

# Mutation profiles of synchronous colorectal cancers from a patient with Lynch syndrome suggest distinct oncogenic pathways

Scott R. Wheeler<sup>1</sup>, Chanjuan Shi<sup>1</sup>, Jonathan A. Holt<sup>2</sup>, Cindy L. Vnencak-Jones<sup>1</sup>

<sup>1</sup>Department of Pathology, Microbiology and Immunology, <sup>2</sup>Department of Bioinformatics, Vanderbilt University Medical Center, Nashville, TN 37211, USA

Correspondence to: Scott R. Wheeler, PhD. PathGroup, 601 Grassmere Park Suite 5, Nashville, TN 37211, USA. Email: swheeler@pathgroup.com.

**Abstract:** Patients with Lynch syndrome often present with multiple synchronous or metachronous colorectal cancers (CRCs). The presence of multiple CRCs with distinct genetic profiles and driver mutations could complicate treatment as each cancer may respond differently to therapy. Studies of sporadic CRCs suggested that synchronous tumors have distinct etiologies, but could not rule out differences in genetic background. The presence of multiple cancers in a patient with a predisposing mutation provides an opportunity to profile synchronous cancers in the same genetic background. Here, we describe the case of a patient with Lynch syndrome that presented with six synchronous CRCs. Microsatellite instability (MSI) and genomic profiling indicated that each lesion had a unique pattern of instability and a distinct profile of affected genes. These findings support the idea that in Lynch syndrome, synchronous CRCs can develop in parallel with distinct mutation profiles and that these differences may inform treatment decisions.

**Keywords:** Colorectal neoplasms; hereditary nonpolyposis; medical oncology; neoplasms; multiple primary; molecular targeted therapy

Submitted Nov 23, 2015. Accepted for publication Dec 18, 2015.

doi: 10.21037/jgo.2016.01.07

View this article at: <http://dx.doi.org/10.21037/jgo.2016.01.07>

## Introduction

Due to the high risk of synchronous and metachronous colorectal cancers (CRCs), the recommended treatment for CRC in patients with Lynch syndrome is colectomy (1). However, as the use of targeted therapies increases, including VEGF and EGFR inhibitors, an understanding of the molecular profiles of synchronous CRCs will be crucial in determining the most effective courses of treatment. If synchronous cancers are clonally derived, treatment may be relatively uniform. However, if they are distinct, each with a different oncogenic history, a treatment plan may need to include additional agents or therapeutic regimens.

In unselected patient populations with synchronous or metachronous CRCs, it is thought that different cancerous lesions in the same patient may follow distinct pathways of progression (2-4). Studies of matched synchronous CRCs suggested that exposure to different toxins at various locations within the colon were associated with distinct progression profiles (5,6). However, the effects of genetic

background and environmental variation are confounding aspects in studying initiation and progression in sporadic cancers from different patients. Patients with Lynch syndrome frequently present with multiple synchronous or metachronous cancers (1,7,8). The presence of synchronous cancers arising in a single patient with a predisposing genetic mutation presents an opportunity to study the oncogenic pathways that lead to CRC while minimizing the effects of genetic background. In this study, we examined microsatellite instability (MSI) and genomic profiling in synchronous CRCs from a Lynch syndrome patient.

## Materials and Methods

### *Pathology review and DNA extraction*

Specimens were formaldehyde fixed, paraffin embedded and hematoxylin and eosin (H&E) stained. Representative sections were reviewed by a pathologist (CS) for histologic subtype, grade, staging and percent tumor burden. Sections

were macrodissected to increase the representation of tumor DNA in the total DNA extracted from the specimen. DNA was isolated following cell lysis and proteinase K treatment using the QiaQuick extraction method (Qiagen, Hilden Germany).

### *Analysis of microsatellite instability (MSI)*

DNA from normal and tumor specimens was amplified using the MSI analysis system version 1.2 (Promega, Madison WI, USA) according to the manufacturer's instructions. Amplicons were detected using capillary electrophoresis on an ABI 3130xl Genetic Analyzer (Life Technologies, Carlsbad CA, USA) and the results were analyzed using GeneMapper V3.7 software (Life Technologies, Carlsbad CA, USA). The presence of instability in two, or more, of the five loci (>30%) was considered MSI-high (MSI-H).

### *SNaPshot mutation profiling*

A SNaPshot single base extension assay was used to assess the mutation status of 62 loci in 7 genes (*AKT1*, *BRAF*, *KRAS*, *NRAS*, *PIK3CA*, *PTEN* and *SMAD4*) associated with CRC prognosis and treatment (9). SNaPshot products were separated using an ABI 3130xl Genetic Analyzer (Life Technologies, Carlsbad CA, USA) and compared to positive and normal controls for interpretation.

### *Massively parallel DNA sequencing*

Multiplex amplicon-based sequencing libraries were prepared using the GeneRead DNA-seq Human Comprehensive Cancer Panel NGHS-501X (Qiagen, Hilden Germany) following the manufacturer's instructions. This panel targets coding and UTR regions of 124 commonly mutated genes in multiple cancer types. Once prepared, libraries were sequenced using a MiSeq 300 cycle V2 reagent kit (Illumina, San Diego CA, USA) with MiSeq Control software V2.3.0.3 and RTA software V1.18.42.0. Variant analysis was performed using the CLCbio genomics workbench (Qiagen, Hilden Germany).

## **Results**

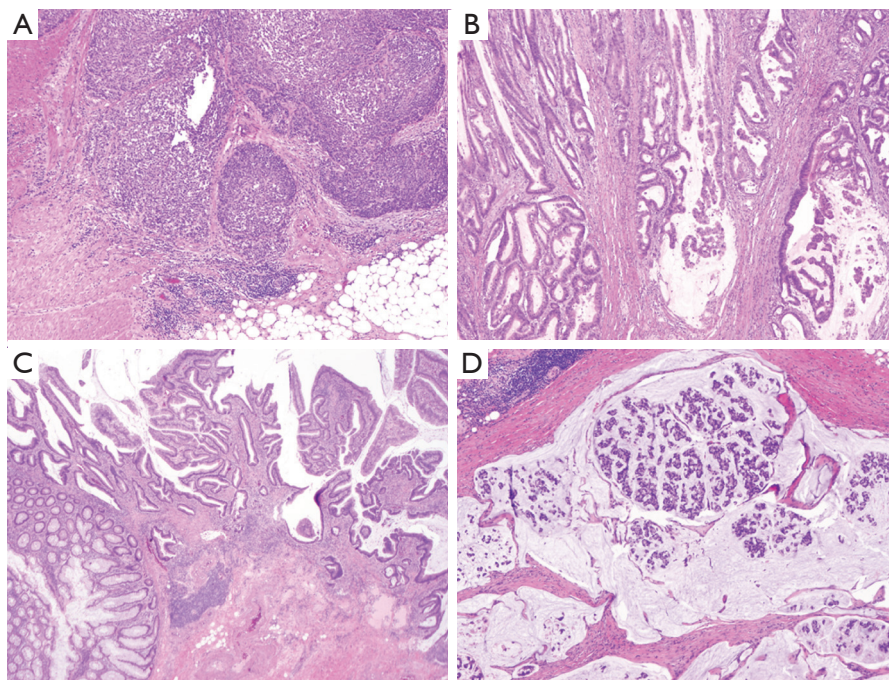
An 81-year-old Egyptian male presented with weight loss, upper quadrant abdominal and rectal pain and blood streaked stools. The medical history was significant for

cancer including a tumor of unreported origin removed by a partial small bowel resection in his 50's, renal cell carcinoma removed by nephrectomy in his 60's and prostate cancer treated with implantation of radioactive seeds in his 70's. The family history was also significant for cancer. Two first-degree relatives had colon cancer, 2 first-degree relatives had kidney cancer and a first-degree relative had an unspecified lymph node/head and neck cancer.

