Appendices

Table of Contents

List of Figures

List of Tables

Appendix 2.1 Quantitative Polymerase Chain Reaction (qPCR)

This study used Taqman technology (ThermoFisher Scientific) for qPCR. Prior to performing qPCR, the optimal RNA input was determined, and the reaction efficiencies calculated to ensure that the ΔΔCt method (Livak and Schmittgen 2001) could be used for data analysis.

Appendix 2.1.1 Determination of Optimal RNA Input

RNA from a healthy control (2.3.16) was extracted from blood (2.3.3), and its concentration quantified using the Qubit 2.0 Fluorometer (Life Technologies). A series of 10-fold dilutions was performed. Each dilution underwent reverse transcription (RT) in triplicate, following the standard protocol (2.3.3.2). The resultant cDNA was amplified by qPCR using Taqman technology (Applied Biosystems), using the *ACTB* assay (HS99999903_m1). qPCR was performed using standard reagents and reaction conditions (2.3.15). Results were interpreted using the QuantStudio 12K Flex Software (Life Technologies). The dilution curve and amplification plots were generated by the QuantStudio software and can be seen in Figures A2.1 and A2.2 below.

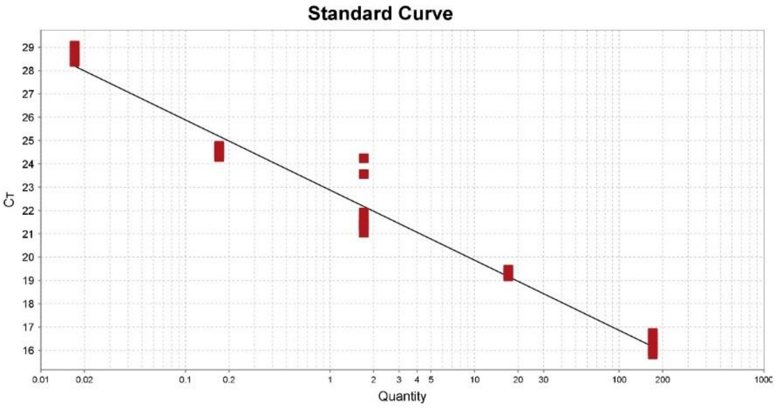


Figure A2.1: RNA dilution curve

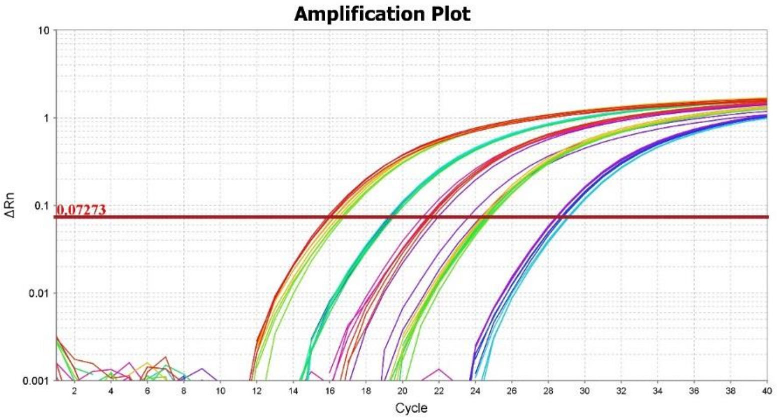


Figure A2.2 The amplification plot illustrates the amplification curves at the variable concentrations of input RNA

The optimum amount of input RNA was determined to be 154ng (the 10-fold dilution of the neat RNA sample), due to its tight amplification plot.

Appendix 2.1.2 qPCR Reaction Efficiencies

This thesis has used the $\Delta\Delta C_T$ method (Livak and Schmittgen 2001) for qPCR analysis. In order for this to be a valid method to use, it was vital that the reaction efficiencies for all the Taqman assays were equivalent. ThermoFisher guarantee that all their assays are suitable for the $\Delta\Delta C_T$ method, but I sought to confirm this claim using three different hydrolysis probes.

The qPCR reaction efficiency was determined using a cDNA dilution curve. The cDNA was from the same healthy control as above (Appendix 2.1.1). The RNA input for RT was 154ng as described above (Appendix 2.1.1). RT was carried out using the standard protocol (2.3.3.2). The resultant cDNA underwent a series of 10-fold dilutions. Each dilution was then subject to qPCR in triplicate. The assays used were *ACTB* HS99999903_m1; *APC* HS01568269_m1; *MUTYH* HS01014856_m1. The standard qPCR reagents and reaction conditions were employed (2.3.15).

The cDNA dilution curves, qPCR reaction efficiencies and correlation coefficients are below (Figures A2.3 and A2.4, Tables A2.1 and A2.2). The qPCR efficiency was determined using Life Technologies qPCR Reaction Efficiency Calculator (https://www.lifetechnologies.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/qpcr-efficiency-calculator.html accessed

30/03/2015). The correlation coefficient was calculated by the QuantStudio software.

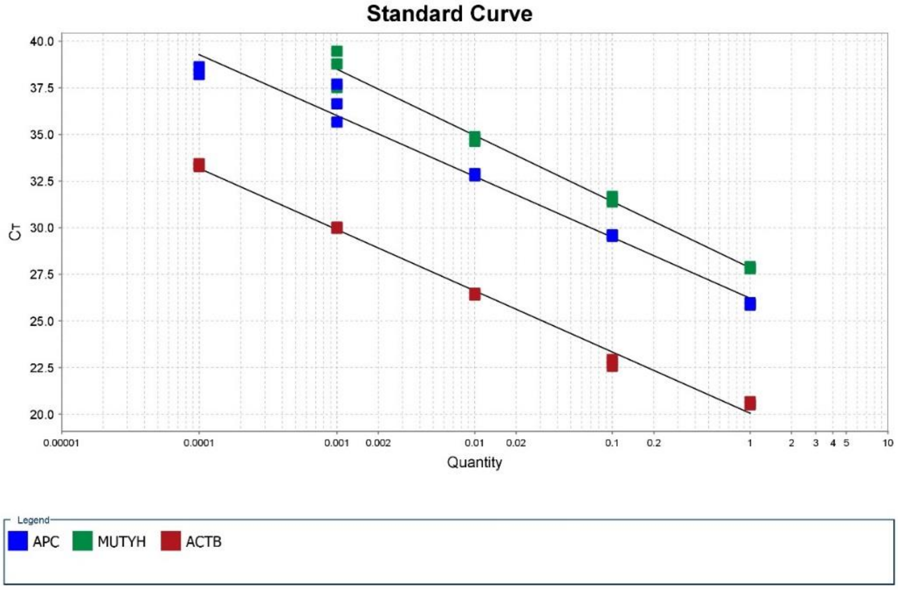


Figure A2.3: cDNA dilution curve

Assay	**Slope of Line**	**Reaction Efficiency**	**Correlation Coefficient**
ACTB	-3.279	101.82%	99.3%
APC	-3.268	102.30%	98.1%
MUTYH	-3.544	91.50%	98.8%

Table A2.1: Reaction efficiencies for 3 Taqman assays

To confirm that the $\Delta\Delta C_T$ method would be valid, graphs of $\log_{10}$ template dilution against ΔC_T were plotted. The linear section of the curve was deemed to lie between dilution points 1, 2 and 3, so these were used in the graphs:

	Neat cDNA	**1:10 dilution**	**1:100 dilution**	**1:1000 dilution**	**1:10000 dilution**
Log_{10} cDNA dilution	**0**	**1**	**2**	**3**	**4**
Mean C_T					
ACTB	20.564	22.749	26.427	30.003	33.331
APC	25.904	29.590	32.854	36.664	38.420
MUTYH	27.849	31.506	34.748	38.581	Undetermined
ΔC_T					
ACTB/APC	-5.34	-6.841	-6.427	-6.661	-5.089
ACTB/ MUTYH	-7.285	-8.757	-8.321	-8.578	N/A

Table A2.2: Table to show the ΔC_T values for different target genes and an endogenous control at different template dilutions

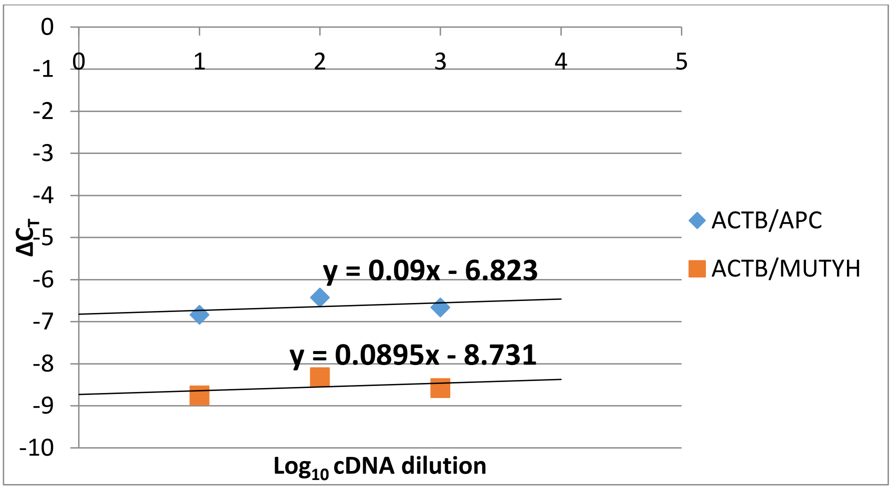


Figure A2.4: Graph to show the ΔCT values for different target genes and an endogenous control at different template dilutions

For the *ACTB/APC* graph, the slope of the line was 0.09. For *ACTB/MUTYH* the slope was 0.0895, meaning that the $\Delta\Delta C_T$ method could be employed.

Appendix 3.1 Primers for *APC* 1A Promoter

	Forward	**Reverse**	**Amplicon Size (bp)**
Methylation-specific primer	TATTGCGGAGTGCGGGTC	TCGACGAACTCCCGACGA	149
Unmethylated-specific primer	GTGTTTTATTGTGGAGTGTGGGTT	CCAATCAACAAACTCCCAACAA	151

Appendix 3.2 Primers for *APC* 1B Promoter

	Forward	**Reverse**	**Amplicon Size (bp)**
Methylation-specific primer	TAGAATAGCGAACGAGTGTTC	TCCGACGACCACACCCCG	190
Unmethylated-specific primer	GATAGAATAGTGAATGAGTGTTT	CTTCCAACAACCACACCCCA	195

Appendix 3.3 *APC* and *MUTYH* SNPs in AI Studies

Gene	**SNP**	**Normal Allele**	**Alternate Allele**	**Exon**
APC	rs2229992	T	C	11
APC	rs351771	G	A	13
APC	rs41115	G	A	15
APC	rs42427	G	A	15
APC	rs866006	T	G	15
APC	rs465899	G	A	15
MUTYH	rs3219489	G	C	12
MUTYH	rs3219497	G	A	16

Appendix 3.4 gDNA Primers for AI

Gene	SNP	Exon	Primer	Primer sequence	Amplicon Size (bp)
APC	rs2229992	11	Forward: APC-Ex11-F	ACCATATATTCTCATTGATTGAGTTTT	396
			Reverse: APC-Ex11-R	CCTGTGCTTCACATTCGCT	
APC	rs351771	13	Forward: APC_13v3_MF	ACCACGGCTAGCCAGAATTT	500
			Reverse: APC_13v3_MR	CAACCTCCGTCTCCCATGT	
APC	rs41115	15	Forward: APC_Ex15Jv2-F	TGCTGAAAAGAGAGAGAGTGGA	379
			Reverse: APC_Ex15Jv2-R	TTATTTCTGCCATGCCAACA	
APC	rs42427	15	Forward: APC_Ex15Lv2-F	CCACGGGTGTATTGTGTTGA	377
			Reverse: APC_Ex15Lv2-R	TCTGCACCCAACAAAAATCA	
APC	rs866006	15	Forward: APC_Ex15Mv2-F	CCCAAAGGGAAAAGTCACAA	347
			Reverse: APC_Ex15Mv2-R	TCACCTCATCATTACACGCC	

APC	rs465899	15	Forward: APC_Ex15O-F	GCAGCCAATAAATCGAGGTC	376
			Reverse: APC_Ex15O-R	TGAAGATGACCTGTTGCAGG	
MUTYH	rs3219489	12	Forward: MUTYH_12_F	AGAAGCCCCATTCCAGTTCT	379
			Reverse: MUTYH_12_R	CCGATTCCCTCCATTCTCTC	
MUTYH	rs3219497	16	Forward: MUTYH_16_F	CGAAACCAGTCTGAGCAACA	397
			Reverse: MUTYH_16_R	TTGAGGCAGGATGAAAGCTC	

Appendix 3.5 Optimisation of gDNA AI

Sample Identification	**SNP(s)**	**Deviation from Standard Protocol. Ta (^{0}C)**
Halo77	rs465899	51.3
Halo75	rs465899	51.3

Appendix 3.6 RNA Primers for AI

Gene	**SNP**	**Primer Name**	**Sequence**	**Amplicon Size (bp)**
APC	rs2229992	APC4_RNA_F	CCTGCTGTGTGTGTTCTAATGA	500
		APC4_RNA_R	GCTGTAGATGGTGCACTTGC	

APC	rs351771	APC4_RNA_F	CCTGCTGTGTGTGTTCTAATGA	500
		APC4_RNA_R	GCTGTAGATGGTGCACTTGC	
APC	rs41115	APC4_RNA_F	CCTGCTGTGTGTGTTCTAATGA	500
		APC4_RNA_R	GCTGTAGATGGTGCACTTGC	
APC	rs42427	APC5_5_RNA_F	GCAACCCCAAAAGCATGTTA	599
		APC5_5_RNA_R	TGAAAAATAATTCCAAGGTCTTCA	
APC	rs866006	APC5_5_RNA_F	GCAACCCCAAAAGCATGTTA	599
		APC5_5_RNA_R	TGAAAAATAATTCCAAGGTCTTCA	
APC	rs465899	APC5_6_RNA_F	CAACAAAGATTCAAAGAAACAGAA	699
		APC5_6_RNA_R	CTCTGAAGATGACCTGTTGCA	
MUTYH	rs3219489	MUTYH_4_F	ATTGCCTCTATCGCCTTTGG	679
		MUTYH_4_R	TGTGAGAGAAGGTGTGGACA	
MUTYH	rs3219497	MUTYH_5_F	GTGGTCAACTTCCCCAGAAAG	588
		MUTYH_5_R	TCACTGGGCTGCACTGTT	

Appendix 3.7 Optimisation of RNA AI (*APC*)

Sample Identification	**SNP(s)**	**Deviation from Standard Protocol. Ta (^{0}C)**
Halo08	rs866006	55.9
Halo17	rs42427 rs866006	55.9
Halo51	rs866006	55.9
Halo53	rs41115 rs42427 rs866006	55.9
Halo62	rs41115	55.9

Appendix 3.8 Optimisation of RNA AI (*MUTYH*)

Sample Identification	**rs3219489**	
	Reagents for cDNA sequencing	**Ta (^{0}C)**
Halo06 **Halo08** **Halo20** **Halo27** **Halo65** **Halo66** **Halo67** **Halo70** **Halo71**	Water 9.5µl F primer (10µM) 1µl R primer (10µM) 1µl MMG 12.5µl cDNA (5ng/µl) 1µl	61
Halo19 **Halo30** **Halo51** **Halo62** **Halo63** **Halo64**	Water 5.5µl F primer (10µM) 1µl R primer (10µM) 1µl MMG 12.5µl cDNA (5ng/µl) 5µl	61

Halo24 **Halo40**	Water 9.5µl F primer (10µM) 1µl R primer (10µM) 1µl MMG 12.5µl cDNA (5ng/µl) 1µl	58
Halo34	Water 9.5µl F primer (10µM) 1µl R primer (10µM) 1µl MMG 12.5µl cDNA (5ng/µl) 1µl	64.8
Halo45	Water 5.5µl F primer (10µM) 1µl R primer (10µM) 1µl MMG 12.5µl cDNA (5ng/µl) 5µl	63.5
Halo53	Water 5.5µl F primer (10µM) 1µl R primer (10µM) 1µl MMG 12.5µl cDNA (5ng/µl) 5µl	57.8

Appendix 3.9 Primers for *APC* cDNA Sequencing

Primer	**Forward**	**Reverse**	**Amplicon Size (bp)**
APC1_RNA	CTTCCCACCTCCCACAAGAT	GGAACAGGACTGCACTCTCC	581
APC2_RNA	TGGCTTCTTCTGGACAGATTG	TCGTGTAGTTGAACCCTGACC	587

APC3_RNA	TCAACATGGCAACTTCTGGT	TCATTCATTGCATGTCTATGCTC	598
APC4_RNA	CCTGCTGTGTGTGTTCTAATGA	GCAAGTGCACCATCTACAGC	500
APC5_1_RNA	TCAGCACATTGCACTGAGAA	TGTTGCTGGATGGTAGTTGC	816
APC5_2_RNA	GAGAACGCGGAATTGGTCTA	AGGCTGATCCACATGACGTT	968
APC5_3_RNA	GCATGAAGAAGAAGAGAGACCAA	AACATGAGTGGGGTCTCCTG	696
APC5_4_RNA	CCAAAAGTGGTGCTCAGACA	TATCATCCCCCGGTGTAAAA	828
APC5_5_RNA	GCAACCCCAAAAGCATGTTA	TGAAGACCTTGGAATTATTTTTCA	599
APC5_6_RNA	CAACAAAGATTCAAAGAAACAGAA	TGCAACAGGTCATCTTCAGAG	699
APC5_7_RNA	CCTGACTCACAGGGAGAACC	AGTGGTGGCTGTTTGACCTT	831

APC5_8_RNA	CAGCCTCCAAAAGCCCTAGT	GAAGGGAAGGACTTAAAACTGGA	699
APC5_9_RNA	ACTAGGTCCCAGGCACAAAC	GCCGAGCTGAAGTGCTATCT	982
APC5_10_RNA	CCCCTGACCAAAAAGGAACT	ACTTGTTCATCGGGAACCAT	844

Appendix 3.10 Optimisation of *APC* cDNA Sequencing

Sample Identification	**Primer Set (*APC*)**	**Deviation from Standard Protocol**		
		Ta (^{0}C) for PCR	**Volume of cDNA (5ng/µl) for PCR**	**Big Dye Reaction**
Halo08	5_5	55.9		
Halo17	3, 5_1			5µl PCR product added
	5_5	55.9		
Halo35	5_2	61		
Halo45	1	61	5µl	
	3, 5_1, 5_2, 5_4, 5_7, 5_9, 5_10			5µl PCR product added
Halo51	5_5	55.9		
Halo52	5_9	61		

Halo53	5_4	55.9		
	5_5	55.9		
Halo62	5_4	55.9		

Appendix 3.11 Primers for *MUTYH* cDNA Sequencing

	Primer Name	**Primer Sequence**	**Amplicon Size (bp)**
MUTYH_1	MUTYH1F	ATGACACCGCTCGTCTCC	244
	MUTYH1R	GGTATGAGGAGACAGAGGCC	
MUTYH_2	3M2F	GGGATGATTGCTGAGTGTCC	384
	2M2R	GCCACGAGAATAGTAGCCCA	
MUTYH_3	MUTYH3F	TGTGGGTCTCAGAGGTCATG	399
	MUTYH3R	AGAGCTGCTGGGAAACAAGG	
MUTYH_4	1_10_14M3F	ATTGCCTCTATCGCCTTTGG	679
	1_10_14M3R	TGTGAGAGAAGGTGTGGACA	
MUTYH_5	MUTYH5F	GTGGTCAACTTCCCCAGAAAG	588
	21_10_M5R	TCACTGGGCTGCACTGTT	

Appendix 3.12 Optimisation of *MUTYH* cDNA sequencing

Unless otherwise stated, those samples which underwent optimisation required 12.5µl of MMG, and 1µl of Forward and Reverse primers (10µM). The volume of water was that required to make a total volume of 25µl.

	Halo04		**Halo06**		**Halo07**	
	Reagents	**Ta (^{0}C)**	**Reagents**	**Ta (^{0}C)**	**Reagents**	**Ta (^{0}C)**
MUTYH_1		54.5	Standard	58	Standard	58.4

	cDNA (5ng/µl) 5µl					
MUTYH_2	cDNA (5ng/µl) 1µl	55.1 and 60.6	cDNA (5ng/µl) 5µl	62.6	cDNA (5ng/µl) 1µl	55.1, 60.6, 63.4
MUTYH_3	Standard	60.0	cDNA (5ng/µl) 1µl	64.9	Standard	58.4
MUTYH_4	cDNA (5ng/µl) 1µl	61.5	cDNA (5ng/µl) 1µl	60.5	cDNA (5ng/µl) 1µl	58
MUTYH_5	cDNA (5ng/µl) 1µl	64.7	Standard	55.9	cDNA (5ng/µl) 1µl	61

	Halo17		**Halo20**		**Halo23**	
	Reagents	**Ta (^{0}C)**	**Reagents**	**Ta (^{0}C)**	**Reagents**	**Ta (^{0}C)**
MUTYH_1	cDNA (5ng/µl) 5µl	64.7	Standard	64.7	cDNA (5ng/µl) 1µl	64.7
MUTYH_2	cDNA (5ng/µl) 5	64.7	cDNA (5ng/µl) 1µl	60.6	Standard	61.5
MUTYH_3	cDNA (5ng/µl) 5µl	64.7	cDNA (5ng/µl) 1µl	61	Standard	60.0
MUTYH_4	cDNA (5ng/µl) 1µl	64.7	cDNA (5ng/µl) 1µl	61	cDNA (5ng/µl) 1µl	58

MUTYH_5	cDNA (5ng/µl) 1µl	64.7	cDNA (5ng/µl) 1µl	61	cDNA (5ng/µl) 1µl	64.7

	Halo24		**Halo25**		**Halo27**	
	Reagents	**Ta (^{0}C)**	**Reagents**	**Ta (^{0}C)**	**Reagents**	**Ta (^{0}C)**
MUTYH_1	cDNA (5ng/µl) 1µl	64.7	Standard	64.7	Standard	64.7
MUTYH_2	Standard	55.1	Standard	62.6	Standard	54.5
MUTYH_3	Standard	55.1	Standard	64.7	cDNA (5ng/µl) 1µl	55.9, 64.9
MUTYH_4	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	61	cDNA (5ng/µl) 1µl	61
MUTYH_5	Standard	55.9	cDNA (5ng/µl) 1µl	61	cDNA (5ng/µl) 1µl	61

	Halo36		**Halo41**		**Halo44**	
	Reagents	**Ta (^{0}C)**	**Reagents**	**Ta (^{0}C)**	**Reagents**	**Ta (^{0}C)**
MUTYH_1	Standard	58	cDNA (5ng/µl) 1µl	55	Standard	58
MUTYH_2	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	55	cDNA (5ng/µl) 5µl	57.8
MUTYH_3	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	55.9

MUTYH_4	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	58
MUTYH_5	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	58

	Halo63		**Halo71**	
	Reagents	**Ta (^{o}C)**	**Reagents**	**Ta (^{o}C)**
MUTYH_1	Standard	58	Standard	58
MUTYH_2	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	58
MUTYH_3	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	58
MUTYH_4	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	58
MUTYH_5	cDNA (5ng/µl) 1µl	58	cDNA (5ng/µl) 1µl	57.8

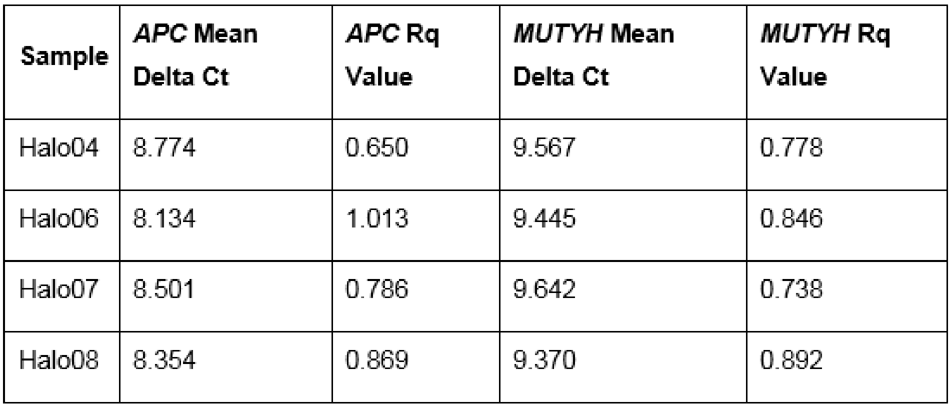
Appendix 3.13 Primers for *APC* and *MUTYH* UDS Validation

Location of Variant	**Forward Primer**	**Reverse Primer**	**Amplicon Size (bp)**
5:112043492	CCCAGGTACTGTTGTTGGCT	GAAGAGAGCCACAGAGTCGG	347
5:112043225	GGACTAGTGTGTGCAGAAGG	CTTCCTCACCAACAGCCAAC	400
5:112095775	TTTAAAAGCTGCCACTTCTCAAA	TTTGCCTTTCCAGTTCAGCC	760

5:112043282	AAGTGCTGCAACTGAGACTC	CTCTTCGTCTCCTGCCGG	498
5:112162474	AGTTAGGCCACAGTGTATGACA	TATGCCAGTTTCTCCACCCC	206
5:112102107	GTGCGTGCTTTGAGAGTGAT	GCTGTACTTGGATCTACACACC	283
5:112102998	TTGTTTTCAGTCATGTATATTTGTGG	CTGGAGTACACAAGGCAATGTT	376
5:112104652	AGTAGCGTAGAACACAGTCCT	AGGCTAGGAAACACAGAGGG	253
5:112177427	GCATCAGGCTATGCTCCTAAA	GCAGCAGCAGCTTGATGTAA	353
5:112090597	TTTCTTTAAAAACAAGCAGCCA	TCTCCATCTTCAGTGCCTCA	144
5:112178658	CATCAAAGAAGCTCCAAGCC	AAACTGGAGTTTGTGCCTGG	122

Appendix 3.14 Results of *APC* and *MUTYH* Gene Transcription Studies

The results highlighted in red indicate those samples with an *APC* or *MUTYH* Rq value <0.6.



