



RESEARCH ARTICLE

# Characterization of an APC Promoter 1B deletion in a Patient Diagnosed with Familial Adenomatous Polyposis via Whole Genome Shotgun Sequencing [v1; ref status: indexed, <http://f1000r.es/5i1>]

Ted Kalbfleisch<sup>1</sup>, Pamela Brock<sup>2</sup>, Angela Snow<sup>3</sup>, Deborah Neklason<sup>3,4</sup>, Gordon Gowans<sup>2</sup>, Jon Klein<sup>5</sup>

<sup>1</sup>Department of Biochemistry and Molecular Genetics, School of Medicine, University of Louisville, Louisville, Kentucky, 40202, USA

<sup>2</sup>Clinical Genetics, Weisskopf Child Evaluation Center, University of Louisville, Louisville, Kentucky, 40202, USA

<sup>3</sup>Huntsman Cancer Institute, Salt Lake City, Utah, 84112, USA

<sup>4</sup>Division of Genetic Epidemiology, Department of Internal Medicine, University of Utah, Salt Lake City, Utah, 84112, USA

<sup>5</sup>Department of Medicine, School of Medicine, University of Louisville, Louisville, Kentucky, 40202, USA

**v1** First published: 26 Jun 2015, 4:170 (doi: [10.12688/f1000research.6636.1](https://doi.org/10.12688/f1000research.6636.1))  
 Latest published: 26 Jun 2015, 4:170 (doi: [10.12688/f1000research.6636.1](https://doi.org/10.12688/f1000research.6636.1))

**Abstract**

Recently, deletions have been identified and published as causal for Familial Adenomatous Polyposis in the 1B promoter region of the APC gene. Those deletions were measured using multiplex ligation-dependent probe amplification. Here, we present and characterize an ~11kb deletion identified by whole genome shotgun sequencing. The deletion occurred in a patient diagnosed with Familial Adenomatous Polyposis, and was located on chr5, between bases 112,034,824 and 112,045,845, fully encompassing the 1B promoter region of the APC gene. Results are presented here that include the sequence evidence supporting the presence of the deletion as well as base level characterization of the deletion site. These results demonstrate the capacity of whole genome sequencing for the detection of large structural variants in single individuals.

**Open Peer Review**

Referee Status:

	Invited Referees	
	1	2
<b>version 1</b> published 26 Jun 2015	 report	 report

- 1** **Nicholas Davidson**, Washington University in St. Louis USA, **Yiing Lin**, Washington University in St. Louis USA
- 2** **Anna Rohlin**, University of Gothenburg Sweden

**Discuss this article**

Comments (0)

**Corresponding author:** Ted Kalbfleisch ([ted.kalbfleisch@louisville.edu](mailto:ted.kalbfleisch@louisville.edu))

**How to cite this article:** Kalbfleisch T, Brock P, Snow A *et al.* **Characterization of an APC Promoter 1B deletion in a Patient Diagnosed with Familial Adenomatous Polyposis via Whole Genome Shotgun Sequencing [v1; ref status: indexed, <http://f1000r.es/511>]** *F1000Research* 2015, 4:170 (doi: [10.12688/f1000research.6636.1](https://doi.org/10.12688/f1000research.6636.1))

**Copyright:** © 2015 Kalbfleisch T *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

**Grant information:** The Next Generation Sequencing work was supported by DOE grant DE-EM0000197 (Kalbfleisch, Rouchka co-PI). Dr. Kalbfleisch received additional financial support from the National Institute of General Medical Sciences of the National Institutes of Health Grant# P20GM103436 (Cooper PI). University of Utah work is supported by PO1CA073992.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Competing interests:** T.K. serves as the CEO of Intrepid Bioinformatics.

**First published:** 26 Jun 2015, 4:170 (doi: [10.12688/f1000research.6636.1](https://doi.org/10.12688/f1000research.6636.1))

**First indexed:** 15 Jul 2015, 4:170 (doi: [10.12688/f1000research.6636.1](https://doi.org/10.12688/f1000research.6636.1))

## Introduction

Familial Adenomatous Polyposis (FAP) is an autosomal dominant condition characterized by the development of hundreds to thousands of polyps in the colon. This condition results in colon cancer in adult individuals in their late 20s to early 30s with nearly 100 percent penetrance. Mutations in two genes, the adenomatous polyposis coli (APC) and mutY homolog (MUTYH) loci, have been identified as causative for this disease. The majority of the mutations occur in the APC locus. The APC mutations often take the form of single nucleotide substitutions or small insertions or deletions in the coding region of the gene that produce premature stop codons, or frame shifts respectively. These result in a change of function. The exact mechanism by which these mutations affect the disease is unknown. However, deletions of APC promoter 1B are known to cause a significant change in transcription levels of the APC RNA marked by allele specific differences in transcription<sup>1,2</sup>. Several mutations have been reported in the promoter region of the APC gene<sup>1-4</sup>, identified either by sequencing, or by multiplex ligation-dependent probe amplification (MLPA).

