HNF1A and RB1, which have been reported to be related to CRC (Table 2). A detailed annotation of each variant detected was shown in (Table 3). The number of tumor – associated mutations in the tongue coating samples was correlated to TNM stages of CRC patients (Kruskal-Wallis test, $p < 0.005$, Figure 2).

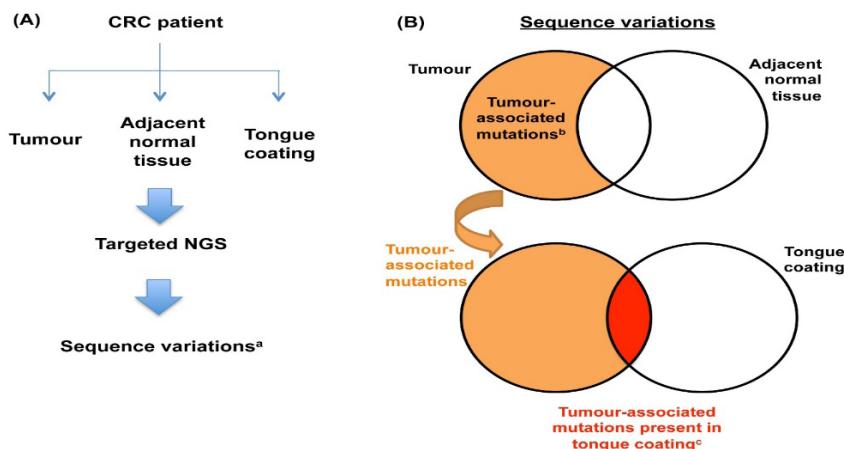


Figure 1: Illustration of project workflow and data analysis.

Tongue coating samples, tumor tissue, adjacent normal tissue were collected for targeted NGS for each CRC patient. Sequence variations, i.e., genomic sites in which alleles alternative to those in the reference human genome (hg19 of UCSC database), were detected.

Table 1. Summary of sequence variations detected by targeted sequencing

CRC patients	Number of sequence variations ^a			Tumor -associated mutations ^b	Tumor - associated mutations present in tongue coating ^c		
	Tumors	Adjacent normal tissues			Before surgery		
						After surgery ^d	
1	428	194	36	234	5	2	
2	128	46	22	82	3	1	
3	105	51	25	54	2	0	
4	135	37	25	98	4	1	
5	289	78	34	211	5	3	
6	65	24	23	41	4	0	
7	71	32	18	39	3	1	
8	74	27	28	47	6	1	
9	554	444	106	110	5	1	
10	512	267	87	245	4	0	
11	426	117	29	309	5	2	
12	146	52	28	94	4	0	
13	157	37	24	120	3	1	
14	234	58	41	176	5	2	

15	315	152	77	163	4	1
16	117	43	13	85	3	0
17	56	18	7	74	2	0
18	319	162	68	157	4	1
19	83	24	10	59	3	1
20	162	69	26	93	3	0
21	267	134	64	133	4	0
22	202	126	53	76	5	2
23	148	54	19	94	2	0

^aIllumina TruSeq Amplicon Cancer Panel, which simultaneously detected somatic mutation hotspots in 48 cancer-associated genes, was utilized to detect the presence of alternative alleles in the targeted genomic regions.

^bSequence variations that were present in the tumors but absent in paired adjacent normal tissues.

^cTumor - associated mutations that were present in tongue coating of the corresponding patients.

^dTongue coating samples collected at day 6 after surgical tumor resection.

Table 2. The genes in which the tongue coating tumor - associated mutations are located and their reported relationships with CRC

V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (KIT)	A proto-oncogene that contributes to L1-mediated metastasis. KIT over-expression was suggested to be a prognostic factor for TNM stage II CRC
Cadherin 1, type 1, E-cadherin, epithelial (CDH1)	A CRC susceptibility locus. Its promoter variation (-347GàGA) may be a prognostic factor of sporadic CRC. Down-regulation of the protein is important for invasion and metastasis of CRC, and a prognostic factor of sporadic CRC.
Catenin (cadherin-associated protein) beta 1 (CTNNB1)	The protein is a key player in the Wnt pathway that is critical in sustaining cancer initiating cells in the colon. The nuclear over-expression of β -catenin is associated with disease progress and worse prognosis in CRC patients
GNAS complex locus (GNAS)	<i>GNAS</i> mutations are not frequently observed but are functionally significant in CRC. <i>GNAS</i> mutations are a characteristic genetic feature of colorectal villous adenoma
Ret proto-oncogene (RET)	<i>RET</i> is a tumor suppressor in CRC as it induces apoptosis and suppress anchorage independent growth in CRC cancer cells. Mutational inactivation and aberrant methylation of <i>RET</i> promote CRC formation and progression
SMAD family member 4 (SMAD4)	The gene is located in human chromosome 18q21, a region with frequent genetic loss of heterozygosity in CRC. <i>SMAD4</i> mutations are found in approximately 10% of sporadic CRC. Loss of SMAD4 protein is strongly correlated with CRC progression and is a prognostic marker in CRC
Retinoblastoma 1 (RB1)	<i>RB1</i> is a tumor suppressor gene and the protein is a negative regulator of cell cycle. Over-expression of <i>RB1</i> mRNA was observed in CRC tissues, and it is a prognostic biomarker in tumorigenesis of sporadic CRC
HNF1 homeobox A (HNF1A)	Defect of the gene causes maturity-onset diabetes of the young. The gene was hypermethylated in colon cancer cell lines