Sample	*APC* Mean Delta Ct	*APC* Rq Value	*MUTYH* Mean Delta Ct	*MUTYH* Rq Value
Halo04	8.774	0.650	9.567	0.778
Halo06	8.134	1.013	9.445	0.846
Halo07	8.501	0.786	9.642	0.738
Halo08	8.354	0.869	9.370	0.892

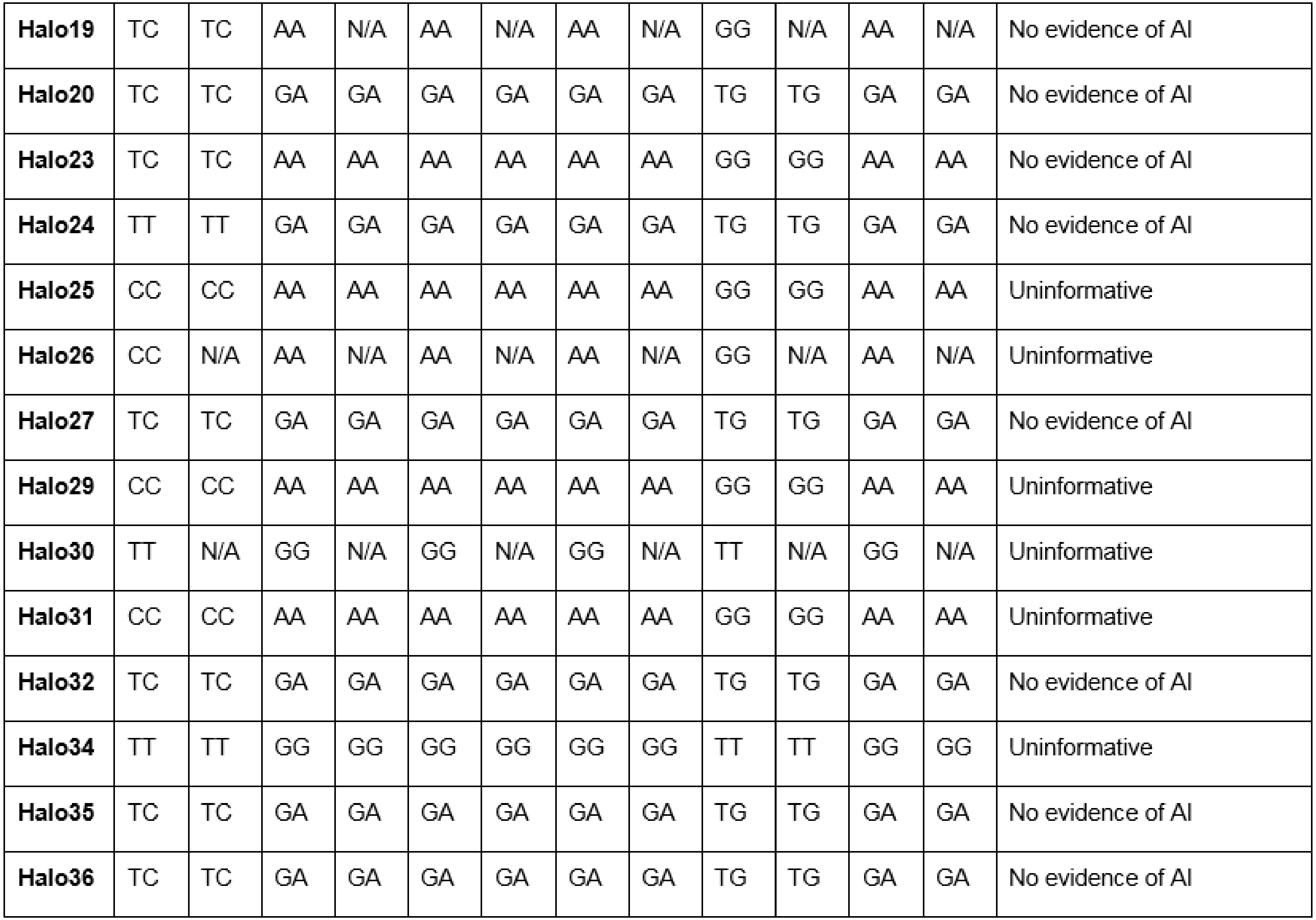
Sample	***APC* Mean Delta Ct**	***APC* Rq Value**	***MUTYH* Mean Delta Ct**	***MUTYH* Rq Value**
Halo13	8.591	0.738	9.439	0.850
Halo14	8.810	0.634	9.361	0.898
Halo17	7.927	1.169	9.546	0.790
Halo19	7.773	1.3	9.673	0.723
Halo20	8.260	0.928	9.114	1.065
Halo23	8.466	0.804	8.882	1.250
Halo24	8.427	0.826	9.837	0.645
Halo25	8.277	0.917	9.519	0.804
Halo26	8.437	0.821	9.665	0.727
Halo27	8.078	1.053	9.040	1.121
Halo29	8.478	0.798	9.445	0.847
Halo30	8.498	0.787	9.765	0.678
Halo31	8.540	0.765	9.899	0.618
Halo32	8.494	0.789	9.454	0.841
Halo34	8.481	0.796	9.155	1.036
Halo35	8.315	0.894	9.560	0.782
Halo36	8.758	0.657	10.493	0.41
Halo40	8.728	0.671	9.753	0.684
Halo41	8.371	0.860	9.117	1.063
Halo44	8.161	0.994	8.683	1.436
Halo45	8.202	0.966	9.551	0.787

Sample	***APC* Mean Delta Ct**	***APC* Rq Value**	***MUTYH* Mean Delta Ct**	***MUTYH* Rq Value**
Halo46	8.981	0.563	9.503	0.813
Halo47	7.070	2.118	7.507	3.245
Halo48	8.328	0.886	9.878	0.627
Halo51	8.428	0.826	9.846	0.641
Halo52	9.594	0.368	9.552	0.786
Halo53	9.028	0.545	8.738	1.382
Halo55	7.776	1.298	9.147	1.041
Halo56	8.487	0.793	9.677	0.721
Halo58	8.276	0.918	9.735	0.693
Halo61	8.561	0.753	9.660	0.729
Halo62	8.537	0.766	9.833	0.647
Halo63	8.829	0.626	10.179	0.509
Halo64	9.415	0.417	9.574	0.774
Halo65	8.209	0.961	8.890	1.244
Halo66	7.969	1.135	9.493	0.819
Halo67	8.289	0.909	9.666	0.726
Halo70	8.569	0.749	9.311	0.929
Halo71	8.296	0.905	10.182	0.508
Halo72	8.171	0.987	9.274	0.954
Halo80	8.699	0.685	9.724	0.698

Appendix 3.15 Results of *APC* AI

The results highlighted in red indicate any sample with evidence of AI.

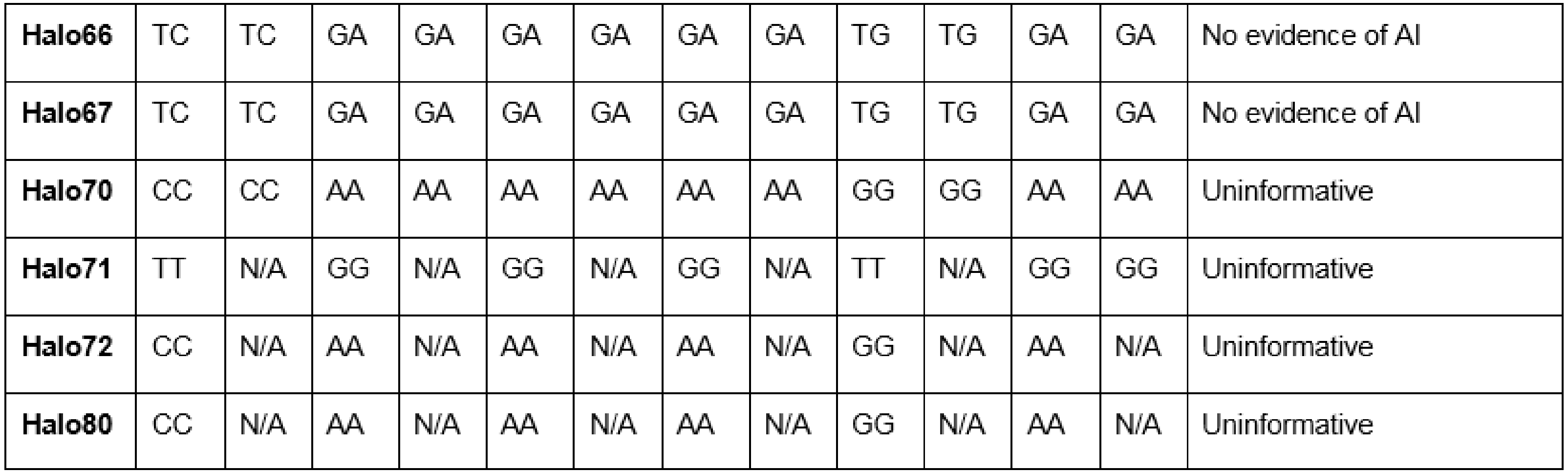
	rs2229992		**rs351771**		**rs41115**		**rs42427**		**rs866006**		**rs465899**		**Conclusion**
	T>C		**G>A**		**G>A**		**G>A**		**T>G**		**G>A**		
Sample ID	**DNA**	**RNA**	**DNA**	**RNA**	**DNA**	**RNA**	**DNA**	**RNA**	**DNA**	**RNA**	**DNA**	**RNA**	
Halo04	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo06	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo07	TT	N/A	GG	N/A	GG	N/A	GG	N/A	TT	N/A	GG	N/A	Uninformative
Halo08	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	A/-	AI at final SNP, with abnormal ratio at some preceding SNPs
Halo13	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo14	CC	CC	AA	AA	AA	AA	AA	AA	GG	GG	AA	AA	Uninformative
Halo17	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI



Halo19	TC	TC	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	No evidence of AI
Halo20	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo23	TC	TC	AA	AA	AA	AA	AA	AA	GG	GG	AA	AA	No evidence of AI
Halo24	TT	TT	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo25	CC	CC	AA	AA	AA	AA	AA	AA	GG	GG	AA	AA	Uninformative
Halo26	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo27	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo29	CC	CC	AA	AA	AA	AA	AA	AA	GG	GG	AA	AA	Uninformative
Halo30	TT	N/A	GG	N/A	GG	N/A	GG	N/A	TT	N/A	GG	N/A	Uninformative
Halo31	CC	CC	AA	AA	AA	AA	AA	AA	GG	GG	AA	AA	Uninformative
Halo32	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo34	TT	TT	GG	GG	GG	GG	GG	GG	TT	TT	GG	GG	Uninformative
Halo35	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo36	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI

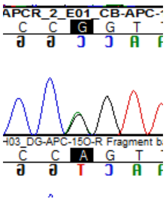
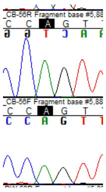
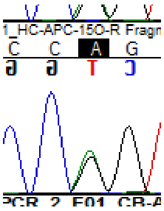
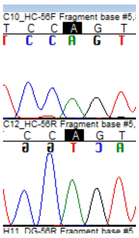
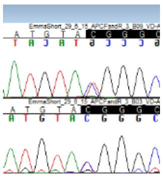
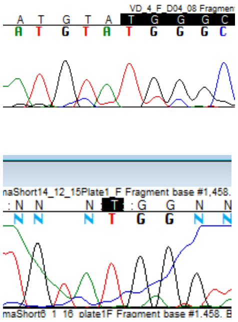
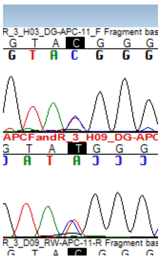
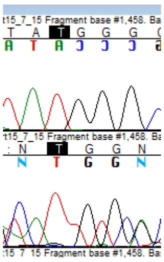
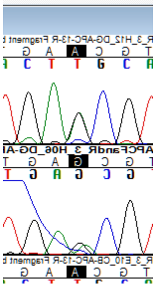
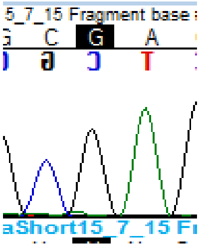
Halo40	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo41	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo44	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo45	TT	N/A	GG	N/A	GG	N/A	GG	N/A	TT	N/A	GG	N/A	Uninformative
Halo46	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo47	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo48	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo51	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	A/-	AI at final SNP, with one allele having a weaker signal at some preceding SNPs
Halo77 (Halo51 mother)	TT	N/A	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo78 (Halo51 father)	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI

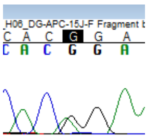
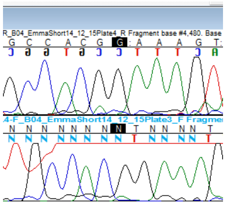
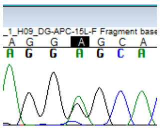
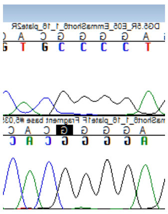
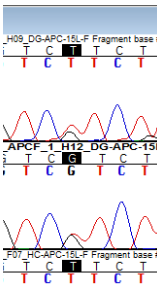
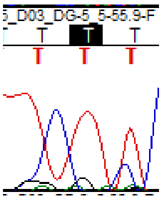
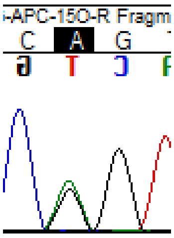
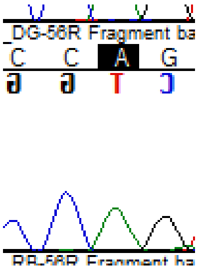
Halo52	TC	T/-	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	AI at first SNP
Halo53	TC	T/-	GA	G/-	GA	G/-	GA	G/-	TG	T/-	GA	A/-	AI throughout
Halo75 (Halo53 mother)	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo76 (Halo53 father)	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo55	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo56	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo58	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo61	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo62	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo63	CC	CC	AA	AA	AA	AA	AA	AA	GG	GG	AA	AA	Uninformative
Halo64	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo65	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	N/A	GA	No evidence of AI



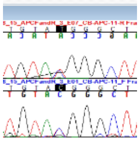
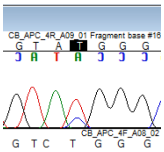
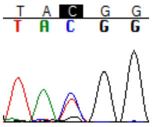
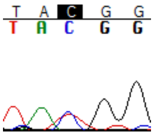
Halo66	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo67	TC	TC	GA	GA	GA	GA	GA	GA	TG	TG	GA	GA	No evidence of AI
Halo70	CC	CC	AA	AA	AA	AA	AA	AA	GG	GG	AA	AA	Uninformative
Halo71	TT	N/A	GG	N/A	GG	N/A	GG	N/A	TT	N/A	GG	GG	Uninformative
Halo72	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative
Halo80	CC	N/A	AA	N/A	AA	N/A	AA	N/A	GG	N/A	AA	N/A	Uninformative

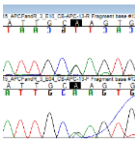
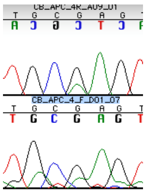
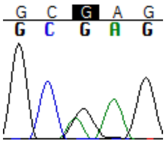
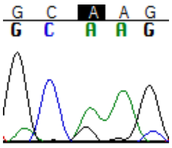
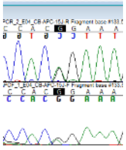
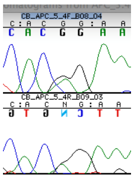
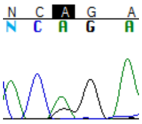
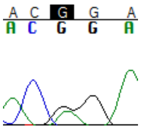
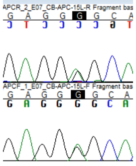
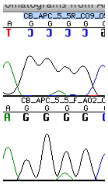
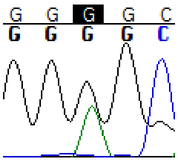
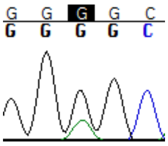
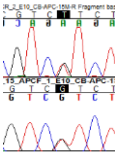
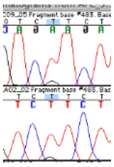
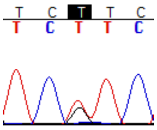
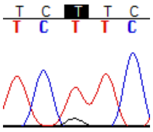
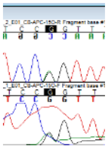
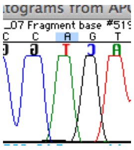
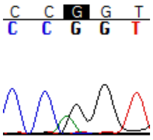
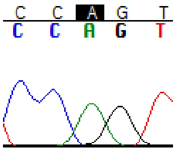
Appendix 3.16 Sequencing Results for *APC* SNPs Exhibiting AI

Sample	Gene	SNP	Heterozygous State	gDNA	cDNA
Halo08	*APC*	rs465899	GA		
Halo51	*APC*	rs465899	GA		
Halo52	*APC*	rs2229992	TC		
Halo53	*APC*	rs2229992	TC		
	APC	rs351771	GA		

	APC	rs41115	GA		
	APC	rs42427	GA		
	APC	rs866006	TG		
	APC	rs465899	GA		

Appendix 3.17 Sequencing Results for Halo08 and Halo51

SNP	**Halo08**		**Halo51**	
	DNA	**cDNA**	**DNA**	**cDNA**
rs2229992	TC 	TC Reduced C 	TC 	TC 

rs351771	GA 	GA Reduced A 	GA 	GA Reduced G 
rs41115	GA 	GA 	GA 	GA 
rs42427	GA 	GA Reduced A 	GA 	GA Reduced A 
rs866006	TG 	TG Reduced G 	TG 	TG Reduced G 
rs465899	GA 	A/- 	GA 	A/- 

Appendix 3.18 Results for *MUTYH* AI

	rs3219497		**rs3219489**		
	G/A		**G/C**		
Sample ID	**DNA**	**RNA**	**DNA**	**RNA**	**Conclusion**
Halo04	GG	GG	GG	GG	Uninformative

Halo06	GG	GG	GC	GC	No evidence of AI
Halo07	GG	GG	GG	GG	Uninformative
Halo08	GG	N/A	GC	GC	No evidence of AI
Halo13	GG	N/A	GG	N/A	Uninformative
Halo14	GG	N/A	GG	N/A	Uninformative
Halo17	GG	GG	GG	GG	Uninformative
Halo19	GG	N/A	GC	GC	No evidence of AI
Halo20	GG	GG	GC	GC	No evidence of AI
Halo23	GG	GG	GG	GG	Uninformative
Halo24	GG	GG	GC	GC	No evidence of AI
Halo25	GG	GG	CC	CC	Uninformative
Halo26	GG	N/A	GG	N/A	Uninformative
Halo27	GG	GG	GC	GC	No evidence of AI
Halo29	GG	N/A	GG	N/A	Uninformative
Halo30	GG	N/A	GC	GC	No evidence of AI
Halo31	GG	N/A	GG	N/A	Uninformative
Halo32	GG	N/A	GG	N/A	Uninformative
Halo34	GG	N/A	GC	GC	No evidence of AI
Halo35	GG	N/A	GG	N/A	Uninformative
Halo36	GG	N/A	GG	N/A	Uninformative
Halo40	GG	N/A	GC	GC	No evidence of AI
Halo41	GG	N/A	GG	N/A	Uninformative
Halo44	GG	N/A	GC	GC	No evidence of AI
Halo45	GG	N/A	GC	GC	No evidence of AI

Halo46	GG	N/A	GG	N/A	Uninformative
Halo47	GG	N/A	GG	N/A	Uninformative
Halo48	GG	N/A	GG	N/A	Uninformative
Halo51	GG	N/A	GC	GC	No evidence of AI
Halo52	GG	N/A	GG	N/A	Uninformative
Halo53	GG	N/A	GC	GC	No evidence of AI
Halo55	GG	N/A	GG	N/A	Uninformative
Halo56	GG	N/A	GG	N/A	Uninformative
Halo58	GG	N/A	GG	N/A	Uninformative
Halo61	GG	N/A	GG	N/A	Uninformative
Halo62	GG	N/A	GG	GG	No evidence of AI
Halo63	GG	N/A	GG	GG	Uninformative
Halo64	GG	N/A	GC	GC	No evidence of AI
Halo65	GG	N/A	GG	GG	Uninformative
Halo66	GG	N/A	GG	GG	Uninformative
Halo67	GG	N/A	GC	GC	No evidence of AI
Halo70	GG	N/A	GG	GG	Uninformative
Halo71	GG	N/A	GG	GG	Uninformative
Halo72	GG	N/A	GG	N/A	Uninformative
Halo80	GG	N/A	CC	N/A	Uninformative

Appendix 3.19 *APC* Coverage with UDS

Haloplex ID	**Min. reads**	**Max. reads**	**Mean reads**	**Median reads**	**0x (%)**	**1x (%)**	**10x (%)**	**100x (%)**	**1000x (%)**	**2000x (%)**	**3000x (%)**
Halo01	0	7326	723	574	2.9	97.1	95.9	88.6	26.5	4.3	0.4
Halo02	0	42381	5928	4598	2.1	97.9	97.2	96	86.8	75.6	65.9
Halo03	0	4010	664	536	2.7	97.3	96	88.1	22.7	2.8	0.2
Halo04	0	1674	258	190	3.4	96.6	94	70.7	1.4	0	0
Halo05	0	3276	477	346	3	97	95.6	81.7	12.1	1.2	0
Halo06	0	5629	771	550	3	97	96	86.9	28.8	7.1	1.5
Halo07	0	2641	295	202	3.3	96.7	93.9	70.7	3.9	0	0
Halo08	0	2203	306	219	3	97	94	73.5	3.5	0	0
Halo09	0	2715	366	256	3	97	94.5	76.3	6.5	0.3	0
Halo10	0	2670	328	231	2.8	97.2	94.7	74.7	4.8	0.1	0
Halo11	0	2172	264	185	3.2	96.8	94.1	69.4	2.1	0	0
Halo12	0	1390	182	125	3.3	96.7	92.2	58.1	0.3	0	0
Halo13	0	4121	625	490	3	97	95.8	86.5	20.5	2.6	0.1
Halo14	0	1658	230	160	3.2	96.8	93.1	65.7	1.3	0	0
Halo15	0	15172	2109	1480	2.5	97.5	96.7	93.2	62.8	39.5	24.8
Halo16	0	1324	236	183	3	97	94.9	71.1	0.3	0	0
Halo17	0	5434	819	643	2.8	97.2	96.3	90.8	30.8	7	0.8
Halo11	0	5457	881	692	2.6	97.4	96.5	91.2	34.2	8.9	1.3

Halo19	0	3345	474	377	2.9	97.1	95.6	85.1	10.8	0.3	0
Halo20	0	4557	628	501	3	97	96	88.6	20.8	1.7	0.1
Halo22	0	10766	2277	2010	2.7	97.3	96.7	94.2	74.5	50.2	30
Halo23	0	16724	1763	1203	2.9	97.1	95.9	89.9	55.9	33.2	19.4
Halo24	0	3846	716	596	2.8	97.2	96.4	90.8	25.9	3	0.2
Halo25	0	4350	806	659	2.6	97.4	96.5	90.6	31.3	5.9	0.5
Halo26	0	4515	884	727	2.7	97.3	96.3	91.3	35.7	8	1.1
Halo27	0	5115	820	668	2.8	97.2	96.5	91.2	31.4	6	0.5
Halo28	0	3111	520	419	3	97	95.4	86.4	13.3	0.5	0
Halo29	0	7058	1236	972	2.6	97.4	96.4	91.6	49	20.1	7.3
Halo30	0	50401	8389	6734	2.1	97.9	97	96.4	90.3	83.7	76.4
Halo31	0	4598	842	685	2.7	97.3	96.4	90.7	32.9	6.9	0.8
Halo32	0	2650	466	373	2.9	97.1	95.9	85	9.7	0.3	0
Halo33	0	1446	235	192	3.2	96.8	94.3	72.2	0.3	0	0
Halo34	0	2996	476	384	2.9	97.1	95.9	85.4	10.2	0.3	0
Halo35	0	3642	607	482	2.8	97.2	96.1	87.6	19.5	1.8	0.1
Halo36	0	25880	4653	3746	2.4	97.6	96.9	96.1	85.5	72.4	59.4
Halo37	0	18619	2739	2174	2.6	97.4	96.8	94.6	74.6	53.9	36.4
Halo38 (same sample as Halo33)	0	4574	817	655	2.8	97.2	96.2	89.8	32.1	6.6	0.7
Halo39	0	2849	409	335	3.1	96.9	95.1	83.2	6.4	0	0
Halo40	0	9148	1404	1063	3	97	95.4	88.1	52.1	26.6	11.9