The patient analyzed in this work is a 50 year-old Caucasian female who has a personal and maternal family history of FAP. She developed colon polyps at 14 years of age and underwent a partial colectomy at 16 years. The patient had a complete colectomy and a Whipple procedure in her 20's. Her mother and multiple avunculars and cousins on the maternal side are affected. One sibling has a clinical diagnosis of FAP and three siblings are unaffected. The patient's maternal grandfather died of colon cancer later in life, but a diagnosis of FAP was not confirmed.

Previously, DNA testing in family members had failed to identify a causative mutation. Therefore, the patient and her family participated in a linkage analysis project through the Mayo Clinic in Rochester, Minnesota to identify at-risk family members. The FAP in the family showed linkage to the APC locus on chromosome 5. The patient underwent molecular testing of the APC gene (sequence analysis and Southern blot) and MUTYH gene (analysis for 2 common mutations) in 2008. No mutations were detected. A variant of unknown significance (referred to as Glu1317Gln) was found in the APC gene. However, this variant was absent in other affected family members and was present in the patient's unaffected child. It was later classified as likely benign<sup>5</sup>. The multiplex ligation-dependent probe amplification (MLPA) assays for the APC locus in use at the time did not characterize the APC promoters, and was negative for APC mutations for this patient.

In an effort to comprehensively search for potential mutations, the patient's genomic DNA was sent to Illumina whole genome sequencing. A deletion of ~11kb encompassing the APC promoter 1B was identified, and is consistent with the deletion identified recently by Snow *et al.*<sup>2</sup> via an updated MLPA assay for APC that now includes promoter 1B and by Lin *et al.*<sup>4</sup>.

In this work, we present a comprehensive characterization of this deletion using Illumina short reads, including base level resolution of the deletion site. Further, it is demonstrated that this deletion is detectable using the MLPA assay for the APC locus current at the submission of this article, and would be ambiguous if this, or any single patient were analyzed solely via whole exome sequencing.

## Methods

### Sequencing and alignment

The whole blood sample for this study was collected under a protocol approved by the University of Louisville IRB (IRB tracking number 11.0659, approval date 1/30/2012). Written informed consent for publication of clinical details was obtained from the patient/next of kin. The blood was sent to the Illumina Clinical Services Laboratory for paired end sequencing of 100 bp reads from fragments with a target length of 300bp. The reads produced were mapped via CASAVA (CASAVA-1.9.0a1\_110909) to the human reference genome build 37.1 at an average depth of coverage 37.51X.

### Remapping and variant detection

The pipeline employed in our lab for read mapping and variant detection uses the Burrows-Wheeler Alignment<sup>6</sup> algorithm, and the Genome Analysis Toolkit<sup>7</sup> respectively. To be consistent with other work in our lab, reads for the regions of interest were extracted from the bam file produced by Illumina, and run through our pipeline.

Mapped reads were extracted from the binary alignment map file for remapping using Samtools<sup>8</sup> version 0.1.18 from the individual's full binary alignment map file (provided by Illumina) corresponding to 50,000 bases upstream and downstream of the APC, and MUTYH loci defined respectively by the mapping of accession NM\_001127511.1 and NM\_001293192.1 to human genome build 37.1 (chr5:111,993,219-112,231,936 and chr1:45,744,915-45,856,143). Reads mapping to other chromosomes, or positions on chromosomes 1 and 5 outside of the target region would have also been extracted if their mate mapped within the target regions. Reads in these extraneous regions were not considered in variant detection. To be consistent with the remainder of our work, the FASTQ files corresponding to the first and second reads of the pair (R1 and R2) were re-derived via BEDTools<sup>9</sup> from the BAM file provided by Illumina, and remapped using the BWA algorithm for short read alignment. Duplicates were marked, indels were realigned, base quality scores recalibrated, and variants identified and simultaneously genotyped for our trace data by applying the GATK MarkDuplicates, IndelRealigner, BaseRecalibrator, and HaplotypeCaller algorithms respectively<sup>10,11</sup>.

The deletion was identified by visual inspection within the Integrative Genomics Viewer (IGV)<sup>12</sup> of the mapped next generation sequence data set as well as the variation reported in the accompanying variant call format file. This deletion is characterized by a loss of heterozygosity of variants measured relative to the reference, a cluster of 11 paired end reads (target length 500 bases) whose mates map in excess of 11kb from one another, as well as 15 reads that span the junction of the deletion that were soft trimmed by the mapping algorithm. The option "Show soft-clipped bases" within View/Preferences/Alignments was turned on and revealed soft trimming that began in several reads at positions 112,034,824 and 112,045,845 on chromosome 5. Bases from these reads were copied from within the IGV user interface for subsequent analysis in BLAT<sup>13</sup> to confirm the position of the deletion.

### PCR and Sanger Sequencing Confirmation

Primers were designed to specifically interrogate this deletion with one primer pair flanking the deletion, and one primer pair with one primer located in the deleted region. The primer located 3' to the

deleted region was common to both pairs. Full description of the primers is in provided in Table 1.