After total proctocolectomy, 6 lesions were identified. These included 5 colonic and 1 rectal lesion. Lesion 1 (medullary carcinoma, 5 cm in greatest dimension, T4N0M0) was located in the left colon near to the splenic flexure (*Figure 1A*). Lesion 2 (moderately differentiated invasive adenocarcinoma, 6 cm, T3N0M0) was located in the right colon near to the hepatic flexure (*Figure 1B*). The colon also contained 3 early invasive carcinomas (T1N0M0) arising from tubulovillous adenomas that, from proximal to distal, were 1cm (lesion 3, *Figure 1C*), 1.3 cm (lesion 4) and 1.2 cm (lesion 5). Lesion 6 was identified in the rectum (invasive mucinous adenocarcinoma with clusters of signet ring cells, 6.5 cm, 8 of 15 lymph nodes involved, T3N2bM0) (*Figure 1D*).

MSI was detected in all lesions. However, the character of MSI was different in each (*Figure 2*). Differences included the number of unstable loci and the pattern and extent of instability. Each lesion was also screened for a panel of 62 hot-spot variants in 7 genes related to prognosis and treatment in colon cancer (9). *BRAF* V600E was not detected in any of the lesions (data not shown) suggesting Lynch syndrome rather than sporadic CRC (10). Other identified variants included *KRAS* G12D in lesion 1 and *KRAS* G13D in lesion 6 (data not shown). To identify potential Lynch syndrome-related germline variants, *MLH1*, *MSH2*, *MSH6* and *PMS2* were examined in DNA isolated from peripheral blood. A frameshift variant was detected in *MSH2*, c.2082delT (p.Phe694Leufs\*16). This variant was previously reported in patients with CRC and established a diagnosis of Lynch syndrome in this patient (11). Taken together, these findings suggest that each lesion arose from a unique event in a background predisposing to cancer.

To understand better how these cancers developed, sequencing of 124 cancer-related genes was performed. Only lesions 1, 2 and 3 had an adequate amount of high quality DNA for this analysis. The majority of variants (48.7%, 186/382) including substitutions, insertions and deletions were identified in all 3 lesions. In the absence of normal DNA to use for comparison, the common variants are a good estimate of the patient's germline



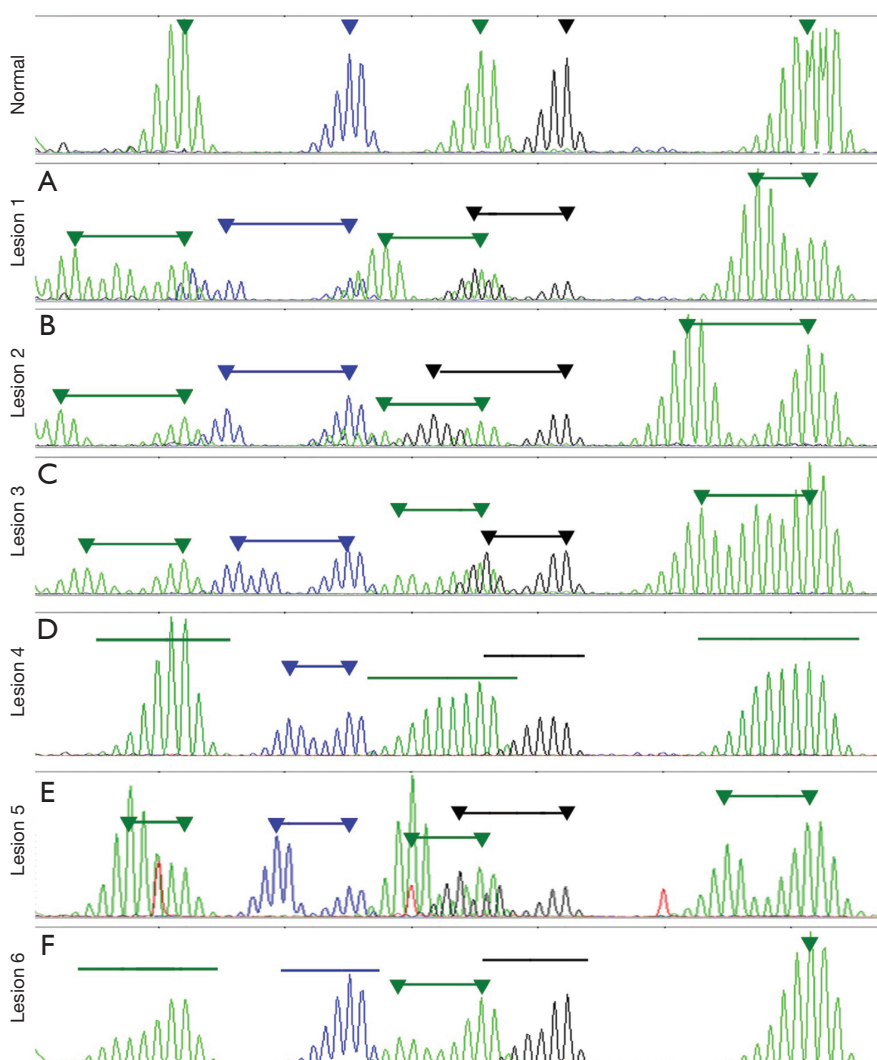
**Figure 1** The morphology of synchronous CRCs (H&E, 40 $\times$ ). (A) Lesion 1: left-sided colonic lesion showing poorly differentiated medullary carcinoma (T4N0M0); (B) lesion 2: right-sided colonic lesion showing moderately differentiated invasive adenocarcinoma (T3N0M0); (C) lesion 3: a representative section of one of 3 tubulovillous adenoma with early invasion in the submucosa (T1N0M0); (D) lesion 6 is a rectal lesion consisting of invasive mucinous adenocarcinoma with clusters of signet ring cells (T3N2bM0). CRCs, colorectal cancer.

variation. However, some of these variants may also be hot-spot mutations that occurred independently in each lesion. Germline variants have an expected frequency of approximately 50% (heterozygous) or 100% (homozygous). Of the presumed germline variants identified in the patient's CRCs, 89.2% have a variant frequency that is indicative of a heterozygous (58.6% with a variant frequency of 40–60%) or homozygous (30.6% with a variant frequency of 90–100%) state. Presumed germline variants were found in 61 genes, all with low or moderate impact predictions (Tables S1,S2). The single high impact variant observed in all lesions was the previously identified c.2082delT in the *MSH2* gene.

