Module: molecular

Module Contents

1	**CENTER_NO**	number (2,0)	Required: true
	Center identification number.		

Allowable Values	
11	Sinai Health Systems (formerly Cancer Care Ontario)
12	Cedars-Sinai & Cleveland Clinic (formerly USC Consortium)
13	University of Melbourne
14	University of Hawaii Cancer Center
15	Mayo Clinic
16	Fred Hutch, Seattle
17	UCSF: University of California at San Franscisco (formerly CPIC, originally Northern California (NCCC))

2	**PERSON_ID**	string (12)	Required: true
	Number that Uniquely Identifies an Individual.		

3	**MSI_TEST_LAB**	number (1,0)	Required: true
	Code for type of lab that did the MSI testing.		

Allowable Values	
1	CFR Lab
2	Clinical Lab

4	**COLLECTION_CID**	string (30)	Required: true
	Identifier used internally by centers to denote a unique surgical event.		

| 5 | **TEST_TISSUE** | number (1,0) | Required: true |
| | Tissue on which test was performed. | | |

	Allowable Values
1	Cancer
2	Contiguous (Adjacent) adenoma
3	Colon Polyp
4	Gastric polyp
5	Sebaceous adenoma

| 6 | **TUMOR_NO** | number (2,0) | Required: true |
| | Numbers labeling tumors of the individual. Numbers are not necessarily sequential. | | |

	Allowable Values
1 to 99	Range
-9	NA/Out of scope: Tissue does not currently have a TUMOR_NO and is not cancer or contiguous adenoma
-15	Information Unknown

| 7 | **POLYP_NO** | number (2,0) | Required: true |
| | Sequential number over range of 1 to 99 to distinguish a polyp removed on a particular date. The tuple PERSON_ID, DATE_COLLECTED, and POLYP_NO uniquely identify a physical polyp globally within the CFR database. | | |

	Allowable Values
1 to 99 or -9, -1	Range
-9	NA/Out of scope. Tissue is not a polyp.
-1	Polyp has IHC/MSI result but center is currently unable to locate polyp pathology information. Center review to obtain this information is currently underway.

| 8 | **BLOCK_SPEC_CID** | string (40) | Required: true |
| | Unique local identifier used at a center to uniquely identify a block tissue specimen. | | |

| 9 | **NUC_ACID_CID_TUMOR** | string (40) | Required: true |
| | Center's unique identifier for the tumor DNA used in this MSI test. Each DNA extraction has a unique identifier. | | |

| 10 | **NUC_ACID_CID_NORMAL** | string (40) | Required: true |
| | Center's unique identifier for the normal DNA used in this MSI test. Each DNA extraction has a unique identifier. | | |

| 11 | **METASTASIS** | number (1,0) | Required: true |
| | Primary or metastatic tumor. | | |

	Allowable Values
1	Primary
2	Metastatasis

	-9	Test is not performed on cancerous tissue

12	**MSI_DETECTION_METHOD**	number (2,0)	Required: true
	MSI Testing method.		

Allowable Values	
1	PCR non-radiolabeled
2	PCR radiolabeled
3	Fluorescent PCR capillary electrophoresis
4	Fluorescent PCR slab gel electrophoresis
5	Silver-stained gel electrophoresis
-13	Not tested: Quantity of DNA or tissue not sufficient
-15	Method not specified

13	**MSI_ACTC**	number (2,0)	Required: true
	MSI test result at ACTC locus		

Allowable Values	
1	Stable
2	Unstable
6	Equivocal (Inconclusive)
7	Stable - Normal DNA not used in test
8	Unstable - Normal DNA not used in test
9	Equivocal (Inconclusive) - Normal DNA not used in test
-11	No Amplification
-12	Not tested, reason not specified
-13	Quantity of DNA or tissue not sufficient
-14	Tested, outcome not specified
-15	Unknown if tested

14	**MSI_BAT25**	number (2,0)	Required: true
	MSI test result at BAT25 locus		

Allowable Values	
1	Stable
2	Unstable
6	Equivocal (Inconclusive)
7	Stable - Normal DNA not used in test
8	Unstable - Normal DNA not used in test
9	Equivocal (Inconclusive) - Normal DNA not used in test

	-11	No Amplification
	-12	Not tested, reason not specified
	-13	Quantity of DNA or tissue not sufficient
	-14	Tested, outcome not specified
	-15	Unknown if tested

15	**MSI_BAT26**	number (2,0)	Required: true
	MSI test result at BAT26 locus		

	Allowable Values	
	1	Stable
	2	Unstable
	6	Equivocal (Inconclusive)
	7	Stable - Normal DNA not used in test
	8	Unstable - Normal DNA not used in test
	9	Equivocal (Inconclusive) - Normal DNA not used in test
	-11	No Amplification
	-12	Not tested, reason not specified
	-13	Quantity of DNA or tissue not sufficient
	-14	Tested, outcome not specified
	-15	Unknown if tested

16	**MSI_BAT34C4**	number (2,0)	Required: true
	MSI test result at BATC34C4 locus		

	Allowable Values	
	1	Stable
	2	Unstable
	6	Equivocal (Inconclusive)
	7	Stable - Normal DNA not used in test
	8	Unstable - Normal DNA not used in test
	9	Equivocal (Inconclusive) - Normal DNA not used in test
	-11	No Amplification
	-12	Not tested, reason not specified
	-13	Quantity of DNA or tissue not sufficient
	-14	Tested, outcome not specified
	-15	Unknown if tested

	MSI_BAT40	number (2,0)	Required: true

17	MSI test result at BAT40 locus		

Allowable Values	
1	Stable
2	Unstable
6	Equivocal (Inconclusive)
7	Stable - Normal DNA not used in test
8	Unstable - Normal DNA not used in test
9	Equivocal (Inconclusive) - Normal DNA not used in test
-11	No Amplification
-12	Not tested, reason not specified
-13	Quantity of DNA or tissue not sufficient
-14	Tested, outcome not specified
-15	Unknown if tested