Table 3: Annotation of each variant detected in the tongue coating DNA samples

Gene	Variant	Chromosome	Coordination	Transcript	Consequence	HGVSc	HGVSp
KIT	G>C	4	55602765	NM_000222.2	synonymous variant	NM_000222.2:c.2586G>C	NM_000222.2:c.2586G>C(p.=)
KIT	A>C	4	55593464	NM_000222.2	missense variant	NM_000222.2:c.1621A>C	NP_000213.1:p.Met541Leu
CDH1	G>A	16	68847302	NM_004360.3	synonymous variant	NM_004360.3:c.1224G>A	NM_004360.3:c.1224G>A(p.=)
RET	G>T	10	43615633	NM_020975.4	synonymous variant	NM_020975.4:c.2712C>G	NM_020975.4:c.2712C>G(p.=)
RB1	C>T	13	48919219	NM_000321.2	synonymous variant	NM_000321.2:c.384C>T	NM_000321.2:c.384C>T(p.=)
CTNNB1	A>C	3	41266037	NM_001098210.1	missense variant	NM_001098210.1:c.34A>C	NP_001091680.1:p.Met12Leu
GNAS	T>C	20	57484417	NM_080425.2	missense variant	NM_080425.2:c.2527T>C	NP_536350.2:p.Cys843Arg
SMAD4	A>T	18	48584609	NM_005359.5	missense variant	NM_005359.5:c.782A>T	NP_005350.1:p.His261Leu
HNF1A	C>T	12	121431446	NM_000545.5	missense variant	NM_000545.5:c.650C>T	NP_000536.5:p.Ala217Val

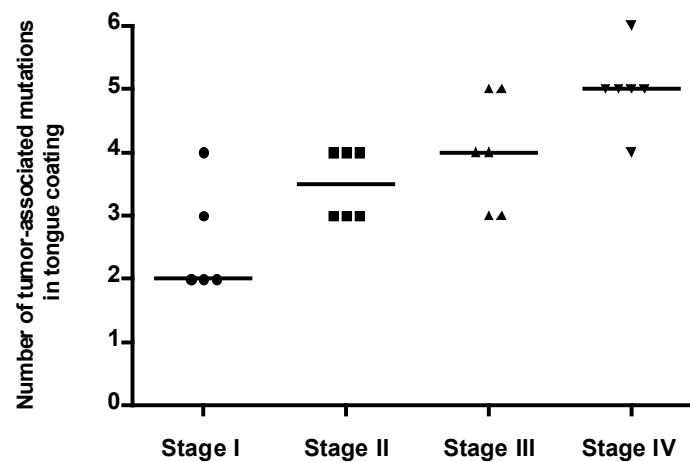


Figure 2: Number of tumor – associated mutations in tongue coating samples from various TNM stages of CRC patients.

To validate the NGS result, we performed direct sequencing for the identified tumor -associated tongue coating mutations in the respective tumors, adjacent normal tissues and tongue coating samples. Consistent genotyping results as determined using NGS and direct sequencing were obtained (Figure 3).