Halo41	0	36595	5928	4730	2	98	97	95.8	87.8	77.4	67.2
Halo42	0	2528	408	319	2.9	97.1	95.6	82.2	7.2	0.1	0
Halo43	0	6970	1013	812	2.7	97.3	96.6	91.5	41.1	12.4	2.6
Halo44	0	3919	601	488	2.8	97.2	96.1	87.9	18.4	1.4	0.1
Halo45	0	4295	906	825	3	97	96.2	90.4	40.1	5.8	0.3
Halo46	0	17407	2458	2005	2.3	97.7	96.8	94.6	73.4	50.1	32
Halo47	0	3162	477	372	3	97	95.3	82.7	11.4	0.4	0
Halo48	0	10914	2188	1907	2.7	97.3	96.5	93	72.6	47.6	28
Halo49	0	8702	1680	1544	2.7	97.3	96.6	93.6	68.6	35.6	13.5
Halo50	0	12430	1882	1447	3.3	96.7	95.6	88	60.2	38.5	22.4
Halo51	0	12399	1887	1287	3.4	96.6	95.6	87.3	57	36.5	23.3
Halo52	0	17027	2342	1633	2.8	97.2	95.8	89.5	63.3	43.5	29.4
Halo53	0	10476	1911	1640	2.9	97.1	96.2	91.4	68.3	41.6	21.9
Halo54	0	10038	1765	1482	2.8	97.2	96.3	92	64.8	36.5	18.1
Halo55	0	7887	1622	1398	2.8	97.2	96.6	91.9	63.3	33	14.3
Halo56	0	10550	2378	2103	2.7	97.3	96.7	94.7	76.4	52.1	31.5
Halo57	0	13493	2701	2317	2.5	97.5	96.8	94.5	77.1	56.5	37.8
Halo58	0	10885	2349	2110	2.7	97.3	96.7	94.2	75.8	52.2	31.1
Halo59	0	11045	2382	2100	2.8	97.2	96.9	94.4	75.3	52.6	32.4
Halo61	0	9568	2041	1746	2.7	97.3	96.7	93.5	70.2	44	24.8
Halo62	0	3835	775	678	3	97	95.5	88.6	30.6	3.6	0.2
Halo63	0	6959	1282	1097	2.9	97.1	96	90.4	54.2	22.4	6.4
Halo64	0	17690	3418	2855	2.6	97.4	96.4	93.5	79	63.1	48.1

Halo65	0	16879	3870	3548	2.4	97.6	96.8	95.3	84.8	72.1	57.8
Halo66	0	8857	1893	1639	2.7	97.3	96.2	92.4	68.9	40.6	21
Halo67	0	22867	4601	3949	2.7	97.3	96.5	94.5	83.8	72.7	60.3
Halo68	0	15566	2827	2202	2.8	97.2	96.2	92.2	73.1	53.7	38.1
Halo69	0	20422	3806	3115	2.5	97.5	96.3	93.8	79.6	64.9	51.4
Halo70	0	4692	849	715	3.2	96.8	95	86.6	35.3	7.4	0.8
Halo71	0	28764	5110	4091	2.7	97.3	96.5	94.3	81.9	71.3	59.9
Halo72	0	3283	626	535	3.1	96.9	94.5	86.2	20.9	1.3	0
Halo73	0	12970	2079	1553	2.9	97.1	95.9	91.1	63.1	40.9	25.8
Halo74	0	11446	1770	1296	3	97	96	90.8	58.8	34.5	19.3
Halo76	0	29960	3178	2007	2.7	97.3	96.4	92.9	69.5	50.2	37.1
Halo77	0	9663	1055	634	3.1	96.9	95.5	85.6	36.3	15.9	7.6
Halo78	0	4908	1051	940	3	97	96.1	91.4	46.9	11.8	1.6
Halo79	0	5223	613	415	3.1	96.9	94.8	81.7	20.6	4.4	0.8
Halo80	0	19145	3899	3474	2.6	97.4	96.8	95	84.3	71	56.6

Appendix 3.20 *MUTYH* Coverage with UDS

Haloplex ID	Min. reads	Max. reads	Mean reads	Median reads	0x (%)	1x (%)	10x (%)	100x (%)	1000x (%)	2000x (%)	3000x (%)
Halo01	0	16968	982	746	2.8	97.2	96.7	92.9	38.4	9.6	2.4
Halo02	0	99444	9281	7504	1.5	98.5	96.2	95.5	91.4	84.9	77.7
Halo03	0	9019	972	788	2.7	97.3	96.7	93.9	40.5	9.7	2.3
Halo04	0	2835	388	305	3.8	96.2	95.3	79.7	5.5	0.2	0
Halo05	0	5006	827	663	2.8	97.2	96.3	90.1	33.1	5.8	1.4
Halo06	0	9577	1391	1073	2.8	97.2	97	93.3	52.6	24	9.7
Halo07	0	3572	520	385	2.8	97.2	96.3	82	16.1	1.7	0.2
Halo08	0	4548	542	414	2.7	97.3	96.3	86	15	1.7	0.2
Halo09	0	4684	661	491	2.8	97.2	96.4	86.6	23.1	3.9	0.4
Halo10	0	4438	676	509	2.7	97.3	96.4	90.3	22.6	3.8	0.7
Halo11	0	3828	451	331	2.8	97.2	96.3	82.5	9.5	0.4	0.2
Halo12	0	2154	346	262	3	97	96.2	78.8	3.9	0.2	0
Halo13	0	10062	998	817	2.8	97.2	96.9	93.1	42.1	11	1.9
Halo14	0	2957	470	361	2.9	97.1	96.3	84.3	10.6	0.5	0
Halo15	0	24576	4139	3231	2.5	97.5	97.2	96.2	80.4	67.5	52.2
Halo16	0	3105	367	292	2.8	97.2	96.5	81.6	4.1	0.2	0.1
Halo17	0	8691	1055	854	2.8	97.2	96.6	94.2	42.2	14.1	4.1
Halo18	0	12096	1374	1135	2.7	97.3	97.1	95.4	54.4	22.5	9.1
Halo19	0	6821	628	517	2.8	97.2	96.6	90.7	19.6	1.8	0.4

Halo20	0	10709	911	736	2.8	97.2	97.1	93.4	34.7	7.3	1.7
Halo22	0	9587	2699	2386	2.8	97.2	97.1	96.3	81.7	59.4	38.9
Halo23	0	15466	3517	2711	3.2	96.8	96.1	94.7	78	59.8	45.6
Halo24	0	7965	989	828	2.8	97.2	97.1	94.8	43	9.5	2.4
Halo25	0	10143	1159	927	2.8	97.2	96.7	95.5	47.4	16.1	4.1
Halo26	0	10447	1258	1046	2.8	97.2	96.6	95.4	51.8	18.4	5.7
Halo27	0	11381	1168	985	2.2	97.8	97.2	95.7	49.2	16.6	4.9
Halo28	0	6620	650	528	2.8	97.2	96.4	89.9	21	2.3	0.4
Halo29	0	15951	1988	1654	2.5	97.5	96.1	95.2	67.5	41.1	22.7
Halo30	0	92335	12917	10725	1.5	98.5	97.2	97.1	95.9	90.1	85.5
Halo31	0	11237	1208	1014	2.8	97.2	97.1	95.9	50.4	17.4	3.9
Halo32	0	5754	708	597	2.8	97.2	96.6	91.5	24.6	3	0.4
Halo33	0	3153	354	292	3.6	96.4	95.3	83.4	2.8	0.2	0.2
Halo34	0	7554	719	588	2.8	97.2	96.6	91.4	25.3	3.3	0.4
Halo35	0	8504	902	734	2.4	97.6	95.5	92.2	34.7	6.9	1.6
Halo36	0	65078	6726	5543	2	98	96.8	96	90.1	81.4	72.7
Halo37	0	42653	3994	3267	2.4	97.6	97.2	96.4	84.1	68.8	53.9
Halo38 (same sample as Halo033)	0	11213	1169	1000	3.1	96.9	95.5	93.7	50	15.9	4.2
Halo39	0	6560	638	528	2.8	97.2	96.9	90.3	19.9	1.7	0.4
Halo40	0	11303	2443	2096	2.8	97.2	97.2	94.7	74.9	52.3	33
Halo41	0	85725	7636	6043	1.8	98.2	97.2	96.6	91.4	83.4	74.8

Halo42	0	5770	604	487	2.8	97.2	96.6	88.9	19.4	1.7	0.2
Halo43	0	16591	1541	1252	2.8	97.2	97.2	95.9	59.4	28.5	11
Halo44	0	8613	901	727	2.8	97.2	96.6	93.5	36	6.4	1.5
Halo45	0	4334	1178	1060	2.8	97.2	97.1	95.2	54.2	12.9	1.9
Halo46	0	40330	3539	2917	2.2	97.8	97.1	96.6	81.4	65.2	49.5
Halo47	0	7628	755	615	3.3	96.7	95.5	90.3	26.8	4.1	0.6
Halo48	0	12727	3354	3107	3.1	96.9	96.1	95.2	86.5	69.7	51.1
Halo49	0	7098	1787	1627	2.3	97.7	97.2	96.3	73.2	38.7	14.3
Halo50	0	16551	3697	3147	1.4	98.6	96.5	95.3	86.9	70.9	52.9
Halo51	0	15447	3618	3038	2.7	97.3	96.5	94.9	82.5	64.7	50.8
Halo52	0	23847	4408	3587	2.8	97.2	96.4	95.6	85.2	69.7	55.6
Halo53	0	11679	2908	2646	2.8	97.2	97.2	95.8	85.2	63.2	41.5
Halo54	0	11315	2582	2344	2.8	97.2	97.2	96.2	81.8	57.8	34.3
Halo55	0	8985	2434	2221	2.8	97.2	97.2	96.1	80	57.3	33.2
Halo56	0	12406	3279	2976	2.5	97.5	97.2	96.4	85.9	71.1	49
Halo57	0	17287	4239	3860	2.2	97.8	97.2	96.9	88.4	76.9	64.8
Halo58	0	14405	3846	3553	2.5	97.5	97	95.9	89.3	76.5	60.1
Halo59	0	12597	3387	3226	2.7	97.3	97.2	96.9	84.9	69.4	54.3
Halo61	0	11441	3095	2831	2.2	97.8	97.2	96.9	83.4	67.4	46.6
Halo62	0	4262	1354	1238	2.8	97.2	97	95.5	61.3	22.2	3.2
Halo63	0	7278	1803	1655	2.6	97.4	95.5	94.7	73.1	40.1	14.9
Halo64	0	20947	5170	4490	2.7	97.3	97.1	96.4	89.9	81	70.1
Halo65	0	17526	5615	5393	2.5	97.5	96.1	96	92.8	86.1	75.8

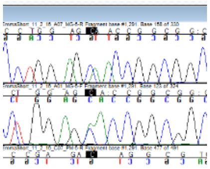
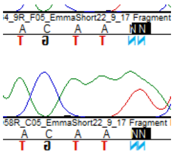
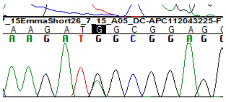
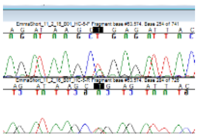
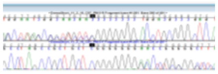
Halo66	0	10181	2795	2530	3.3	96.7	96.1	95.2	84.1	61	41.3
Halo67	0	25114	6585	5767	2.8	97.2	97.2	96.3	92	86.1	78.1
Halo68	0	15964	4228	3630	2.8	97.2	97.2	96.2	86.4	72.7	58.8
Halo69	0	18551	5126	4461	2.8	97.2	97.1	96.2	89.5	78.7	69.1
Halo70	0	4804	1332	1202	2.8	97.2	95.7	94.2	59.2	21.3	4.5
Halo71	0	30259	7685	6568	1.8	98.2	97.2	96.2	91.3	85.2	79.3
Halo72	0	3667	961	889	2.8	97.2	96.3	94.6	42.7	6.4	0.2
Halo73	0	17999	3890	3481	2.8	97.2	96.6	95.8	85.7	70.4	55.6
Halo74	0	21707	3474	2949	2.7	97.3	96.1	95.1	84.3	67.3	49.7
Halo76	0	36609	5364	4038	2.4	97.6	97.2	96.3	85.7	74.9	62
Halo77	0	14129	2399	1887	2.8	97.2	96.7	95.2	69.2	47.2	30.6
Halo78	0	5274	1802	1659	2.8	97.2	97.2	96.3	73.9	40.7	14.4
Halo79	0	5030	1337	1192	2.8	97.2	96.4	92.3	57.6	23.8	5
Halo80	0	21305	6395	5933	1.3	98.7	97.2	96.3	91.7	86.1	77.2

Appendix 3.21 *APC* and *MUTYH* Variants for Validation

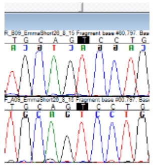
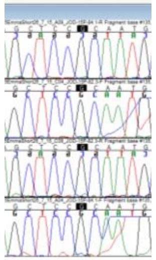
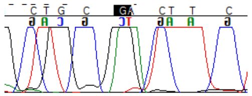
Sample ID	**Gene**	**Chr.**	**Position**	**Variant**	**Location**	**dbSNP ID**	**Variant Frequency (% of reads)**	**Effect**	**Additional Information**	**CADD Score**
Halo23	*APC*	5	112043492	C>A	5’UTR/ Exon1	rs113782655	2257/4689 (48%)	N/A	None	21.1
Halo25	*APC*	5	112102998	T>G	Exonic	None	13/857 (1.52%)	Missense	None	20.6
Halo30	*APC*	5	112102960	C>T	Exonic	rs139196838	2141/5656 (37.9%)	Missense	Previously validated by NHS	34
Halo31	*APC*	5	112177427	G>T	Exonic	None	8/635 (1.26%)	Missense	None	26.8
Halo35	*APC*	5	112104652	T>A	Intronic	None	4/17 (23.5%)	N/A	None	18.02
Halo45	*APC*	5	112090597	G>T	Exonic	None	40/945 (4%)	Missense	None	24.2
Halo46	*APC*	5	112043225	G>A	5'UTR	None	2845/5546 (51.30%)	N/A	None	22.4

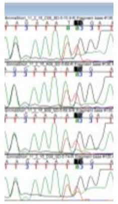
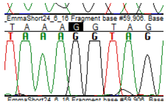
Sample ID	**Gene**	**Chr.**	**Position**	**Variant**	**Location**	**dbSNP ID**	**Variant Frequency (% of reads)**	**Effect**	**Additional Information**	**CADD Score**
Halo51	*APC*	5	112095775	T>A	Intronic	rs4705624	2575/5124 (50%)	N/A	None	16.36
Halo55	*APC*	5	112178658	T>G	Exonic	None	28/2697 (1%)	Missense	None	25.3
Halo56	*APC*	5	112043282	C>CG	5’UTR	None	615/1292 (48%)	N/A	None	16.52
Halo62	*APC*	5	112162474	C>T	Intron	None	472/953 (50%)	N/A	None	18.98
Halo63	*APC*	5	112102107	G>T	Exonic	None	2/11 (18%)	Stop gain	None	37
Halo66	*APC*	5	112163697	A>C	Exonic	None	502/991 (51%)	Missense	Previously identified by NHS	25.4
Halo71	*APC*	5	112102960	C>T	Exonic	rs139196838	369/879 (40.8%)	Missense	Previously validated by NHS	34
Halo40	*MUTYH*	1	45797851	C>T	Exonic	rs140156029	1551/3105 (50%)	Missense	Previously identified by NHS. Patient only has one pathogenic mutation	20.4

Appendix 3.22 UDS Validation Sequencing Results

Sample ID	Gene	Chr.	Position	Variant	Validated
Halo23	*APC*	5	112043492	C>A	Yes 
Halo35	*APC*	5	112104652	T>A	Yes. 
Halo46	*APC*	5	112043225	G>A	Yes 
Halo51	*APC*	5	112095775	T>A	Yes 
Halo56	*APC*	5	112043282	C>CG	Yes 

Appendix 3.23 UDS Validation Sequencing Results: COLD-PCR

Sample	Gene	Location	Variant	Effect	Protocol for COLD PCR	Frequency	Denat. Temp (°C)	Validated
Halo25	*APC*	5:112102998	T>G	Tm increasing	Full	1.52% (13/857)	83.0	No 
Halo31	*APC*	5:112177427	G>T	Tm reducing	Fast	1.26% (8/635)	82.3	No 
Halo45	*APC*	5:112090597	G>T	Tm reducing	Fast	4% (40/945)	84.8	Possibly 

Sample	**Gene**	**Location**	**Variant**	**Effect**	**Protocol for COLD PCR**	**Frequency**	**Denat. Temp (°C)**	**Validated**
Halo55	*APC*	5:112178658	T>G	Tm increasing	Full	1% (28/2697)	As low as 68.0*	No 
Halo63	*APC*	5:112102107	G>T	Tm reducing	Fast	18% (2/11)	81.0	No 

Appendix 4.1 Primers for *NTHL1* c.268A>T Screening

Mutation	**Forward Primer**	**Reverse Primer**	**Amplicon Size (bp)**
NTHL1 c.268A>T	CCCCTACCCTACCTTCACCT	AAAAGCCACCGGGTAGAAAG	374

Appendix 4.2 Primers for Targeted Exome Validation

Location of Variant	**Forward Primer**	**Reverse Primer**	**Amplicon Size (bp)**
2:47612347	CATGAGCCACTCCACTGG	AGGAAGGTAGAGTGATTAGGAGT	294
2:47639633	CATCATATCAGTGTCTTGCACA	CAAAAGTGAAACTCTGTCTCAAAAA	470
2:47639637	AAAGAGGAGGAATTCTGATCACA	ATCCATGTACCTGATTCTCCATT	150
2:47630458	AACAGCTTAGTGGGTGTGGG	CACTGGAGAGGCTGCTCAC	371
2:48010592	AGCTTCTTCCCCAAGTCTCC	TCATTCAAGCCAACTCTGCG	384
2:48026228	AGCTTTCTCTGCCCCTCAAA	AATCTGCCACCACTTCCTCA	235
2:48026861	AGTGTGCTGGAAGGTGATCC	CGGGTATCAGACCTTCCTGA	298

2:48030603	CTGATAAAACCCCCAAACGA	GGTATCTTCCGGCAACAGAA	135
2:48030669	GATGTTTTACTGTGCCTGGCT	CATAAGCGTAGACTTGCCCC	261
3:37067140	TCGGGCAGAATTGCTTCTAT	ATCCTCTGTGACAATGGCCT	373
7:6045634	TTTTGGGTACATGTTTAAGGATTG	GGCACACCGTAAGAACACAA	539
10:88677029	GGAGAAGTATGGATGGGCAA	ATGGCGCATTAGCACAGTTT	120
10:89720678	TCATGTGAATGAAAATGCAACA	AGTCAACAACCCCCACAAAA	472
10:89720870	ACCAGGACCAGAGGAAACCT	CCAGATTTCAGCAAACAACG	425
12:133202816	CTGTTGCCAATCCATGTGAG	AAGGAAGGGAGGAAAGGACA	455
12:133245032	AACAGAAAGTGGGAGCAGGA	ATAAGCCTGGAGCAAACTGC	395
12:133245452	GCAGTTTGCTCCAGGCTTAT	GTTGGTGAACTGCACACGAG	333
12:133249812	GAATTCTACTGCCGCCTCTG	TCTCCTTACTGTGTGTGCGG	352

12:133253974	GGGGAAAAGCAGCAAACATA	GATGTTCCCTACCACATCCG	374
16:68855966	GCCAAGCTGCCACATTTTCT	ACAGAAGGGACAAGGAAGCA	400
16:68867388	GAAAGCGGCTGATACTGACC	TTTTTCCCCAGAAACTCATCTC	312
17:63534353	ACAAGCACATATCCACACGC	CTCCTCCCTTTCCTTTCCAC	385
17:63533512	TGCCGCCCTCTTAGAAACTA	ACCATTCGCAGTACCACTCC	409
17:7577117	TGCTAGGAAAGAGGCAAGGA	CAAGGGTGGTTGGGAGTAGA	331
17:7578245	TTGCACATCTCATGGGGTTA	AGTCACAGCACATGACGGAG	389
17:7578388	ACACGCAAATTTCCTTCCAC	GTTTCTTTGCTGCCGTCTTC	397
18:48603114	GGCATTGGTTTTTAATGTATGGA	ATGCAAACAGGGTCATAGGC	396
19:50918229	GCTCCCTCCCAGGTCTACTT	CTGTGTTCCCAGAGGAGGAG	404
19:50919693	TGGTGACCTGAGAGCCCTAC	GGAGAGGAGGCAGGTGGT	403

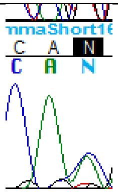
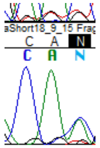
22:29091740	AAAATTAGCTGGGCATGGTG	ACGTGTCTTCTTGGACTGGC	455
22:29107974	TTCATCCATCTAAGCAGGGG	TGTTTTTGGCTGTTAAGCCC	413
22:29090061	AGCAGGGCTTCCCATGTATT	GATTTGCCCAATTGTTGCTT	159
22:29121042	CTCCCAAAGTGCTGGGATTA	CAGCAAGAAACACTTTCGGA	450

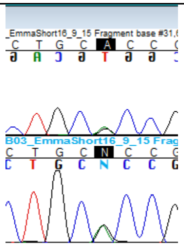
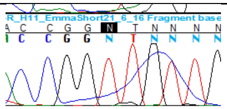
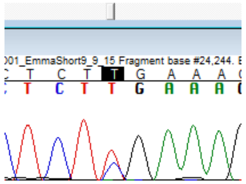
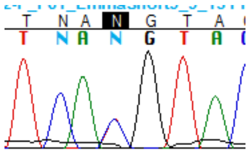
Appendix 4.3 cDNA Variant Validation Primers

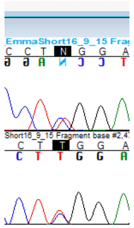
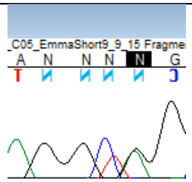
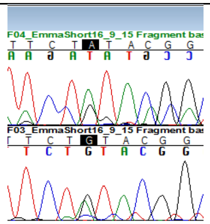
Location of Variant	**Forward Primer**	**Reverse Primer**	**Amplicon Size (bp)**
2:47612347	GGGGAACAACTGGATCTGGA	TTCCCTATGCATCTCACCCA	225
2:47630458	CTTCGTGCGCTTCTTTCAGG	GGATGCCTTATTTCCAGCTCT	283
2:48026228	CTCAGCCACCAAACAAGCAA	AATCTGCCACCACTTCCTCA	289
2:48030669	GATGTTTTACTGTGCCTGGCT	GGTCCAGTAACAAGCACACA	233
12:133245032	GTGCTGCCTGTGACTTCAAT	GTGAGACGCTCTTCCACCTT	280
12:133245452	AGGAGAAAGTGCCTGTGGAG	ATTGAAGTCACAGGCAGCAC	218
16:68855966	TCACATCCTACACTGCCCAG	CACACACGCTGACCTCTAAG	400

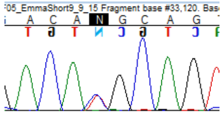
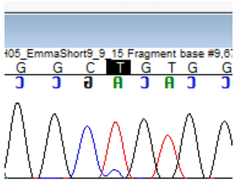
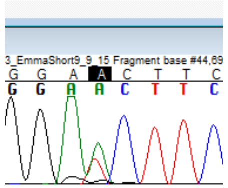
Location of Variant	**Forward Primer**	**Reverse Primer**	**Amplicon Size (bp)**
16:68867388	CCTGCCAATCCCGATGAAAT	CCACCATGTAAAGAGTGATGGG	492
17:63533512	TGCCGCCCTCTTAGAAACTA	ACCATTCGCAGTACCACTCC	409
17:7578245	GCAGCTGTGGGTTGATTCC	GCACAAACACGCACCTCA	399
17:7578388	GCAGCTGTGGGTTGATTCC	GCACAAACACGCACCTCA	399
19: 50918229	GGCTGGAGTTTGAGAAGGTC	CTTCCTCATCCTCTCGGCC	356
19:50919693	GAGAGGATGAGGAAGCGGG	GATACAGCTCAGACTCCCGG	400
22: 29091740	GAGAACCTTATGTGGAACCCC	TGCCCAGACTTCAGGAATGA	217
22:29107974	TGCACTGTCACTAAGCAGAA	TGATGATGCAAGGATGATTTAGC	293

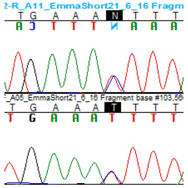
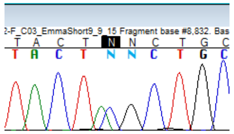
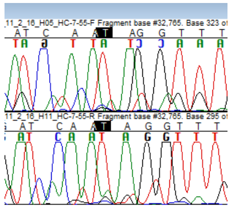
Appendix 4.4 Variant Validation: Targeted Exome UDS

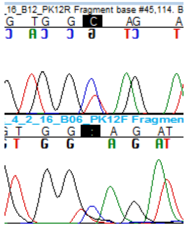
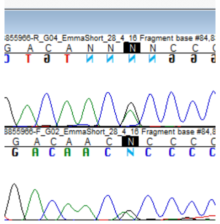
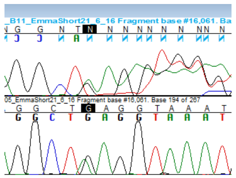
Sample	Gene	Location	Variant	Frequency	Validated	Image	CADD Score	Amino Acid Change	Genomic Annotation
Halo05	*CHEK2*	22: 29091740	C>G	1335/3023 (44%)	Yes	  Alt. allele appears to be A, although sequencing is repeatedly untidy. The true C>G variant was confirmed in cDNA	26.9	p.Arg406Pro	nonsynonymous SNV