**DNA extraction, PCR, and Sanger Sequencing**

Whole blood was fractionated by spinning at 5,000 rpm for 10 minutes at room temperature. White cells were transferred to sterile, nuclease free microcentrifuge tubes and stored at -20°C until processing. Genomic DNA was isolated from 250uL buffy coat with Genra Puregene Genomic DNA purification buffers (Qiagen, Valencia, CA). Separate amplification of the wild type or deletion APC fragments were performed in a 20uL reaction containing 0.4uL Phusion HF DNA Polymerase (Thermo Fisher Scientific, Pittsburg, PA), 1x Phusion Reaction Buffer, 200uM dNTP's (Promega Corporation, Madison, WI), 200ng gDNA, and 0.5uM each primer. The cycling conditions were as follows: 98°C for 30s followed by 35 cycles of 98°C for 10s, 60°C for 30s, and 72°C for 60s, ending with a final extension of 72°C for 7min.

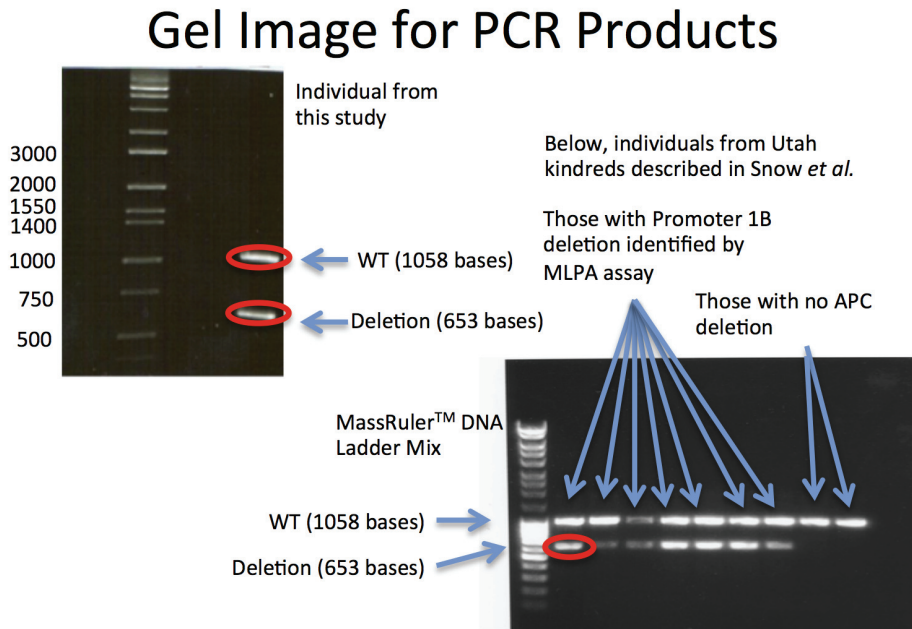
The amplicons were sequenced with BigDye® Terminator v3.1 (Life Technologies Corporation, Carlsbad, CA) utilizing the PCR primers and standard sequencing conditions. The sequence reactions were purified with Performa DTR Ultra 96-well filtration plates (Edge Biosystems, Gaithersburg, MD) and processed on the ABI 3130xl Genetic Analyzer (Life Technologies Corporation, Carlsbad, CA).

The resulting gel for the PCR products is shown in Figure 1, and the sequencing results are shown in Figure 2, rendered in Geospiza's FinchTV, (<http://www.geospiza.com/Products/finchtv.shtml>).