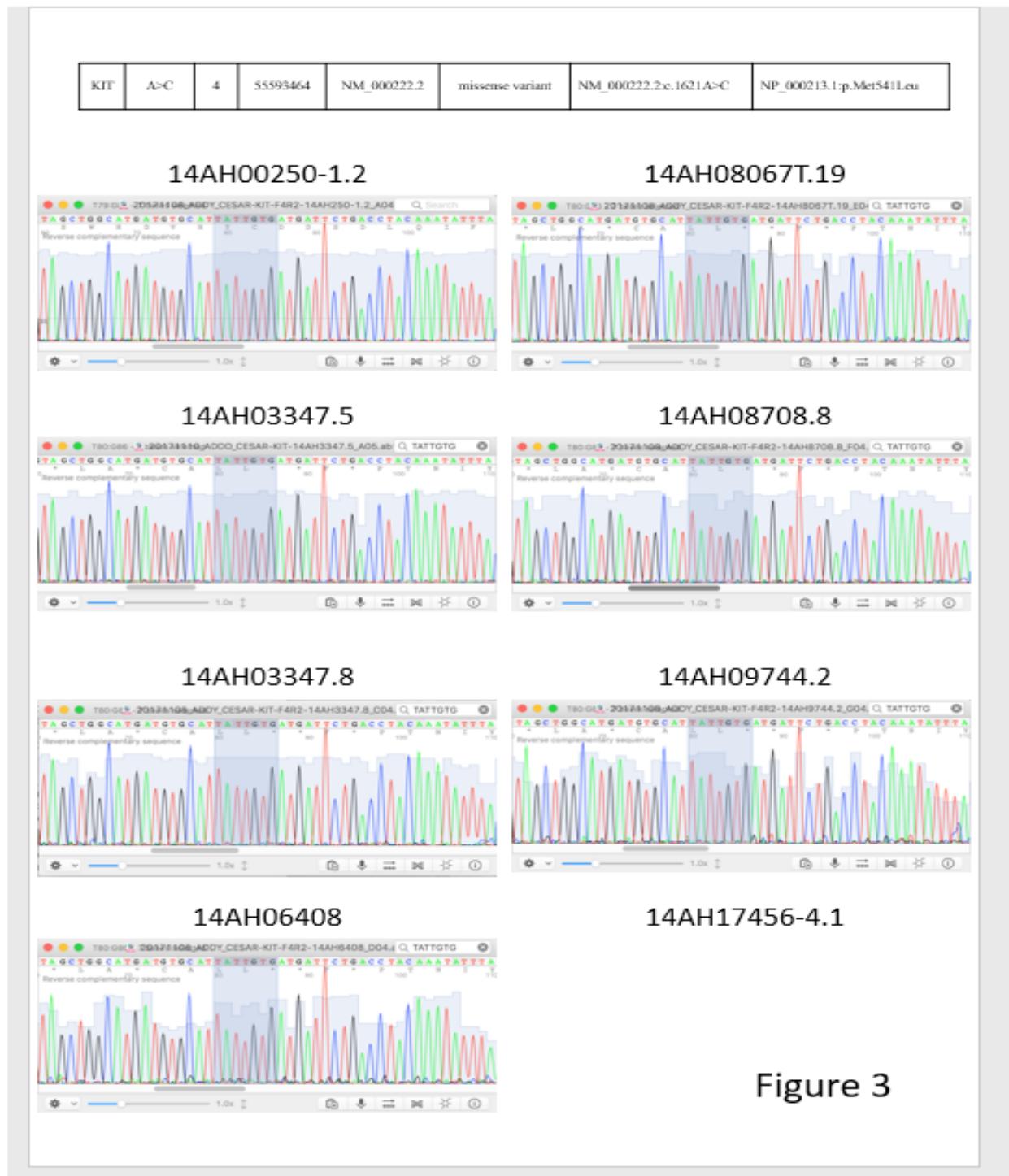
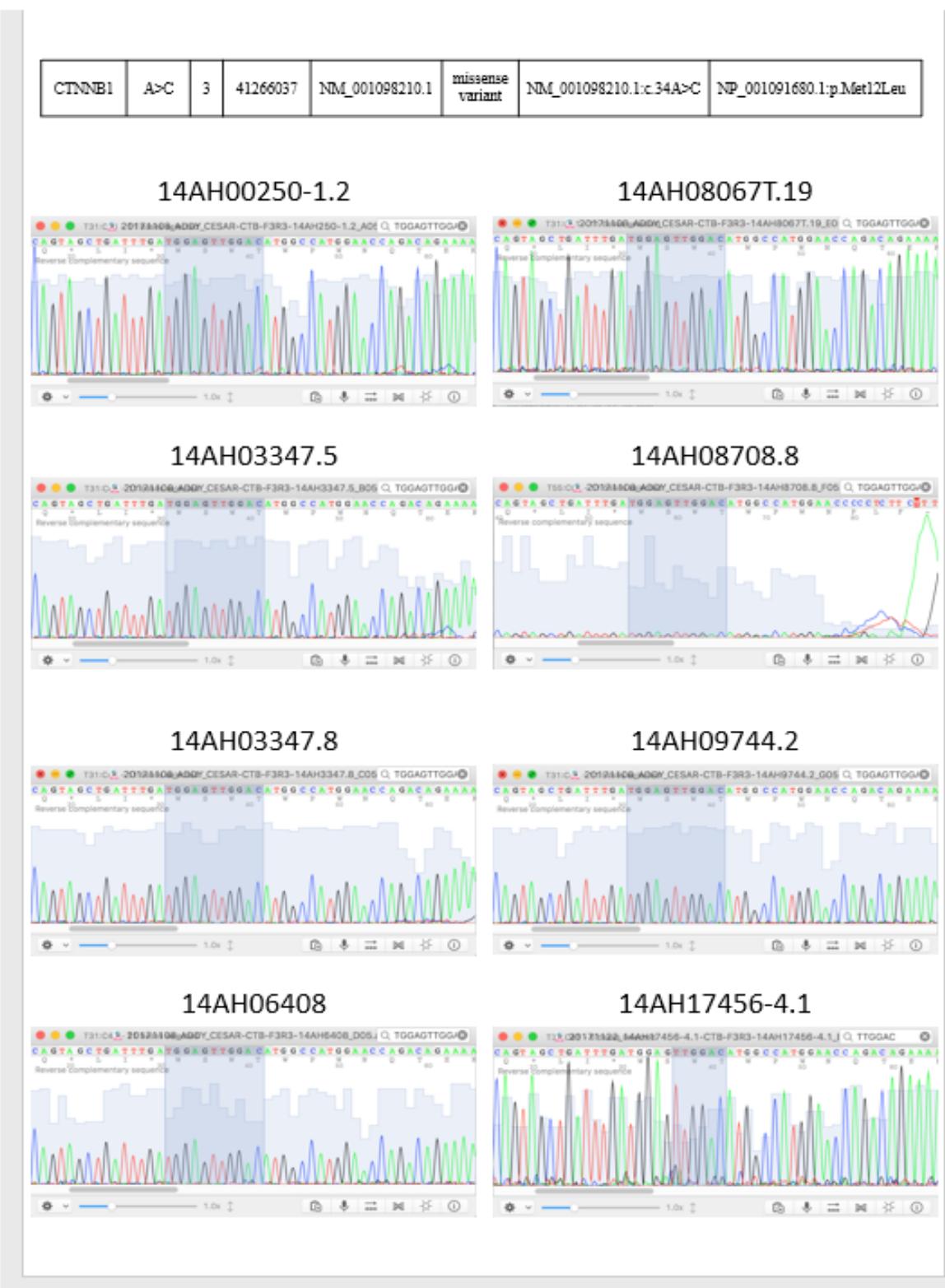