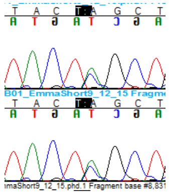
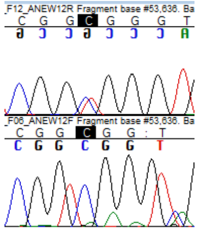
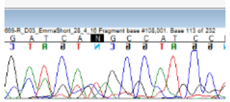
Sample	Gene	Location	Variant	Frequency	Validated	Image	CADD Score	Amino Acid Change	Genomic Annotation
Halo06	*POLD1*	19: 50918229	G>A	1716/3706 (46.4%)	Yes		22.4	p.Arg849His	nonsynonymous SNV
	MSH2	2:47639637	T>A	3/3 (100%)	Absent		24.5	p.Leu244Met	nonsynonymous SNV
Halo14	*CHEK2*	22:29107974	C>T	277/632 (43.8%)	Yes		24.8	p.Glu239Lys	nonsynonymous SNV
Halo15	*POLE*	12:133249812	T>C	22813/45964 (49.6%)	Yes		24.4	p.Met471Val	nonsynonymous SNV

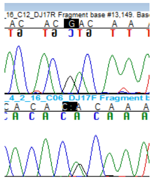
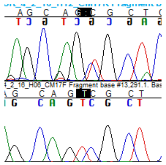
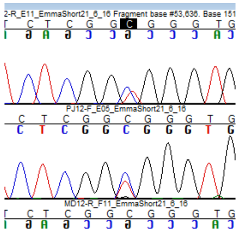
Sample	**Gene**	**Location**	**Variant**	**Frequency**	**Validated**	**Image**	**CADD Score**	**Amino Acid Change**	**Genomic Annotation**
Halo18	*POLE*	12:133202816	C>T	310/609 (50.9%)	Yes		23.8	p.Glu2140Lys	nonsynonymous SNV
Halo25	*CDH1*	16:68867388	G>A	1856/3910 (47.5%)	Yes		27.8	p.Gly879Ser	nonsynonymous SNV
Halo26	*MSH2*	2:47630458	A>G	439/858 (51.2%)	Yes		26.2	p.Tyr43Cys	nonsynonymous SNV

Sample	**Gene**	**Location**	**Variant**	**Frequency**	**Validated**	**Image**	**CADD Score**	**Amino Acid Change**	**Genomic Annotation**
Halo27	*POLD1*	19:50919693	C>T	1081/2232 (48.5%)	Yes		34	p.Thr954Met	nonsynonymous SNV
Halo28	*AXIN2*	17:63534353	T>C	1147/3499 (32.8%)	Yes		15.42	p.Ser390Gly	nonsynonymous SNV
Halo31	*POLE*	12:133245032	A>T	2428/4862 (49.9%)	Yes		23	p.Phe695Ile	nonsynonymous SNV

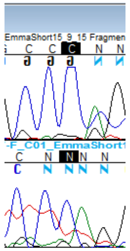
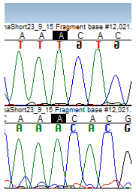
Sample	Gene	Location	Variant	Frequency	Validated	Image	CADD Score	Amino Acid Change	Genomic Annotation
Halo45	*MSH6*	2:48026228	C>T	580/1532 (37.9%)	Yes		29.5	p.Thr369Ile	nonsynonymous SNV
Halo47	*AXIN2*	17:63533512	C>A	357/742 (48.2%)	Yes		37	p.Glu548*	stopgain
Halo51	*PMS2*	7:6045634	T>C	936/2788 (34%)	No		25.6	p.Ile18Val	nonsynonymous SNV

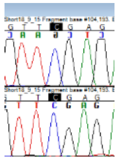
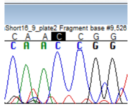
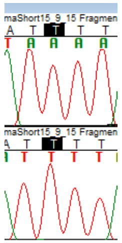
Sample	Gene	Location	Variant	Frequency	Validated	Image	CADD Score	Amino Acid Change	Genomic Annotation
Halo61	*POLE*	12:133245452	T>C	3617/6785 (53%)	Yes		26.0	p.Tyr623Cys	nonsynonymous SNV
Halo62	*CDH1*	16:68855966	G>A	609/1192 (51%)	Yes		23.9	p.Ala592Thr	nonsynonymous SNV
	EPCAM	2:47612347	G>A	267/922 (29%)	Yes		29.1	p.Glu301Lys	nonsynonymous SNV

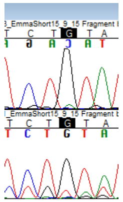
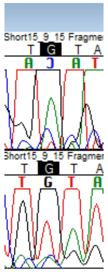
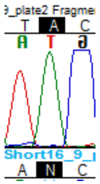
Sample	**Gene**	**Location**	**Variant**	**Frequency**	**Validated**	**Image**	**CADD Score**	**Amino Acid Change**	**Genomic Annotation**
Halo68	*AXIN2*	17:63533512	C>A	1176/2351 (50%)	Yes		37	p.Glu548*	stopgain
Halo69	*POLE*	12:133253974	C>T	1611/3091 (52%)	Yes		24.4	p.Arg259His	nonsynonymous SNV
Halo70	*MSH6*	2:48030669	C>T	398/997 (40%)	Yes	 Poor quality sequencing trace – the variant was later confirmed in cDNA	35	p.Arg1095Cys	nonsynonymous SNV

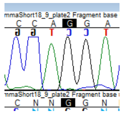
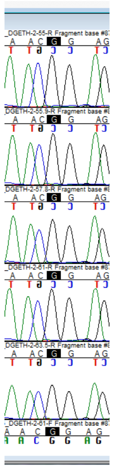
Sample	Gene	Location	Variant	Frequency	Validated	Image	CADD Score	Amino Acid Change	Genomic Annotation
Halo72	*TP53*	17:7578245	G>A	763/1656 (46%)	Yes		16.71	p.Arg202Cys	nonsynonymous SNV
Halo80	*TP53*	17:7578388	C>T	56/162 (35%)	Yes		28.5	p.Arg181His	nonsynonymous SNV
Halo81	*POLE*	12:133253974	C>T	1851/3784 (48.9%)	Yes		24.4	p.Arg259His	nonsynonymous SNV

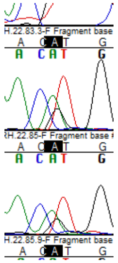
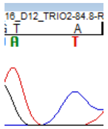
Standard PCR

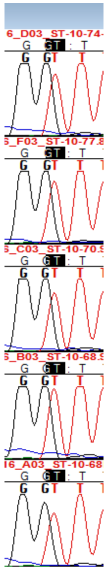
Sample	Gene	Location	Variant	Effect	Protocol for COLD PCR	Freq. (%)	Denat. Temp (°C)	Validated	Amino Acid Change	Genomic annotation	CADD Score
Halo07	*SMAD4*	18:48603114	C>A	Tm reducing	Fast	9/890 (1.01%)	87.7	No 	p.Pro472Gln	nonsynonymous SNV	24.7
Halo17	*TP53*	17:7577117	A>T	Tm neutral	Full	55/3547 (1.56%)	Variant not validated with standard PCR. Full COLD protocol failed to amplify the fragment.	No 	p.Val274Asp	nonsynonymous SNV	29.8

Sample	**Gene**	**Location**	**Variant**	**Effect**	**Protocol for COLD PCR**	**Freq. (%)**	**Denat. Temp (°C)**	**Validated**	**Amino Acid Change**	**Genomic annotation**	**CADD Score**
Halo20	*MSH6*	2:48026861	C>T	Tm reducing	Fast	7/669 (1.05%)	86.0	No 	p.Ser580Leu	nonsynonymous SNV	32
Halo28	*MSH2*	2:47639633	C>A	Tm reducing	Fast	8/767 (1.04%)	84.0 Double the amount of DNA required for PCR	No 	p.Asn242Lys	nonsynonymous SNV	25
Halo29	*PTEN*	10:89720870	T>G	Tm reducing	Fast	264/2436 (10.84%)	80.0	No 	p.Phe341Val	nonsynonymous SNV	27.9

Sample	Gene	Location	Variant	Effect	Protocol for COLD PCR	Freq. (%)	Denat. Temp (°C)	Validated	Amino Acid Change	Genomic annotation	CADD Score
Halo35	*CHEK2*	22:29121042	G>A	Tm reducing	Fast	152/3145 (4.83%)	84.0	No  	p.Thr215Ile	nonsynonymous SNV	16.85
Halo43	*PTEN*	10:89720678	A>T	Tm neutral	Full	12/848 (1.41%)	78.3	No 	p.Thr277Ser	nonsynonymous SNV	25.1

Sample	**Gene**	**Location**	**Variant**	**Effect**	**Protocol for COLD PCR**	**Freq. (%)**	**Denat. Temp (°C)**	**Validated**	**Amino Acid Change**	**Genomic annotation**	**CADD Score**
Halo46	*MLH1*	3:37067140	G>T	Tm reducing	Fast	7/619 (1.13%)	85.0	No 	p.Gly351*	stopgain	39
Halo49	*MSH6*	2:48010592	G>T	Tm reducing	Fast	8/734 (1%)	COLD PCR protocol did not amplify the fragment. No evidence of the variant with standard PCR protocol at varying Ta	No 	p.Gly74*	stopgain	36

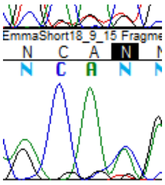
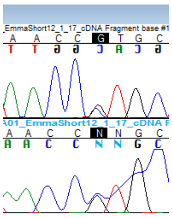
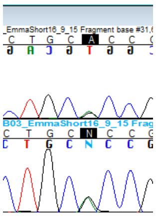
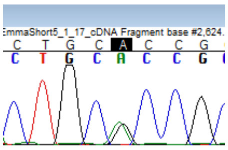
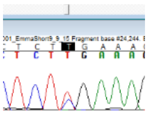
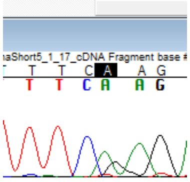
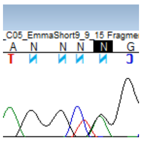
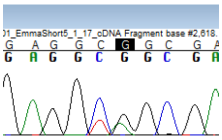
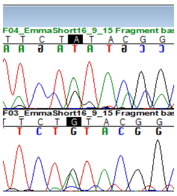
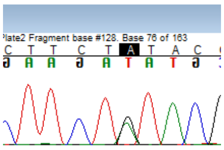
Sample	**Gene**	**Location**	**Variant**	**Effect**	**Protocol for COLD PCR**	**Freq. (%)**	**Denat. Temp (°C)**	**Validated**	**Amino Acid Change**	**Genomic annotation**	**CADD Score**
Halo54	*CHEK2*	22:29090061	G>A	Tm reducing	Fast	71/4620 (1.5%)	Mutation visible at multiple denaturation temps (83.3, 85.0, 85.9)	Yes 	p.Arg474Cys	nonsynonymous SNV	34
Halo58	*MSH6*	2:48030603	C>T	Tm reducing	Fast	44/4148 (1%)	84.8	No. Poor quality trace 	p.Pro771Ser	nonsynonymous SNV	23.8

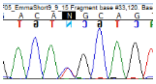
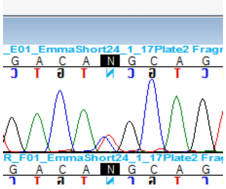
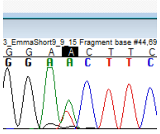
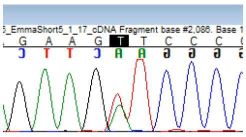
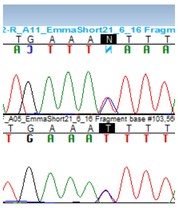
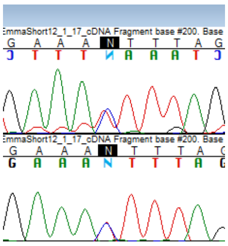
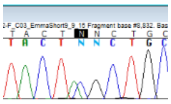
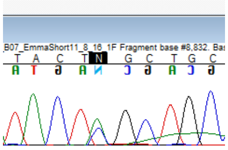
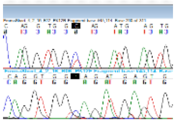
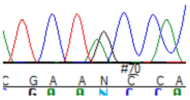
Sample	**Gene**	**Location**	**Variant**	**Effect**	**Protocol for COLD PCR**	**Freq. (%)**	**Denat. Temp (°C)**	**Validated**	**Amino Acid Change**	**Genomic annotation**	**CADD Score**
Halo64	*BMPR1A*	10:88677029	T>G	Tm increasing	Full	19/1240 (1.5%)	As low as 68.0*	No 	p.Phe272Val	nonsynonymous SNV	29.7

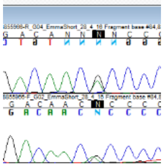
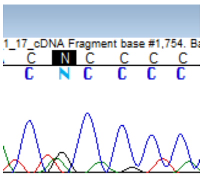
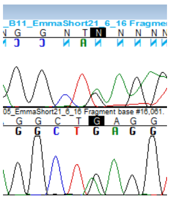
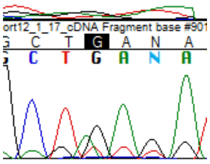
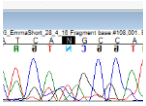
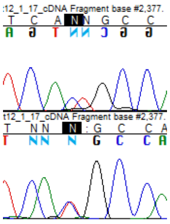
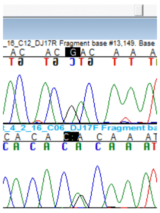
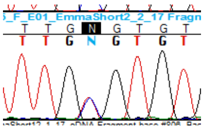
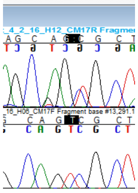
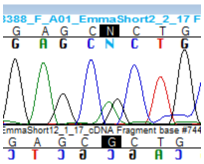
COLD PCR

*The full COLD PCR protocol appeared to work at denaturation temperatures as low as 68°C. However, this would be extremely unlikely, if not impossible. It is most probable that the hybridisation step did not work, therefore the effective denaturation temperature would have been 95°C.

Appendix 4.5 cDNA Validation

Sample	Gene	Location	Variant	Validation in DNA	Variant Present in cDNA
Halo05	*CHEK2*	22: 29091740	C>G		Read in reverse 
Halo06	*POLD1*	19: 50918229	G>A		
Halo14	*CHEK2*	22:29107974	C>T		Read in reverse 
Halo25	*CDH1*	16:68867388	G>A		
Halo26	*MSH2*	2:47630458	A>G		

Sample	**Gene**	**Location**	**Variant**	**Validation in DNA**	**Variant Present in cDNA**
Halo27	*POLD1*	19:50919693	C>T		
Halo31	*POLE*	12:133245032	A>T		Read in reverse 
Halo45	*MSH6*	2:48026228	C>T		
Halo47	*AXIN2*	17:63533512	C>A		
Halo61	*POLE*	12:133245452	T>C		Read in reverse 

Sample	**Gene**	**Location**	**Variant**	**Validation in DNA**	**Variant Present in cDNA**
Halo62	*CDH1*	16:68855966	G>A		
	EPCAM	2:47612347	G>A		
Halo70	*MSH6*	2:48030669	C>T		
Halo72	*TP53*	17:7578245	G>A		Read in reverse 
Halo80	*TP53*	17:7578388	C>T		Read in reverse 

Appendix 4.6 CRC In-House Candidate Gene List

A2LD1	*DCLRE1C*	*FNBP1L*	*MBTPS1*	*RAB2A*	*TGIF1*
ACOT9	*DDB1*	*FOSL1*	*MCC*	*RABEP1*	*THRAP3*
ACVR1	*DDI2*	*FOXE1*	*MECOM*	*RAC1*	*TJP2*
ACVR1B	*DEAF1*	*FOXP4*	*MED1*	*RAD17*	*TKT*
ACVR2A	*DHH*	*FRAT1*	*METAP2*	*RAD18*	*TLE1*
ACVRL1	*DHX15*	*FRMD6*	*MEX3C*	*RAD23A*	*TLK1*
ADAM10	*DIDO1*	*FRS2*	*MFNG*	*RAD23B*	*TM9SF2*
ADAM17	*DIEXF*	*FRYL*	*MGA*	*RAD50*	*TM9SF3*
ADAMTS6	*DIP2B*	*FRZB*	*MGMT*	*RAD51*	*TMBIM6*
ADCY6	*DISP1*	*FSHB*	*MICAL2*	*RAD51B*	*TMED10*
ADIPOR2	*DISP2*	*FST*	*MID1*	*RAD51C*	*TMEM132B*
ADSSL1	*DISP3*	*FUBP1*	*MKL1*	*RAD51D*	*TMPRSS2*
AEN	*DIXDC1*	*FZD1*	*MKL2*	*RAD52*	*TNKS*
AES	*DKK1*	*FZD10*	*MKLN1*	*RAD54B*	*TNKS2*
AGO1	*DLG1*	*FZD2*	*MLH1*	*RAD9A*	*TNPO1*
AGO2	*DLL1*	*FZD3*	*MLH3*	*RAI1*	*TNPO3*
AGO3	*DLL3*	*FZD4*	*MLL3*	*RALGAPA2*	*TNRC6A*
AGO4	*DLL4*	*FZD5*	*Mlxip*	*RANBP9*	*TNRC6B*
AGPAT6	*DLX2*	*FZD6*	*MMP7*	*RAP1B*	*TNS3*
AHCYL1	*DMC1*	*FZD7*	*MMS19*	*RASEF*	*TOMM70A*
AKAP9	*DMXL1*	*FZD8*	*MNAT1*	*RBBP6*	*TOP1*
ALKBH1	*DNAJC5*	*FZD9*	*MOB1B*	*RBM14*	*TP53*
ALKBH2	*DNMT1*	*G3BP1*	*MOB3B*	*RBM15*	*TPCN2*
ALKBH3	*DOCK9*	*GAB1*	*MOBKL1A*	*RBM25*	*TPD52*
AMD1	*DOT1L*	*GAS1*	*MOBKL2B*	*RBM26*	*TRA2A*
AMER1	*DSC1*	*GATA4*	*MPG*	*RBM39*	*TRA2B*
AMH	*DSC2*	*GATA6*	*MPP5*	*RBM47*	*TREX1*
AMHR2	*DSC3*	*GATAD2A*	*MRE11A*	*RBM5*	*TREX2*
ANKHD1	*DSG1*	*GATAD2B*	*MRGPRD*	*RBPJL*	*TRIM2*
ANKRD11	*DSG2*	*GDF2*	*MSH2*	*RDM1*	*TSC22D1*
ANKRD17	*DSG3*	*GDF3*	*MSH3*	*RECQL*	*TSC22D2*
ANP32B	*DSG4*	*GDF5*	*MSH4*	*RECQL4*	*TSHB*
ANXA11	*DSTN*	*GDF6*	*MSH5*	*RECQL5*	*TTC39A*
AP1AR	*DTX1*	*GDF7*	*MSH6*	*RERE*	*TUBA1B*
APC	*DUSP1*	*GGACT*	*MSL2*	*REV1*	*TXLNG*
APEX1	*DUSP10*	*GDI2*	*MTDH*	*REV3L*	*U2AF2*
APEX2	*DUSP11*	*GGNBP2*	*MTSS1*	*RFC1*	*UBAP2*
APTX	*DUSP12*	*GLI1*	*MTUS1*	*RFNG*	*UBE2D2*
AQP4	*DUSP13*	*GLI2*	*MUS81*	*RFT1*	*UBE2D3*
ARF1	*DUSP14*	*GLI3*	*MUTYH*	*RFX1*	*UBE2G1*
ARF4	*DUSP15*	*GMDS*	*MYH9*	*RHOU*	*UBE2G2*
ARGLU1	*DUSP16*	*GNA13*	*MYNN*	*RHPN2*	*UBE2N*

ARHGAP20	*DUSP18*	*GNAI3*	*MYO10*	*RLF*	*UBE2V2*
ARHGAP21	*DUSP19*	*GNB1*	*MYO1D*	*RLIM*	*UBE2Z*
ARHGAP5	*DUSP2*	*GPAT4*	*MYO5B*	*RMND1*	*UBE3A*
ARHGEF26	*DUSP21*	*GPBP1*	*NAA15*	*RNF111*	*UBQLN1*
ARHGEF38	*DUSP22*	*GPC6*	*NAA50*	*RNF4*	*UBR4*
ARID1A	*DUSP23*	*GRAMD4*	*NBAS*	*RNF43*	*UHRF2*
ARID1B	*DUSP26*	*GREM1*	*NBL1*	*ROCK2*	*USP46*
ARID2	*DUSP27*	*GSC*	*NBN*	*ROD1*	*USP6NL*
ARID5B	*DUSP28*	*GSK3A*	*NCOA1*	*RPA1*	*USP7*
ARIH1	*DUSP3*	*GSK3B*	*NCOA5*	*RPA2*	*USP9X*
ASH1L	*DUSP4*	*GTF2H1*	*NCOR1*	*RPA4*	*UTP1*
ASXL1	*DUSP5*	*GTF2H2*	*NCOR2*	*RPL14*	*UTP10*
ATAD2B	*DUSP6*	*GTF2H3*	*NCSTN*	*RPL18*	*UTP11*
ATF2	*DUSP7*	*GTF2H4*	*NEDD4*	*RPL18A*	*UTP11L*
ATL2	*DUSP8*	*GTF2H5*	*NEDD9*	*RPL19*	*UTP12*
ATM	*DUSP9*	*GTF2IRD1*	*NEIL1*	*RPL21*	*UTP13*
ATP1A1	*DUT*	*H6PD*	*NEIL2*	*RPL27*	*UTP14A*
ATP2A2	*DVL1*	*HDAC1*	*NEIL3*	*RPL28*	*UTP14C*
ATP6V0A2	*DVL2*	*HDHD3*	*NEURL*	*RPL29*	*UTP15*
ATP8B1	*DVL3*	*HEATR1*	*NEURL1*	*RPL35*	*UTP17*
ATR	*DYRK1A*	*HECTD1*	*NF2*	*RPL37*	*UTP18*
ATRX	*E2F3*	*HELQ*	*NFATC3*	*RPL4*	*UTP19*
ATXN2	*ECE1*	*HES1*	*NFIA*	*RPL7A*	*UTp2*
AXIN1	*EEF2K*	*HES5*	*Nfib*	*RPLP2*	*UTP20*
AXIN2	*EHF*	*HEY1*	*NFKB1*	*RPRD1B*	*UTP21*
B4GALNT2	*EHMT1*	*HEY2*	*NHEJ1*	*RPS11*	*UTP22*
BAMBI	*EIF1*	*HEYL*	*NIPBL*	*RPS13*	*UTP23*
BAZ1B	*EIF1AD*	*HHAT*	*NKD1*	*RPS14*	*UTP24*
BCL9	*EIF1AX*	*HHIP*	*NLK*	*RPS15*	*UTP25*
BCLAF1	*EIF1AY*	*HIPK2*	*NOC4L*	*RPS16*	*UTP3*
BGLAP	*EIF1B*	*HMGB1*	*NODAL*	*RPS26*	*UTP4*
BIRC6	*EIF2A*	*HNF1A*	*NOG*	*RPS28*	*UTP5*
BLM	*EIF2AK1*	*HNF1B*	*NOL6*	*RPS3*	*UTP6*
BMP1	*EIF2AK2*	*HNRNPA1*	*NONO*	*RPS4X*	*UTP7*
BMP2	*EIF2AK3*	*HNRNPA2B1*	*NOP14*	*RPS4Y1*	*VAPA*
BMP3	*EIF2AK4*	*HNRNPA3*	*NOTCH1*	*RPS5*	*VDAC2*
BMP4	*EIF2B1*	*HNRNPC*	*NOTCH2*	*RPS6*	*VEGFA*
BMP5	*EIF2B2*	*HNRNPD*	*NOTCH2NL*	*RPS9*	*VEGFD*
BMP6	*EIF2B3*	*HNRNPF*	*NOTCH3*	*RRBP1*	*VGLL4*
BMP7	*EIF2B4*	*HNRNPM*	*NOTCH4*	*RREB1*	*VPS13D*
BMP8B	*EIF2B5*	*HNRNPR*	*NPC1*	*RRM2B*	*VWCE*
BMPER	*EIF2C1*	*HOOK1*	*NR0B1*	*RRP1B*	*WAC*
BMPR1A	*EIF2C2*	*HOXB4*	*NR1I2*	*RSPO2*	*WAPAL*