18	**MSI_D5S346**	number (2,0)	Required:true
	MSI test result at D5S346 locus		

Allowable Values	
1	Stable
2	Unstable
6	Equivocal (Inconclusive)
7	Stable - Normal DNA not used in test
8	Unstable - Normal DNA not used in test
9	Equivocal (Inconclusive) - Normal DNA not used in test
-11	No Amplification
-12	Not tested, reason not specified
-13	Quantity of DNA or tissue not sufficient
-14	Tested, outcome not specified
-15	Unknown if tested

19	**MSI_D10S197**	number (2,0)	Required:true
	MSI test result at D10S197 locus		

Allowable Values	
1	Stable
2	Unstable
6	Equivocal (Inconclusive)
7	Stable - Normal DNA not used in test

8	Unstable - Normal DNA not used in test
9	Equivocal (Inconclusive) - Normal DNA not used in test
-11	No Amplification
-12	Not tested, reason not specified
-13	Quantity of DNA or tissue not sufficient
-14	Tested, outcome not specified
-15	Unknown if tested

20	**MSI_D17S250**	number (2,0)	Required: true
	MSI test result at D17S250 locus		

Allowable Values

1	Stable
2	Unstable
6	Equivocal (Inconclusive)
7	Stable - Normal DNA not used in test
8	Unstable - Normal DNA not used in test
9	Equivocal (Inconclusive) - Normal DNA not used in test
-11	No Amplification
-12	Not tested, reason not specified
-13	Quantity of DNA or tissue not sufficient
-14	Tested, outcome not specified
-15	Unknown if tested

21	**MSI_D18S55**	number (2,0)	Required: true
	MSI test result at D18S55 locus		

Allowable Values

1	Stable
2	Unstable
6	Equivocal (Inconclusive)
7	Stable - Normal DNA not used in test
8	Unstable - Normal DNA not used in test
9	Equivocal (Inconclusive) - Normal DNA not used in test
-11	No Amplification
-12	Not tested, reason not specified
-13	Quantity of DNA or tissue not sufficient
-14	Tested, outcome not specified
-15	Unknown if tested

22	**MSI_D2S123**	number (2,0)	Required: true
	MSI test result at D2S123 locus		

	Allowable Values
1	Stable
2	Unstable
6	Equivocal (Inconclusive)
7	Stable - Normal DNA not used in test
8	Unstable - Normal DNA not used in test
9	Equivocal (Inconclusive) - Normal DNA not used in test
-11	No Amplification
-12	Not tested, reason not specified
-13	Quantity of DNA or tissue not sufficient
-14	Tested, outcome not specified
-15	Unknown if tested

23	**MSI_MYCL**	number (2,0)	Required: true
	MSI test result at MyCL locus		

	Allowable Values
1	Stable
2	Unstable
6	Equivocal (Inconclusive)
7	Stable - Normal DNA not used in test
8	Unstable - Normal DNA not used in test
9	Equivocal (Inconclusive) - Normal DNA not used in test
-11	No Amplification
-12	Not tested, reason not specified
-13	Quantity of DNA or tissue not sufficient
-14	Tested, outcome not specified
-15	Unknown if tested

Module: molecular

Module Contents

1	**CENTER_NO**	number (2,0)	Required: true
	Center identification number.		

	Allowable Values	
11	Sinai Health Systems (formerly Cancer Care Ontario)
12	Cedars-Sinai & Cleveland Clinic (formerly USC Consortium)
13	University of Melbourne
14	University of Hawaii Cancer Center
15	Mayo Clinic
16	Fred Hutch, Seattle
17	UCSF: University of California at San Franscisco (formerly CPIC, originally Northern California (NCCC))

2	**PERSON_ID**	string (12)	Required: true
	Number that Uniquely Identifies an Individual.		

3	**IHC_TEST_LAB**	number (1,0)	Required: true
	Lab that conducted the results		

	Allowable Values
1	CFR lab
2	Clinical lab

4	**COLLECTION_CID**	string (40)	Required: true
	The label assigned by the center to all tissue associated with a single surgical event - resection or biopsy/endoscopy procedure.		

5	**TEST_TISSUE**	number (1,0)	Required: true
	Type of tissue tested.		

	Allowable Values
1	Cancer
2	Contiguous (Adjacent) adenoma

	3	Colon Polyp
	4	Gastric polyp
	5	Sebaceous adenoma

| 6 | **TUMOR_NO** | number (2,0) | Required: true |

Numbers labeling tumors of the individual. Numbers are not necessarily sequential.

Allowable Values	
1 to 99 or -9, -15	Range
-9	NA/Out of scope: Tissue does not currently have a TUMOR_NO and is not cancer or contiguous adenoma
-15	Information Unknown

| 7 | **POLYP_NO** | number (2,0) | Required: true |

Sequential number over range of 1 to 99 to distinguish a polyp removed on a particular date. The tuple PERSON_ID, DATE_COLLECTED, and POLYP_NO uniquely identify a physical polyp globally within the CFR database.

Allowable Values	
1 to 99 or -1, -9	Range
-9	NA/Out of scope. Tissue is not a polyp.
-1	Polyp has IHC/MSI result but center is currently unable to locate polyp pathology information. Center review to obtain this information is currently underway.

| 8 | **BLOCK_SPEC_CID** | string (40) | Required: true |

Unique local identifier used at a center to uniquely identify a block tissue specimen.

| 9 | **METASTASIS** | number (1,0) | Required: true |

Primary or metastatic tumor.