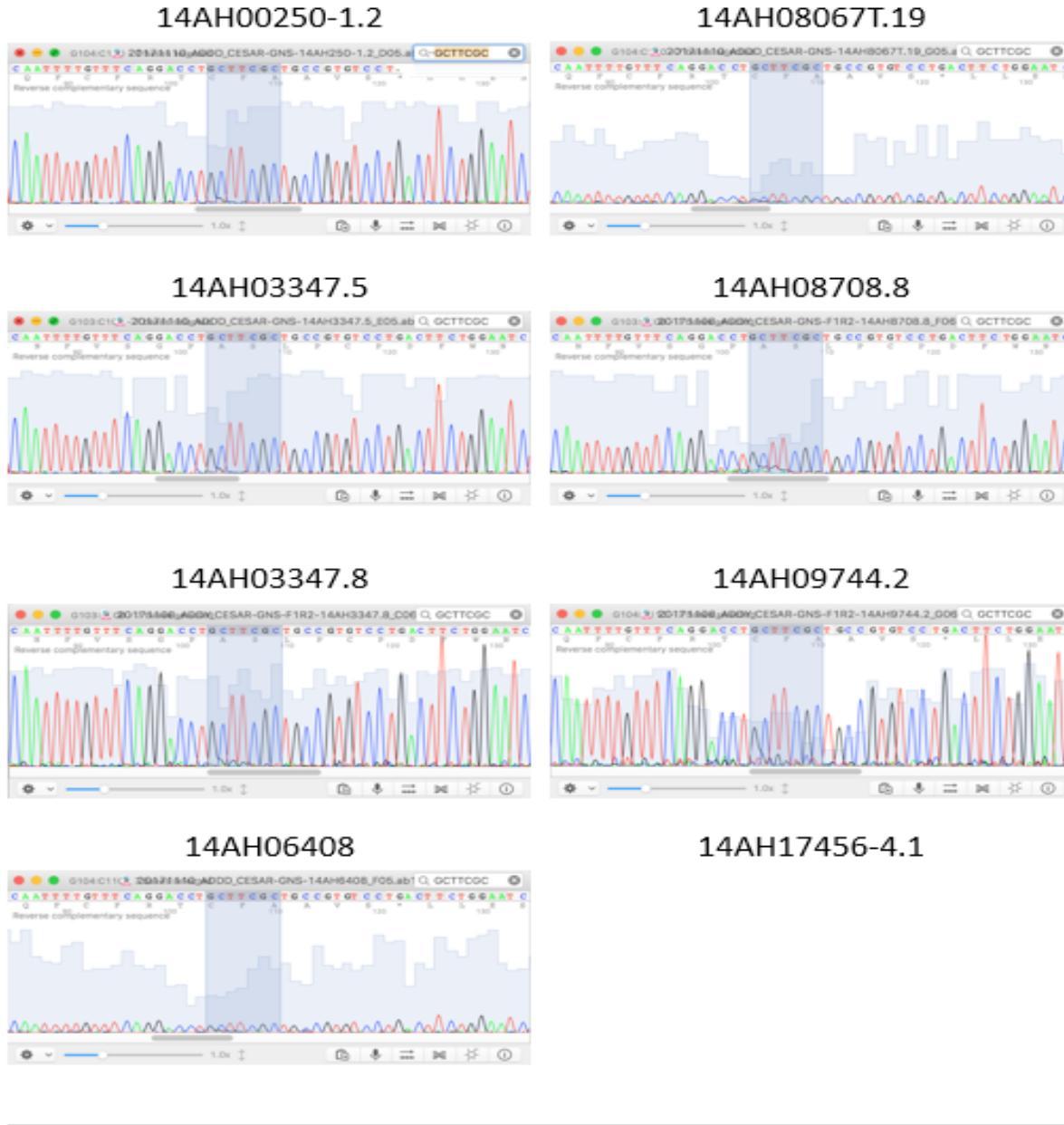