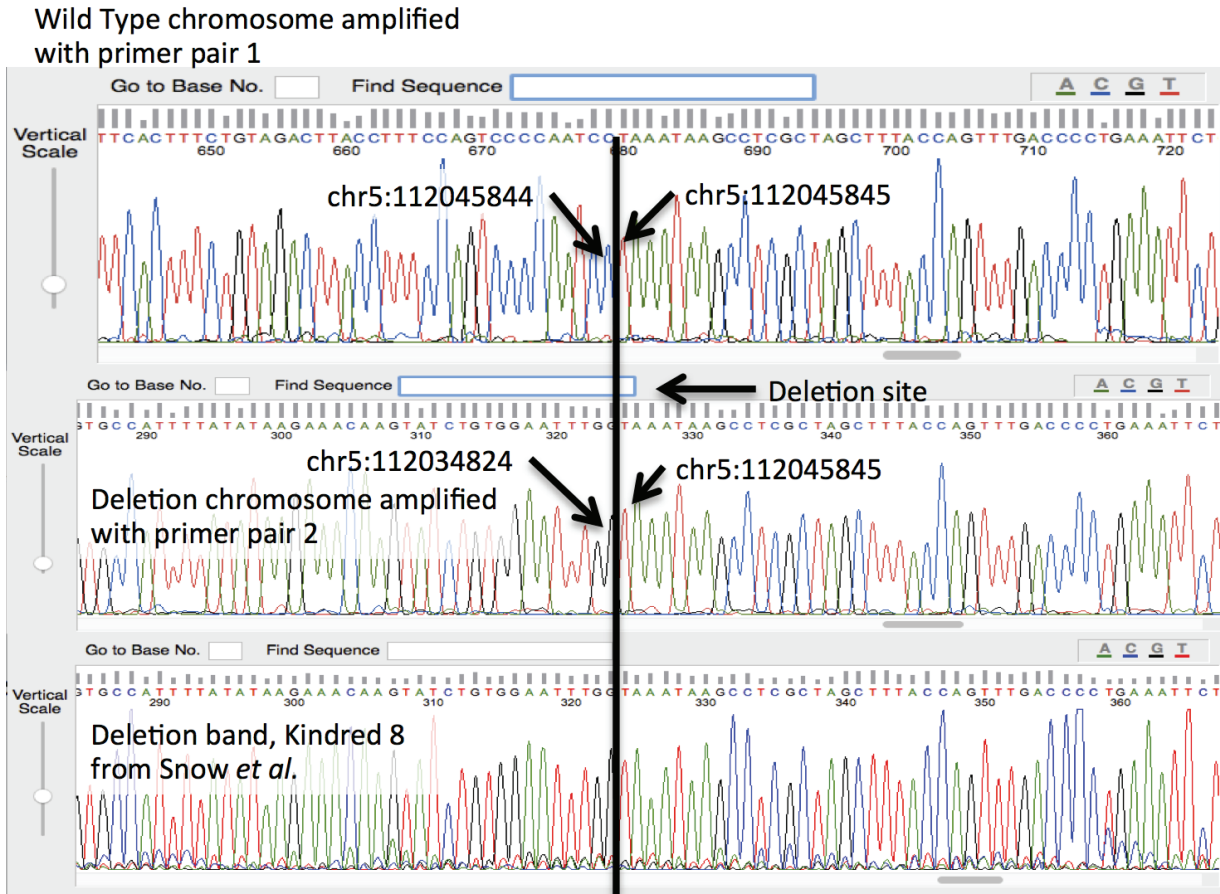
**Dataset 1. Raw Gel electrophoresis image for Figure 1**  
<http://dx.doi.org/10.5256/f1000research.6636.d50276>  
 The gel image represented in Figure 1, showing in lane 1 a control human sample that was not part of this work which was cropped from the in-text figure<sup>15</sup>.

**Table 1. Primers used to genotype samples for the presence or absence of the deletion described in this work.**

	Primer Pair 1	Primer Pair 2
Left Primer	GGGCTAGTTCATTCGTTGCT	CACACCTACCATTGTGTTACCATT
Right Primer	GAGGGGGTTGCTCTTGAAA	GAGGGGGTTGCTCTTGAAA
Product Length-No Deletion	1058 Bases	11653 Bases
Product Length-With Deletion	No Product	653 Bases



**Figure 1. Gel images for the PCR products produced in the Louisville Index case, as well as the seven kindreds reported by Snow et al., using the two primer pairs described in Table 1.** The relationship between the lanes and the nine kindreds defined in Snow et al. are Lane 1: Ladder, Lane 2: Kindred 8, Lane 3: Kindred 43, Lane 4: Kindred 44, Lane 5: Kindred 256, Lane 6: Kindred 509, Lane 7: Kindred 685, Lane 8: Kindred 691, Lane 9: Kindred 353 (APC c.426\_427delAT) And Lane 10: Kindred 6699 (APC c.532-941G>A). These images demonstrate heterozygous deletions in eight of the samples analyzed. PCR products corresponding to the bands circled in red were sequenced using Sanger technology. Those results are shown in Figure 2.



**Figure 2.** Results from Sanger sequencing for both alleles in the Louisville patient, as well as the deleted allele in a patient from the study of Snow *et al.* The top two traces indicate the nucleotide sequence of the wild type APC locus in the Promoter 1B region, and the deletion site. The bottom trace demonstrates that the deletion detected by Snow *et al.* is identical to the deletion to the individual described in this work.

## Results

Paired end whole genome sequence data was generated at ~40X coverage for the patient, and mapped to the human reference assembly Build-37.1. Given the clinical phenotype our initial analysis of the data was limited to the APC and MUTYH loci. Variation analysis was performed in the region defined by the 5' and 3' most exons of the longest reported transcript for APC and MUTYH, plus and minus 50,000 bases respectively (described in detail in Methods). The resulting counts of single nucleotide variations (SNVs) and small indels are shown in Table 2–Table 4. The corresponding VCF file, along with mapped reads for these regions are available for download or visualization at <http://dx.doi.org/10.13013/J6QN64N8>. When viewing in IGV, navigate to the APC locus by entering APC in the text box at the top of the frame.