Allowable Values	
1	Primary
2	Metastatasis
-9	Test is not performed on cancerous tissue

| 10 | **IHC_MLH1** | number (2,0) | Required: true |

IHC result for MSI test at locus MLH1

Allowable Values	
0	Negative - Expression absent and abnormal
1	Positive - Expression present and normal
2	Positive - Technically poor, probably positive
3	Positive - Technically equivocal, not truly interpretable but if forced would be positive

4	Negative - Technically poor, probably negative
5	Negative - Technically equivocal, not truly interpretable but if forced would be negative
7	Heterogeneous tumor - some areas lack expression and some have expression
8	Reduced - expression reduced but not absent
-11	Technical failure
-12	Not tested
-13	Quantity of tissue not sufficient

11	**IHC_MSH2**	number (2,0)	Required:true
	IHC result for MSI test at locus MSH2		

	Allowable Values
0	Negative - Expression absent and abnormal
1	Positive - Expression present and normal
2	Positive - Technically poor, probably positive
3	Positive - Technically equivocal, not truly interpretable but if forced would be positive
4	Negative - Technically poor, probably negative
5	Negative - Technically equivocal, not truly interpretable but if forced would be negative
7	Heterogeneous tumor - some areas lack expression and some have expression
8	Reduced - expression reduced but not absent
-11	Technical failure
-12	Not tested
-13	Quantity of tissue not sufficient

12	**IHC_MSH6**	number (2,0)	Required:true
	IHC result for MSI test at locus MSH6		

	Allowable Values
0	Negative - Expression absent and abnormal
1	Positive - Expression present and normal
2	Positive - Technically poor, probably positive
3	Positive - Technically equivocal, not truly interpretable but if forced would be positive
4	Negative - Technically poor, probably negative
5	Negative - Technically equivocal, not truly interpretable but if forced would be negative
7	Heterogeneous tumor - some areas lack expression and some have expression
8	Reduced - expression reduced but not absent
-11	Technical failure
-12	Not tested
-13	Quantity of tissue not sufficient

13	**IHC_PMS2**		number (2,0)	Required: true
	IHC result for MSI test at locus PMS2			

Allowable Values	
0	Negative - Expression absent and abnormal
1	Positive - Expression present and normal
2	Positive - Technically poor, probably positive
3	Positive - Technically equivocal, not truly interpretable but if forced would be positive
4	Negative - Technically poor, probably negative
5	Negative - Technically equivocal, not truly interpretable but if forced would be negative
7	Heterogeneous tumor - some areas lack expression and some have expression
8	Reduced - expression reduced but not absent
-11	Technical failure
-12	Not tested
-13	Quantity of tissue not sufficient

Module: molecular

Module Contents

1	**CENTER_NO (*PK)**	number (2,0)	Required:true
	Center identification number. *CENTER_NO & NUC_ACID CID are the primary key for the table.		

Allowable Values	
11	Sinai Health Systems (formerly Cancer Care Ontario)
12	Cedars-Sinai & Cleveland Clinic (formerly USC Consortium)
13	University of Melbourne
14	University of Hawaii Cancer Center
15	Mayo Clinic
16	Fred Hutch, Seattle
17	UCSF: University of California at San Franscisco (formerly CPIC, originally Northern California (NCCC))

2	**PERSON_ID**	string (12)	Required:true
	Number that Uniquely Identifies an Individual.		

3	**GENE**	string (20)	Required:true
	Gene on which testing has been performed.		

Allowable Values	
BRCA1	
BRCA2	
MLH1	
MSH2	
MSH6	
PMS2	
EPCAM	

4	**NUC_ACID_CID (*PK)**	string (40)	Required:true
	Identifier used internally by centers for a nucleic acid sample from a single extraction. *CENTER_NO & NUC_ACID_CID are the primary key for the table.		

5	**MLPA_DATE_TESTED**	string (8)	Required:true
	Date of test.		

	Date Value Check
	The date must follow to the following format: Format YYYYMMDD. Must consist of valid date. Components of date should be right justified and zero filled. MM = 01 - 12, 88, 99 DD = 01 - 31, 88, 99 YYYY = **Minimum year** - system date year, 8888, 9999 Use 88, 8888 for not currently known, in progress to obtain information. Use 99, 9999 for not known. If century is known, but year is unknown then give an estimate of year or code YYYY = 9999. If MM = 99 then DD must = 99. If century is known, but year is unknown then give an estimate of year or code YYYY = 9999. If YYYY = 9999 then MM and DD must = 99. The following special parameters are used:

1700	Minimum year

6	**MLPA_INVEST_ID**	number (1,0)	Required:true
	Laboratory investigator conducting the tests on samples.		

Allowable Values	
1	Dr. Steve Thibodeau
2	Dr. Melissa Southey
3	GMP Genetics (Corporation)
4	Joanne Young
5	Clinical Laboratory
6	Dan Buchanan

7	**MLPA_DETECTION_METHOD**	number (1,0)	Required:false
	Testing method.		

Allowable Values	
1	Multiplex ligation-dependent probe amplification (MLPA)
2	Conversion analysis

8	**MLPA_RESULT**	number (1,0)	Required:true
	Summary of test findings.		

Allowable Values	
0	Failed test
1	Negative (no deletion or duplication)
2	Deletion

3	Duplication
4	Equivocal
5	Inversion

9	**MLPA_RESULT_SUB**	number (2,0)	Required:false
	Additional specific findings of test.		

Allowable Values	
0	Failed test
1	Confirmed by second MLPA test only
2	Confirmed by Southern blot
3	Confirmatory test not done for positive result
4	Confirmed by other method or other information
5	Deletion artifact. DNA sequence change under primer found by sequencing (scored as negative)
6	Deletion artifact. Results not consistent, normal on repeat (scored as negative)
7	Deletion artifact. Southern did not confirm deletion
8	Deletion artifact. Sequencing and/or Southern not done to determine cause for discordant result (assumed to be negative based on analysis of other similar cases)
9	Deletion artifact, Other
10	Duplication artifact. Results not consistent, normal on MLPA repeat (scored as negative)
11	Duplication artifact. Southern did not confirm duplication
12	Duplication artifact. Sequencing and/or Southern not done to determine cause for discordant result (assumed to be negative based on analysis of other similar cases)
13	Duplication artifact, Other
14	Equivocal. Single exon deletion, unclear whether artifact; Southern could not be performed
15	Equivocal. Multiple exons deleted, not enough sample to repeat test
16	Equivocal. Single exon duplication, unclear whether artifact; Southern could not be performed
17	Equivocal. Multiple exons duplicated, not enough sample to repeat test
-7	NA One assay only for negative result

Error Description
If MLPA_RESULT = 0, then MLPA_RESULT_SUB = 0
If MLPA_RESULT=1 then MLPA_RESULT_SUB must be in (-7,1,2) or in range 4-13
If MLPA_RESULT is 2 or 3, then MLPA_RESULT_SUB must be in range 1-4
If MLPA_RESULT=4, then MLPA_RESULT_SUB must be in range 14-17

10	**MLPA_EXON_FROM**	number (10,0)	Required:false
	Starting exon in range.		