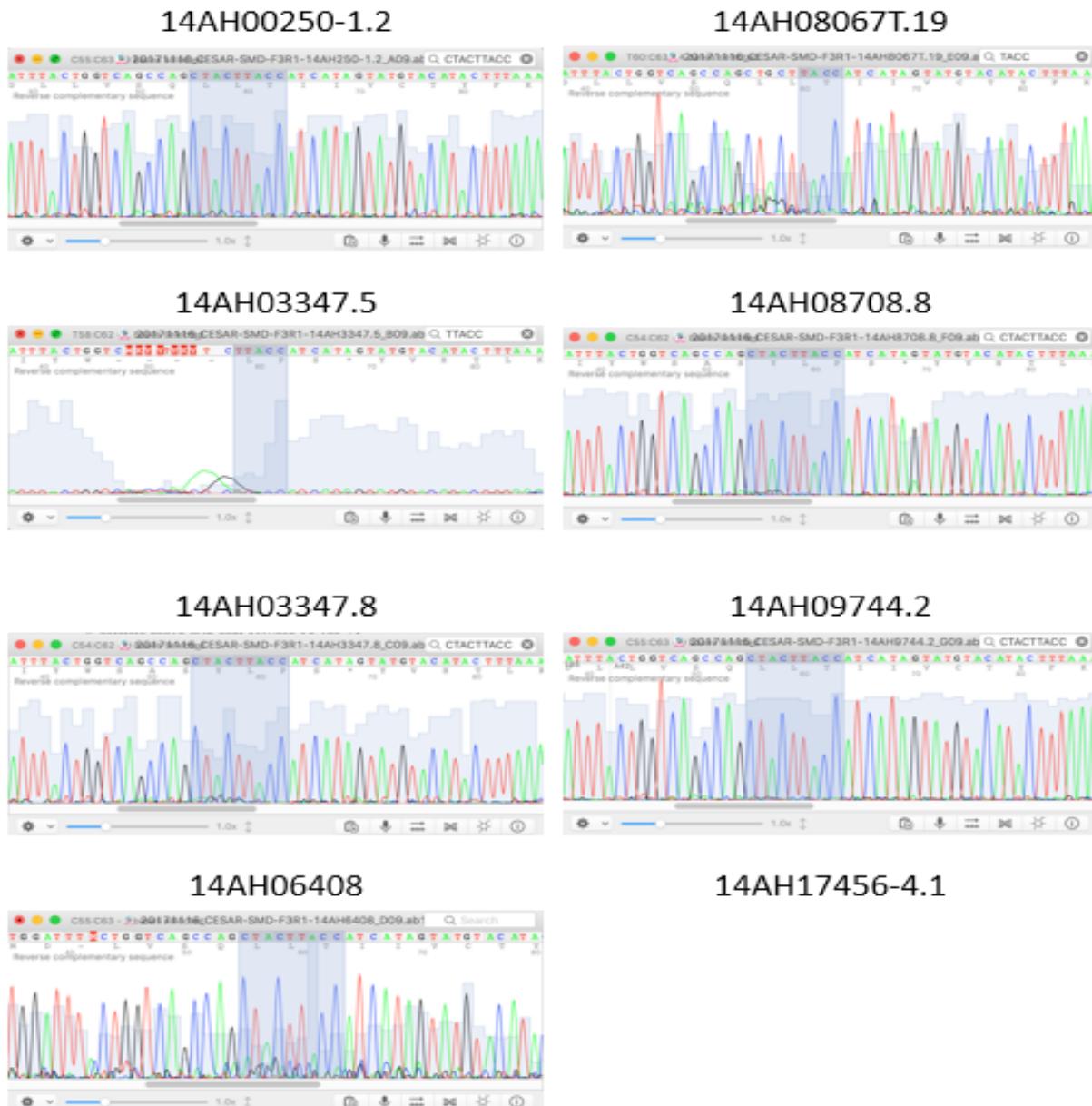
Figure 3



GNAS	T>C	20	57484417	NM_080425.2	missense variant	NM_080425.2:c.2527T>C	NP_536350.2:p.Cys843Arg
------	-----	----	----------	-------------	------------------	-----------------------	-------------------------



SMAD4	A>T	18	48584609	NM_005359.5	missense variant	NM_005359.5:c.782A>T	NP_005350.1:p.His261Leu
-------	-----	----	----------	-------------	------------------	----------------------	-------------------------



HNF1A	C>T	12	121431446	NM_000545.5	missense variant	NM_000545.5:c.650C>T	NP_000536.5:p.Ala217Val
-------	-----	----	-----------	-------------	------------------	----------------------	-------------------------



Figure 3: 2 targeted sequencing results were validated by sanger sequencing.

In order to prove that the identified tumor - associated tongue coating mutations were not originated in the mouth, buccal swab DNA samples from 8 of the recruited CRC patients were collected and were analysed by targeted NGS as described above. We found that for each patient, 80-85% of the sequence variations found in the tongue coating samples were not detected in the corresponding buccal swab samples, confirming that a majority of these identified mutations were indeed tumor - derived.

For the 23 recruited CRC patients, we further collected their tongue coating samples at day 6 after surgical operation. The samples were analysed by targeted NGS as described above. We found that for each patient, 60 – 100% of the identified tumor - derived mutations in tongue coating disappeared, i.e., with no mutant alleles observed, after surgical operation (Table 4). The data hence further demonstrated the tumor - specificity of the identified mutations in the tongue coating samples.

Table 4. Tumor - associated somatic mutations in the tongue coating DNA samples

CRC patients	Tumor - associated mutations in tongue coating								
	<i>KIT</i> ^a chr4: 55593464 A>C	<i>KIT</i> chr4: 55602765 G>C	<i>CDH1</i> chr16: 68847302 G>A	<i>CTNNB1</i> chr3: 41266037 A>C	<i>GNAS</i> chr20: 57484417 T>C	<i>RET</i> chr10: 43615633 G>T	<i>SMAD4</i> chr18: 48584609 A>T	<i>RBI</i> chr13: 48919219 C>T	<i>HNF1A</i> chr12: 121431446 C>T
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
Mutation rates (%)^b	47.8%	43.5%	43.4%	56.5%	47.8%	30.4%	60.9%	34.8%	30.4%

Squares in red: mutations that present in tongue coating before surgical operations, and cleared from tongue coating after operations.

Squares in blue: mutations that remained in tongue coating after surgeries.

^aHg19 genome coordinates of the mutations and the cancer - associated genes that the mutations located.

^b(number of patients in whom the mutation was observed) / (total number of patients)

Tumor – associated mutations in yellow: synonymous variants with no change in protein sequences.

From the NCBI database, KIT chr4: 55602765 G>C, CDH1 chr16: 68847302 G>A, RET chr10: 43615633 G>T and RB1 chr13: 48919219 C>T are synonymous variants (Table 4). Therefore, their protein sequences have not been changed and we shall not discuss those results.