BMPR1B
BMPR2
BOC
BRCA1
BRCA2
BRD4
BRD8
BRIP1
BTBD3
BTBD7
BTRC
C14orf43
C19ORF40
CABLES2
CALM1
CAPN7
CAPRIN1
CCNH
CCNL1
CD2AP
CD44
CDC25A
CDC42
CDH1
CDH10
CDH11
CDH12
CDH13
CDH15
CDH16
CDH17
CDH18
CDH19
CDH2
CDH20
CDH22
CDH24
CDH26
CDH3
CDH4
CDH5
CDH6
CDH7
EIF2C3
EIF2C4
EIF2D
EIF2S1
EIF2S2
EIF2S3
EIF3A
EIF3B
EIF3C
EIF3CL
EIF3D
EIF3E
EIF3F
EIF3G
EIF3H
EIF3I
EIF3J
EIF3K
EIF3L
EIF3M
EIF4A1
EIF4A2
EIF4A3
EIF4B
EIF4E
EIF4E1B
EIF4E2
EIF4E3
EIF4EBP1
EIF4EBP2
EIF4EBP3
EIF4ENIF1
EIF4G1
EIF4G2
EIF4G3
EIF4H
EIF5
EIF5A
EIF5A2
EIF5AL1
EIF5B
EIF6
ELAC1
HR
HSD17B12
HSPA4
HUS1
HUWE1
ID1
ID2
IFNG
IFT52
IGF1
IGFBP3
IHH
IL17B
IL2RA
IL6
INHA
INHBA
INHBB
INTS6
IPO7
IQGAP1
IREB2
IRF2
ITCH
ITGA6
ITGAM
ITGB1
ITGB5
ITGB7
ITPR1
JADE3
JAG1
JAG2
JUNB
KAT6A
KCMF1
KCNQ1
KCTD11
KCTD20
KDM3B
KDM6A
KHDRBS1
KLF3
NR3C1
NR4A2
NR6A1
NRG3
NRIP1
NSD1
NSD2
NSUN5
NTHL1
NUDT1
NUDT3
NUMB
NUP98
Ocln
OGG1
OGT
ONECUT2
OTOA
OTX2
PABPC1
PAFAH1B1
Pak2
PALB2
PAN3
PANK2
PAPOLA
PARD3
PARP1
PARP2
PAWR
PAX5
PCBP2
PCF11
PCNA
PCNP
PDCD4
PDCD6IP
PDE4D
PDE4DIP
PDGFA
PDGFRA
PDGFRB
PDLIM1
RSRC2
RUNX1
RUNX2
RYBP
SAMD12
SARNP
SBNO2
SCAF11
SCAF4
SCAF8
SDC1
SDC3
SEC63
SEL1L
SEM1
SEMA4B
SENP2
SERBP1
SERPINE1
SETD2
SETD5
SETDB1
SF3B3
SFI1
SFPQ
SFRP1
SFRP2
SFRP4
SGMS1
SH2D1A
SH3BGRL2
SHFM1
SHH
SHROOM3
SIK3
SIPA1L1
SLC12A2
SLC16A10
SLC16A5
SLC25A25
SLC35A3
SLC6A19
SLC9A1
WAPL
WARS
WASF2
WDR3
WDR33
WDR36
WDR43
WDR46
WDR75
WDR82
WHSC1
WIF1
WISP1
WNT1
WNT10A
WNT10B
WNT11
WNT16
WNT2
WNT2B
WNT3
WNT3A
WNT4
WNT5A
WNT5B
WNT6
WNT7A
WNT7B
WNT8A
WNT8B
WNT9A
WNT9B
WRN
WTAP
XAB2
XPA
XPC
XPO1
XPO7
XRCC1
XRCC2
XRCC3
XRCC4

CDH8	*ELAVL1*	*KMT2C*	*PDS5A*	*SLC9A3R1*	*XRCC5*
CDH9	*ELF1*	*KPNA4*	*PDXDC1*	*SMAD1*	*XRCC6*
CDK12	*ELMSAN1*	*KPNB1*	*PHACTR4*	*SMAD2*	*YAP1*
CDK16	*EMCN*	*KREMEN1*	*PHF12*	*SMAD3*	*YBX1*
CDK7	*EME1*	*KRT1*	*PHF16*	*SMAD4*	*YTHDC1*
CDK8	*EME2*	*KRT19*	*PICALM*	*SMAD5*	*YTHDF2*
CDKN1A	*EML4*	*LAMA1*	*PIGL*	*SMAD6*	*YTHDF3*
CDKN2B	*ENDOV*	*LAMA2*	*PIK3C2A*	*SMAD7*	*YWHAB*
CDON	*ENG*	*LAMA3*	*PIK3R1*	*SMAD9*	*YWHAQ*
CELF1	*EP300*	*LAMA4*	*PIP5K1B*	*SMARCA5*	*YY1*
CELSR1	*EP400*	*LAMA5*	*PLAU*	*SMEK1*	*ZBTB14*
CELSR2	*EPC2*	*LAMB1*	*PLS3*	*SMOC2*	*ZC3H14*
CELSR3	*EPCAM*	*LAMB2*	*PMS1*	*SMUG1*	*ZC3H4*
CER1	*EPHB2*	*LAMB3*	*PMS2*	*SND1*	*ZCCHC7*
CERS1	*EPHB4*	*LAMB4*	*PMS2P3*	*SNHG1*	*ZER1*
CERS2	*EPS8*	*LAMC1*	*PMS2P4*	*SNRNP70*	*ZFAND3*
CERS3	*ERBB2IP*	*LAMC2*	*PNKP*	*SNW1*	*ZFAND6*
CERS4	*ERBB4*	*LAMC3*	*POFUT1*	*SNX24*	*ZFHX3*
CERS5	*ERBIN*	*LARP1*	*POLB*	*SON*	*ZFP1*
CERS6	*ERCC1*	*LARP1B*	*POLD1*	*SP1*	*ZFP106*
CFLAR	*ERCC2*	*LARP4*	*POLE*	*SP3*	*ZFP112*
CFTR	*ERCC3*	*LARP4B*	*POLG*	*SPATA5*	*ZFP14*
CHAF1A	*ERCC4*	*LARP6*	*POLH*	*SPEN*	*ZFP161*
CHD1	*ERCC5*	*LARP7*	*POLI*	*SPINT2*	*ZFP2*
CHD2	*ERCC6*	*LATS1*	*POLK*	*SPO11*	*ZFP28*
CHD7	*ERCC8*	*LATS2*	*POLL*	*SPRED1*	*ZFP3*
CHEK1	*ERG*	*LCOR*	*POLM*	*SREBF2*	*ZFP30*
CHEK2	*ERICH6*	*LCORL*	*POLN*	*SRFBP1*	*ZFP36L1*
CHL1	*ESCO1*	*LEF1*	*POLQ*	*SRRM1*	*ZFP36L2*
CHRD	*ETF1*	*LEFTY1*	*PORCN*	*SRRM2*	*ZFP37*
CHUK	*ETS2*	*LEKR1*	*PPFIBP1*	*SRSF1*	*ZFP41*
CLDN15	*EXO1*	*LEMD3*	*PPM1B*	*SRSF10*	*ZFP42*
CLINT1	*EXOC2*	*LFNG*	*PPP1CA*	*SRSF11*	*ZFP57*
CLOCK	*FAAP24*	*LGR4*	*PPP1CB*	*SRSF3*	*ZFP62*
CLSTN2	*FADS2*	*LHFPL4*	*PPP1CC*	*SRSF4*	*ZFP64*
CLTC	*FAM107B*	*LIG1*	*PPP1R12A*	*SSB*	*ZFP82*
CNBP	*FAM120A*	*LIG3*	*PPP1R13B*	*SSBP3*	*ZFP90*
CNOT1	*FAM123B*	*LIG4*	*PPP2CA*	*STAG1*	*ZFP91*
CNOT2	*FAM166A*	*LLGL2*	*PPP2R1A*	*STAT1*	*ZFPL1*
COL1A1	*FAM193A*	*LMO7*	*PPP2R2A*	*STAT6*	*ZFPM1*
COL1A2	*FAM194A*	*LOR*	*PPP4R2*	*STIL*	*ZFPM2*
COL3A1	*FAM19A5*	*LPAR5*	*PPP4R3A*	*STRIP2*	*ZFR*
CPNE3	*FAM40B*	*LRBA*	*PPP6C*	*STRN3*	*ZIC1*

CPSF6	*FAM83B*	*LRCH1*	*PPP6R3*	*STT3B*	*ZIC2*
CREBBP	*FANCA*	*LRIG1*	*PRDM16*	*STXBP1*	*ZMIZ1*
CRKL	*FANCB*	*LRP2*	*PRKACA*	*STXBP5*	*ZMYM5*
CSDE1	*FANCC*	*LRP5*	*PRKACB*	*SUFU*	*ZMYND8*
CSF1R	*FANCD2*	*LRP6*	*PRKCA*	*SYMPK*	*ZNRF3*
CSNK1A1	*FANCE*	*LRRC41*	*PRKDC*	*SYNCRIP*	
CSNK1D	*FANCF*	*LRRK1*	*PROM1*	*TAB2*	
CSNK1E	*FANCG*	*LSM14A*	*PSEN1*	*TAF15*	
CSNK1G1	*FANCL*	*LTBP1*	*PSEN2*	*TAOK1*	
CSNK1G2	*FANCM*	*LTBP2*	*PSENEN*	*TAOK3*	
CSNK1G3	*Fat1*	*LTBP4*	*PSMA3*	*TAX1BP1*	
CSNK2A1	*FAT4*	*LUC7L*	*PTBP3*	*TBCA*	
CSNK2A2	*FBXO11*	*LUC7L2*	*PTCH1*	*TBL1X*	
CSNK2B	*FBXW11*	*LUC7L3*	*PTCH2*	*TBL2*	
ainCST3	*FBXW2*	*MACROD2*	*PTCHD1*	*TCERG1*	
CSTF3	*FBXW4*	*MAD1L1*	*PTCHD2*	*TCF12*	
CTBP1	*FBXW7*	*MAEA*	*PTCHD3*	*TCF4*	
CTNNA1	*FCF1*	*MAGT1*	*PTCRA*	*TCF7L2*	
CTNND1	*FCHSD2*	*MAML1*	*PTEN*	*TDG*	
CUL1	*FEN1*	*MAML3*	*PTP4A2*	*TDP1*	
CXXC4	*FGD6*	*MAN1A2*	*PTPN2*	*TELO2*	
CYB5B	*FGF9*	*MAPK1*	*PTPRJ*	*TFRC*	
D17Wsu92e	*FGFR1OP2*	*MARK2*	*PTPRK*	*TGFB1*	
DAAM1	*FGFR3*	*MARK3*	*PUM1*	*TGFB2*	
DACH1	*FIGF*	*MATR3*	*PUM2*	*TGFB3*	
DAG1	*FKBP1B*	*MAU2*	*PWP2*	*TGFBR1*	
DCBLD1	*FKBP4*	*MAX*	*RAB10*	*TGFBR2*	
DCLRE1A	*FKBP8*	*MBD4*	*RAB14*	*TGFBR3*	
DCLRE1B	*FLT3*	*MBNL1*	*RAB23*	*TGFBRAP1*	

Appendix 4.7 WES Validation Primers

Chr.	**Position**	**Forward Primer**	**Reverse Primer**	**Amplicon Size (bp)**
1	1273404	CCAAGTACACAGCAGGAGCA	ACCCTGAACCTCAACAGTGG	384
1	11854085	TCCAATCCATGCTCATTCAA	GGCAAAGGGAGAAGGGTAAG	450

1	26515380	CCCACAGACAGCAGTGAAGA	AGTCTGGAGTCTGGTGCCC	288
1	27190196	GACGACAAGAAGCGCATCAT	AGGGTGCTGTCTTTGTAGGA	242
1	32688188	GACCCATAATCCGGAAGTGA	TGGCCTCTCCCTACTTCTCA	393
1	36757052	GCTTATTGGCTCAGGAACGC	CCCAAAACTCCTTATCCACCC	248
1	40228846	AGAGCTGTGGGAAGTTAGCG	GGGGACGCAAAGTGAAGATA	455
1	45267346	CTCAGGGAGCACAGATGGAG	CTCAGGAGTCTCACCCAACT	215
1	76345823	TGTGCATTTATCTTGACATTTGTTT	TTGCAATTAGATCTTCCCCG	334
1	109812092	CCCACATCAGAGTCCCCAAA	AAGACACAGATGGGCTTGGT	160
1	226564855	CAAGGAGGATGCCATTGAGC	AGCAGACAGTGTAAGGGCAT	155
1	230907799	AGTTCTGGATGGAGCTGGTC	GAGAGGCAGGATAGGGTGAC	211
2	9633092	CAGGCCAAACTTTGCTAAGAA	GCAATCTCTTCCTGTCAGCC	419
2	20870532	CAGGCAGCTGGACTTCTCC	GGATCTCCCGGAATAAGCTC	342
2	32706513	TGCACCTCCTCTGTCCTCTT	CCACCACGTCCAGCTAATTT	313
2	32726929	GGAGCAAGTGGATTACATCTCA	CCACAGCCCTCTTGTACTGA	314
2	203407059	AGGATTTCCAAATGTGCCTG	TGATTTGTGGCATTAGGCAA	411

2	208633009	TAGAGCGGGTGTGACTCCTT	GACCTGCGCTTCTTCCTATG	380
2	217026733	GCAAGGTGAGTGAAAGCACA	CCGCCAGTAAAGAAAAATGC	406
2	219754822	TCCTTGTGCCAGACTCTCCT	ATTCTCGCGTGGATGTCTCT	474
3	33883492	CTGGGATTACACGCGTGAG	ACTACACTGTACTAGCATGAGGA	472
3	50155887	CATACCCTACCTCCCAGCTG	TGTGCTCCTTGTCATCTCTCT	201
3	121207509	TCTCCTTCTACACCAGTGGGA	TGCCAAACTAGGAGCAAAGG	405
3	121207520	TCTCCTTCTACACCAGTGGGA	TGCCAAACTAGGAGCAAAGG	405
3	142272170	CATGCTGCAATATACCCCTG	TTTCCGTTTGTTTCTGGGAT	423
3	142272170	TTTTAGGGCCGCAAAAGGAG	GCCTATAGTCCAGACAAACGC	209
4	84374567	TCCCTTGAATTGGGCATAAA	GAACTGCCCCATAATTGCAT	297
4	96070060	GAGTTGGCACCAAACGCTAT	AAAAGCAGGGTAGGCTGACA	376
5	60194107	GGTAGTGGGTAAGGGTGGGT	TTGCAGCAAACACTGCTCAT	312
5	80063896	CCCAAGAGTTCTTCTTGATTGTC	AACACTTTTAGGGTCCAATCACA	335
5	80063899	CCCAAGAGTTCTTCTTGATTGTC	AACACTTTTAGGGTCCAATCACA	335
6	28219686	CGCCCAGTCTGCAGAAGA	CATTGTCCGCAGAGTTCTCG	207

6	39864686	CTGACCTGTGCCTCCTCTC	GGGTGAGATAGAGGCCAAGG	238
6	133789765	GACTCAGAAACAAATGGGGC	ATTATCTGGGAGCTTTCCGC	390
6	136599544	GGTGAAGGACTATGGGGTGA	GTCGTTCACGTTCCAGGAGT	370
7	2284301	TGACTCTGCCCTCTCACCTT	CCTGCTCCAGGTCACTTAGC	304
7	34125622	CTCTGTGGTCTTTGTGGCAA	AATGGGGCTGACTCAAGATG	435
7	55273086	CAGAGACCCACACTACCAGG	TGGAGCCCTTAAAGATGCCA	212
7	98591187	CTAGAACCAGGGCCTTCTCC	TATGGGCCCTGGATCTGATG	233
7	91623985	GAAGTTGATTTGGAAGCACCA	TTCATACATGATATAATACGAAATGGC	449
7	100421340	CTTCCCTCCCCAGACCTAAG	GCTGACCTCCTTCTCTGCAC	458
7	116381047	GAAGCTCTTTCCACCCCTTC	GCCAAGAGAGAGATGGTGCT	381
8	20107358	TCGTGTGGAAGGAGGAGAAG	AGTGGCTATGATGTCCTCGT	194
8	22049596	TTTGGGAGCCCCTTAACTCT	CCATACCAGGTTGTCATCCC	353
8	6479113	TCGCCTACGCTATGGAGACT	AGGCTGAGACATCAACCCAG	427
8	28420428	TCTCATGACAGGTACACGCC	TCAGTCATGCAAAAGGCAAA	298
8	33361016	AGGTGTGTGAAAATGGGAGG	AGGCCCTGGGCTTTTATTTA	447

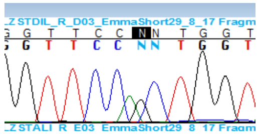
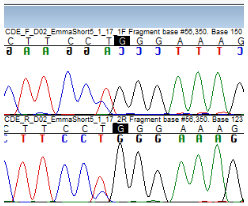
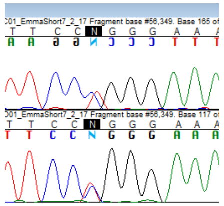
8	53030923	GCAAGGGTGGCACTTTTCTA	TGGAATCTAGAAACACTCAGCA	455
8	61734439	TGTATGTGGTCAAATGAATCCAA	ACTCCCTGGAACTCTCCGAT	332
8	134232908	TTTGTGCCTCTGTTCCTCCT	CCTCACATACCCACTGCTCA	348
9	84228372	GAGATAGGAAGGAGGGTGGC	GCTAGAAGGTGCCCAGTGTC	374
9	130588091	GGCTTTATAAGGGACCGGAG	CGAACCTAGGTCCTCTGACG	402
9	140069828	GAGCAAGGCGGTGAAGATG	CGTCACTGTCAATGAGCACC	158
10	18795447	TGAGGATGCGAAATAAGTGTCAG	ACTGGATAGAGCTGGGGAAAG	194
10	50690821	CCTGGCCATCTTTCTCACAT	GAATGGCGATTGCTAAAAGG	342
10	88259879	CCAGAACCTCCTTCATCACC	GCCTTAGCCAAAGCAAACAG	371
10	99218456	AGGCCTAGCATTTGTGCTGT	CCCCTGGATGACAAGAAGAG	300
11	810305	GTATTTTTGGGACGGAGGC	AGGCAGAAGAATCGCCTGTA	313
11	17140810	ATTCAAATGGGGCAGATGAC	TCAGGACCAGATTGAGATTGTG	366
11	44626709	CCCTAGAAACCTCCTCCAGC	TGGATGAGAGGAAATGGGCA	178
11	46918516	CCTCAACCCAACTGGAAAGA	GTGATCCCGGTGTCAAGAGT	337
11	65482096	GTATTTTCCACTGGCCCTGG	CACAGGTAGAGGACAGGCAA	221

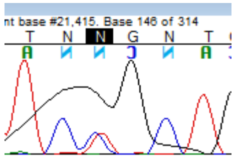
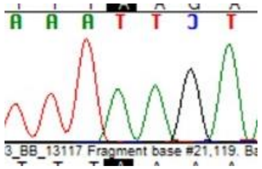
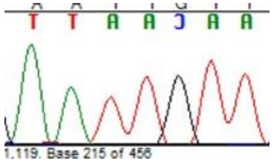
11	68183958	CCAGCTCCTCTGTGGCTTAC	TCCTCCCTCTGCTAAGGACA	352
11	108114679	TCTGCGACCTGGCTCTTAAA	TGAGTCTAAAACATGGTCTTGCA	481
11	108114684	TCTGCGACCTGGCTCTTAAA	TGAGTCTAAAACATGGTCTTGCA	481
11	108114689	TCTGCGACCTGGCTCTTAAA	TGAGTCTAAAACATGGTCTTGCA	481
11	108186610	TCAAACTCGTGTTGTTTGAACTG	GTTGATGAGGGGATTGCTGT	450
11	108196797	AGACAGATAGGCAGACGTGG	CACTCAGAGACTCCACAGCT	245
11	119148958	GGACCCAGACTAGATGCTTTCT	GGCCACCCCTTGTATCAGTA	386
11	119169085	CTGCAGTGTTGGTGCAGA	TCACCATCCAGAGACAACCA	366
12	46233172	TGTTCACCAGTGATGGCATT	TGAAAACACTAACCTGGGGC	249
12	56982155	CCCCTCAGCTGCTTCAAATG	CCTGCCTCCGTTCATTTTCC	201
12	50146761	ATTTTGGGCATTTGGTCTTC	TTGTACCACAGCAACCAGGA	362
12	57850383	CCTTATGTTGCAGGCGAGAC	ATGGAAAGAGGCAGCAATGC	152
12	124824869	ACAAAACCAACCACCACGTC	TCTGCTCTGTACCTGGTGAC	234
12	124911260	CTCTGTGTCTTCCCTGGGTC	CTTCTCCTGCTCACTCCACA	244

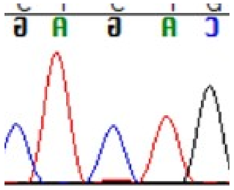
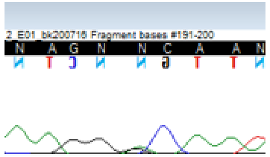
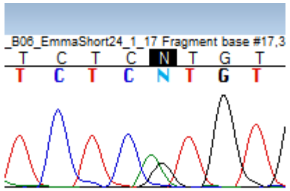
12	132471141	ATTTGTCTGCTTGCATCGTG	AGGACTCTGTGGTTTGGTGG	368
13	38156538	AATGAGCTTTACAACGGGCA	TTCACCCAAGCTACCCAGTT	217
13	45008887	GGTGACTGTGGAGGCAGATT	AGCCATTTGATGTATGCGGT	334
13	49047524	TGATTAGACGGGCACTGTTAGA	GCAATATGCCTGGATGAGGT	330
14	68353893	TTGCTTAGCCTTTGCAATACC	GATGATGGCGATGTTTGCTA	427
14	104205127	AATCTGTCCACCCCACACTC	AAGCGGACCAACTTGAAGAA	375
14	104245134	GCTGATCTCAAACTCCTGCC	GCGCAGCCTGAACAACTTTA	397
15	42028820	ATGCCATCATGTCAAGACCA	ATTTTGTCACCCAGTTCCCA	367
15	42058553	AGCAAGCCATCTGACCATCT	TCACTGGAATCAAGAGCTGC	392
16	22269048	GGGATGGCGCTCTTCTTCTA	CAGCCCTGGTATTCATCAGTT	243
16	23635370	GCCCAGCCTAGGTTCACAT	TGAGCCTTCAAATGATGAAAA	414
16	86602293	CCTGAGCGCTCTCAACCTC	TCTGCAGCCCCTTAATTGTC	486
17	27239701	GGTGAGCTGAATCGTTGGTT	GCTAATGGCCCACACCTCTA	437
17	42171169	CTTCCTGTCTCTTCTGCCCA	GGTGGTCATGCTGTTTCTGG	242
17	44953675	GGGTCTCTTTCCCATTCTCC	ATGTACACCAGGTCCCCAGA	364

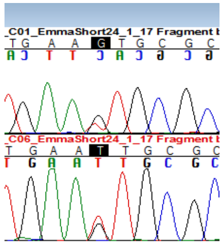
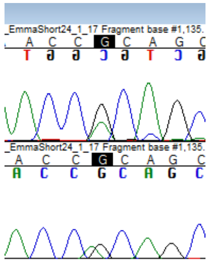
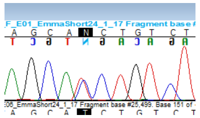
18	28605748	TTGCTTCATTTCAGGCTTACC	TTTTCTACTTTTAGGTTGAATCTGATG	373
18	28647999	TGAGAGAAAAACCCCCACAA	ACGTTGTGTTTTTCCCCTTG	332
18	28672114	ATGTGACACAGCCCTTTTCC	TAAGTTTTTGGCTCTCCCCA	460
18	50961517	GGGGTGATCAAAGTAATCTGTTC	TGCCCCAGACACTCTTGAAT	250
19	7694391	GACATGACGATCCGGTCTCT	AAGTCAATAACGTGGGCCTG	337
19	36831616	AAGCCTTCGTATGTGGTCCA	TCTCTGATGGCGAACTAGTTGT	209
19	48620943	GTCCCCTCCCTACTCTGCTC	TGTTTACTTCAAGGGGGTGC	340
19	55697712	ACGCCTGGAGAGATAAGAGC	GACTCAGATCTCTCACCCGG	220
20	30193351	CCCATTCTGTTTCAGCCAGT	TAGTCGATGACGTGCTGGAG	333
21	32496789	AGACAGCTCCTCAAAACCGA	AACCGTCCTCCTACATCAGC	238
X	70342603	ACAGGCAGTTGAAGTTCGCT	ATGCTCTGCTTTCAATGCCT	415

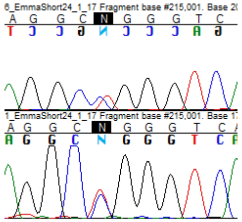
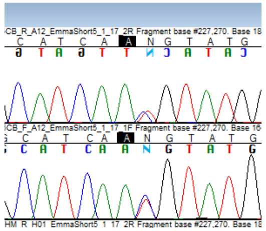
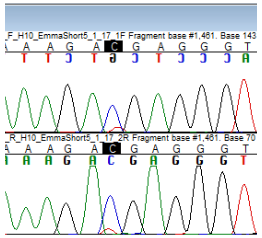
Appendix 4.8 Validation of WES Variants

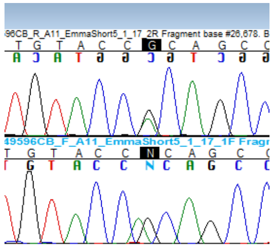
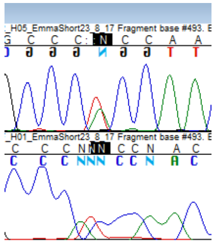
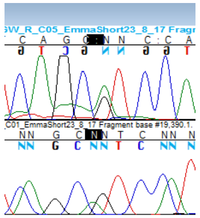
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo05	*LZTS1*	8	20107358	G>A	34	Yes			Missense
Halo05	*ERCC6*	10	50690821	G>A	34	No	 Repeat: 	Adjacent variant is a synonymous change	Missense