Allowable Values	
1 to 9999999999 or -7	Range

	-7	NA: Negative result or failed test Otherwise number of starting exon.

Error Description
If MLPA_RESULT is in (0,1), then MLPA_EXON_FROM must be -7
If MLPA_RESULT is in (2,3,4), then MLPA_EXON_FROM must not be -7
If GENE = MLH1 then MLPA_EXON_FROM must be -7 or in range 1-19
If GENE = MSH2 then MLPA_EXON_FROM must be -7 or in range 1-16
If GENE = MSH6 then MLPA_EXON_FROM must be -7 or in range 1-10
If GENE = PMS2 then MLPA_EXON_FROM must be -7 or in range 1-15

11	**MLPA_EXON_TO**	number (10,0)	Required:false
	Ending exon in range.		

Allowable Values	
1 to 9999999999 or -7	Range
-7	NA: Negative result or failed test Otherwise number of starting exon.

Error Description
If MLPA_RESULT is 0 or 1, then MLPA_EXON_TO must be -7
If MLPA_RESULT is in (2,3,4), then MLPA_EXON_TO must not be -7
If GENE = MLH1 then MLPA_EXON_TO must be -7 or in range 1-19
If GENE = MSH2 then MLPA_EXON_TO must be -7 or in range 1-16
If GENE = MSH6 then MLPA_EXON_TO must be -7 or in range 1-10
If GENE = PMS2 then MLPA_EXON_TO must be -7 or in range 1-15

12	**MLPA_OTHER_TEST_COMMENT**	string (400)	Required:false
	Text field containing comments regarding other test findings.		

13	**GENOMIC_REGION_TESTED**	string (100)	Required:false
	Indicates what part(s) of the gene or other genomic region(s) was tested.		

14	**VARIANT_NAME**	string (100)	Required:false
	Proper variant name using HGVS nomenclature.		

Module: molecular

Module Contents

1	CENTER_NO (*PK)	number (2,0)	Required:true
	Center identification number. * CENTER_NO & GENOMIC_CID are the primary key for the table.		

	Allowable Values	
11	Sinai Health Systems (formerly Cancer Care Ontario)	
12	Cedars-Sinai & Cleveland Clinic (formerly USC Consortium)	
13	University of Melbourne	
14	University of Hawaii Cancer Center	
15	Mayo Clinic	
16	Fred Hutch, Seattle	
17	UCSF: University of California at San Franscisco (formerly CPIC, originally Northern California (NCCC))	

2	PERSON_ID	string (12)	Required:true
	Number that Uniquely Identifies an Individual.		

3	GENOMIC_CID (*PK)	string (40)	Required:true
	Center identifier for the test. *CENTER_NO & GENOMIC_CID are the primary key for the table.		

4	GENE	number (1,0)	Required:true
	Gene on which testing has been performed.		

Allowable Values	
1	MLH1
2	MSH2
3	MSH6
4	MUTYH
5	PMS2
8	EPCAM
10	APC
11	TP53

5	**VARIANT_ORIGIN**	number (1,0)	Required:true
	Describes if the variant was tested as an acquired or a germline variant.		

Allowable Values	
1	Presumed somatic
2	Known somatic
3	Germline

Error Description
If GENE is in (6,7) AND SOURCE_NUC_ACID is in (7,8), then VARIANT_ORIGIN = 1
If GENE is in (6,7) AND SOURCE_NUC_ACID is not in (7,8), then VARIANT_ORIGIN = 3

6	**SOURCE_NUC_ACID**	number (1,0)	Required:true
	Specifies source of tissue used for testing		

Allowable Values	
1	blood
2	Mouth wash/saliva
3	lymphoblastoid cells
4	normal fresh frozen tissue
5	other non-tumour tissue
6	polyp
7	tumour (paraffin embedded tumour- PET)
8	tumor- fresh frozen
9	unknown/private lab

7	**NUC_ACID_CID**	string (12)	Required:true
	Identifier used internally by centers for a single extraction on which the test was carried out		

Error Description

| | | If SOURCE_NUC_ACID=9, then NUC_ACID_CID must be -9 | | |

8	**DETECTION_METHOD**		number (2,0)	Required:true
	Germline or somatic DNA testing method. (Mass Spec = Sequenom MassARRAY system, MALDI-TOF)			

	Allowable Values	
	1	DNA Sequencing
	10	DHPLC (Denaturing High Performance Liquid Chromatography)
	11	Phase 1 protocol DHPLC+SEQ (MLH1 and MSH2)
	12	12 SNP panel genotyping by Mass Spec
	13	12 SNP panel genotyping, dHPLC with sequencing (MUTYH)
	14	9 SNP panel genotyping by Mass Spec
	15	9 SNP panel genotyping with sequencing (MUTYH)
	16	PTT
	17	SSCP
	18	Fluorescent ASP (BRAF)
	19	HRM with sequencing (KRAS)
	33	Sequencing from Conversion Analysis
	99	Unknown/private genetic test results

9	**GENOMIC_REGION_TESTED**		string (100)	Required:true
	Indicates what part(s) of the gene or specific mutation or other genomic region(s) was TESTED			

10	**TEST_RESULT**		number (1,0)	Required:true
	Result of the test carried out and this is at the test level (e.g. entire gene sequencing or MSH2 exon 2 etc.). This is NOT at person level or at variant level.			

	Allowable Values	
	1	Change detected
	2	No change detected
	3	Test failed
	9	Equivocal

11	**VARIANT_NAME_RESULT**		string (100)	Required:false
	Proper variant name using Human Genomic Variant Society (HGVS) nomenclature.			