All missense variants identified had corresponding records in dbSNP and are listed in Table 4. None are reported as deleterious. There were no non-sense SNVs or frame shifting small insertions or deletions identified. The search was then turned toward larger structural variants. Visual inspection of the VCF file for the APC locus revealed a region of approximately 10kb with 17 measured SNVs or small insertions relative to the reference. None of their

respective genotypes were classified as heterozygous. This loss of heterozygosity suggested a deletion. Upon further inspection, there were other signatures characteristic of a deletion, that included a cluster of paired end reads whose mate mapped ~11kb from their respective start, and several mates that were soft trimmed because they spanned the deletion site. These soft trimmed mates were identified (described in methods), and aligned via BLAT to hsBuild-37.1, revealing the deleted region to be of length 11,020 bases, located on chr5, between bases 112,034,824 and 112,045,845, spanning the annotated APC promoter 1B. This deletion is illustrated in Figure 3, along with the positions of commercial probe sets, and other annotation relevant to this work. Given that this deletion was consistent with the deletion reported by Snow *et al.*, the primers used for verification in this work, were run on the kindreds studied in that work. It was verified that the deletion reported there was identical to the one reported here. Also, this deletion is identical to a deletion published by Lin *et al.*,<sup>4</sup> identified in kindreds from Missouri, Illinois, and Idaho not known to be related to each other.

The Illumina paired end short read data that provides evidence for the deletion relative to the reference has been isolated from the

**Table 2.** Single nucleotide variants (SNVs) identified in our patient for the APC and MUTYH loci.

Gene	SNVs	Intronic	5' UTR	In Coding Region			3' UTR
				Silent	Missense	Nonsense	
APC	154	143	2	6	2	0	1
MUTYH	7	5	0	1	1	0	0

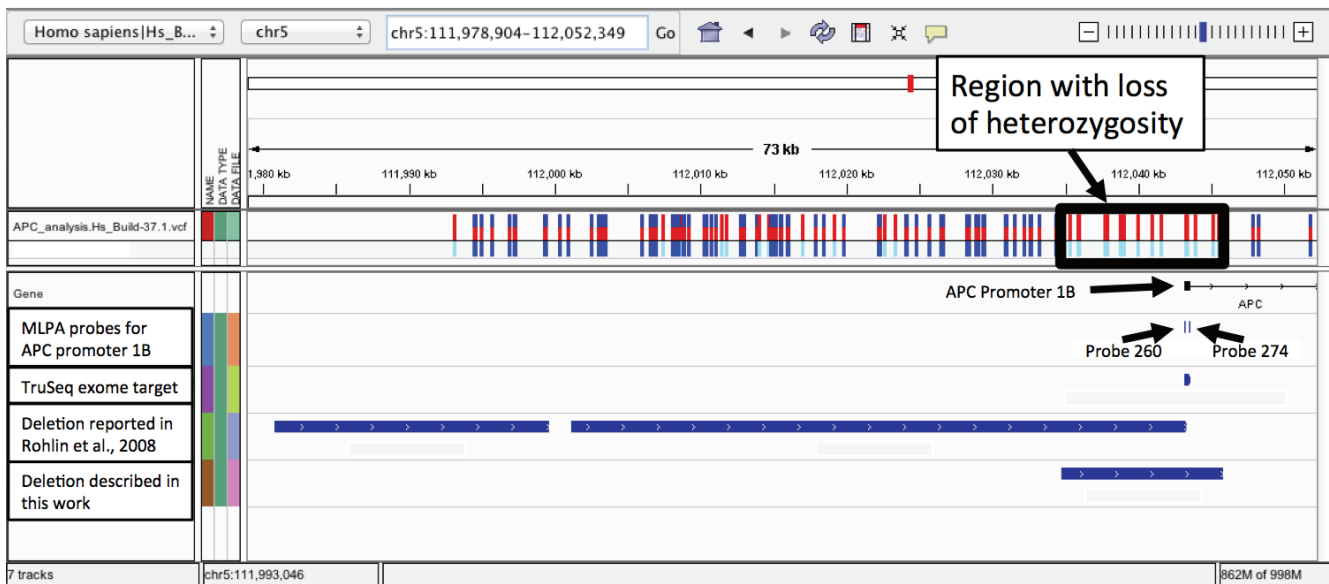
**Table 3.** Insertions and Deletions identified in our patient for the APC and MUTYH loci.

Gene	In/Dels	Intronic	5' UTR	Frame Shifting	3' UTR
APC	26	25	1	0	0
MUTYH	3	3	0	0	0

**Table 4.** Positions of the missense variants detected in the MUTYH and APC loci.

Single Nucleotide Polymorphism database (dbSNP) accession numbers and Human Genome Variation Society (HGVS) names for the gene, including the amino acid change and position are also listed.

Gene	Chr	Coordinate	Reference Allele	Variant Allele	dbSNP	HGVS Names
MUTYH	1	45800156	C	T	rs3219484	NP_001041636.1 Val22Met
APC	5	112175240	G	C	rs1801166	NP_000029.2 Glu1317Gln
APC	5	112176756	T	A	rs459552	NP_000029.2 Val1822Asp



**Figure 3.** A rendering of human chromosome 5 with features between bases 111,978,904 and 112,052,349 for human genome build 37.1 in the Integrative Genomics Viewer. Records from the VCF file for the patient described here are displayed in the top track indicating a region with a loss of heterozygosity consistent with a deletion. We also render the exon identified as APC promoter 1B, the MLPA probes used commercially to analyze this locus, the region selected for pull-down in the TruSeq exome capture kit, the deletion reported by Rohlin *et al.* in 2008, and the position of the deletion described here relative to all these features.



larger dataset, and is made available in its own binary alignment map file for inspection at the DOI included above.