Discussion

In traditional Chinese medicine (TCM), the tongue is considered to link to a number of critical internal organs, and the appearance of tongue coating is one of the most important indicators in TCM diagnosis. The biological basis of diagnosis using tongue coating, however, is rarely investigated. Thus far, there are only a few studies on tongue coating which include the microbiome and glucose metabolites in gastritis patients, and keratin expression in hepatitis B patients [18-21]. Recently, a study by Han et al. has demonstrated the first time that tongue diagnosis using the images of tongue and tongue coating analysis may provide potential screening and early diagnosis strategies for CRC, gastric cancer and lung cancer [22].

In this study, we hypothesise that tongue coating may contain DNA originated from other internal organs of the body. Our hypothesis is based on the fact that the tongue and the large intestine are physically linked in the long tube of digestive system. Also, the tongue is richly supplied with blood and lymphatic vessels, which may provide another route of passage of non-oral-derived DNA. Therefore, it is logical for us to search for tumor - derived DNA in the tongue coating of CRC patients. This study involved the profiling of over 212 cancer-related somatic mutations in tongue coating and tumor tissues of CRC patients by targeted sequencing. A subset of mutation markers that are prevalent in the local CRC population are identified, which may be useful for the future development of an inexpensive non - invasive screening test for CRC using tongue coating.

In this preliminary study, we have been able to detect tumor - associated somatic mutations (Range: 2 to 6), i.e. mutations that were present in tumor tissues but absent in adjacent normal tissues, in the tongue coating of the same CRC patient. Moreover, those somatic mutations were reduced after surgery for 6 days (Table 1). On the other hand, those mutations were found from 30.4% to 60.9% in the tongue coating of CRC patients (Table 4).

The description of each tumor -associated somatic mutation is shown as follows:

KIT Gene (chr4:55593464, A>C, missense mutations M537L or M541L). This gene encodes the human homolog of the proto-oncogene c-kit [23]. C-kit was first identified as the cellular homolog of the feline sarcoma viral oncogene v-kit [23]. This protein is a type 3 transmembrane receptor for MGF (mast cell growth factor, also known as stem cell factor) [23]. Mutations in this gene are associated with gastrointestinal stromal tumors, mast cell disease, acute myelogenous leukemia, and piebaldism [24]. Multiple transcript variants encoding different isoforms have been found for this gene [24].

According to the ClinVar, NCBI database, the clinical significance of this mutation is likely to be benign. CTNNB1 (chr3:41266037, A>C, missense mutation M12L)

The protein encoded by this gene is part of a complex of proteins that constitute adherens junctions (AJs) [25]. AJs are necessary for the creation and maintenance of epithelial cell layers by regulating cell growth and adhesion between cells [25]. The encoded protein also anchors the actin cytoskeleton and may be responsible for transmitting the contact inhibition signal that causes cells to stop dividing once the epithelial sheet is complete [25]. Finally, this protein binds to the product of the APC gene, which is mutated in adenomatous polyposis of the colon [26]. Mutations in this gene are a cause of colorectal cancer (CRC), pilomatricoma (PTR), medulloblastoma (MDB), and ovarian cancer [26]. Alternative splicing results in multiple transcript variants [26].

There has been no report in the ClinVar, NCBI database on the position of this missense mutation. GNAS (chr20:57484417, T>C, missense mutation C185R, C186R, C200R, C201R)

This locus has a highly complex imprinted expression pattern [27]. It gives rise to maternally, paternally, and biallelically expressed transcripts that are derived from four alternative promoters and 5' exons [27]. Some transcripts contain a differentially methylated region (DMR) at their 5' exons, and this DMR is commonly found in imprinted genes and correlates with transcript expression [27]. An antisense transcript is produced from an overlapping locus on the opposite strand [27]. One of the transcripts produced from this locus, and the antisense transcript, are paternally expressed noncoding RNAs, and may regulate imprinting in this region [27]. In addition, one of the transcripts contains a second overlapping ORF, which encodes a structurally unrelated protein – Alex [28]. Alternative splicing of downstream exons is also observed, which results in different forms of the stimulatory G-protein alpha subunit, a key element of the classical signal transduction pathway linking receptor-ligand interactions with the activation of adenylyl cyclase and a variety of cellular responses [28]. Multiple transcript variants encoding different isoforms have been found for this gene [28]. Mutations in this gene result in pseudohypoparathyroidism type 1a, pseudohypoparathyroidism type 1b, Albright hereditary osteodystrophy, pseudopseudohypoparathyroidism, McCune-Albright syndrome, progressive osseus heteroplasia, polyostotic fibrous dysplasia of bone, and some pituitary tumors [28].

There has been no report in the ClinVar, NCBI database on the position of this missense mutation. SMAD4 (chr18:48584609, A>T, missense mutation H261L)

This gene encodes a member of the Smad family of signal transduction proteins [29]. Smad proteins are phosphorylated and activated by transmembrane serine-threonine receptor kinases in response to TGF-beta signaling [29]. The product of this gene forms homomeric complexes and heteromeric complexes with other activated Smad proteins, which then accumulate in the nucleus and regulate the transcription of target genes [29].