To determine the lesion-specific variants, the presumed germline variants were subtracted from each lesion. After subtraction, the percent of variants in each lesion that met the criteria for homozygous or heterozygous state was drastically reduced: 8.1% in lesion 1, 11.8% in lesion 2 and 2.2% in lesion 3. These values are significantly different from the values in the presumed germline variation suggesting that the majority of variants ascribed to the lesions are the result of somatic mutation. Respectively,

lesions 1, 2 and 3 had 96, 74 and 46 somatic variants with similar distributions of types (Table 1). To compare the mutational landscape between lesions, similarities and differences in affected genes were examined (Figure 3, Table 2). All 3 lesions had somatic variants in two genes, *SMARCA4* and *ALK*. For *SMARCA4*, each lesion contained distinct high impact variants: lesion 1 had a non-synonymous coding variant, lesion 2 had a frameshift and lesion 3 had two distinct non-synonymous coding variants. In *ALK*, lesion 2 contained a C-terminal deletion while lesions 1 and 3 contained the same N-terminal deletion. These data suggest an important role for loss of function of these genes in the development of these lesions. Additionally, as expected, all 3 lesions had different combinations of variants in WNT pathway genes including *APC*, *ARID1A*, *CTNNB1*, and *FBXW7*.

There were also significant differences between the lesions. Thirteen genes had variants in 2 lesions and 34 genes had variants in only a single lesion indicating significant differences between lesions. Lesion 1 had variants in 18 unique genes involved in Notch, RAS, and NF2 signaling (Table 2). In addition, a variant in *MSH2*

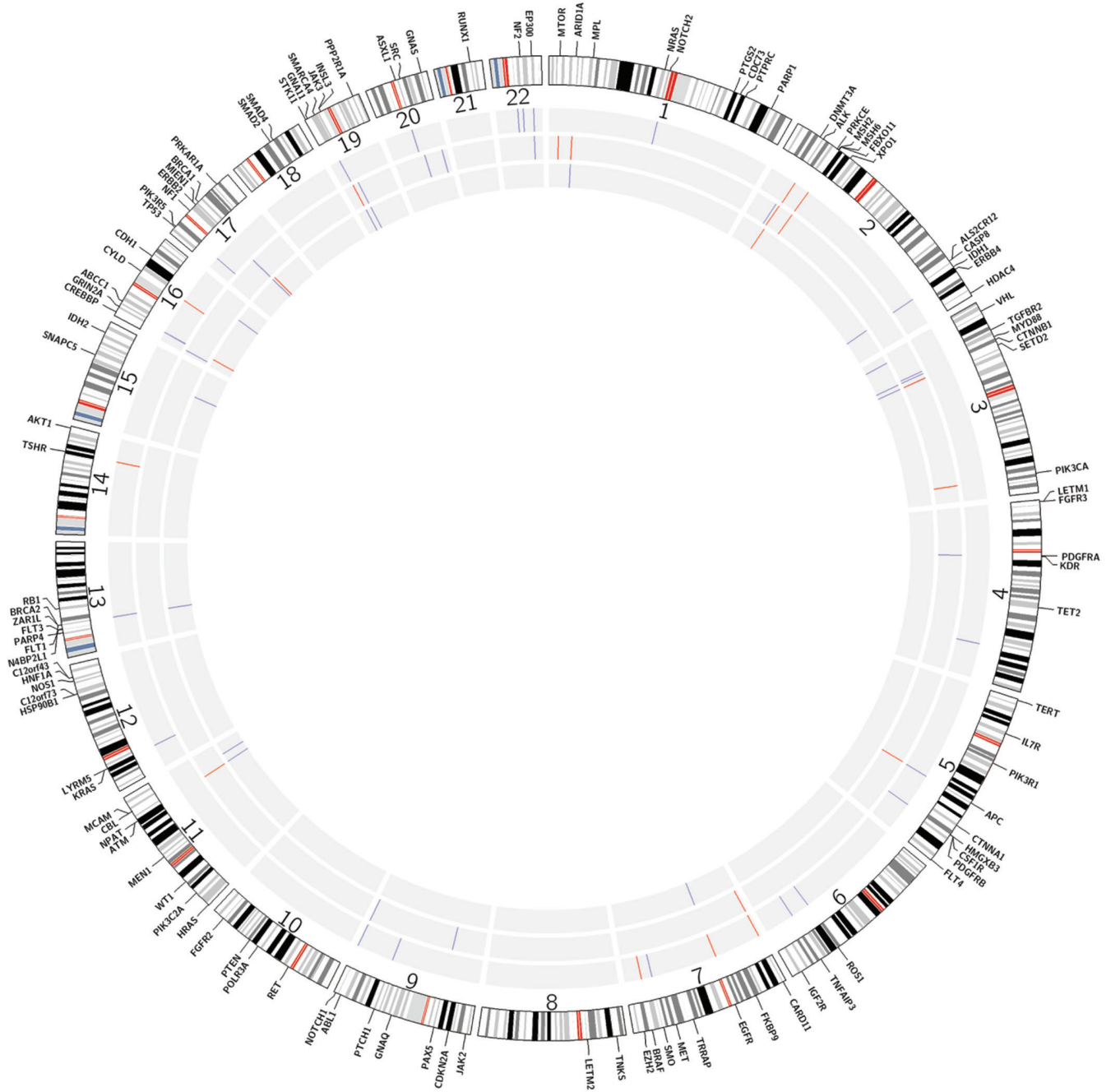


**Figure 2** MSI patterns. Each lesion showed an MSI-H phenotype with a unique pattern of peaks indicating differences in repeat number or deletion size for each locus. Arrowheads connected by a bar indicate the size difference between the highest peak in the spread for the normal and the instability alleles for each locus. Bars without arrowheads indicate larger spreads of peaks that are considered instability alleles, but do not have an obvious second allele size. Lesions 1-6, show instability at all 5 loci compared to the normal. MSI, microsatellite instability; MSI-H, MSI-high.

**Table 1** Identified variants by location and predicted effect

Variant type	Germline	Lesion 1	Lesion 2	Lesion 3
Frameshift	0.5% (1/186)	7.3% (7/96)	10.8% (8/74)	6.5% (3/46)
Non-synonymous coding	16.7% (31/186)	29.2% (28/96)	23.8% (25/74)	32.6% (15/46)
Splice site	0.0% (0/186)	3.1% (3/96)	0.0% (0/74)	0.0% (0/46)
Stop codon gained	0.0% (0/186)	0.0% (0/96)	2.7% (2/74)	0.0% (0/46)
Intron	38.7% (72/186)	43.8% (42/96)	33.8% (25/74)	37.0% (17/46)
Synonymous coding	37.6% (70/186)	13.5% (13/96)	17.6% (13/74)	17.4% (8/46)
Upstream	1.1% (2/186)	0.0% (0/96)	0.0% (0/74)	2.2% (1/46)
UTR	5.4% (10/186)	3.1% (3/96)	1.4% (1/74)	4.3% (2/46)
Total	186	96	74	46





**Figure 3** Comparison of variants by lesion. The outer wheel of this Circos plot depicts each chromosome with sequenced genes labeled. The inner wheels depict the variants identified in lesion 1 (outermost of inner wheels), lesion 2 (middle) and lesion 3 (innermost). Colored lines in the inner wheels indicate variants identified in the corresponding gene. Colors indicate high (red) and moderate (blue) predicted effect on gene function as categorized by SnpEff (12).