In order to confirm the deletion, PCR primers were designed to specifically interrogate it. These primers produce a product of approximately 1kb for individuals with no deletion, and a second pair of primers was designed that flank the deletion site. This placement produces a product of 0.6kb from chromosomes with the deletion, and 11.7kb in chromosomes without. As the NGS data suggests a heterozygous deletion, the expectation was a single band with the first primer pair, and two bands, one strong from the 0.6kb amplicon, and one weak (if detectable) for the 12kb amplicon. This was confirmed in the gel represented in [Figure 1](#). The ~1kb and .6kb bands were cut from the gel and sequenced using Sanger technology. The trace images are shown for the two different alleles in [Figure 2](#). One read shows the deletion, and the second allele is consistent with the reference. The deletion is confirmed by the Sanger sequence data, and the primers are provided as a definitive Sanger sequencing assay for it. The second PCR image in [Figure 1](#), and third read included in [Figure 2](#) confirmed that our respective kindred shares the same deletion as the seven families reported by Snow *et al.* We predict that all families descend from a common founder.

Although this deletion was identified by visual inspection, the binary alignment map file for the region was analyzed by the application BreakDancer<sup>14</sup> to determine if the deletion could be identified algorithmically from whole genome sequence data. BreakDancer identifies putative deletions by identifying read pairs, clustered by genomic coordinate, that have similar inferred insert sizes which are either much larger or smaller than the standard distribution of insert sizes measured for mapped pairs. Using this algorithm, a deletion was identified on chr5 and was approximated to lie between bases 112,034,793 and 112,045,844, corroborating the finding presented here.

The methods of Snow *et al.* used multiplex ligation-dependent probe amplification (MLPA) assays. These are described in a document from MRC-Holland, available at the time of publication at (<http://www.mlpa.com/WebForms/WebFormDBData.aspx?FileOID=McLO2Mc0V%5Cc%7C>). Information for those probes, including the partial sequence adjacent to the ligation site, as well as the genomic coordinate derived from a BLAT search using the partial sequence information is reproduced in [Table 5](#), and rendered in [Figure 3](#) relative to the deletion identified in this work. These coordinates are contained within the region deleted for this patient, and as such result in a deletion of the signals corresponding to these probes. The next probe in the set, APC 142, which is outside the deleted region, did not indicate a deletion.

**Table 5. Probe sequence and genomic coordinate information for the MLPA probes that interrogate APC Promoter 1B.**

APC Probe	Partial sequence adjacent to the ligation site	Genomic Coordinate
260	GCATTGTAGTCT-TCCCACCTCCCA	chr5:112,043,195–112,043,218
274	TACTTCTGGCCA-CTGGGCGAGCGT	chr5:112,043,549–112,043,572

## Discussion/Conclusion

Several years ago, a female patient of the University of Louisville Weisskopf Child Evaluation Center presented with Familial Adenomatous Polyposis (FAP). Whole genome shotgun sequencing on the Illumina platform revealed a deletion on chromosome 5 between bases 112,034,824 and 112,045,845, fully encompassing promoter 1B of the APC locus. Deletions that include this promoter have been demonstrated to affect the expression of the full length APC transcript.

In other work by Snow *et al.*, a deletion was identified via MLPA that is consistent with the deletion characterized here. An investigation via PCR of their seven kindreds with the primers used in this work establishes that the deletion is identical to the deletion reported here. Furthermore, this deletion is also reported by Lin *et al.*, in three kindred not known to be related to each other, or these families. It is likely that this mutation descends from an ancestor common to each of these reported families.

Exome capture has become a popular tool for mutation screening in clinical genetics. The deletion reported here extends several kilobases beyond the region captured by one of the more popular exome capture products ([Figure 3](#)). This deletion would have been very difficult to identify by exome capture since the only practical measurements that could have been employed would have been read density and loss of heterozygosity in the captured region.

The whole genome sequencing approach taken here produces an information rich dataset capable of resolving large deletions in individuals. These structural variants result in a number hallmarks that are easily detected. Specifically, the loss of heterozygosity over a large region, a collection of read pairs whose mates consistently map much further apart than the majority of the read pairs, and soft trimmed reads all pinpoint the deletion site unequivocally. We have demonstrated that whole genome sequencing is both a sensitive and accurate approach for the detection and characterization of deletions of this size.