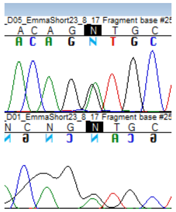
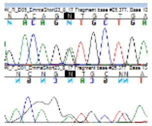
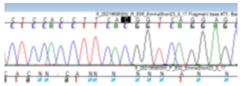
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo05	*LRP4*	11	46918516	G>A	29.3	Yes	 Read in reverse		Missense
Halo05	*ATM*	11	108114679	G>T	24.4	No		Only present in 2/5 reads (NGS)	Splice site acceptor
Halo05	*ATM*	11	108114684	G>T	23.3	No		Only present in 2/7 reads (NGS)	Missense

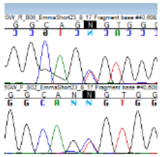
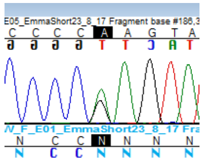
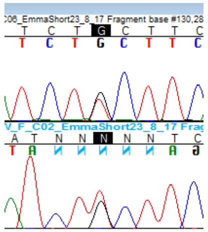
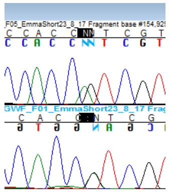
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo05	*ATM*	11	108114689	C>T	23.4	No		Only present in 2/10 reads (NGS)	Missense
Halo05	*ATM*	11	108196797	G>A	34	Yes			Missense
Halo05	*PALB2*	16	23635370	C>T	26.6	Yes	 Read in reverse		Missense

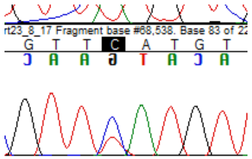
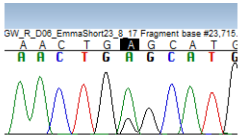
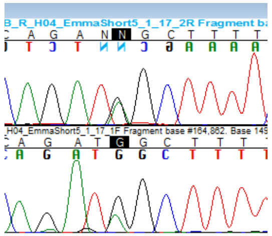
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo05	*LIG1*	19	48620943	C>A	29	Yes	 Read in reverse		Missense
Halo05	*FZD5*	2	208633009	C>T	22.5	Yes	 Read in reverse		Missense
Halo05	*ATR*	3	142272170	A>G	20.5	Yes	 Read in reverse		Missense

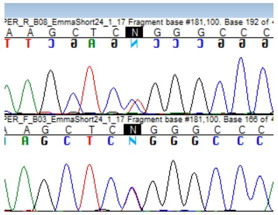
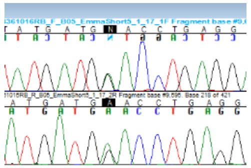
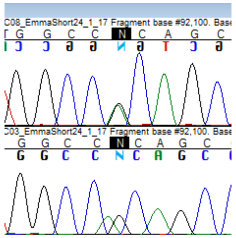
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo05	*MCPH1*	8	6479113	C>T	24.6	Yes			Missense
Halo08	*EYA4*	6	133789765	C>T	24.8	Yes			Missense
Halo08	*BCLAF1*	6	136599544	G>A	36	Yes	 Read in reverse		Stopgain

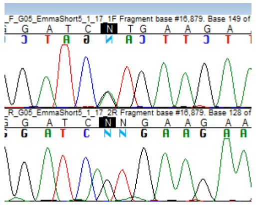
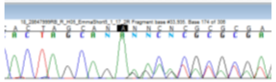
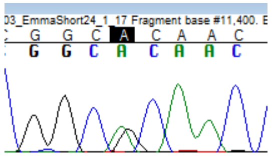
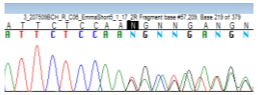
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo08	*BMP1*	8	22049596	G>A	33	Yes			Missense
Halo15	*SFN*	1	27190196	A>T	26.1	Yes			Missense
Halo15	*CELSR2*	1	109812092	G>A	22	Yes			Missense

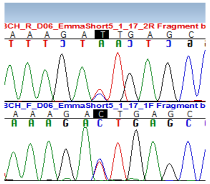
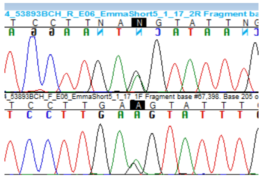
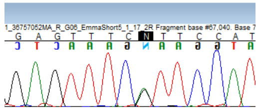
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo15	*CAPN9*	1	230907799	C>T	35	Yes			Missense
Halo15	*ATR*	3	142272170	A>G	20.5	Yes			Missense
Halo15	*ZKSCAN 4*	6	28219686	CGGTCA >C	24.4	No		There was low coverage at this locus. The variant was called in 2/3 reads	Frameshift

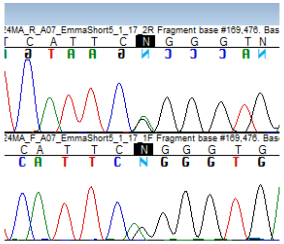
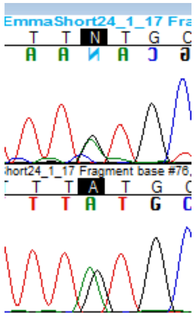
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo15	*DAAM2*	6	39864686	C>T	34	Yes			Missense
Halo15	*EGFR*	7	55273086	G>A	35	Yes			Missense
Halo15	*ST18*	8	53030923	G>T	31	Yes			Missense
Halo15	*NCOR2*	12	124824869	G>A	24.3	Yes			Missense

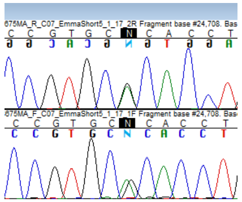
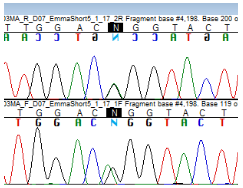
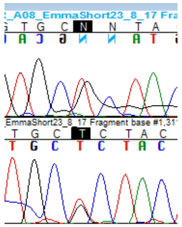
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo15	*NCOR2*	12	124911260	C>T	24.6	Yes			Missense
Halo15	*HDAC5*	17	42171169	G>A	25.6	Yes			Missense
Halo17	*BMPR2*	2	203407059	G>A	23.1	Yes			Missense

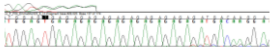
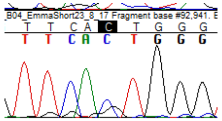
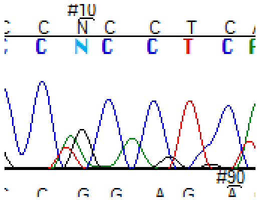
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo17	*BMPER*	7	34125622	C>T	34	Yes			Missense
Halo17	*TTI2*	8	33361016	C>T	24.6	Yes			Missense
Halo17	*CBL*	11	119169085	G>A	21.6	Yes			Missense

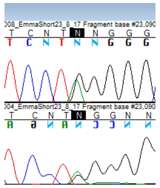
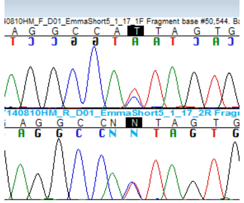
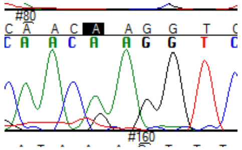
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo17	*DSC3*	18	28605748	C>T	35	Yes	 Read in reverse		Missense
Halo17	*DSC2*	18	28647999	T>TTC	35	Yes	 Read in reverse		Frameshift
Halo18	*CNKSR1*	1	26515380	G>A	33	Yes			Missense
Halo18	*POLQ*	3	121207509	CAATAGT A>C	34	Yes	 Read in reverse		Frameshift

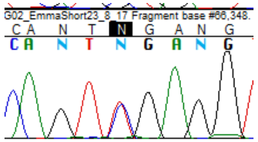
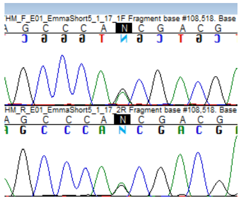
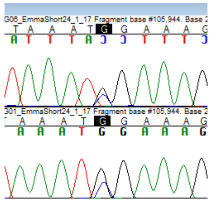
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo18	*CBL*	11	119148958	T>C	23.9	Yes			Missense
Halo18	*RAD51B*	14	68353893	A>G	27.2	Yes			Missense
Halo19	*THRAP3*	1	36757052	G>A	24.3	Yes			Missense

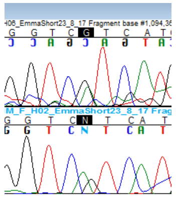
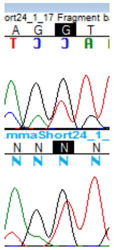
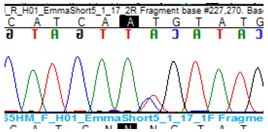
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo19	*RB1*	13	49047524	G>A	28	Yes			Missense
Halo19	*MGA*	15	42028820	A>G	22.2	Yes			Missense

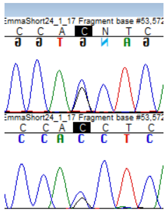
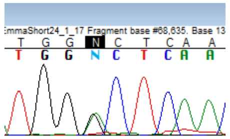
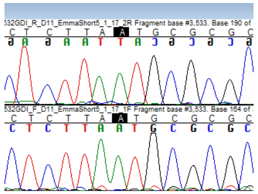
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo19	*WNT9B*	17	44953675	G>A	31	Yes			Missense
Halo19	*MED12*	X	70342603	G>A	33	Yes			Missense
Halo27	*PLK3*	1	45267346	G>T	33	Yes			Missense

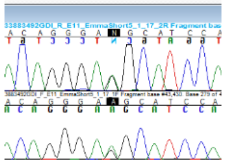
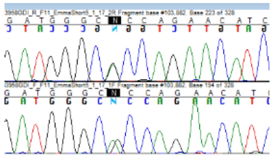
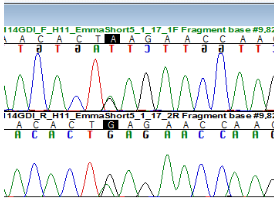
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo27	*RBM5*	3	50155887	TGA>T	35	No		There was relatively low coverage at this locus and the variant was only called in 5/22 reads	Frameshift
Halo27	*TRRAP*	7	98591187	G>C	21.6	Yes			Missense
Halo27	*KAT5*	11	65482096	G>A	28.1	Yes			Missense

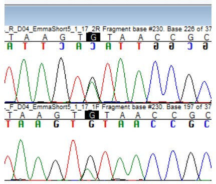
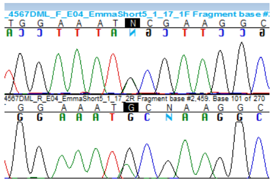
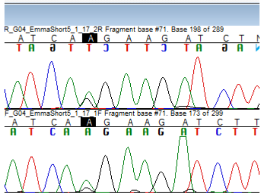
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo27	*PTPRH*	19	55697712	G>A	36	Yes			Stopgain
Halo28	*PIK3C2A*	11	17140810	A>G	23.5	Yes	 Read in reverse		Missense
Halo28	*CD82*	11	44626709	G>A	33	Yes			Missense

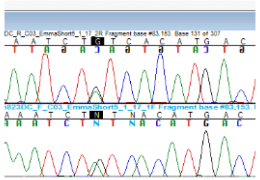
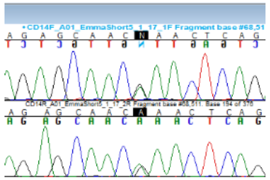
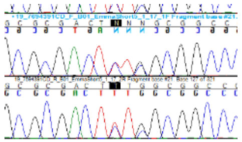
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo28	*RBMS2*	12	56982155	C>T	21.3	Yes			Missense
Halo28	*PPP1R13B*	14	104205127	T>C	23.1	Yes	 Read in reverse		Missense
Halo28	*MGA*	15	42058553	G>C	22	Yes			Missense

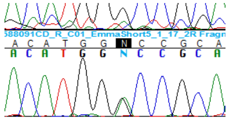
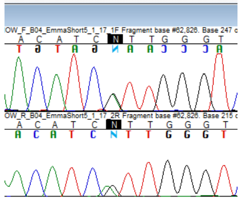
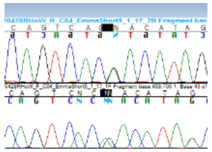
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo28	*DCC*	18	50961517	G>A	28.2	Yes			Missense
Halo28	*BIRC6*	2	32706513	G>T	32	Yes			Missense
Halo28	*EYA4*	6	133789765	C>T	24.8	Yes			Missense

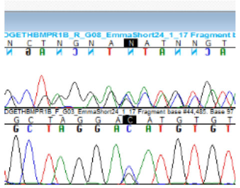
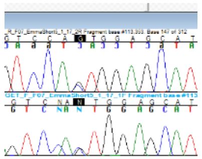
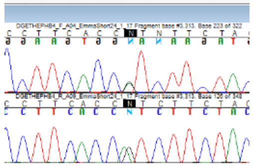
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo28	*AKAP9*	7	91623985	G>C	26	Yes			Missense
Halo28	*MET*	7	116381047	A>G	23.1	Yes			Missense
Halo40	*GDF7*	2	20870532	C>A	23.6	Yes		Variant at position 20870531 (G>A) is a synonymous change	Missense

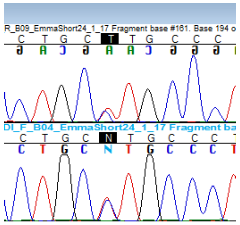
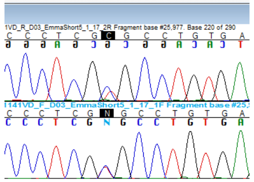
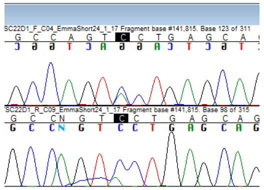
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo40	*PDCD6IP*	3	33883492	G>A	30	Yes			Missense
Halo40	*LRP5*	11	68183958	G>A	23.9	Yes			Missense
Halo40	*DSC2*	18	28672114	C>T	21.7	Yes	 Read in reverse		Missense

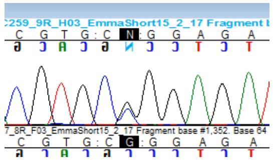
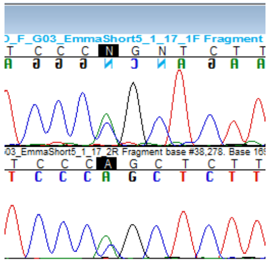
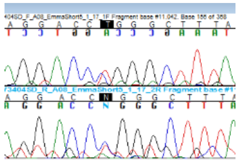
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo45	*EIF3I*	1	32688188	A>G	28.2	Yes			Missense
Halo45	*HELQ*	4	84374567	C>T	25.5	Yes	 Read in reverse		Missense
Halo45	*RPLP2*	11	810305	A>G	22.6	Yes			Missense

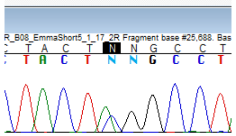
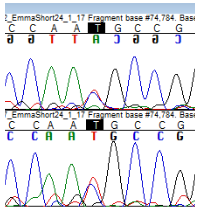
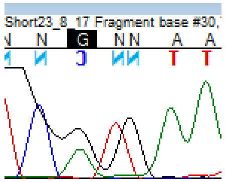
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo46	*MSH4*	1	76345823	A>G	28.6	Yes			Missense
Halo47	*PPP1R13B*	14	104245134	C>T	23.5	Yes			Missense
Halo47	*XAB2*	19	7694391	G>C	23.5	Yes	 Read in reverse	The variant at position 7694393 is a synonymous change	Missense

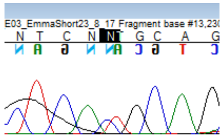
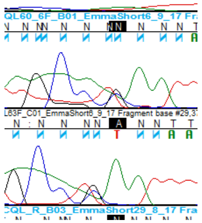
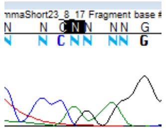
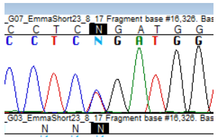
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo47	*ENG*	9	130588091	C>T	23.9	Yes			Missense
Halo48	*ADAM17*	2	9633092	C>T	34	Yes			Missense
Halo48	*FZD3*	8	28420428	G>A	23	Yes			Missense

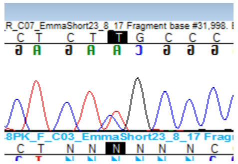
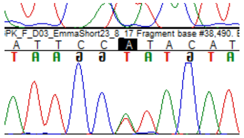
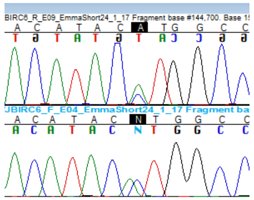
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo49	*BMPR1B*	4	96070060	G>C	33	Yes			Missense
Halo49	*MSH3*	5	80063899	G>C	23.7	Yes			Missense
Halo49	*EPHB4*	7	100421340	C>T	25.2	Yes			Missense

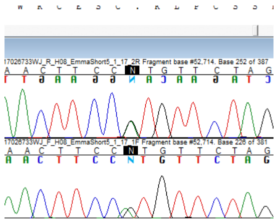
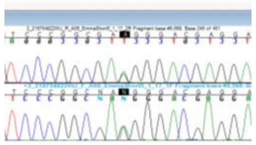
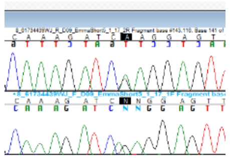
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo49	*ID1*	20	30193351	C>T	29	Yes			Missense
Halo52	*EP400*	12	132471141	C>T	15.34	Yes			Missense
Halo52	*TSC22D1*	13	45008887	G>T	24.6	Yes	 Read in reverse		Missense

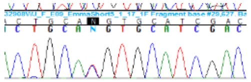
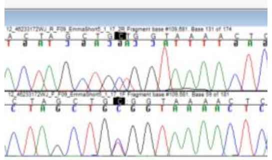
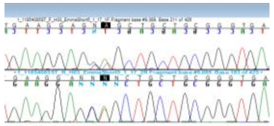
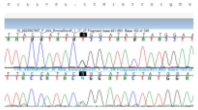
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo52	*FOXC2*	16	86602293	G>C	28.3	Yes			Missense
Halo52	*PHF12*	17	27239701	T>G	18.87	Yes			Missense
Halo55	*DVL1*	1	1273404	G>A	22.8	Yes	 Read in reverse		Missense

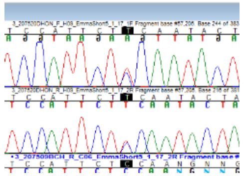
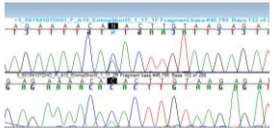
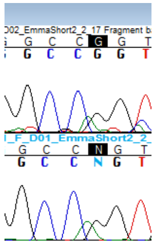
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo55	*BMP8B*	1	40228846	G>T	39	Yes			Stopgain
Halo55	*TLE1*	9	84228372	G>A	29.5	Yes	 Read in reverse		Missense
Halo61	*PARP1*	1	226564855	G>A	26.2	Yes			Missense

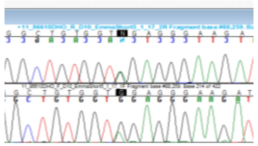
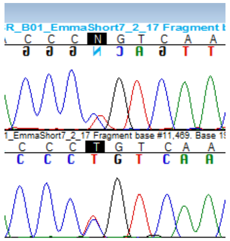
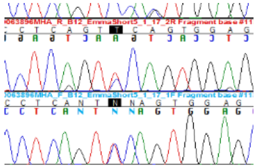
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo61	*ANAPC2*	9	140069828	C>T	31	Yes			Missense
Halo61	*RECQL*	12	21623219	G>C	38	Yes			Stopgain
Halo61	*INHBE*	12	57850383	G>A	33	Yes			Missense
Halo61	*POSTN*	13	38156538	C>T	34	Yes			Missense

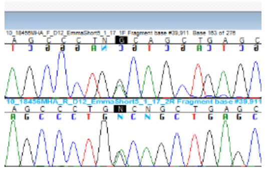
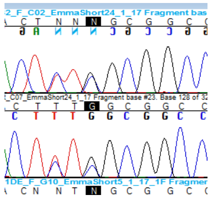
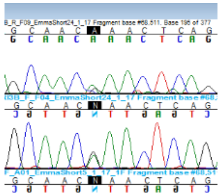
Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo61	*EEF2K*	16	22269048	C>T	26.8	Yes			Missense
Halo61	*ZFP14*	19	36831616	T>A	25.9	Yes			Missense
Halo63	*BIRC6*	2	32726929	A>C	22.3	Yes			Missense

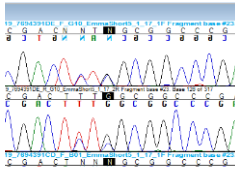
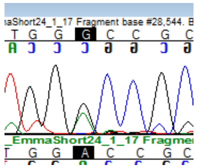
Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo63	*XRCC5*	2	217026733	G>A	24.1	Yes			Missense
Halo63	*WNT10A*	2	219754822	G>A	24	Yes			Missense
Halo63	*CHD7*	8	61734439	G>A	25.7	Yes			Missense

Patient ID	Gene	Chr.	Location	Variant	CADD	Validated	Image	Notes	Genomic Annotation
Halo63	*WISP1*	8	134232908	C>T	24.3	Yes			Missense
Halo63	*ARID2*	12	46233172	C>T	26.4	Yes			Missense
Halo64	*MTHFR*	1	11854085	T>A	23.3	Yes	 Read in reverse	The adjacent variant has a CADD score of 10.80	Missense
Halo64	*WAPAL*	10	88259879	T>C	25	Yes	 Read in reverse		Missense

Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo65	*POLQ*	3	121207520	G>A	24.2	Yes	 Read in reverse		Missense
Halo65	*ERCC8*	5	60194107	G>T	24.4	Yes	 Read in reverse		Missense
Halo65	*NUDT1*	7	2284301	G>A	23.7	Yes			Missense

Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo65	*ATM*	11	108186610	G>A	31	Yes			Missense
Halo65	*TMBIM6*	12	50146761	C>T	24.7	Yes			Missense
Halo66	*MSH3*	5	80063896	C>T	22.8	Yes			Missense

Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo66	*MMS19*	10	99218456	C>T	32	Yes			Missense
Halo66	*XAB2*	19	7694391	G>C	23.5	Yes	 Read in reverse	The variant at position 7694393 is a synonymous change	Missense
Halo68	*PPP1R13B*	14	104245134	C>T	23.5	Yes	 Read in reverse		Missense

Patient ID	**Gene**	**Chr.**	**Location**	**Variant**	**CADD**	**Validated**	**Image**	**Notes**	**Genomic Annotation**
Halo68	*XAB2*	19	7694391	G>C	23.5	Yes	 Read in reverse	The variant at position 7694393 is a synonymous change	Missense
Halo68	*ENG*	9	130588091	C>T	23.9	Yes	 Read in reverse		Missense

Appendix 4.9 Database Interrogation for Validated WES Variants

Please note that the column headings refer to the following:

Exac: Allele frequency in germline exome sequencing data

1000G: Allele frequency in 1000 Genomes Project data: germline

HGMD: The clinical scenario in which an inherited variant has been identified, and its assumed clinical significance (FP: functional polymorphism; ?DM: Likely pathological mutation reported to be disease causing in the corresponding report, but where the author has indicated that there may be some degree of doubt, or subsequent evidence has come to light in the literature, calling the deleterious nature of the variant into question; DM: pathological mutation reported to be disease-causing)

LOVD: The clinical scenario in which an inherited variant has been described, and its concluded pathogenicity (if reported)

COSMIC: The cancer in which the variant has been identified as a somatic mutation. The numbers in brackets refer to the number of samples with the mutation

CBioPortal: The cancer in which the variant has been identified as a somatic mutation. The numbers in brackets refer to the number of samples with the mutation

ClinVar: The clinical scenario in which an inherited variant has been described, and its presumed pathogenicity

CanVar: Allele frequency in patients with colorectal cancer (NSCCG: National Study of Colorectal Cancer Genetics)

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
ADAM17	2:9633092	C>T	34	ENST00000310823	p.Val673Ile	rs61754177	0.009146	0.004	-	-	-	-	-	0.006481 NSCCG
AKAP9	7:91623985	G>C	26	ENST00000356239	p.Gln209His	-	1.669e-05	-	-	-	-	-	-	-
ANAPC2	9:140069828	C>T	31	ENST00000323927	p.Arg706His	rs144294113	0.001547	0.000196805	-	-	-	-	-	0.002857 NSCCG
ARID2	12:46233172	C>T	26.4	ENST00000334344	p.Ala464Val	rs376273452	9.899e-05	-	-	-	-	-	-	-
ATM	11:108186610	G>A	31	ENST00000278616	p.Gly2023Arg	rs11212587	0.001575	0.000199681	?DM breast cancer CM092585	Ataxia telangiectasia	COSM24635 lymphoid neoplasm (2), large	Breast carcinoma (2)	Ataxia telangiectasia syndrome,	-

											intestine (1)		hereditary cancer predisposing syndrome. Conflicting interpretations of pathogenicity	
Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
ATM	11:108196797	G>A	34	ENST00000278616	p.Ala2274Thr	rs567060474	0.0001075	0.000199681	?DM breast cancer CM016181	-	-	-	. Hereditary cancer predisposing syndro	-

													me, ataxia-telangiectasia syndrome NOS. Uncertain significance/ likely benign	
Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
ATR	3:142272170	A>G	20.5	ENST00000350721	p.Ser902Pro	rs146202702	0.0006848	0.00199681	-	-	-	-	-	0.001608 NSCCG
BCLAF1	6:136599544	G>A	36	ENST00000353331	p.Arg157*	rs140155621	0.001105	-	-	-	-	-	-	0.001988 NSCCG