**Table 2** Genes in each lesion with at least one identified variant

Lesion 1	Lesion 2	Lesion 3
ALK	ALK	ALK
AMER1	x	x
APC	APC	x
x	ARID1A	ARID1A
ASXL1	x	x
x	ATM	ATM
BRAF	x	x
x	x	BRCA1
BRCA2	x	BRCA2
CARD11	CARD11	x
x	x	CBL
x	x	CDH1
x	CREBBP	x
CSF1R	x	x
x	CTNNB1	CTNNB1
CYLD	x	x
x	DNMT3A	x
EGFR	x	EGFR
EP300	x	x
x	ERBB2	x
ERBB4	x	ERBB4
FBXW7	x	x
x	GNAS	x
GRIN2A	x	GRIN2A
x	JAK3	x
x	KIT	x
KRAS	x	x
x	x	MAP2K1
MAP2K4	x	x
MET	x	x
MSH2	x	x
x	MSH6	x
x	MTOR	x
x	MYD88	x
NF2	x	x
NOTCH1	NOTCH1	x
NOTCH2		x
PAX5	PAX5	x
x	PIK3CA	x
PTCH1	x	x
ROS1	x	x
x	SETD2	SETD2
SMARCA4	SMARCA4	SMARCA4
SMARCB1	x	x
SMO	x	x
x	SRC	x
TNFAIP3	x	x
TSHR	x	TSHR
x	x	VHL

(c.2634+1G>A) is described as pathogenic in ClinVar because it interrupts a canonical splice site and is likely the second MSH2 inactivating event (13). Lesion 2 had variants in 11 unique genes involved in the EGFR, KIT, MTOR and SRC signaling pathways and transcriptional regulation (*CREBBP*, *DNMT3A*, *MYD88*, *PAX5*) (Table 2). Of note, a frameshift mutation was identified in *MSH6* (c.3205delG, p.G1070fs\*9). Heterozygosity of both partners in the MSH2/MSH6 heterodimer may result in reduced function. Lesion 3 was the least advanced of the cancers studied (T1N0M0) and had the smallest number genes containing variants. Identified variants did not correspond to obvious signaling molecules, but were present in genes involved in DNA repair, gene expression and proteasome function. These data, like the MSI data, indicate distinct oncogenic histories for each lesion.

## Discussion

The presence of multiple synchronous cancers in this patient allowed for a unique analysis of genetic diversity among CRCs without the confounding effects of genetic background. Taken together, these results provide a diverse picture of colon carcinogenesis with distinct histology, MSI patterns and gene mutation profiles.

Two genes had variants identified in all 3 sequenced lesions. Both *SMARCA4* and *ALK* are involved in other cancer types, but neither was identified at significant levels in the survey of CRCs performed by The Cancer Genome Atlas (14). *SMARCA4* encodes a subunit of the SWI/SNF complex and mutations could alter transcriptional regulation through that mechanism (15). Outside of childhood neuroblastoma, *ALK* mutations have not been widely identified, although *ALK* fusions are involved in the development of multiple cancer types (16). It is possible that the development of variants in these genes is unique to this patient or to patients with Lynch syndrome. Comprehensive DNA sequencing studies of cancers in patients with Lynch syndrome are needed to identify genes that may be particularly susceptible to mutation in this patient population.

The analysis of MSI is usually focused on the interpretation: stable, low or high. However, when studying the relatedness of cancerous lesions, the pattern of errors observed can act as a genetic fingerprint. This is because MSI assesses the accumulation of replication errors during the clonal expansion of cancer cells. Therefore, independent cancers should have different patterns of errors owing to their unique oncogenic

history and the randomness of the errors. In this study, each synchronous lesion had a distinct pattern of errors indicating distinct oncogenic histories. The genetic differences implied by the MSI analyses were observed in the SNaPshot and sequencing data as each lesion had a unique set of variants affecting a unique set of genes. Therefore, it may be possible to use MSI analysis as a screening tool to determine the relatedness of synchronous or metastatic lesions.

It is estimated that approximately 4% of unselected patients presenting with CRCs have a synchronous cancer and the 5-year survival rate for both synchronous and single lesion CRC was approximately 50% (17,18). These studies, while comprehensive, examined data from patients collected up until 2004 prior to the wide-spread use of targeted agents in CRC therapy. Targeted therapies have shown positive effects on progression free and overall survival, but the presence of multiple cancers complicates the choice of targeted therapies (19). In the current study, 2 of 6 lesions harbored activating KRAS mutations and all 3 sequenced lesions had either EGFR, ERBB2 or ERBB4 variants identified. EGFR antibody therapy is unlikely to be effective for the lesions with KRAS mutations and the variants in ERBB2 and ERBB4 may affect the efficiency of agents targeted to those molecules. These findings suggest that for patients with synchronous cancers, surgical resection may still be the best option as each tumor can have a different response to targeted agents. Further, when targeted therapies are considered, multiple lesions should be profiled, to reveal the complexity of disease and to optimize treatment.

### Acknowledgements

*Funding:* This work was supported by The National Institutes of Health and National Cancer Institute (P50CA095103) to C Shi. The authors would like to acknowledge technical assistance from Travis Clark, PhD; Thomas Stricker, MD, PhD and Vanderbilt's VANTAGE genomics core facility, which is supported by the Clinical and Translation Science Award (5UL1 RR024975-03), the Vanderbilt Ingram Cancer Center (P30 CA68485), the Vanderbilt Vision Center (P30 EY08126) and the National Institutes of Health and The National Center for Research Resources (G20 RR030956).

### Footnote

*Conflicts of Interest:* The authors have no conflicts of interest to declare.