## Data availability

*F1000Research*: Dataset 1. Raw Gel electrophoresis image for [Figure 1](#), [10.5256/f1000research.6636.d50276](https://doi.org/10.5256/f1000research.6636.d50276)<sup>15</sup>

## Author contributions

TK and JK conceived of and led the project. TK performed primary and secondary data analyses, and wrote the manuscript. PB and GG served as clinical liaisons. AS and DN performed PCR and background on their sample sets. All provided input during the preparation of the manuscript. All authors have seen and agreed to the final content of the manuscript.

## Competing interests

T.K. serves as the CEO of Intrepid Bioinformatics.

## Grant information

The Next Generation Sequencing work was supported by DOE grant DE-EM0000197 (Kalbfleisch, Rouchka co-PI). Dr. Kalbfleisch received additional financial support from the National Institute

of General Medical Sciences of the National Institutes of Health Grant# P20GM103436 (Cooper PI). University of Utah work is supported by PO1CA073992.

*I confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

## Acknowledgements

The PCR and Sanger Sequencing work were performed in the UofL Center for Genetics and Molecular Biology core facility by Ms. Elizabeth Hudson. The alignment and analysis work for the next generation sequencing data was performed on the University of Louisville Cardinal Research Cluster.

## References

- Rohlin A, Engwall Y, Fritzell K, *et al.*: **Inactivation of promoter 1B of APC causes partial gene silencing: evidence for a significant role of the promoter in regulation and causative of familial adenomatous polyposis.** *Oncogene*. 2011; **30**(50): 4977–89.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Snow AK, Tuohy TM, Sargent NR, *et al.*: **APC promoter 1B deletion in seven American families with familial adenomatous polyposis.** *Clin Genet*. 2014.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kadiyska TK, Todorov TP, Bichev SN, *et al.*: **APC promoter 1B deletion in familial polyposis—implications for mutation-negative families.** *Clin Genet*. 2014; **85**(5): 452–7.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lin Y, Lin S, Baxter MD, *et al.*: **Novel APC promoter and exon 1B deletion and allelic silencing in three mutation-negative classic familial adenomatous polyposis families.** *Genome Med*. 2015; **7**(1): 42.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kerr SE, Thomas CB, Thibodeau SN, *et al.*: **APC germline mutations in individuals being evaluated for familial adenomatous polyposis: a review of the Mayo Clinic experience with 1591 consecutive tests.** *J Mol Diagn*. 2013; **15**(1): 31–43.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics*. 2009; **25**(14): 1754–60.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- McKenna A, Hanna M, Banks E, *et al.*: **The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.** *Genome Res*. 2010; **20**(9): 1297–303.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li H, Handsaker B, Wysoker A, *et al.*: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics*. 2009; **25**(16): 2078–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics*. 2010; **26**(6): 841–2.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- DePristo MA, Banks E, Poplin R, *et al.*: **A framework for variation discovery and genotyping using next-generation DNA sequencing data.** *Nat Genet*. 2011; **43**(5): 491–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Van der Auwera GA, Carneiro MO, Hartl C, *et al.*: **From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.** *Curr Protoc Bioinformatics*. 2013; **11**(1110): 11.10.1–11.10.33.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Robinson JT, Thorvaldsdóttir H, Winckler W, *et al.*: **Integrative genomics viewer.** *Nat Biotechnol*. 2011; **29**(1): 24–6.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kent WJ: **BLAT—the BLAST-like alignment tool.** *Genome Res*. 2002; **12**(4): 656–64.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chen K, Wallis JW, McLellan MD, *et al.*: **BreakDancer: an algorithm for high-resolution mapping of genomic structural variation.** *Nat Methods*. 2009; **6**(9): 677–81.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kalbfleisch T, Brock P, Angela S, *et al.*: **Dataset 1 in: Characterization of an APC Promoter 1B deletion in a Patient Diagnosed with Familial Adenomatous Polyposis via Whole Genome Shotgun Sequencing.** *F1000Research*. 2015.  
[Data Source](#)



# Open Peer Review

Current Referee Status:



Version 1

Referee Report 15 July 2015

doi:10.5256/f1000research.7129.r9224



**Anna Rohlin**

Department of Medical and Clinical Genetics, University of Gothenburg, Gothenburg, Sweden

Kalbfleisch *et al.* present an interesting article on identification and characterization of an 11kb *APC* promoter deletion from whole genome sequencing (WGS) data. The design of the study, the methods used and the presentation of the results are presented in a conclusive and way suitable to the investigation. They have analyzed the WGS data only for the *APC* and *MUTYH* locus and identified the deletion by visual inspection of the data. The identified deletion is confirmed and breakpoints identified with sanger sequencing. The deletion is also found to be identical to the previously identified deletions by Snow *et al.* and Lin *et al.*, and they all descent from a common founder.