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
BIRC6	2: 32706513	G>T	32	ENST00000421745	p.Val2512Phe	rs146185570	0.001195	0.0001997	-	-	-	-	-	0.001496 NSCCG
BIRC6	2: 32726929	A>C	22.3	ENST00000421745	p.Met3061Leu	-	-	-	-	-	-	-	-	-
BMP1	8:22049596	G>A	33	ENST00000306349	P.Arg371His	rs145284541	0.003708	0.003	-	-	-	NHL (1)	Osteogenesis imperfecta. Uncertain significance	0.006329 NSCCG
BMP8B	1: 40228846	G>T	39	ENST00000372827	p.Ser326*	rs147323607	4.977e-05	-	-	-	-	-	-	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
BMPER	7:34125622	C>T	34	ENST00000297161	p.Arg555Trp	rs10249320	0.004901	0.002	-	-	COSM1755259 urinary tract malignancy (2)	Bladder cancer (1)	Diaphanospondylodysostosis. Uncertain significance	0.01064 NSCCG
BMPR1B	4: 96070060	G>C	33	ENST00000440890	p.Arg443Thr	-	8.239e-06	-	-	-	-	-	-	-
BMPR2	2:203407059	G>A	23.1	ENST00000374574	p.Met434Ile	-	1.65e-05	-	-	-	-	-	Primary pulmonary hypertension. Uncertain significance	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
CAPN9	1:230907799	C>T	35	ENST00000271971	p.Arg277Trp	rs28359655	0.001689	0.0003993610	-	-	-	-	-	0.003083 NSCCG
CBL	11: 119148958	T>C	23.9	ENST00000264033	p.Ile393Thr	-	8.243e-06	-	-	-	-	Pancreatic cancer (1)	-	-
CBL	11:119169085	G>A	21.6	ENST00000264033	p.Ala757Thr	rs146517083	0.001293	0.001	-	-	COSM3687124 large intestine (2)	Penile cancer (1)	Rasopathy, Noonan-like syndrome. Benign/ likely benign	0.001594 NSCCG
CD82	11: 44626709	G>A	33	ENST00000227155	p.Glu80Lys	rs145881169	0.0009743	0.000196805	-	-	-	-	-	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
CELSR2	1: 109812092	G>A	22	ENST00000271332	p.Val2287Ile	rs141489111	0.0008612	-	-	-	COSM6120119 lung carcinoma (1)	CRC (1)	-	-
CHD7	8: 61734439	G>A	25.7	ENST00000423902	p.Glu930Lys	rs377330239	6.334e-05	0.000199681	-	-	-	-	Likely benign ?phenotype	-
CNKSR1	1: 26515380	G>A	33	ENST00000361530	p.Arh603His	-	1.659e-05	-	-	-	-	-	-	-
DAAM2	6: 39864686	C>T	34	ENST00000274867	p.Arg814Cys	rs146966805	0.0003089	0.001	-	-	COSM1444408 large intestine carcinoma (1)	-	-	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
DCC	18:50961517	G>A	28.2	ENST00000442544	p.Arg1056His	rs200519902	0.0001236	0.00018681	-	-	-	-	-	-
DSC2	18:28647999	T>TTC	35	ENST00000280904	p.Ala897Lysfs*4	rs200056085	0.008334	0.004	?DM in ARVD C1065824	-	-	RCC (1)	ARVD/ primary familial hypertrophic cardiomyopathy. Conflicting interpretations of pathogenicity	0.01383 NSCCG

Gene	Posn.	Variant	CADD	Transcript	AA Change	dbSNP ID	Exac	1000G	HGMD	LOVD	Cosmic	CBio Portal	ClinVar	CanVar
DSC2	18: 28672114	C>T	21.7	ENST00000251081	p.Glu102Lys	rs144799937	0.0007352	-	DM ARVD CMO74775	ARVD, pathogenicity unknown	-	-	ARVD. Conflicting interpretations of pathogenicity	-
DSC3	18:28605748	C>T	35	ENST00000360428	p.Arg203His	rs183968347	0.0001901	0.00039936	-	-	-	-	-	0.001008 NSCCG
DVL1	1: 1273404	G>A	22.8	ENST00000378891	p.Pro531Leu	-	9.253e-06	-	-	-	-	-	-	-
EEF2K	16: 22269048	C>T	26.8	ENST00000263026	p.Ser329Leu	-	0.0001403	-	-	-	-	-	-	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
EGFR	7: 55273086	G>A	35	ENST00000275493	p.Glu1137Lys	rs780967013	8.257e-06	-	-	-		Metastatic melanoma (1)	-	-
EIF3I	1: 32688188	A>G	28.2	ENST00000355082	p.Tyr18Cys	rs140164523	4.943e-05	-	-	-	-	-	-	-
ENG	9: 130588091	C>T	23.9	ENST00000344849	p.Gly191Asp	rs41322046	0.01522	0.003	?DM gastrointestinal polyposis CM135239		COSM3763658 haematopoietic and lymphoid neoplasm (2), large intestine carcinoma (1)	-	Osler haemorrhagic telangiectasia syndrome, Juvenile Polyposis. Benign/	0.01896 NSCCG

													Likely benign	
Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
EP400	12: 132471141	C>T	15.34	ENST00000330386	p.Ala671Val	-	7.417e-05	-	-	-	-	-	-	-
EPHB4	7: 100421340	C>T	25.2	ENST00000358173	p.Val113Ile	rs55866373	0.0003605	0.000199681	-	-	-	Carcinoma of the urinary tract (1)	-	-
ERCC8	5: 60194107	G>T	24.4	ENST00000265038	p.Thr280Lys	rs61754098	0.001905	0.000199681	-	-	-	-	Cockayne Syndrome. Likely benign/uncertain	0.00199 NSCCG

													significance	
Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
EYA4	6:133789765	C>T	24.8	ENST00000355167	p.Thr289Met	rs41286200	0.00039936	0.0006274	-	-	-	Cutaneous melanoma (1)	Dilated cardiomyopathy, non-syndromic hearing loss. Uncertain significance	0.001505 NSCCG
FOXC2	16:86602293	G>C	28.3	ENST00000320354	p.Arg451Pro	-	-	-	-	-	-	-	-	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
FZD3	8:28420428	G>A	23	ENST00000240093	p.Arg634Gln	rs147574227	0.0002308	0.0001997	-	-	-	-	-	0.0004985 NSCCG
FZD5	2:208633009	C>T	22.5	ENST00000295417	p.Arg152His	rs61733651	0.004263	0.001	-	-	-	-	-	-
GDF7	2:20870532	C>A	23.6	ENST00000272224	p.Leu234Met	rs184953707	0.01126	0.002	-	-	-	-	-	-
HELQ	4:84374567	C>T	25.5	ENST00000295488	p.Ala277Thr	-	3.323e-05	-	-	-	-	-	-	0.000497 NSCCG
HDAC5	17:42171169	G>A	25.6	ENST00000393622	p.Pro43Leu	rs200315067	1.4e-05	-	-	-	-	-	-	-
ID1	20:30193351	C>T	29	ENST00000376105	p.Pro54Leu	-	-	-	-	-	COSM4989109 rhabdomyosarcoma (1)	-	-	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
INHBE	12: 57850383	G>A	33	ENST00000266646	p.Glu269Lys	rs571755879	1.648e-05	0.000196805	-	-	-	Bladder urothelial carcinoma (1), head and neck SCC (1)	-	-
KAT5	11: 65482096	G>A	28.1	ENST00000341318	p.Arg274Trp	-	-	-	-	-	-	-	-	-
LIG1	19:48620943	C>A	29	ENST00000263274	p.Lys845Asn	rs145821638	0.0008951	-	-	-	-	-	-	0.002677 NSCCG

Gene	Posn.	Variant	CADD	Transcript	AA Change	dbSNP ID	Exac	1000G	HGMD	LOVD	Cosmic	CBio Portal	ClinVar	CanVar
LRP4	11:46918516	G>A	29.3	ENST00000378623	p.Arg276Cys	rs557300271	2.471e-05	0.000199681	-	-	-	-	-	0.0004975 NSCCG
LRP5	11: 68183958	G>A	27	ENST00000294304	p.Arg997His	-	2.49e-05	-	-	-	-	-	-	-
LZTS1	8: 20107358	G>A	34	ENST00000265801	p.Arg556Trp	rs748585946	3.295e-05	-	-	-	COSM3898704 stomach carcinoma (1)	-	-	-
MCPH1	8:6479113	C>T	24.6	ENST00000344683	p.Arg785Trp	-	0.0001077	-	-	-	-	-	-	-
MED12	X: 70342603	G>A	33	ENST00000333646	p.Arg455Gln	-	-	-	-	-	COSM289289 skin (1) and large intestine	CRC (1)	-	-

											carcinoma (1)			
Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
MET	7: 116381047	A>G	23.1	ENST00000318493	p.Thr557Ala	rs374733251	6.625e-05	-	-	-	-	-	RCC, hereditary cancer predisposing syndrome. Uncertain significance	-
MGA	15: 42028820	A>G	22.2	ENST00000219905	p.Tyr1453Cys	rs2695167	0.008707	0.005	-	-	-	-	-	0.02143 NSCCG

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
MGA	15: 42058553	G>C	22	ENST00000219905	p.Trp2758Ser	rs201447485	0.000988	0.000399	-	-	-	-	-	0.004482 NSCCG
MMS19	10: 99218456	C>T	32	ENST00000438925	p.Gly1029Asp	rs36023427	0.007004	0.002	-	-	-	-	-	0.004532 NSCCG
MSH3	5: 80063896	C>T	22.8	ENST00000265081	p.Pro681Ser	rs115198722	0.0007755	0.000199681	?DM colon cancer CM015300	-	-	-	-	0.002991 NSCCG
MSH3	5: 80063899	G>C	23.7	ENST00000265081	p.Val682Leu	rs145657887	0.0001815	-	-	-	-	-	-	-
MSH4	1: 76345823	A>G	28.6	ENST00000263187	p.Tyr589Cys	rs5745459	0.009125	0.004	FP variant associa	-	-	NSCLC (1)	-	0.006474 NSCCG

									ted with genome wide recombination CM140046					

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
MTHFR	1: 11854085	T>A	23.3	ENST00000376590	p.Glu470Val	rs142617551	0.00144	0.00039936	-	-	-	-	Neural tube defects, folate sensitive. Uncertain significance	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
NCOR2	12:124824869	G>A	24.3	ENST00000405201	p.Thr1820Met	rs61755988	0.004692	0.001	-	-	-	NHL (1)	-	0.009967 NSCCG
NCOR2	12:124911260	C>T	24.6	ENST00000405201	p.Met412Leu	-	-	-	-	-	-	-	-	-
NUDT1	7:2284301	G>A	23.7	ENST00000397048	p.Arg54Gln	rs139825597	0.0001274	0.000199681	-	-	-	Mesothelioma (1)	-	-
PALB2	16:23635370	C>T	26.6	ENST00000261584	p.Val932Met	rs45624036	0.005972	0.001	?DM breast cancer CM11210	Breast Ca and prostate Ca. Concluded pathogenicity unknown	-	-	Familial breast cancer, Pancreatic cancer, hereditary cancer predisp	0.006958 NSCCG

													osing syndrome. Benign/ likely benign.	
Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
PARP1	1: 226564855	G>A	26.2	ENST00000366794	p.Thr632Met	rs138228205	0.000766	0.000196805	-	-	COSM1219296 large intestine carcinoma (1) and rhabdomyosarcoma (1)	CRC (1)	-	-
PDCD6IP	3: 33883492	G>A	30	ENST00000307296	p.Gly434Ser	rs148256302	0.007635	0.005	-	-	-	-	-	0.01594 NSCCG

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
PHF12	17: 27239701	T>G	18.87	ENST00000268756	p.Ser630Arg	rs116895969	0.007401	0.004	-	-	-	-	-	0.008037 NSCCG
PIK3C2A	11: 17140810	A>G	23.5	ENST00000265970	p.Ile966Thr	-	8.277e-06	-	-	-	-	-	-	-
PLK3	1: 45267346	G>T	33	ENST00000372201	p.Arg163Leu	rs142872152	0.0006378	0.001	-	-	-	-	-	0.003264 NSCCG
POLQ	3: 121207509	CAATAGTA>C	34	ENST00000264233	p.Ile1421Argfs*8	-	0.005876	-	-	-	-	-	-	0.00996 NSCCG
POLQ	3: 121207520	G>A	24.2	ENST00000264233	p.Pro1420Ser	-	-	-	-	-	-	-	-	-
POSTN	13: 38156538	C>T	34	ENST00000379747	p.Gly453Arg	rs117103342	0.004362	0.003	-	-	-	NHL (1)	-	0.0111 NSCCG

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
PPP1R13B	14: 104205127	T>C	23.1	ENST00000202556	p.Asn918Ser	rs368408684	8.281e-06	-	-	-	-	-	-	-
PPP1R13B	14: 104245134	C>T	23.5	ENST00000202556	p.Arg101Gln	-	-	-	-	-	COSM284182 liver carcinoma (1), endometrial carcinoma (1), large intestine carcinoma (1)	CRC (1), uterine carcinoma (1)	-	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
PTPRH	19: 55697712	G>A	36	ENST00000376350	p.Gln887*	rs147881000	0.002689	0.002	?DM Parkinson's Disease CM171953	-	-	-	-	0.002654 NSCCG
RAD51B	14: 68353893	A>G	27.2	ENST00000390683	p.Lys243Arg	rs34594234	0.007347	0.003	-	-	-	Glioma (1)	-	0.01212 NSCCG
RB1	13: 49047524	G>A	28	ENST00000267163	p.Gly840Arg	rs374157786	8.285e-06	-	-	-	-	-	Retinoblastoma. Uncertain significance	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
RBMS2	12: 56982155	C>T	21.3	ENST00000262031	Miscoded in exome results. Synonymous change p.Val380Val	rs143825451	0.002512	0.002	-	-	-	-	-	0.005473 NSCCG
RECQL	12: 21623219	G>C	38	ENAT00000444129	p.Ser620*	rs142038240	0.001073	0.000196805	-	-	-	-	-	-
RPLP2	11: 810305	A>G	22.6	ENST00000321153	p.Lys24Arg	rs111850318	0.0008147	-	-	-	-	-	-	-
TSC22D1	13: 45008887	G>T	24.6	ENST00000261489	p.Pro1033Thr	-	1.657e-05	-	-	-	-	-	-	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
THRAP3	1: 36757052	G>A	24.3	ENST00000354618	p.Arg608His	rs41303413	0.0006344	-	-	-	-	-	-	0.001992 NSCCG
TLE1	9: 84228372	G>A	29.5	ENST00000376499	p.Thr328Met	rs144027618	0.003419	0.001	-	-	-	-	-	0.004886 NSCCG
TMBIM6	12: 50146761	C>T	24.7	ENST00000552699	p.Pro79Leu	rs58919844	6.595e-05	-	-	-	-	-	-	-
TTI2	8:33361016	C>T	24.6	ENST00000360742	p.Gly397Gln	rs150984360	0.0005271	0.000199681	-	-	-	-	-	0.000998 NSCCG
TRRAP	7: 98591187	G>C	21.6	ENST00000446306	p.Val3267Leu	rs370895367	8.25e-05	-	-	-	-	-	-	-

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
SFN	1: 27190196	A>T	26.1	ENST00000339276	p.Thr165Ser	rs77755255	0.0002473	-	-	-	-	NHL (1)	-	-
ST18	8: 53030923	G>T	31	ENST00000276480	p.Ala954Glu	rs117471862	0.00358	0.001	-	-	-	-	-	0.005473 NSCCG
WAPAL	10: 88259879	T>C	25	ENST00000298767	p.Gln374Arg	rs140780773	0.0005768	0.000199681	-	-	-	-	-	-
WISP1	8: 134232908	C>T	24.3	ENST00000250160	p.Thr145Met	rs139669488	0.00149	0.001	-	-	COSM1165967 large intestine adenoma (1), stomach carcinoma (1), liver	Oesophagogastric carcinoma (1)	-	0.001805 NSCCG

											carcinoma (1)			
Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
WNT9B	17: 44953675	G>A	31	ENST00000290015	p.Arg222His	rs138314634	0.001119	0.000399	DM in Mayer-Rokitansky-Küster Hauser syndrome CM166073	-	-	-	-	-
WNT10A	2: 219754822	G>A	24	ENST00000258411	p.Gly165Arg	rs77583146	0.007335	0.002	?DM in tooth agenesis CM138364	-	-	NHL (1)	Odontoonychodermal dysplasia, Schopf-Schulz-Passar	0.009958 NSCCG

													ge syndrome, Selective tooth agenesis. Likely benign	
Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
XAB2	19: 7694391	G>C	23.5	ENST00000358368	p.Ser8Trp	rs146660753	0.004467	0.000399	-	-	-	-	-	0.002198 NSCCG
XRCC5	2: 217026733	G>A	24.1	ENST00000392132	p.Arg599His	rs55943434	4.121e-05	0.00199681	-	-	-	-	-	-
ZFP14	19: 36831616	T>A	25.9	ENST270001	p.Lys371Met	rs139225970	0.003222	0.001	-	-	-	-	-	0.01441 NSCCG

Gene	**Posn.**	**Variant**	**CADD**	**Transcript**	**AA Change**	**dbSNP ID**	**Exac**	**1000G**	**HGMD**	**LOVD**	**Cosmic**	**CBio Portal**	**ClinVar**	**CanVar**
ZKSCAN4	6: 28219686	CGGTCA>C	24.4	ENST00000377294	p.Leu23Argfs*83	-	-	-	-	-	-	-	-	-

Appendix 5.1 Protocol for Microsatellite Instability Testing

For MSI testing, the Promega MSI Analysis System allowed co-amplification and detection of a panel of microsatellite markers that have been shown to be sensitive and specific for detection of MSI samples with mismatch repair deficiencies. The AWMGS protocol is outlined as follows:

1. Generate a worksheet in Shire for the samples to be tested. This worksheet automatically includes normal and NTC samples.
2. In Shire look up the DNA concentration for each patient and record this in the Con/Dil box against the corresponding patient on the worksheet.
3. Calculate from the DNA concentration the dilution required to achieve a final concentration of 5ng/µl and record this on the worksheet.
4. In QPulse print off a copy of the MSI backing sheet (LF-GEN-MSIwksbk) and record the worksheet number in the space provided.
5. In pre-PCR remove the MSI v1.2 kit from the freezer and remove the Gold* Buffer, Primer Pair Mix, H_20, Ampitaq Gold and the K562 Normal Control and allow to defrost. Vortex these reagents for 5-10 seconds before each use. A precipitate may form in the Gold ST*R Buffer. If this occurs, warm the buffer briefly at 37°C, then vortex until it is in solution. Please note the AmpliTaq Gold® DNA polymerase is not included in the MSI Promega kit.
6. In a pre-PCR Lab choose an appropriate rack and place the samples to be tested in worksheet order.
7. Label 0.5ml eppendorf tubes with the sample number on the top and sample number and initials on the side and place next to the corresponding sample in the rack.
8. Label a 96-well plate with the worksheet number and highlight the wells to be used.
9. At this point a DNA dilution and pre-PCR tube check is required before continuing.
10. Dilute the samples according to the calculations on the worksheet and dilute the K562 Nomal Control 1/10.
11. From the MSI backing sheet calculate the amounts required for the amplification mix (no. of samples +1) and make up this mix.

Component	**Volume per sample (µl)**
Nuclease-free water	5.85
Gold ST*R 10X Buffer	1.00
MSI 10X Primer Pair Mix	1.00
AmpliTaq Gold® DNA polymerase (5u/µl)	0.15
DNA (1-10ng/µl)	2.0
Total reaction volume	10.0

12. Add 8ml of the amplification mix to each well of the labelled plate and 2µl of the DNA dilution to the correct well. Seal well with a sticky lid, vortex and pulse spin.
13. Run the plate according to the following PCR conditions.

	Temperature (°C)	**Time**	**Cycles**
	95°C	**11 min**	
	96°C	**1 min**	
ramp 29% (0.5°C/sec) ramp 23% (0.2°C/sec)	**94°C** **to 58°C** **to 70°C**	**30 sec** 30 sec 1 min	**11**
ramp 29% (0.5°C/sec) ramp 23% (0.2°C/sec)	**90°C** **to 58°C** **to 70°C**	**30 sec** 30 sec 1 min	**22**
	60°C	30 min	
	4°C	∞	

14. Record the lot number of each reagent on the MSI backing sheet.
15. After completion of the thermal cycling protocol, store the samples at -20°C**,** protected from light.

16. The 3730-plate record can be made either directly on the 3730-data collection software or using the Excel spreadsheet template in: S:\MedGen\SHARED\MOLEC\Services\3730_Fragment_Analysis. Each injection will take approximately an hour.
17. Prepare the loading mix as follows:
    - 7.6µl HiDi
    - 0.4µl ILS 600 (Internal Lane Standard 600).
    - The size standard is included in the MSI kit and it is stored at -20°C.
18. Denature at 95°C for 5 mins and store on ice for at least 3mins.
19. Analyse on ABI 3730 using POP7 and the following conditions

Instrument protocol	Results group
MSI_2kV_6s_1800s	FRAGMENT_ANALYSIS

20. After completion of the run transfer the raw data from the 3730 to a PC with PeakScanner software using a pen drive. 3730 raw data are stored in M:\ Fragment Sizing/MSI RESULTS/ Year and save in the Raw Data folder.
21. Open PeakScanner software and add each patient into a new project. Select size standard *ILS600-1* (1) and analysis method *Sizing Default –PP-1* (2) then click the green Analysis button (3).
22. The NTC and NORMAL can be added in the same project.
23. Next select the edit table tab (4) and select ‘MSI Analysis Sep2015’ from the drop-down menu. DO NOT SAVE, just close the box. This setting aligns the trace to enable correct sizing.
24. Select the dye colours blue, green and yellow (black) and zoom in on the region of interest. Click above the top axis where you want to start and drag the box right. The y axis can be pulled down to just above the highest peak.
25. Save each patients file in M/Fragment Sizing/MSI ANALYSIS/Year/Analysed Data/Worksheet Number/Patient Name.