Error Description
If TEST_RESULT is not 1 or 9 then VARIANT_NAME_RESULT must be null
If TEST_RESULT is 1 and DETECTION_METHOD is not 10, then VARIANT_NAME_RESULT must not be null

12	**EXON_NO**		number (2,0)	Required:false
	Number of exon containing variant. In the case the variant spans more than one exon, then this field should indicate the first exon			

affected by the variant.

Allowable Values	
1 to 99	Range

Error Description
If TEST_RESULT is not 1 or 9 and DETECTION_METHOD is not 10, then EXON_NO must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then EXON_NO must be null
If DETECTION_METHOD = 10 AND CENTER_NO = 11 then must EXON_NO be null
If INTRON_NO is null AND OTHER_REGION is null and TEST_RESULT = 1 AND LAB_ID not in(998,999), then EXON_NO must not be NULL

13	**INTRON_NO**	number (2,0)	Required:false
	Number of intron containing variant. In the case the variant spans more than one intron, then this field should indicate the first intron affected by the variant		

Allowable Values	
1 to 99	Range

Error Description
If TEST_RESULT is not 1, INTRON_NO then must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then INTRON_NO must be null
If DETECTION_METHOD = 10 AND CENTER_NO = 11 then INTRON_NO must be null
If EXON_NO is null AND OTHER_REGION is null and TEST_RESULT = 1 AND LAB_ID not in(998,999), then INTRON_NO must not be NULL

14	**OTHER_REGION**	number (1,0)	Required:false
	Region, other than coding exon and intron, containing the variant.		

Allowable Values	
1	5 UTR
2	3 UTR
3	Other

Error Description
If TEST_RESULT is not 1 or 9, then OTHER_REGION must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then OTHER_REGION must be null
If DETECTION_METHOD = 10 AND CENTER_NO = 11 then must OTHER_REGION be null
If EXON_NO is null AND INTRON_NO is null and TEST_RESULT = 1 AND LAB_ID not in(998,999), then OTHER_REGION must not be NULL

15	**NUC_POS_START**	string (20)	Required:false
	For exonic mutations, the numeric value of the nucleotide first affected by the mutation relative to the reference sequence. For intronic mutations, the numeric value of the first affected nucleotide relative to the exon number; a (-) indicates the number of nucleotides upstream (5) of the exon, and a (+) indicates the number of nucleotides downstream (3) from the exon. The 5 UTR and 3 UTR are considered to be part of the first andlLast Exon respectively but have special notation here. Positions in the 5-UTR		

have (-) sign. Positions in the 3-UTR have (*) sign.

Error Description
If TEST_RESULT is not 1 or 9, then NUC_POS_START must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then NUC_POS_START must be null
If TEST_RESULT=1 and LAB_ID not in(998,999), then NUC_POS_START must not be null
If NUC_POS_START is not null AND if OTHER_REGION = 1, then NUC_POS_START must begin with (-)
If NUC_POS_START is not null AND TYPE_VARIANT = 3 (substitution) AND EXON_NO is not null AND OTHER_REGION is null, then NUC_POS_START must not begin with (+,-)
If NUC_POS_START is not null AND If TYPE_VARIANT = 3 (substitution) AND INTRON_NO is not null AND OTHER_REGION is null then NUC_POS_START must begin with in (+, -)
If NUC_POS_START is not null AND If TYPE_VARIANT = 3 (substitution) AND INTRON_NO is not null AND OTHER_REGION is null then NUC_POS_START must begin with in (+, -)
If NUC_POS_START is not null AND if TYPE_VARIANT = 3, then NUC_POS_START = NUC_POS_END
If NUC_POS_START is not null AND if TYPE_VARIANT not in (1,3,9), then NUC_POS_START must not equal NUC_POS_END

16	**NUC_POS_END**	string (20)	Required:false
	For exonic mutations, the numeric value of the nucleotide last affected by the mutation relative to the reference sequence. For intronic mutations, the numeric value of the first affected nucleotide relative to the exon number; a (-) indicates the number of nucleotides upstream (5) of the exon, and a (+) indicates the number of nucleotides downstream (3) from the exon. The 5 UTR and 3 UTR are considered to be part of the first andlLast Exon respectively but have special notation here. Positions in the 5-UTR have (-) sign. Positions in the 3-UTR have (*) sign		

Error Description
If TEST_RESULT is not 1 or 9, then NUC_POS_END must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then NUC_POS_END must be null
If TEST_RESULT=1 and LAB_ID not in(998,999), then NUC_POS_END must not be null
If NUC_POS_END is not null AND if OTHER_REGION = 1, then must begin with (-)
If NUC_POS_END is not null AND If TYPE_VARIANT = 3 (substitution) AND EXON_NO is not null AND OTHER_REGION is null, then NUC_POS_END must not begin with (+,-)
If NUC_POS_END is not null AND If TYPE_VARIANT = 3 (substitution) AND INTRON_NO is not null AND OTHER_REGION is null then NUC_POS_END must begin with in (+, -)
If NUC_POS_END is not null AND if TYPE_VARIANT = 3, then NUC_POS_START = NUC_POS_END
If NUC_POS_END is not null and TYPE_VARIANT not in (1,3,9), then NUC_POS_START must not equal NUC_POS_END

17	**NUCLEOTIDE_FROM**	string (1)	Required:false
	Expected nucleotide with respect to the reference sequence.		