## Minor points for revision and comments:

1. When a coding DNA reference sequence is used the following recommendations can be followed according to HGVS; the coding DNA reference sequence should be complete and preferably derived from the *RefSeq database (format NM\_033337.2)*. In general the longest transcripts are used, and for the *MUTYH* gene this is NM\_001128425.1 (16 exon 549 aa). It would be more appropriate to at least name the variants found according to the NM reference sequence that has been used in table 4. For example the variant rs3219484 if the NM\_001128425.1 would have been used the variant would preferable be named **NM\_001128425.1:c.64A>G, p.Val22Met**. For the *APC* gene NM\_000038.5 is usually used. In the tables and text when discussing genomic coordinates, the genome build used would be good too include for example (GRCh37/hg19).
2. When discussing MLPA, it would be appropriate to mentioned the versions of the MLPA kit, like P043 (version C1), this is a common way to present which kit has been used, a link to the document describing the assay is not necessary. Probes for the *APC* promoter 1B were included in the MLPA kit in 2011.
3. In the result section findings of missense and nonsense variants in coding region and small indels are mentioned, what about variants in splice acceptor or donor-sites are these included in the analyses done? It would be nice to mention this in the text and also in the tables (2 and 4) if anyone's are found as these variants are important and often constitute disease-causing mutations.
4. Regarding the deleteriousness of the variants found how was this interpreted? It would be nice to mentioned which databases and/or prediction tools has been used since all variants in dbSNP are not benign. I would be recommended to use several prediction tools like for example SIFT, Polyphen-2, Mutation taster, Condel and Combined Annotation-Dependent Depletion (CADD)

among others for missense interpretation and also looking at conservation between species. Other important tools used for interpretation of variants found are databases where variants are reported and sometimes also classified often in a 1-5 scale (1 is benign and 5 pathogenic). The InSiGHT database (<http://insight-group.org/variants/database/>) is commonly used for variants in coloncancer genes, where the *APC* and *MUTYH* gene can be found among others, ClinVar and HGMD professional (Human Genome Mutation Database) are also databases too use. Information about the minor allele frequency of the variant found can also be found in ExAc (Exome Aggregation Consortium) and ESP (Exome Sequencing Project) and can be included when reporting variants.

5. This deletion was found by manual inspection of the region where the reads map further apart than expected and by looking at the soft-clipped bases as well as identifying the region having LOH. It would be interesting to discuss the limitation with this visual method regarding sizes and type of rearrangement that can be detected. Some discussion comparing (pros and cons) regarding this method with different algorithms/methods like BreakDancer (read-pair methods) and others including read-depth methods and split reads methods for example would also be valuable for an interesting discussion useful for readers trying to find methods to analyze for these types of mutations.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.

Referee Report 08 July 2015

doi:10.5256/f1000research.7129.r9222



**Nicholas Davidson<sup>1</sup>, Yiing Lin<sup>2</sup>**

<sup>1</sup> Division of Gastroenterology, Washington University in St. Louis, St. Louis, MO, USA

<sup>2</sup> Department of Surgery, Washington University in St. Louis, St. Louis, MO, USA

Mutations in the *APC* gene have been established as a cause of classical FAP. However, in a small subset (~20%) of affected families, mutations in the coding regions of *APC* or other polyposis-associated genes such as *MUTYH* cannot be identified. In this study, the authors perform whole-genome sequencing on one such subject and identify a heterozygous 11kb deletion in the exon 1B / promoter region of *APC*. The three non-synonymous mutations identified in *APC* and *MUTYH* were not felt to be causal mutations. The authors confirmed the presence of the heterozygous deletion through differential PCR and Sanger sequencing.

The authors then went on to investigate nine *APC* mutation-negative kindreds which were previously found to have exon 1B / promoter deletions using multiplex ligation-dependent probe amplification assays. In all nine kindreds, the same coordinates of the 11kb deletion were inferred.

The study design, presentation of results and conclusions are straightforward. The authors find an *APC* promoter deletion through whole-genome sequencing and establish the deletion coordinates through Sanger sequencing. Ideally, further investigation of a number of affected and unaffected members in the kindred of this study would provide convincing evidence that this 11kb deletion is associated with the affected state. However, considered in conjunction with the results from Lin *et al.* and Snow *et al.*, this

study shows that whole-genome sequencing is a suitable method for the detection of non-coding mutations in *APC*-mutation negative FAP individuals. This builds upon mounting evidence that associates *APC* promoter deletions with FAP. Interestingly, these studies all show that affected members of eleven FAP kindreds in the United States share the promoter deletion with identical coordinates. It might be worth noting in the discussion that the original Snow *et al* paper using MPLA identified a deletion that was thought to be much larger (>33kb) but that approach did not map the exact coordinates.

Two minor points: (1) In the Results section (p 6 left column second paragraph), it is stated: “these primers produce a product of approximately 1kb for individuals with no deletion.” More accurately, the 1kb product is produced from chromosomes without the deletion (as the authors later state); the 1kb product is also produced in all individuals with the deletion (Fig. 1).

(2) In the discussion, the authors discuss potential difficulties in using targeted capture techniques to discover or assess larger deletions such as the one described. However, we successfully used such a targeted capture assay to discover the promoter deletion described in our study. The analysis methods for detecting individual reads and read pairs that straddle the deletion are applicable and to targeted sequencing strategies.

**We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.

---