### References

- Schlüssel AT, Gagliano RA Jr, Seto-Donlon S, et al. The evolution of colorectal cancer genetics-Part 2: clinical implications and applications. *J Gastrointest Oncol* 2014;5:336-44.
- Aslanian HR, Burgart LJ, Harrington JJ, et al. Altered DNA mismatch repair expression in synchronous and metachronous colorectal cancers. *Clin Gastroenterol Hepatol* 2008;6:1385-8.
- Lawes DA, Pearson T, Sengupta S, et al. The role of MLH1, MSH2 and MSH6 in the development of multiple colorectal cancers. *Br J Cancer* 2005;93:472-7.
- Bae JM, Cho NY, Kim TY, et al. Clinicopathologic and molecular characteristics of synchronous colorectal cancers: heterogeneity of clinical outcome depending on microsatellite instability status of individual tumors. *Dis Colon Rectum* 2012;55:181-90.
- Zauber P, Huang J, Sabbath-Solitare M, et al. Similarities of molecular genetic changes in synchronous and metachronous colorectal cancers are limited and related to the cancers' proximities to each other. *J Mol Diagn* 2013;15:652-60.
- Yamauchi M, Morikawa T, Kuchiba A, et al. Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. *Gut* 2012;61:847-54.
- Fitzgibbons RJ Jr, Lynch HT, Stanislav GV, et al. Recognition and treatment of patients with hereditary nonpolyposis colon cancer (Lynch syndromes I and II). *Ann Surg* 1987;206:289-95.
- Box JC, Rodriguez-Bigas MA, Weber TK, et al. Clinical implications of multiple colorectal carcinomas in hereditary nonpolyposis colorectal carcinoma. *Dis Colon Rectum* 1999;42:717-21.
- Mikhitarian K, Pollen M, Zhao Z, et al. Epidermal growth factor receptor signaling pathway is frequently altered in ampullary carcinoma at protein and genetic levels. *Mod Pathol* 2014;27:665-74.
- Cushman-Vokoun AM, Stover DG, Zhao Z, et al. Clinical utility of KRAS and BRAF mutations in a cohort of patients with colorectal neoplasms submitted for microsatellite instability testing. *Clin Colorectal Cancer* 2013;12:168-78.
- Wu Y, Nyström-Lahti M, Osinga J, et al. MSH2 and MLH1 mutations in sporadic replication error-positive colorectal carcinoma as assessed by two-dimensional DNA electrophoresis. *Genes Chromosomes Cancer*

- 1997;18:269-78.
12. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6:80-92.
  13. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014;42:D980-5.
  14. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487:330-7.
  15. Roskoski R Jr. Anaplastic lymphoma kinase (ALK): structure, oncogenic activation, and pharmacological inhibition. *Pharmacol Res* 2013;68:68-94.
  16. Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* 2011;11:481-92.
  17. Latournerie M, Jooste V, Cottet V, et al. Epidemiology and prognosis of synchronous colorectal cancers. *Br J Surg* 2008;95:1528-33.
  18. Lam AK, Carmichael R, Gertraud Buettner P, et al. Clinicopathological significance of synchronous carcinoma in colorectal cancer. *Am J Surg* 2011;202:39-44.
  19. Luu C, Arrington AK, Schoellhammer HF, et al. Targeted therapies in colorectal cancer: surgical considerations. *J Gastrointest Oncol* 2013;4:328-36.

**Cite this article as:** Wheeler SR, Shi C, Holt JA, Vnencak-Jones CL. Mutation profiles of synchronous colorectal cancers from a patient with Lynch syndrome suggest distinct oncogenic pathways. *J Gastrointest Oncol* 2016;7(3):E64-E71. doi: 10.21037/jgo.2016.01.07