Allowable Values	
A	Adenine
C	Cytosine
G	Guanine
T	Thymidine

Error Description

If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then NUCLEOTIDE_FROM must be null

18	NUCLEOTIDE_TO	string (1)	Required:false
	Observed (mutated) nucleotide.		

Allowable Values	
A	Adenine
C	Cytosine
G	Guanine
T	Thymidine

Error Description
If TEST_RESULT is not 1 or 9, then NUCLEOTIDE_TO must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then NUCLEOTIDE_TO must be null

19	DELETE_SEQ	string (50)	Required:false
	Deletion nucleotide sequence		

Error Description
If TEST_RESULT is not 1, then NUCLEOTIDE_TO must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then NUCLEOTIDE_TO must be null
If TEST_RESULT=1 and TYPE_VARIANT in (1,7), then NUCLEOTIDE_TO must not be null

20	INSERT_SEQ	string (50)	Required:false
	Insertion nucleotide sequence		

Error Description
If TEST_RESULT is not 1, then INSERT_SEQ must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then INSERT_SEQ must be null
If TEST_RESULT=1 and TYPE_VARIANT in (2,4,7), then INSERT_SEQ must not be null

21	TYPE_VARIANT	number (1,0)	Required:false
	Type of genetic change		

Allowable Values	
1	Deletion
2	Insertion
3	Substitution
4	Duplication
5	Translocation
6	Inversion

7	Complex Rearrangement
9	Uncertain

Error Description
If TEST_RESULT is not 1 or 9, then TYPE_VARIANT must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then TYPE_VARIANT must be null
If TEST_RESULT is 1 and LAB_ID is not 998 or 999, then TYPE_VARIANT must not be null
If VARIANT_NAME_RESULT is null, AND LAB_ID is not 998 or 999 then TYPE VARIANT must not be null

22	**CODON**	number (4,0)	Required:false
	Numeric location of first mutated amino acid of reference sequence. A numeric value of 1 to the total number of codons on the reference sequence		

Allowable Values	
1 to 9999	Range

Error Description
If TEST_RESULT is not 1, then CODON must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then CODON must be null
If TEST_RESULT=1 and EXON_NO=null, then CODON must be null
If TEST_RESULT=1 AND EXON_NO is not null AND TYPE_VARIANT =3 , then CODON must not be null

23	**AA_FROM**	string (1)	Required:false
	Wild type amino acid value of first affected amino acid on reference sequence		

Allowable Values	
A	Ala
C	Cys
D	Asp
E	Glu
F	Phe
G	Gly
H	His
I	Ile
K	Lys
L	Leu
M	Met
N	Asn
P	Pro
Q	Gln
R	Arg

S	Ser
T	Thr
V	Val
W	Trp
Y	Tyr
X	Stop

Error Description
If TEST_RESULT is not 1, then AA_FROM must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then AA_FROM must be null
If TEST_RESULT=1 and EXON_NO=null, then AA_FROM must be null
If TEST_RESULT=1 AND EXON_NO is not null AND TYPE_VARIANT =3 , then AA_FROM must not be null

24	**AA_TO**	string (1)	Required:false
	Expressed amino acid value of last affected amino acid on reference sequence		

Allowable Values	
A	Ala
C	Cys
D	Asp
E	Glu
F	Phe
G	Gly
H	His
I	Ile
K	Lys
L	Leu
M	Met
N	Asn
P	Pro
Q	Gln
R	Arg
S	Ser
T	Thr
V	Val
W	Trp
Y	Tyr
X	Stop

Error Description
If TEST_RESULT is not 1, then AA_TO must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then AA_TO must be null
If TEST_RESULT=1 and CODON=null, then AA_TO must be null
If TEST_RESULT=1 AND EXON_NO is not null AND TYPE_VARIANT =3 , then AA_TO must be not null

25	**CONSEQUENCE**	number (2,0)	Required:false
	Category of mutation- defining outcome of the variant		

Allowable Values	
1	Missense
2	Premature termination
4	Aberrant splicing
7	Nonstop
8	Other
99	Uncertain

Error Description
If TEST_RESULT is not 1 or 9, then CONSEQUENCE must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then CONSEQUENCE must be null
If TEST_RESULT is 1, then CONSEQUENCE must NOT be null

26	**ZYGOSITY**	number (1,0)	Required:false
	Indicates the heterozygote/homozygote state of a variation, regardless of whether it is a polymorphism, unclassified variant or well established mutation.		

Allowable Values	
1	Homozygous wildtype
2	Heterozygous
3	Homozygous change
9	unknown

Error Description
If VARIANT_ORIGIN does not equal 3 (germline), then ZYGOSITY must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then ZYGOSITY must be null
If VARIANT_ORIGIN = 3 and TEST_RESULT = 3 (failed), then ZYGOSITY must be null
If VARIANT_ORIGIN = 3 and TEST_RESULT not equals 3, then ZYGOSITY must be not null
If VARIANT_ORIGIN = 3 and TEST_RESULT=9, then ZYGOSITY must be 9

27	**SEVERITY**	number (1,0)	Required:false
	Type of mutation		

Allowable Values

1	Deleterious
2	Polymorphism, assumed neutral
3	Unclassified variant
4	Unclassified variant, possibly pathogenic. DNA missense alteration with corresponding loss of expression by tumor IHC; alternatively, missense alteration that cosegregates with the disease in family.

Error Description
If TEST_RESULT is not 1 or 9, then SEVERITY must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then SEVERITY must be null
If VARIANT_ORIGIN in (1,2) AND TEST_RESULT = 1, then SEVERITY must be null

28	**INSIGHT_SEVERITY**	number (1,0)	Required:false
	This will be the 5-point severity code which InSiGHT use.		

Allowable Values	
5	Pathogenic
4	Likely Pathogenic
3	Unclassified
2	Likely Neutral
1	Neutral
9	Unknown

Error Description
If TEST_RESULT is not 1 or 9, then INSIGHT_SEVERITY must be null
If DETECTION_METHOD = 10 AND CENTER_NO in (11,16) then INSIGHT_SEVERITY must be null
If TEST_RESULT=1 and SEVERITY=null, INSIGHT_SEVERITY then must be null
If GENE in (6,7), then INSIGHT_SEVERITY must be null

29	**LAB_ID**	number (3,0)	Required:true
	Lab or individual who performed the testing.		