Appendix 5.2 Tumours Undergoing MSI Testing

Haloplex ID	**Diagnosis**	**Reference**
Halo26	TVA LGD	Halo26T1
Halo26	TVA LGD	Halo26T2
Halo26	VA LGD	Halo26T3
Halo26	TA LGD	Halo26T4
Halo26	TA LGD	Halo26T5
Halo45	TA LGD	Halo45T1
Halo45	HPP	Halo45T2
Halo45	TA LGD	Halo45T3
Halo70	HPP	Halo70T1
Halo70	TA LGD	Halo70T2
Halo70	TA LGD	Halo70T3
Halo70	TA LGD	Halo70T4
Halo70	TA LGD	Halo70T5
Halo70	TA LGD	Halo70T6
Halo70	TA LGD	Halo70T7
Halo70	TA LGD	Halo70T8
Halo70	TA LGD	Halo70T9
Halo70	TVA LGD	Halo70T10
Halo70	TVA LGD	Halo70T11
Halo70	TA LGD	Halo70T12

Appendix 5.3a Human POLD1 Sequence Alignment Onto Model 3IAY

Length: 908 **E-value:** 0.0 **Score:** 864.759bits

(2233) **Identities:** 437/908(48%) **Positives:** 589/908(65%) **Gaps:** 27/908(3%)

```
81      90       100      110      120      130      140      150      160      170      180
    190
RPTPPALDPQTEPLIFQQLEIDHYXXXXXXXXXXXXXXXXXXXXLRAFGVTDEGFS
VCCHIHGFAPYFYTPAPPGFGPEHMGDLQRELNLAISRDSRGGRELTGPAVLAV
ELCSRESM+  P   DP    + FQQ++ +                  +R FGVT EG SV C++ GF  Y Y PA
P        + + ++         E    A+ ++E+ S++S+
KKLPTDFDPSLYDISFQQIDAEQ-------
SVLNGIKDENTSTVVRFFGVTSEGHSVLCNVTGFKNYLYVPAPNSSDANDQEQIN
KFVHYL--------NETFDHAIDSIEVVSKQSI
34   40      50           60       70       80       90       100      110           120
130
```

```
 200      210      220      230       240      250      260      270       280      290
300      310
FGYHGHGPSPFLRITXXXXXXXXXXXXXXEQG-
IRVAGLGTPSFAPYEANVDFEIRFMVDTDIVGCNWLELPAGKYAL-
RLKEKATQCQLEADVLWSDVVSHPPEGPWQRIAPLRVLS
+GY G    PF +I            E+G +      +     Y+ N+ + +R MVD  IVG +W+ LP GKY++     +
+ CQLE  + + ++++HP EG W   APLR++S
WGYSGDTKLPFWKIYVTYPHMVNKLRTAFERGHLSFNSWFSNGTTTYD-
NIAYTLRLMVDCGIVGMSWITLPKGKYSMIEPNNRVSSCQLEVSINYRNLIAHPAE
GDWSHTAPLRIMS
  140      150      160      170      180       190      200      210      220      230
240      250
```

```
   320      330      340      350      360      370      380      390      400      410
 420      430
FDIECAGRKGIFPEPERDPVIQICSLGLRWGEPEPFLRLALTLRPCAPILGAKVQS
YEKEEDLLQAWSTFIRIMDPDVITGYNIQNFDLPYLISRAQTLKVQTFPFLGRVAGL
CSNIR
```

```
FDIECAGR G+FPEPE DPVIQI ++    G  +PF+R   TL  C+PI G+ + S+  EE++L  W  FI
+DPDVI GYN  NFD+PYL++RA+ LKV  FP+ GR+  +   I+
FDIECAGRIGVFPEPEYDPVIQIANVVSIAGAKKPFIRNVFTLNTCSPITGSMIFSHA
TEEEMLSNWRNFIIKVDPDVIIGYNTTNFDIPYLLNRAKALKVNDFPYFGRLKTVKQ
EIK
    260       270       280       290       300       310       320       330       340       350
  360       37


     440       450       460       470       480       490       500       510       520       530
  540       5
DSSFQSKQTGRRDTKVVSMVGRVQMDMLQVLLREYKLRSYTLNAVSFHFLGEQ
KEDVQHSIITDLQNGNDQTRRRLAVYCLKDAYXXXXXXXXXXXXXNAVEMARVT
GVPLSYLLSRG+S F SK  G R+TK V++ GR+Q+D+LQ + REYKLRSYTLNAVS HFLG
EQKEDV +SII+DLQNG+ +TRRRLAVYCLKDAY          N  EMARVTGVP SYLL+RG
ESVFSSKAYGTRETKNVNIDGRLQLDLLQFIQREYKLRSYTLNAVSAHFLGEQKE
DVHYSIISDLQNGDSETRRRLAVYCLKDAYLPLRLMEKLMALVNYTEMARVTGVP
FSYLLARG
0      380       390       400       410       420       430       440       450       460       470
   480
```

Appendix 5.3b   Human POLE Sequence Alignment Onto Model 4M8O

56.9% identity in 1112 residues overlap; Score: 3181.0; Gap frequency: 2.5%

```
 80 SAVDYYFIQDDGSRFKVALPYKPYFYIATRKGCE-
REVSSFLSKKFQGKIAKVETVPKED
115
SGVDFYFLDEEGGSFKSTVVYDPYFFIACNDESRVNDVEELVKKYLESCLKSLQII
RKED
   * ** **    *  **    * *** **        *     *            ***


139
LDLPNHLVGLKRNYIRLSFHTVEDLVKVRKEISPAVKKNREQDHASDAYTALLSS
VLQRG
175 LTMDNHLLGLQKTLIKLSFVNSNQLFEARKLLRPIL-----
QDNANNNVQRNIYNVAANG
```

```
   *   *** **    * ***     *   **    *       ** *          *    *

199
GVITDEEETSKKIADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWYNVRYRGNAFP
VEIT
230 SEKVDAKHL----------IEDIREYDVPYHVRVSIDKDIRVGKWYKVTQQGF---IEDT
      *            * * ******** * ***  * *  ** *   *     * *

259
RRDDLVERPDPVVLAFDIETTKLPLKFPDAETDQIMMISYMIDGQGYLITNREIVSE
DIE
277 RK---
IAFADPVVMAFAIATTKPPLKFPDSAVDQIMMISYMIDGEGFLITNREIISEDIE
   *       **** ** * *** ******   ************ * ******* *****

319
DFEFTPKPEYEGPFCVFNEPDEAHLIQRWFEHVQETKPTIMVTYNGDFFDWPFVE
ARAAV
334
DFEYTPKPEYPGFFTIFNENDEVALLQRFFEHIRDVRPTVISTFNGDFFDWPFIHNR
SKI
   *** ****** * *  *** **  * ** ***     **    * *********   *

379
HGLSMQQEIGFQKDSQGEYKAPQCIHMDCLRWVKRDSYLPVGSHNLKAAAKAK
LGYDPVE
394
HGLDMFDEIGFAPDAEGEYKSSYCSHMDCFRWVKRDSYLPQGSQGLKAVTQSK
LGYNPIE
   *** *  ****  *  ****   * **** ********** **  ***    **** * *

439
LDPEDMCRMATEQPQTLATYSVSDAVATYYLYMKYVHPFIFALCTIIPMEPDEVLR
KGSG
454
LDPELMTPYAFEKPQHLSEYSVSDAVATYYLYMKYVHPFIFSLCTIIPLNPDETLR
KGTG
```

```
    **** *   * * ** *  ********************** ******  *** **** *

499
TLCEALLMVQAFHANIIFPNKQEQEFNKLTDDGHVLDSETYVGGHVEALESGVFR
SDIPC
514 TLCEMLLMVQAYQHNILLPNKHTDPIERFYD-
GHLLESETYVGGHVESLEAGVFRSDLKN
    **** ******   **  ***        * ** * ********** ** ******

559
RFRMNPAAFDFLLQRVEKTLRHALEEEEKVPVEQVTNFEEVCDEIKSKLASLKDV
PSRIE
573
EFKIDPSAIDELLQELPEALKFSVEVENKSSVDKVTNFEEIKNQITQKLLELKENNI
RNE
     *   * * * ***    *    * * *  *  ******    *  **  **    * *

619
CPLIYHLDVGAMYPNIILTNRLQPSAMVDEATCAACDFNKPGANCQRKMAWQW
RGEFMPA
633
LPLIYHVDVASMYPNIMTTNRLQPDSIKAERDCASCDFNRPGKTCARKLKWAWR
GEFFPS
     ***** **  *****  ******    *  ** **** **  * **  * ***** *

679 SRSEYHRIQHQLESEKFPPL--
FPEGPARAFHELSREEQAKYEKRRLADYCRKAYKKIHI
693
KMDEYNMIKRALQNETFPNKNKFSKKKVLTFDELSYADQVIHIKKRLTEYSRKVY
HRVKV
       **  *   *  * **    *      * ***   *    * **  * ** *

737
TKVEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKLSAAVEVGDAAE
VKRCKN
753 SEIVEREAIVCQRENPFYVDTVKSFRDRRYEFKGLAKTWKGNLSK-
IDPSDKHARDEAKK
```

```
      **    ***** ******  *********** * **  **     *     *

797
MEVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFTGANIITQARELIE
QIGRP
812
MIVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGITCLTGATIIQMARALVE
RVGRP
   * *********** ************* ********** * *** **  ** * *  ***

857
LELDTDGIWCVLPNSFPENFVFKTTNVKKPKVTISYPGAMLNIMVKEGFTNDQYQ
ELAEP
872 LELDTDGIWCILPKSFPETYFFTLENGKK--
LYLSYPCSMLNYRVHQKFTNHQYQELKDP
   ********** ** ****   *   * **     ***  ***  *   *** *****  *

917
SSLTYVTRSENSIFFEVDGPYLAMILPASKEEGKKLKKRYAVFNEDGSLAELKGF
EVKRR
930
LNYIYETHSENTIFFEVDGPYKAMILPSSKEEGKGIKKRYAVFNEDGSLAELKGFE
LKRR
      * * *** ********* ***** ******  ******************** ***

977
GELQLIKIFQSSVFEAFLKGSTLEEVYGSVAKVADYWLDVLYSKAANMPDSELFE
LISEN
990
GELQLIKNFQSDIFKVFLEGDTLEGCYSAVASVCNRWLDVLDSHGLMLEDEDLVS
LICEN
   ******* ***  *  ** * ***  *  ** *  ***** *     *  *  ** **

1037
RSMSRKLEDYGEQKSTSISTAKRLAEFLGDQMVKDAGLSCRYIISRKPEGSPVTE
RAIPL
```

1050
RSMSKTLKEYEGQKSTSITTARRLGDFLGEDMVKDKGLQCKYIISSKPFNAPVTERAIPV
**** * * ****** ** ** *** **** ** * **** ** ********

1097
AIFQAEPTVRKHFLRKWLKSSSLQDFDIRAILDWDYYIERLGSAIQKIITIPAALQQVKN
1110
AIFSADIPIKRSFLRRWTLDPSLEDLDIRTIIDWGYYRERLGSAIQKIITIPAALQGVSN
*** * *** * ** * *** * ** ** ****************** * *

1157 PVPRVKHPDWLHKKLLEKNDVYKQKKISELFT
1170 PVPRVEHPDWLKRKIATKEDKFKQTSLTKFFS
***** ***** * * * ** *

* indicates identity.

Appendix 5.4 Tumour Material for Carriers of *POLE/POLD1* Variants

Haloplex Number	**Case Reference**	**Diagnosis**	**% of Epithelium that is Dysplastic**
Halo15	1	TA LGD	20
Halo15	Halo15Pa	TA LGD	10
Halo15	Halo15Pb	TA LGD	70
Halo15	Halo15Pc	TA LGD	20
Halo15	2	TA LGD	30
Halo15	3	TA LGD	40
Halo15	4	TA LGD	30

Haloplex Number	**Case Reference**	**Diagnosis**	**% of Epithelium that is Dysplastic**
Halo15	5	TA LGD	40
Halo15	6	TA LGD	25
Halo15	7	TA LGD	60
Halo15	8	TA LGD	75
Halo31	9	TA LGD	90
Halo31	10	TA LGD	50
Halo31	11	TA LGD	50
Halo31	13	TA LGD	70
Halo31	14	TA LGD	20
Halo31	15	HPP	10
Halo31	16	TA LGD	30
Halo31	17	TA LGD	40
Halo31	18	HPP	90
Halo31	19	TA LGD	30
Halo31	20	TA LGD	80
Halo31	21	TA LGD	30
Halo61	22	TA LGD	10
Halo61	23	TA LGD	90
Halo61	24	TA LGD	30
Halo61	25	TA LGD	40
Halo69	26	TA LGD	30

Haloplex Number	**Case Reference**	**Diagnosis**	**% of Epithelium that is Dysplastic**
Halo69	27	HPP	60
Halo69	28	HPP	90
Halo69	29	HPP	80
Halo69	30	HPP	80
Halo69	31	HPP	95
Halo69	32	HPP	50
Halo69	33	TA LGD	40
Halo69	34	TA LGD	30
Halo69	36	HPP	20
Halo69	37	HPP	30
Halo69	38	HPP	40
Halo69	39	TA LGD	40
Halo69	40	TA LGD	90
Halo69	41	TA LGD	100
Halo69	42	HPP	30
Halo69	43	Adenocarcinoma	100
Halo81	44	HPP	100
Halo81	45	TA LGD	30
Halo81	46	TA LGD	40
Halo81	47	HPP	90
Halo81	48	TA LGD	50

Haloplex Number	**Case Reference**	**Diagnosis**	**% of Epithelium that is Dysplastic**
Halo81	49	HPP	100
Halo81	50	TVA LGD	50
Halo81	51	TVA LGD	95
Halo81	52	HPP	95
Halo81	53	HPP	50
Halo81	54	HPP	30
Halo81	55	HPP	60
Halo81	56	HPP	100
Halo81	58	TVA arising in a SSL	80% adenoma, 20% SSL
Halo81	59	HPP	95
Halo81	60	TVA LGD	>95

Appendix 5.5 Primers for *APC* and *KRAS* Mutation Hotspot Screening

Location of Variant	**Forward Primer**	**Reverse Primer**	**Amplicon Size (bp)**
APC codon 1114	TTTGGACAGCAGGAATGTGTTT	TCTTCTTGACACAAAGACTGGC	122
APC codon 1338	CACAGCACCCTAGAACCAAATC	TAGTGTTCAGGTGGACTTTTGG	148
KRAS codon 146	ACAGGCTCAGGACTTAGCAA	GAAGCAATGCCCTGTCAAGA	218

Appendix 5.6 Tumours Available for MMR IHC and MSI Testing

Haploplex ID	Case Reference	Diagnosis	% of Epithelium that is Dysplastic
Halo15	1	TA LGD	20
Halo15	Halo15Pa	TA LGD	10
Halo15	Halo15Pb	TA LGD	70
Halo15	Halo15Pc	TA LGD	20
Halo15	2	TA LGD	30
Halo15	3	TA LGD	40
Halo15	4	TA LGD	30
Halo15	5	TA LGD	40
Halo15	6	TA LGD	25
Halo15	7	TA LGD	60
Halo15	8	TA LGD	75
Halo27	Halo27T1	TA LGD	50
Halo27	Halo27T2	TA LGD	50
Halo27	Halo27T3	TA LGD	15
Halo27	Halo27T4	TA LGD	60

Haloplex ID	Case Reference	Diagnosis	% of Epithelium that is Dysplastic
Halo27	Halo27T5	TA LGD	70
Halo27	Halo27T6	TA LGD	100
Halo27	Halo27T7	TA LGD	30
Halo27	Halo27T8	TA LGD	5
Halo27	Halo27T9	TA LGD	15
Halo27	Halo27T10	TA LGD	70
Halo27	Halo27T11	TA LGD	60
Halo27	Halo27T12	SSL (no dysplasia)	80
Halo27	Halo27T13	TA LGD	30
Halo27	Halo27T14	TA LGD	25
Halo27	Halo27T15	TA LGD	Dysplastic regions marked
Halo27	Halo25T16	TA LGD	20
Halo27	Halo27T17	TA LGD	30
Halo27	Halo27T18	TA LGD	40
Halo61	24	TA LGD	30
Halo61	22	TA LGD	30

Halo61	23	TA LGD	50
Halo61	Halo61T1	TA LGD	80
Halo61	Halo61T2	TA LGD	80

Appendix 5.7 Primers for cDNA Validation of *POLE* c.1411A>G:p.Met471Val

Forward Primer	**Reverse Primer**	**Amplicon Size (bp)**
GCTAGGCTATGATCCCGTGG	CATCAGCAAGGCCTCACAC	226

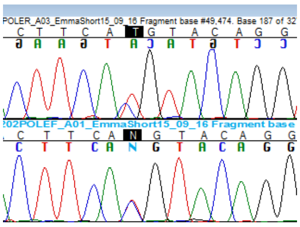
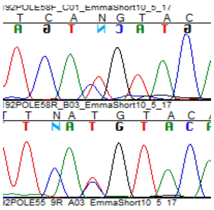
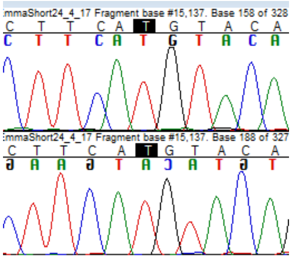
Appendix 5.8 Results of MMR IHC and MSI Testing in Tumours Arising in Carriers of *Pol* Gene Variants

Patient ID	**Case Reference**	**Diagnosis**	**MMR IHC**	**MSI Status**
Halo15	1	TA LGD	Normal	MSS 0/5
Halo15	Halo15Pa	TA LGD	Normal	MSS 0/5
Halo15	Halo15Pb	TA LGD	Normal	MSS 0/5
Halo15	Halo15Pc	TA LGD	Normal	MSS 0/5
Halo15	2	TA LGD	Normal	MSS 0/5
Halo15	3	TA LGD	Normal	MSS 0/5
Halo15	4	TA LGD	Normal	MSS 0/5
Halo15	5	TA LGD	Normal	MSS 0/5

Halo15	6	TA LGD	Normal	MSS 0/5
Halo15	7	TA LGD	Normal	MSS 0/5
Halo15	8	TA LGD	Normal	MSS 0/5
Halo27	Halo27T1	TA LGD	Normal	MSS 0/5
Halo27	Halo27T2	TA LGD	Normal	MSS 0/5
Halo27	Halo27T3	TA LGD	Scanty staining of all antibodies	MSS 0/5
Halo27	Halo27T4	TA LGD	Normal	MSS 0/5
Halo27	Halo27T5	TA LGD	Patchy staining across all samples	MSS 0/5
Halo27	Halo27T6	TA LGD	Normal	MSS 0/5
Halo27	Halo27T8	TA LGD	Normal	MSS 0/5
Halo27	Halo27T9	TA LGD	Normal	-
Halo27	Halo27T10	TA LGD	Normal	-
Halo27	Halo27T11	TA LGD	Normal	MSS 0/5
Halo27	Halo27T12	TA LGD	Normal	MSS 0/5
Halo27	Halo27T13	SSL (no dysplasia)	Normal	-

Halo27	Halo27T14	TA LGD	Normal	-
Halo27	Halo27T15	TA LGD	Normal	-
Halo27	Halo25T16	TA LGD	Normal	MSS 0/5
Halo27	Halo27T17	TA LGD	Normal	-
Halo27	Halo27T18	TA LGD	Normal	MSS 0/5
Halo27	Halo27T19	TA LGD	Normal	MSS 0/5
Halo61	24	TA LGD	Normal	MSS 0/5
Halo61	22	TA LGD	Normal	MSS 0/3 (2 markers repeatedly failed)
Halo61	23	TA LGD	Normal	MSS 0/4 (1 marker repeatedly failed)
Halo61	Halo61T1	TA LGD	Normal	MSS 0/5
Halo61	Halo61T2	TA LGD	Normal	MSS 0/5

Appendix 5.9 Family Testing for *POLE* c.1411A>G:p.Met471Val

Relative	**POLE p.Met471Val Variant Present/ Absent**	**Sequencing Result: DNA**	**Sequencing Result: cDNA**
Halo15 daughter	Present		
Halo15 son	Absent		N/A

Appendix 5.10 Somatic Mutation Screening in Halo15

Data sent from Dr. Claire Palles (30/01/2018)

Tumour	**Case Reference**	**Chr.**	**Pos.**	**Ref.**	**Alt.**	**Genotype**	**Annotation**	**Amino Acid Change**	**Existing Variation**	**Gene**
TA LGD	SmmipHalo15_1	2	47702345	A	T	0/1	missense	E/D	-	*MSH2*
		5	112175531	G	GT	0/1	frameshift	V/VX	CI109904	*APC*
		10	89720679	C	A	0/1	missense	T/K	rs398123329,CM109589,COSM35645,COSM5347081	*PTEN*
		11	108202238	G	T	0/1	missense	G/V	-	*ATM*
		17	70120263	C	T	0/1	missense	P/L	-	*SOX9*
		18	48573419	G	T	0/1	start_lost	M/I	-	*SMAD4*
TA LGD	2	2	203420115	C	A	0/1	missense	P/H	-	*BMPR2*
		5	112128143	C	T	0/1	Stop gained, splice region variant	R/*	rs62619935,CM992133,COSM98420	*APC*

Appendix 5.11 Additional Primers for *AXIN2* Mutation Validation

Forward Primer	**Reverse Primer**	**Amplicon Size (bp)**
CTTACCCAAACTGCTCGCTG	TCCACCACCACTACATCCAC	177

Appendix 5.12 Protocol for Somatic *APC* Screening

A PCR reaction is performed using the following reagents and reaction conditions:

Reagents:

MMG	12.5µl
Primer (5pMoles/ µl)	2µl
Water	8.5µl
DNA (10nmol/ µl)	2µl
Total	25µl

Reaction conditions:

Preheat	95°C	5 minutes
40 cycles of:		
Denaturation	95°C	30 seconds
Annealing	55°C	30 seconds
Extension	72°C	30 seconds
Final extension	72°C	7 minutes
Refrigerate	4°C	

The primers used are:

Name	**Forward Primer**	**Reverse Primer**	**Amplicon Size (bp)**
APC_15N	ACGTCATGTGGATCAGCCTA	TGTGCAGAACTTGGATGGAG	201
APC_15O	CAAGCAGTGAGAATACGTCCA	TGAACTACATCTTGAAAAACATATTGG	191
APC_15P	TTGCAAAGTTTCTTCTATTAACCAA	GCTATTTGCAGGGTATTAGCA	171
APC_15Q	TCAGCTGAAGATGAAATAGGATG	CTTTGTGCCTGGCTGATTCT	208
APC_15R	CCCTAGAACCAAATCCAGCA	AACATGAGTGGGGTCTCCTG	165
APC_15S	GAGCGAAATCTCCCTCCAA	CATGGTTTGTCCAGGGCTAT	224
APC_15T	GTGAACCATGCAGTGGAATG	GCAGCTTGCTTAGGTCCACT	193
APC_15U	CTTCCAGATAGCCCTGGACA	GACCCTCTGAACTGCAGCAT	171
APC_15V	AAAGCACCTACTGCTGAAAAGA	TCTTTCTGTATAAATGGCTCATCG	191
APC_15W	GGTTCTTCCAGATGCTGATACTTT	TTCATTTGATTCTTTAGGCTGCT	202
APC_15X	TGGAATTAAGAATAATGCCTCCA	TGTTGGCATGGCAGAAATAA	197
APC_15Y	GGACCTATTAGATGATTCAGATGATG	ACAGGCAGCTGACTTGGTTT	159

The downstream sequencing protocol follows the AWMGS SOP LP-GEN-sequencing (available on QPulse). Sequence Scanner v1.0 (Applied Biosystems) is used to check the quality of the sequences, and sequence analysis is carried out using Mutation Surveyor v4.0.8 (SoftGenetics LLC).

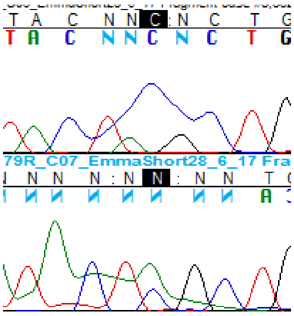
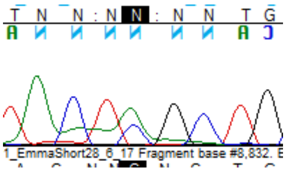
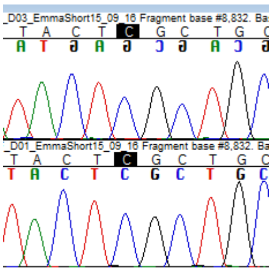
Appendix 5.13 Primers for Site Directed Mutagenesis

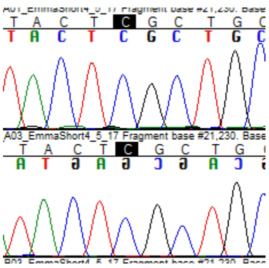
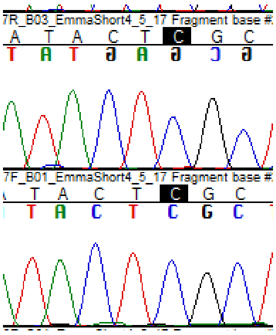
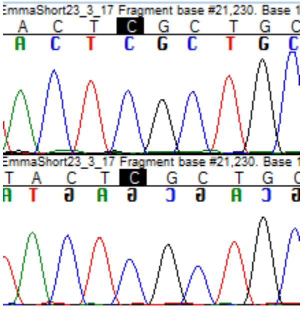
Forward Primer	**Reverse Primer**
AGTAGCAGTAATACTAGCTGCCCCCAGGGC	GCCCTGGGGGCAGCTAGTATTACTGCT

Appendix 5.14 Primers for Mutant *AXIN2* Sequencing

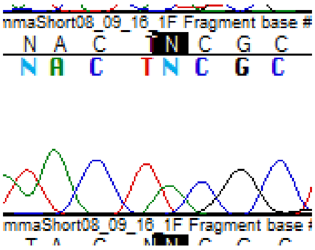
Forward primer	**Reverse primer**
GCTCCGAGCTCACACTCAAT	ACTCCAAGGGGTAGGCCTTT

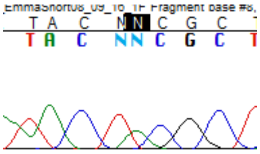
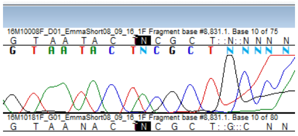
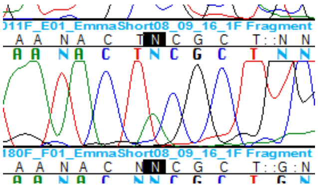
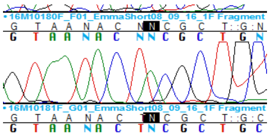
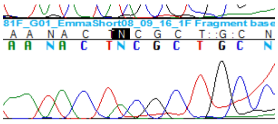
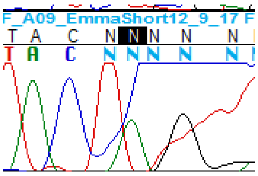
Appendix 5.15 Sequencing Results for *AXIN2* Family Segregation Studies

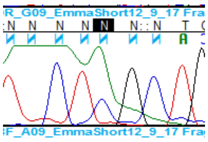
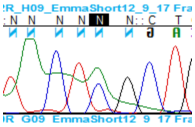
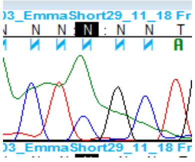
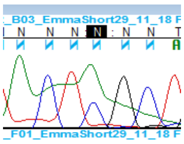
Family Member	***AXIN2* c.1642G>T:p.Glu548 * Mutation**	**Sequencing Result**
Individual 1.1 Maternal aunt of Halo47 and Halo68	Present	
Individual 1.2 Mother of Halo47 and Halo68	Present	
Individual 2.1 Brother of Halo47 and Halo68	Absent	

Individual 3.1 Daughter of Halo68	Absent	
Individual 3.2 Daughter of Halo68	Absent	
Individual 3.3 Niece of Halo47 and Halo68	Absent	

Appendix 5.16 *AXIN2* LOH Sequencing Results

Sample ID	**Case Reference**	**Diagnosis**	**LOH**
Halo47	H47-2	SA	No 

Sample ID	**Case Reference**	**Diagnosis**	**LOH**
Halo47	H47-3	SA	No 
Halo47	H47-4	TA LGD	No 
Halo47	H47-5	TA LGD	No 
Halo47	H47-6 (1: mod. diff) H47-7 (2: poorly diff)	CRC	1: No  2: No 
1.2	1.2-1	VA LGD	No 

Sample ID	**Case Reference**	**Diagnosis**	**LOH**
1.2	1.2-3	VA LGD	No 
1.2	1.2-5	TA LGD	No 
1.1	1.1-1	TA LGD	No 
1.1	1.1-2	TA LGD	No 
1.1	1.1-3	TA LGD	No