Supplementary

Table S1 Identified somatic variants

Variables	Chrom	Pos (GRCh37)	Confidence	Ref	Variant	Gene name	Variant type	Transcript ID	Codon change	AA change	Filtered coverage	Allele coverage	Allele frequency	Variant frequency	P value	SNPEP effect	SNPEP impact	Filter	SFT score	Intron domain	LPInStr	PolyPhen score	COSMIC sample count	Somatic mutation assay	Variant clinical significance
Lesion-1 chr1	1	2140707	High	C	T	MTOR	SNP	NM_004953.3	c.2491	G>T	1750-1735	C-1753A>C-2256	0.226	0.0001	Utr_3_prime	Low	Utr_3_prime	0.18							
Lesion-1 chr1	1	1120920	High	C	T	MTOR	SNP	NM_004953.3	629	A>G	4838-631	A-4832 C-4798	0.298	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr1	1	120456166	High	A	C	NOTCH2	SNP	NM_004408.3	c.61790>A	p.R2098H	1071	C-6179A>G-187	C-6179 T-101	0.01	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0.04	Ankyrin_repeat-containing_domain_C3	Q0721	0.451		NA	
Lesion-1 chr1	1	120464637	High	A	C	NOTCH2	INS	NM_004408.3	207	A>G	ATCTCCG>+140	A-1266 ATCTCCG>-0276	0.676	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr2	2	29446184	High	C	G	ALK	SNP	NM_003044.4	44	C>G	116-28	C-136A G-636	0.636	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr2	2	30143051	High	CG	C	ALK	DEL	NM_003044.4	p.158		300	CG>-225 C-75	CG-175 C-225	0.25	0.0001	Frame_shift	High	Utr_3_prime							
Lesion-1 chr2	2	4730011	High	G	A	MSH2	SNP	NM_001258911.1	30	G>A	10-20	G-1032A>G-627	0.667	0.0001	Splice_site_donor	High	Utr_3_prime								
Lesion-1 chr2	2	4802740	High	ATT	A	MSH2	DEL	NM_001258911.1	236	ATT>A-4>-203 AT>31		ATT-60A C-209T AT>301	0.601	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr2	2	48033551	High	C	G	MSH2	SNP	NM_001258911.1	30	C>G	12-18	C-104 G-152	0.6	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr2	2	212542824	High	A	G	ERBB4	SNP	NM_005252.2	136	A>G	17-19	A-6419 G-1581	0.581	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr2	2	212578279	High	TAAAA	C	ERBB4	DEL	NM_005252.2	58	TAAAA>22 T-18		TAAAA>379 T-318	0.341	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr2	2	212587202	High	C	T	ERBB4	SNP	NM_005252.2	c.1795>A	p.V267I	258	C-1795A>-188	C-1655 T-341	0.31	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0.57	Growth_factor_receptor_(1),Furin-like_cysteine-rich_domain_(1)	G1303-4-G13303-2-G2056G-G1303-3-G13303	0.008,0.013,0.016,0.008,0.016			
Lesion-1 chr2	2	21298960	High	C	T	ERBB4	SNP	NM_005252.2	173	C>107 T-23		C-1089 T-131	0.191	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr2	2	37370427	High	C	T	MLH1	SNP	NM_001262741.1	629	C>438 T-191		C-696 T-304	0.304	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr3	3	37089021	High	G	A	MLH1	SNP	NM_001262741.1	c.17430>A	p.P581	537	G-1743A>-127	G-1764 A-236	0.236	0.0001	Synonymous_coding	Low	Utr_3_prime							
Lesion-1 chr3	3	178921639	High	C	A	PKCDA	SNP	NM_006218.2	26	C>A	8-20	C-231A T-769	0.769	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr4	4	1807894	High	G	A	PRF3	SNP	NM_001163213.1	c.15950>A	p.T63	174	G-1595A>-10	G-10A A-10	1	0.0001	Synonymous_coding	Low	Utr_3_prime							
Lesion-1 chr4	4	153208158	High	C	T	FBXW7	SNP	NM_001091151.1	c.2965>A	p.G69D	2412	C-1732 T-478	G-1718 T-231	0.281	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0.11		B722C-Q096H-Q096H-4-Q096H-9	0.01,0.145,0.003,0.787			
Lesion-1 chr5	5	15047162	High	A	G	APC	DEL	NM_001265.2	c.2965>A	p.G69D	403	A-101 G-193	G-101 G-193	0.193	0.0001	Intron	Low	Utr_3_prime							
Lesion-1 chr5	5	32875203	High	A	C	ADACATG	DEL	NM_002165.3	25	A>C	47&CATG>-17	A-132 ADACATG>-658	0.68	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr5	5	67576303	High	C	T	PKNOX1	DEL	NM_181523.2	76	CTTTT C>106 C-17		CTTTT C>106 C-17	0.356	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr5	5	11217805	High	C	A	APC	SNP	NM_001127510.2	c.75140>A	p.R2500G	499	G-348A>-150	G-1087 A-301	0.301	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime							
Lesion-1 chr5	5	149435031	High	C	T	CFHR1	SNP	NM_000211.3	c.210210>A	p.V638I	811	C-1746 T-164	C-1627 T-1079	0.079	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0.01		P07333	0.476	1	NA	
Lesion-1 chr6	6	117086421	High	T	C	ROSI	SNP	NM_002844.2	33	T>42 C-48		T-4272 C-48	0.273	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr6	6	11708643	High	C	T	ROSI	SNP	NM_002844.2	36	A>15 T-21		A-15 T-21	0.563	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr6	6	117714346	High	C	G	ROSI	SNP	NM_002844.2	24	C>12 G-12		C-12 G-12	0.5	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr6	6	117714448	High	C	T	ROSI	SNP	NM_002844.2	c.12010>A	p.E401K	1002	C-1201A>-280	C-1201A>-280	0.279	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0.23		P0892	0.005	NA		
Lesion-1 chr6	6	138223446	High	G	A	TWAP3	SNP	NM_002900.3	c.22830>A	p.E75K	164	G-118A G-46	G-172 A-278	0.28	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0.26		P12180	0.008	NA		
Lesion-1 chr7	7	296822	High	CG	C	CARD11	DEL	NM_002415.4	p.505		2174	CG>-160 T-C-551	CG-1739 C-223	0.253	0.0001	Frame_shift	High	Utr_3_prime							
Lesion-1 chr7	7	298404	High	C	T	CARD11	SNP	NM_002415.4	c.4460>A	p.R149H	1539	C-1200 T-338	G-174 T-102	0.252	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0.17		Q9EXL7	0.006	5	NA	
Lesion-1 chr7	7	5520053	High	A	T	EGFR	SNP	NM_001284.1	40	A>T	4-10	A-10 T-10	1	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr7	7	5520879	High	AG	A	EGFR	DEL	NM_005253.3	2496	AG>-195 A-541		AG-176 A-217	0.217	0.0001	Splice_site_acceptor	High	Utr_3_prime								
Lesion-1 chr7	7	128850274	High	C	A	SMO	SNP	NM_006614.1	c.15370>A	p.P513T	529	C-373 A-155	G-176 A-293	0.293	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0		A07K5-Q0985	0.999,0.999			
Lesion-1 chr7	7	14034820	High	A	AAAG	BRAF	INS	NM_004333.4	792	A>-388 AAAG>-238		A-503 AAAG>-301	0.301	0.0001	Intron	High	Utr_3_prime								
Lesion-1 chr7	7	140520283	High	T	C	BRAF	SNP	NM_004333.4	363	T>253 C-110		T-253 C-110	0.303	0.0001	Splice_site_acceptor	High	Utr_3_prime								
Lesion-1 chr9	9	2197120	High	A	G	CDKN2A	SNP	NM_008197.4	25	A>-16 A-4		A-16 G-16	1	0.0008	Intron	Low	Utr_3_prime								
Lesion-1 chr9	9	33094203	High	A	A	PAX5	SNP	NM_018734.1	c.11100>T	p.Y370	303	G-10 A-393	G-10 A-110	0.333	0.0001	Synonymous_coding	Low	Utr_3_prime	0.07		ETRFWS,ETFGT0	0.0,0.001			
Lesion-1 chr9	9	8034587	High	G	G	GNAQ	DEL	NM_002072.3	31	GAA&A>-16 G-33		GAA&A>-65 G-33	0.333	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr9	9	9623165	High	G	C	PTCH1	SNP	NM_002064.3	c.21180>G	p.S709R	22	G-18 C-6	G-127 C-423	0.273	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0.27		G13635-4-G13635-3-G13635-2-G13635	0.474,0.394,0.495,0.016			
Lesion-1 chr9	9	9623122	High	G	A	PTCH1	SNP	NM_002064.3	c.19610>C	p.T656M	2472	G-1763 A-708	G-1713 A-286	0.286	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0.02		G13635-4-G13635-3-G13635-2-G13635	0.991,0.983,0.978,0.08			
Lesion-1 chr9	9	13373020	High	G	A	ABL1	SNP	NM_007313.2	c.7770>A	p.T259	2498	G-1691 A-805	G-677 A-322	0.322	0.0001	Synonymous_coding	Low	Utr_3_prime	0						
Lesion-1 chr9	9	133759792	High	C	T	ABL1	SNP	NM_007313.2	c.21720>T	p.G724	51	C>14 T-17	G-1667 T-333	0.333	0.0001	Synonymous_coding	Low	Utr_3_prime							
Lesion-1 chr9	9	13391198	High	C	T	NOTCH1	SNP	NM_017617.3	c.62900>A	p.R2098H	68	C-44 T-24	C-1047 T-353	0.353	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0						
Lesion-1 chr9	9	13391982	High	C	T	NOTCH1	SNP	NM_017617.3	c.51190>A	p.A1707T	403	G-110 T-327	G-107 T-227	0.227	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0.07		Notch_NCD_domain_(1)	P4631	0.999	NA	
Lesion-1 chr9	9	13392192	High	G	T	NOTCH1	SNP	NM_017617.3	c.22657>C	p.W709	2461	A-1147 G-133	G-1083 T-277	0.277	0.0001	Synonymous_coding	Low	Utr_3_prime							
Lesion-1 chr9	9	13941009	High	G	G	GAGAC	NOTCH1	INS	NM_017617.3	53	G>G	GAGAAC>-18	G-1028 GAGAAC>-34	0.34	0.0001	Intron	Low	Utr_3_prime							
Lesion-1 chr9	9	139411890	High	G	A	NOTCH1	SNP	NM_017617.3	73	G>-26 A-47		G-356 A-644	0.644	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr11	11	10813812	High	CT	C	ATM	DEL	NM_000513.3	342	CT>-256 C-86		CT-176 C-285	0.285	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr12	12	25398284	High	C	T	KRAS	SNP	NM_008953.3	c.35G>A	p.G12D	900	C-35 G-19	G-171 T-285	0.285	0.0001	Non_synonymous_coding	Moderate	Utr_3_prime	0						
Lesion-1 chr12	12	33093063	High	CT	C	BRCA2	DEL	NM_002093.3	399	CT>-272 C-109		CT-168 C-123	0.273	0.0001	Intron	Low	Utr_3_prime								
Lesion-1 chr12	12	32910250	High	T	A	BRCA2	SNP	NM_002093.3	c.1037A>C	p.Q404H	249	C-110 T-134	C-110 T-134	0.2											