Allowable Values	
119	Graham Casey
125	Steve Gallinger
128	Peter Laird
129	Joanne Young
141	Steve Thibodeau
228	Robert Haile
300	Dan Buchanan
301	Karen Makar
998	Other (specify, free text)

	999	Unknown	

30	**LAB_OTHER_TEXT**	string (50)	Required:false
	Name of other/private lab		

31	**DATE_TEST**	string (8)	Required:false
	Date on which test was performed. Format: YYYYMMDD		

Date Value Check

The date must follow to the following format:

Format YYYYMMDD. Must consist of valid date.
Components of date should be right justified and zero filled.
MM = 01 - 12, 88, 99
DD = 01 - 31, 88, 99
YYYY = **Minimum year** - system date year, 8888, 9999
Use 88, 8888 for not currently known, in progress to obtain information.
Use 99, 9999 for not known.
If century is known, but year is unknown then give an estimate of year or code YYYY = 9999.
If MM = 99 then DD must = 99.
If century is known, but year is unknown then give an estimate of year or code YYYY = 9999.
If YYYY = 9999 then MM and DD must = 99.

The following special parameters are used:

2002	Minimum year

Error Description

Minimum year 2002

Module: molecular

Module Contents

1	**CENTER_NO (*PK)**	number(2,0)	Required:TRUE
	Center identification number. *CENTER_NO & SOMATIC_CID are the primary key for the table.		

	Allowable Values	
	11	Sinai Health Systems (formerly Cancer Care Ontario)
	12	Cedars-Sinai & Cleveland Clinic (formerly USC Consortium)
	13	University of Melbourne
	14	University of Hawaii Cancer Center
	15	Mayo Clinic
	16	Fred Hutch, Seattle
	17	UCSF: University of California at San Franscisco (formerly CPIC, originally Northern California (NCCC))

2	**SOMATIC_CID (*PK)**	String (40)	Required:TRUE
	Centre identifier for the test. *CENTER_NO & SOMATIC_CID are the primary key for the table.		

3	**PERSON_ID**	string (12)	Required:TRUE
	Number that uniquely identifies an individual.		

4	**TUMOR_NO**	number (2,0)	Required:TRUE
	Number labelling the tumor of the individual. Numbers are not necessarily sequential.		

	Allowable Values	
	-9	NA/Out of scope: Tissue is not cancer or contiguous adenoma.
	-15	Information Unknown

5	**POLYP_NO**	number (2,0)	Required: TRUE

Sequential number over range of 1 to 99 to distinguish a polyp removed on a particular date.

Allowable Values	
-9	NA/Out of scope: Tissue is not a polyp.
-1	Polyp has MSI/IHC result but center is currently unable to locate polyp pathology information

6	**GENE**	number (1,0)	Required: TRUE

Gene on which testing has been performed.

Allowable Values	
6	BRAF
7	KRAS

7	**VARIANT_ORIGIN**	number (1,0)	Required: TRUE

Describes if the variant was tested as an acquired or a germline variant.

Allowable Values	
1	Presumed somatic
2	Known somatic
3	Germline

Error Description
If GENE in (6,7) AND SOURCE_NUC_ACID (5,6,7,8), then VARIANT_ORIGIN must equal 1

8	**SOURCE_NUC_ACID**	number (1,0)	Required: TRUE

Specifies source of tissue used for testing

Allowable Values	
4	normal fresh frozen tissue
5	other non-tumour tissue
6	polyp
7	tumour (paraffin embedded tumour- PET)
8	tumor- fresh frozen
9	unknown/private lab

9	**NUC_ACID_CID**	string (15)	Required: TRUE

Identifier used internally by centers for a single extraction on which the test was carried out

Allowable Values	

-9	unknown

Error Description
If SOURCE_NUC_ACID=9, then NUC_ACID_CID must be -9

10	**DETECTION_METHOD**	number (2,0)	Required:TRUE
	Somatic DNA testing method. (Mass Spec = Sequenom MassARRAY system, MALDI-TOF)		

Allowable Values	
1	DNA Sequencing
18	Fluorescent ASP (BRAF)
19	HRM with sequencing (KRAS)
40	SYBR real time PCR assay
99	Unknown/private genetic test results

11	**GENOMIC_REGION_TESTED**	string (30)	Required:TRUE
	Indicates what part(s) of the gene or specific mutation or other genomic region(s) was TESTED		

Allowable Values	
BRAF c.1799T>A p.Val600Glu	
KRAS codons 12 and 13 (g.10537--g.10647)	
KRAS codon 61	

12	**TEST_RESULT**	number (1,0)	Required:TRUE
	Result of the test carried out and this is at the test level. This is NOT at person level.		

Allowable Values	
1	Change detected
2	No change detected
3	Test failed
9	Equivocal

13	**VARIANT_NAME_RESULT**	string (30)	Required:FALSE
	Proper variant name using Human Genomic Variant Society (HGVS) nomenclature.		

Allowable Values	
BRAF c.1799T>A p.V600E	
KRAS c.34G>A	
KRAS c.34G>C	
KRAS c.34G>T	
KRAS c.35G>A	

| KRAS c.35G>C |
| KRAS c.35G>T |
| KRAS c.37G>A |
| KRAS c.37G>C |
| KRAS c.37G>T |
| KRAS c.38G>A |
| KRAS c.38G>T |

Error Description
If TEST_RESULT does not equal 1, then VARIANT_NAME_RESULT must be null
If TEST_RESULT=1, then VARIANT_NAME_RESULT must not be null

14	**EXON_NO**	number (2,0)	Required:false
	Number of exon containing variant. In the case the variant spans more than one exon, then this field should indicate the first exon affected by the variant.		

Allowable Values	
2	KRAS
15	BRAF

Error Description
If TEST_RESULT does not equal 1, then EXON_NO must be null
If INTRON_NO is null AND OTHER_REGION is null AND LAB_ID not in(998,999), then EXON_NO must not be NULL

15	**NUC_POS_START**	string (20)	Required:false
	For exonic mutations, the numeric value of the nucleotide first affected by the mutation relative to the reference sequence.		