This protein binds to DNA and recognizes an 8-bp palindromic sequence (GTCTAGAC) called the Smad-binding element (SBE) [30]. The Smad proteins are subject to complex regulation by post-translational modifications [30]. Mutations or deletions in this gene have been shown to result in pancreatic cancer, juvenile polyposis syndrome, and hereditary hemorrhagic telangiectasia syndrome [30].

There has been no report in the ClinVar, NCBI database on the position of this missense mutation. HNF1A (chr12:121431446, C>T, missense mutation A217V). The protein encoded by this gene is a transcription factor required for the expression of several liver-specific genes [31]. The encoded protein functions as a homodimer and binds to the inverted palindrome 5'-GTTAATNATTAAC-3' [31]. Defects in this gene are a cause of maturity onset diabetes of the young type 3 (MODY3) and also can result in the appearance of hepatic adenomas [32]. Alternative splicing results in multiple transcript variants encoding different isoforms [32].

There has been no report in the ClinVar, NCBI database on the position of this missense mutation. With these preliminary results, a long term follow up of those 5 mutations in that cohort of CRC patients is being performed. Moreover, a large scale study with more CRC patients is now being arranged. Finally, the functional significance of those 5 mutations will be examined in various CRC cell lines.

This study represents the first demonstration of the presence of colorectal - derived DNA mutations in tongue coating using targeted sequencing. The findings can contribute to the development of a new and non-invasive molecular test for CRC detection, monitoring and treatment. The ultimate outcome is to develop a non-invasive test for screening and close monitoring of CRC patients. The patients will benefit from early detection and treatment of the disease, as well as timely application of adjuvant chemotherapy after surgical operation. Besides CRC, this study has paved the way for the future study of tongue coating DNA relevant to other internal organs.

Materials and Methods

Patients

Twenty-three CRC patients (5 TNM stage I, 6TNM stage II , 6 TNM stage III and 6 TNM stage IV) were recruited and their tumor tissues, adjacent normal tissues and tongue coating samples were collected. The study was approved by the Clinical Research Ethics Committee of the Queen Elizabeth Hospital, Kowloon Central Cluster, Hospital Authority, Hong Kong Special Administrative Region.

DNA extraction

Genomic DNA from the tongue coating was extracted using the QIAamp fast DNA tissue kit (Category no: 51404, Qiagen, Hilden, Germany) according to the manufacturer's instructions. Besides, genomic DNA from the tumor tissues and adjacent normal tissues was extracted using the QIAamp DNA

formalin fixed paraffin embedded (FFPE) Tissue Kit (Category no: 56404, Qiagen) according to the manufacturer's instructions with modifications. Briefly, 15 FFPE sections with 5µm thick were deparaffinized with xylene, followed by overnight proteinase K digestion. The mixture was then loaded into the extraction column. DNA was eluted in 20µl of water, and quantified by the QuantiFluorTM dsDNA system (Promega, Madison, USA).

Targeted NGS using the TruSeq Amplicon-Cancer Panel

The TruSeq Amplicon – Cancer Panel (Illumina) is a highly multiplexed targeted resequencing assay for detecting somatic mutations within important cancer-related genes, including BRAF, KRAS, and EGFR. This panel provides predesigned, optimized oligonucleotide probes for sequencing mutational hotspots in > 35 kilobases (kb) of target genomic sequence. Forty-eight genes are targeted with 212 amplicons in a highly multiplexed, single-tube reaction. This highly targeted approach enables a wide range of applications for discovering, validating, and screening genetic variants in a rapid and efficient manner. The sequencing library was prepared from extracted DNA using the TruSeq Amplicon Cancer Panel Library Preparation Kit (Illumina) according to the manufacturer's protocol. The library of each sample was ligated with different indexed sequence. Libraries from 24 samples were pooled and sequenced on the MiSeq Sequencer (Illumina) using the 2x150 cycle protocol. The sequenced data was processed by the MiSeq Reporter software (Illumina). Sequence variations were detected by the Somatic Variant Caller (Illumina).

Direct sequencing

The targeted NGS results were validated by direct sequencing with the use of the BigDye Terminator v3.1 Cycle sequencing Kit (ThermoFisher Scientific, Waltham, USA) according to the manufacturer's protocol.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgements

The study was supported by the Hong Kong Polytechnic University Central Research Grant PolyU 151037/14M, Research Grants Council HK, HK Innovation and Technology Fund University-Industry Collaborative Programme (grants number UIM/354 and RGCQ71P) and Lim peng Suen Charity foundation (R-ZH5G).

References

1. Sung JJ, Ng SC, Chan FKL, Chiu HM, Kim HS, et al; Asia Pacific Working Group. (2015) An updated Asia Pacific consensus recommendations on colorectal cancer screening. *Gut*.64:121-32.
2. Traverso G, Shuber A, Olsson L, Levin B, Johnson C, et al. (2002) Detection of proximal colorectal cancers through analysis of faecal DNA. *Lancet*.359:403-4.
3. Schoen RE. (2002) The case for population-based screening for colorectal cancer. *Nat Rev Cancer*.2:65-70.