Table S2 Putative genuine variants																
Chrom	Pos GRCh37	Confidence	Ref	Variant	Gene name	Variant type	Transcript ID	Codon change	AA change	Filtered coverage	Allele coverage	Allele frequency	Variant frequency	P Value	SNPEff effect	SNPEff impact
chr1	1120558	chr1_1120558_C_T_MTOR	C	T	MTOR	SNP	NM_004958.3	c.4737G > A	p.A1577						Synonymous_Coding	Low
chr1	1120578	chr1_1120578_G_A_MTOR	G	A	MTOR	SNP	NM_004958.3	c.390C > T	p.Y99						Synonymous_Coding	Low
chr1	1120714	chr1_1120714_A_G_MTOR	A	G	MTOR	SNP	NM_004958.3	c.1423T > C	p.L478						Synonymous_Coding	Low
chr1	12061964	chr1_12061964_G_C_NOTCH2	G	C	NOTCH2	SNP	NM_001200001.1	c.57C > G	p.S19W						Non_synonymous_Coding	Moderate
chr2	29416366	chr2_29416366_G_C_ALK	G	C	ALK	SNP	NM_004304.4	c.4587C > G	p.L1529E						Non_synonymous_Coding	Moderate
chr2	29416481	chr2_29416481_T_C_ALK	T	C	ALK	SNP	NM_004304.4	c.4472A > G	p.K1491R						Non_synonymous_Coding	Moderate
chr2	29416572	chr2_29416572_T_C_ALK	T	C	ALK	SNP	NM_004304.4	c.4381A > G	p.L1464E						Non_synonymous_Coding	Moderate
chr2	29445408	chr2_29445408_G_T_ALK	G	T	ALK	SNP	NM_004304.4	c.3730C > A	p.Y1212S						Synonymous_Coding	Low
chr2	29445819	chr2_29445819_C_T_ALK	C	T	ALK	SNP	NM_004304.4	c.3203G > A	p.Y1012E						Synonymous_Coding	Low
chr2	29450287	chr2_29450287_A_G_ALK	A	G	ALK	SNP	NM_004304.4	c.2203T > C	p.Q64S						Synonymous_Coding	Low
chr2	29454963	chr2_29454963_T_C_ALK	T	C	ALK	SNP	NM_004304.4	c.1500A > G	p.Q50Q						Synonymous_Coding	Low
chr2	29454736	chr2_29454736_G_A_ALK	A	G	ALK	SNP	NM_004304.4	c.1427T > C	p.V476A						Non_synonymous_Coding	Moderate
chr2	29945529	chr2_29945529_A_T_ALK	A	T	ALK	SNP	NM_004304.4	c.702T > A	p.P234						Synonymous_Coding	Low
chr2	47702578	chr2_47702578_GT_G_MSH2	GT	G	MSH2	DEL	NM_002051.2		p.R93						Frame_Shift	High
chr2	48018981	chr2_48018981_A_G_MSH2	A	G	MSH2	SNP	NM_002051.2	c.375A > G	p.P92						Synonymous_Coding	Low
chr3	47125285	chr3_47125285_G_A_SETD2	G	A	SETD2	SNP	NM_014159.6	c.5885C > T	p.P166L						Non_synonymous_Coding	Moderate
chr3	47162961	chr3_47162961_A_G_SETD2	A	G	SETD2	SNP	NM_014159.6	c.3467C > C	p.N115S						Synonymous_Coding	Low
chr4	1803704	chr4_1803704_T_C_FGFR3	T	C	FGFR3	SNP	NM_001424.2	c.882T > C	p.R284						Synonymous_Coding	Low
chr4	1807478	chr4_1807478_G_T_FGFR3	G	T	FGFR3	SNP	NM_01162913.1	c.1653G > T	p.Q55I						Synonymous_Coding	Low
chr4	5514105	chr4_5514105_A_G_PDGFR4	A	G	PDGFR4	SNP	NM_002026.4	c.1701A > G	p.P567						Synonymous_Coding	Low
chr4	55181391	chr4_55181391_T_C_PDGFR4	T	C	PDGFR4	SNP	NM_002026.4	c.3222T > C	p.Y1074						Synonymous_Coding	Low
chr5	1204648	chr5_1204648_C_T_RGS1	C	T	RGS1	SNP	NM_180533.2	c.319G > A	p.S35S						Synonymous_Coding	Low
chr5	35871190	chr5_35871190_G_A_LTR	G	A	LTR	SNP	NM_001185.3	c.413G > A	p.Y138						Non_synonymous_Coding	Moderate
chr5	35871293	chr5_35871293_A_G_LTR	A	G	LTR	SNP	NM_001185.3	c.615A > G	p.E172Q						Non_synonymous_Coding	Moderate
chr5	35876449	chr5_35876449_C_T_LTR	C	T	LTR	SNP	NM_001185.3	c.1241C > T	p.T414M						Non_synonymous_Coding	Moderate
chr5	67588148	chr5_67588148_G_A_PKCRI1	G	A	PKCRI1	SNP	NM_181504.3	c.186G > A	p.M46E						Non_synonymous_Coding	Moderate
chr5	112162854	chr5_112162854_T_C_APC	T	C	APC	SNP	NM_000038.5	c.1488T > C	p.Y48E						Synonymous_Coding	Low
chr5	112172770	chr5_112172770_G_A_APC	G	A	APC	SNP	NM_00127510.2	c.4479G > A	p.T149D						Synonymous_Coding	Low
chr5	112176255	chr5_112176255_G_A_APC	G	A	APC	SNP	NM_00127510.2	c.524G > A	p.S167R						Synonymous_Coding	Low
chr5	112176509	chr5_112176509_T_G_APC	T	G	APC	SNP	NM_00127510.2	c.5268T > G	p.S177E						Synonymous_Coding	Low
chr5	112178756	chr5_112178756_T_A_APC	T	A	APC	SNP	NM_00127510.2	c.5485T > A	p.R182Q						Non_synonymous_Coding	Moderate
chr5	112177171	chr5_112177171_G_A_APC	G	A	APC	SNP	NM_00127510.2	c.5880G > A	p.P196Q						Synonymous_Coding	Low
chr5	149456843	chr5_149456843_C_T_CSF1R	C	T	CSF1R	SNP	NM_005211.3	c.885G > A	p.Y29E						Synonymous_Coding	Low
chr5	149457878	chr5_149457878_G_A_CSF1R	G	A	CSF1R	SNP	NM_005211.3	c.729C > T	p.T24Z						Synonymous_Coding	Low
chr5	149460553	chr5_149460553_G_A_CSF1R	A	G	CSF1R	SNP	NM_005211.3	c.841T > C	p.P78						Synonymous_Coding	Low
chr6	117621744	chr6_117621744_G_C_ROS1	G	C	ROS1	SNP	NM_002944.2	c.686G > G	p.S229Q						Non_synonymous_Coding	Moderate
chr6	117621188	chr6_117621188_T_G_ROS1	T	G	ROS1	SNP	NM_002944.2	c.668A > C	p.K225Q						Non_synonymous_Coding	Moderate
chr6	117622233	chr6_117622233_C_T_ROS1	C	T	ROS1	SNP	NM_002944.2	c.663T > A	p.E221N						Non_synonymous_Coding	Moderate
chr6	117631431	chr6_117631431_T_G_ROS1	T	G	ROS1	SNP	NM_002944.2	c.624T > A	p.R208S						Non_synonymous_Coding	Moderate
chr6	117645537	chr6_117645537_T_C_ROS1	T	C	ROS1	SNP	NM_002944.2	c.559A > G	p.I187V						Non_synonymous_Coding	Moderate
chr6	117708971	chr6_117708971_G_T_ROS1	G	T	ROS1	SNP	NM_002944.2	c.198G > A	p.P62						Synonymous_Coding	Low
chr6	117715073	chr6_117715073_G_C_ROS1	G	C	ROS1	SNP	NM_002944.2	c.169G > C	p.A327V						Non_synonymous_Coding	Moderate
chr6	117754548	chr6_117754548_T_G_ROS1	T	G	ROS1	SNP	NM_002944.2	c.432A > C	p.T145P						Non_synonymous_Coding	Moderate
chr6	117725578	chr6_117725578_T_A_ROS1	T	A	ROS1	SNP	NM_002944.2	c.333A > T	p.I101						Synonymous_Coding	Low
chr7	2957005	chr7_2957005_T_C_CARD11	T	C	CARD11	SNP	NM_032415.4	c.2622A > G	p.P874						Synonymous_Coding	Low
chr7	55214348	chr7_55214348_C_T_EGFR	C	T	EGFR	SNP	NM_005228.3	c.474C > T	p.N158						Non_synonymous_Coding	Low
chr7	55249003	chr7_55249003_G_A_EGFR	G	A	EGFR	SNP	NM_005228.3	c.2361G > A	p.Q787						Synonymous_Coding	Low
chr7	55296417	chr7_55296417_T_C_EGFR	T	C	EGFR	SNP	NM_005228.3	c.2707T > C	p.T903						Synonymous_Coding	Low
chr7	11524269	chr7_11524269_C_T_MET	C	T	MET	SNP	NM_001275823.3	c.1131C > T	p.S377						Synonymous_Coding	Low
chr7	128846328	chr7_128846328_G_C_SMO	G	C	SMO	SNP	NM_005601.4	c.1145G > C	p.S288						Synonymous_Coding	Low
chr7	140449150	chr7_140449150_T_C_BRAF	T	C	BRAF	SNP	NM_004333.4	c.1809A > G	p.G343						Synonymous_Coding	Low
chr7	140478804	chr7_140478804_G_A_BRAF	G	A	BRAF	SNP	NM_004333.4	c.1602C > T	p.Q534						Synonymous_Coding	Low
chr9	133755528	chr9_133755528_A_G_ABL1	A	G	ABL1	SNP	NM_007313.2	c.1554A > G	p.E518						Synonymous_Coding	Low
chr9	133760029	chr9_133760029_C_G_ABL1	C	G	ABL1	SNP	NM_007313.2	c.240G > C	p.F803						Synonymous_Coding	Low
chr9	133791001	chr9_133791001_A_G_ABL1	A	G	ABL1	SNP	NM_007313.2	c.2381A > G	p.P1127						Synonymous_Coding	Low
chr9	133801626	chr9_133801626_G_A_NOTCH1	G	A	NOTCH1	SNP	NM_017617.3	c.655C > T	p.S186E						Non_synonymous_Coding	Low
chr9	139451233	chr9_139451233_C_T_NOTCH1	C	T	NOTCH1	SNP	NM_017617.3	c.383G > A	p.S127H						Non_synonymous_Coding	Moderate
chr9	139418200	chr9_139418200_A_G_NOTCH1	A	G	NOTCH1	SNP	NM_017617.3	c.312T > C	p.N104						Non_synonymous_Coding	Moderate
chr10	43066887	chr10_43066887_G_A_RET	A	G	RET	SNP	NM_020975.4	c.1206A > G	p.A43Z						Synonymous_Coding	Low
chr10	43613843	chr10_43613843_G_T_RET	G	T	RET	SNP	NM_020975.4	c.2307G > T	p.L769						Synonymous_Coding	Low
chr10	12328158	chr10_12328158_T_C_FGFR2	T	C	FGFR2	SNP	NM_022970.2	c.698A > G	p.V23Z						Synonymous_Coding	Low
chr11	32417945	chr11_32417945_T_C_WT1	T	C	WT1	SNP	NM_024265.4	c.1107A > G	p.P389						Synonymous_Coding	Low
chr11	64572578	chr11_64572578_T_C_MEN1	T	C	MEN1	SNP	NM_130790.2	c.1612A > G	p.Y414						Non_synonymous_Coding	Moderate
chr11	64572557	chr11_64572557_A_G_MEN1	A	G	MEN1	SNP	NM_130804.2	c.1314T > C	p.H438						Synonymous_Coding	Low
chr11	108183167	chr11_108183167_G_A_ATM	A	G	ATM	SNP	NM_000051.3	c.5948A > G	p.N1983S						Non_synonymous_Coding	Moderate
chr12	25384462	chr12_25384462_C_T_KRAS	C	T	KRAS	SNP	NM_033360.2	c.483G > A	p.R161						Synonymous_Coding	Low
chr12	121435342	chr12_121435342_C_T_HNF1A	C	T	HNF1A	SNP	NM_000545.5	c.1375C > T	p.L459						Synonymous_Coding	Low
chr12	121435427	chr12_121435427_G_A_HNF1A	G	A	HNF1A	SNP	NM_000545.5	c.1460G > A	p.S487N						Non_synonymous_Coding	Moderate
chr12	121437114	chr12_121437114_G_A_HNF1A	G	A	HNF1A	SNP	NM_000545.5	c.1545G > A	p.Y51S						Synonymous_Coding	Low
chr13	28024294	chr13_28024294_G_A_FLT3	G	A	FLT3	SNP	NM_004119.2	c.690C > T	p.T272M						Non_synonymous_Coding	Moderate
chr13	28030084	chr13_28030084_G_A_FLT3	G	A	FLT3	SNP	NM_004119.2	c.298C > T	p.S96						Synonymous_Coding	Low
chr13	32060729	chr13_32060729_A_C_BRCA2	A	C	BRCA2	SNP	NM_000099.3	c.1114A > C	p.N472H						Non_synonymous_Coding	Moderate
chr13	32091888	chr13_32091888_A_G_BRCA2	A	G	BRCA2	SNP	NM_000099.3	c.3396A > G	p.K1132						Synonymous_Coding	Low
chr13	32092532	chr13_32092532_A_G_BRCA2	A	G	BRCA2	SNP	NM_000099.3	c.7242A > G	p.S							