4. Wong SC, Lo SFE, Cheung MT, Ng KOE, Tse CW, et al. (2004) Quantification of plasma beta-catenin mRNA in colorectal cancer and adenoma patients. *Clin Cancer Res.*10:1613-7.
5. Moertel CG, Fleming TR, Macdonald JS, Haller DG, Laurie JA, et al. (1993) An evaluation of the carcinoembryonic antigen (CEA) test for monitoring patients with resected colon cancer. *JAMA.*270:943-7.
6. Allegra CJ, Paik S, Colangelo LH, Parr AL, Kirsch I, et al. (2003) Prognostic value of thymidylate synthase, Ki-67, and p53 in patients with Dukes' B and C colon cancer: a National Cancer Institute-National Surgical Adjuvant Breast and Bowel Project collaborative study. *J Clin Oncol.*21:241-50.
7. Cascinu S, Staccioli MP, Gasparini G, Giordani P, Catalano V, et al. (2000) Expression of vascular endothelial growth factor can predict event-free survival in stage II colon cancer. *Clin Cancer Res.*6:2803-7.
8. Watanabe T, Wu TT, Catalano PJ, Ueki T, Satriano R, et al. (2001) Molecular predictors of survival after adjuvant chemotherapy for colon cancer. *N Engl J Med.*344:1196-206.
9. Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, et al. (2003) Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *N Engl J Med.*349:247-57.
10. Hong Kong Cancer Registry: Hong Kong Cancer Stat 2004. Hospital Authority of Hong Kong Special Administrative Region. Available at:<http://www3.ha.org.hk/cancereg/eng/canstat2004.pdf> (accessed on August 15, 2008).
11. Rodríguez-Moranta F, Saló J, Arcusa A, Boadas J, Piñol V, et al. (2006) Postoperative surveillance in patients with colorectal cancer who have undergone curative resection: a prospective, multicenter, randomized, controlled trial. *J Clin Oncol.*24:386-93.
12. Chau I, Cunningham D. (2006) Adjuvant therapy in colon cancer—what, when and how? *Annals of Oncology.*19:1347-1359.
13. Gill S, Loprinzi CL, Sargent DJ, Thomé SD, Alberts SR, et al. (2004) Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? *J Clin Oncol.*22:1797-806.
14. André T, Boni C, Mounedji-Boudiaf L, Navarro M, Tabernero J, et al. (2004) Multicenter International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC) Investigators. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med.*350:2343-51.
15. Han SW, Kim HP, Shin JY, Jeong EG, Lee WC, et al. (2013) Targeted sequencing of cancer-related genes in colorectal cancer using next-generation sequencing. *PLoS One.*8:e64271.
16. Shao J, Lou X, Wang J, Zhang J, Chen C, et al. (2013) Targeted re-sequencing identified rs3106189 at the 5'UTR of TAPBP and rs1052918 at the 3'UTR of TCF3 to be associated with the overall survival of colorectal cancer patients. *PLoS One.*8:e70307.
17. Casey G, Conti D, Haile R, Duggan D. (2013) Next generation sequencing and a new era of medicine. *Gut.*62:920-32.
18. Sun ZM, Zhao J, Qian P, Wang YQ, Zhang WF, et al. (2013) Metabolic markers and microecological characteristics of tongue coating in patients with chronic gastritis. *BMC Complement Altern Med.*13:227.
19. Fang F, Liu P, Wang H, Zhang L, Zhang J, et al. (2009) Studies of keratins in tongue coating samples of hepatitis B patients by mass spectrometry. *Rapid Commun Mass Spectrom.*23:1703-9.
20. Jiang B, Liang X, Chen Y, Ma T, Liu L, et al. (2012) Integrating next-generation sequencing and traditional tongue diagnosis to determine tongue coating microbiome. *Scientific Reports.*2:936.
21. Ye J, Cai X, Yang J, Sun X, Hu C, et al. (2016) Bacillus as a potential diagnostic marker for yellow tongue coating. *Sci Rep.*6:32496.
22. Han S, Yang X, Qi Q, Pan Y, Chen Y, et al. (2016) Potential screening and early diagnosis method for cancer: Tongue diagnosis. *Int J Oncol.*48:2257-64.
23. Hirota S, Isozaki K, Moriyama Y, Hashimoto K, Nishida T, et al. (1998) Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science.*279:577-80.
24. Yan L, Zou L, Zhao W, Wang Y, Liu B, et al. (2015) Clinicopathological significance of c-KIT mutation in gastrointestinal stromal tumors: a systematic review and meta-analysis. *Sci Rep.*5:13718.
25. Segditsas S, Tomlinson I. (2006) Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene.*25:7531-7.
26. Clevers H. (2004) Wnt breakers in colon cancer. *Cancer Cell.*5:5-6.
27. Mantovani G, Lania AG, Spada A. (2010) GNAS imprinting and pituitary tumors. *Mol Cell Endocrinol.*326:15-8.
28. Lania AG, Mantovani G, Spada A. (2006) Mechanisms of disease: Mutations of G proteins and G-protein-coupled receptors in endocrine diseases. *Nat Clin Pract Endocrinol Metab.*2:681-93.
29. Miyaki M, Kuroki T. (2003) Role of Smad4 (DPC4) inactivation in human cancer. *Biochem Biophys Res Commun.*306:799-804.
30. Wu M, Chen G, Li YP. (2016) TGF-β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res.*4:16009.
31. Maestro MA, Cardalda C, Boj SF, Luco RF, Servitja JM, et al. (2007) Distinct roles of HNF1beta, HNF1 alpha, and HNF4alpha in regulating pancreas development, beta-cell function and growth. *Endocr Dev.*12:33-45.
32. Ryffel GU. (2001) Mutations in the human genes encoding the transcription factors of the hepatocyte nuclear factor (HNF)1 and HNF4 families: functional and pathological consequences. *J Mol Endocrinol.*27:11-29.