Error Description
If TEST_RESULT does not equal 1, then NUC_POS_START must be null
If TEST_RESULT=1 and LAB_ID not in(998,999), then NUC_POS_START must not be null.
If TYPE_VARIANT =3, then NUC_POS_START must equal NUC_POS_END
If TYPE_VARIANT does not equal 3, then NUC_POS_START must not equal NUC_POS_END

16	**NUC_POS_END**	string (20)	Required:false
	For exonic mutations, the numeric value of the nucleotide last affected by the mutation relative to the reference sequence.		

Error Description
If TEST_RESULT does not equal 1, then NUC_POS_END must be null
If TEST_RESULT=1 and LAB_ID not in(998,999), then NUC_POS_END must not be null
If TYPE_VARIANT =3, then NUC_POS_END must equal NUC_POS_START
If TYPE_VARIANT <> 3, then NUC_POS_END must not NUC_POS_START

17	**NUCLEOTIDE_FROM**		string (1)	Required:false
	Expected nucleotide with respect to the reference sequence.			

Allowable Values	
A	Adenine
C	Cytosine
G	Guanine
T	Thymidine

Error Description
If TEST_RESULT does not equal 1, then NUCLEOTIDE_FROM must be null

18	**NUCLEOTIDE_TO**		string (1)	Required:FALSE
	Observed (mutated) nucleotide.			

Allowable Values	
A	Adenine
C	Cytosine
G	Guanine
T	Thymidine

Error Description
If TEST_RESULT does not equal 1, then NUCLEOTIDE_TO must be null

19	**TYPE_VARIANT**		number (1,0)	Required:false
	Type of genetic change			

Allowable Values	
3	Substitution
9	Uncertain

Error Description
If TEST_RESULT does not equal 1, then TYPE_VARIANT must be null
If TEST_RESULT=1, then TYPE_VARIANT must be in (3,9)

20	**CODON**		number (4,0)	Required:false
	Numeric location of first mutated amino acid of reference sequence. A numeric value of 1 to the total number of codons on the reference sequence			

Error Description
If TEST_RESULT does not equal 1, then CODON must be null
If TEST_RESULT=1 AND EXON_NO does not equal null AND TYPE_VARIANT =3 , then CODON must be not null

	AA_FROM		string (1)	Required:FALSE

21	Wild type amino acid value of first affected amino acid on reference sequence

Allowable Values	
A	Ala
C	Cys
D	Asp
E	Glu
F	Phe
G	Gly
H	His
I	Ile
K	Lys
L	Leu
M	Met
N	Asn
P	Pro
Q	Gln
R	Arg
S	Ser
T	Thr
V	Val
W	Trp
Y	Tyr
X	Stop

Error Description
If TEST_RESULT does not equal 1, then AA_FROM must be null
If TEST_RESULT=1 AND CODON is not null AND TYPE_VARIANT =3 , then AA_FROM must not be null

22	**AA_TO**	string (1)	Required:FALSE
	Expressed amino acid value of last affected amino acid on reference sequence		

Allowable Values	
A	Ala
C	Cys
D	Asp
E	Glu
F	Phe
G	Gly

H	His
I	Ile
K	Lys
L	Leu
M	Met
N	Asn
P	Pro
Q	Gln
R	Arg
S	Ser
T	Thr
V	Val
W	Trp
Y	Tyr
X	Stop

Error Description
If TEST_RESULT does not equal 1, then AA_TO must be null
If TEST_RESULT=1 AND CODON is not null AND TYPE_VARIANT =3 , then AA_TO must not be null

23	**CONSEQUENCE**	number (1,0)	Required:FALSE
	Category of mutation- defining outcome of the variant		

Allowable Values	
1	Missense
2	Premature termination
9	Uncertain

Error Description
If TEST_RESULT does not equal 1, then CONSEQUENCE must be null
If TEST_RESULT=1, then CONSEQUNCE must be not null

24	**LAB_ID**	number (3,0)	Required:TRUE
	Lab or individual who performed the testing.		

Allowable Values	
119	Graham Casey
125	Steve Gallinger
128	Peter Laird
129	Joanne Young

141	Steve Thibodeau
228	Robert Haile
300	Dan Buchanan
301	Karen Makar
998	Other (specify, free text)
999	Unknown

| 25 | **LAB_OTHER_TEXT** | string (30) | Required: false |
| | Name of other/private lab | | |

| 26 | **DATE_TEST** | string (8) | Required: TRUE |
| | Date on which test was performed.
Format: YYYYMMDD | | |

Date Value Check
The date must follow to the following format: Format YYYYMMDD. Must consist of valid date. Components of date should be right justified and zero filled. MM = 01 - 12, 88, 99 DD = 01 - 31, 88, 99 YYYY = **Minimum year** - system date year, 8888, 9999 Use 88, 8888 for not currently known, in progress to obtain information. Use 99, 9999 for not known. If century is known, but year is unknown then give an estimate of year or code YYYY = 9999. If MM = 99 then DD must = 99. If century is known, but year is unknown then give an estimate of year or code YYYY = 9999. If YYYY = 9999 then MM and DD must = 99. The following special parameters are used:

YYYYMMDD :	
MM	01 - 12, 88, 99
DD	01 - 31, 88, 99
YYYY	2002 - system year, 8888, 9999

Module: derived

Module Contents

d_MSI_IHC_PK_MSI

1	**CENTER_NO**	number (2,0)	Required:true
	Center Identification Number.		

	Allowable Values	
	11	Sinai Health Systems (formerly Cancer Care Ontario)
	12	Cedars-Sinai & Cleveland Clinic (formerly USC Consortium)
	13	University of Melbourne
	14	University of Hawaii Cancer Center
	15	Mayo Clinic
	16	Fred Hutch, Seattle
	17	UCSF: University of California at San Franscisco (formerly CPIC, originally Northern California (NCCC))

2	**PERSON_ID**	string (12)	Required:true
	Number that uniquely identifies an individual.		

3	**COLLECTION_CID**	string (30)	Required:true
	Identifier used internally by centers to denote a unique surgical event.		
	This data element is typically used to allow aggregation of all materials, such as blocks, from a single surgery.		

4	**TEST_TISSUE**	number (1,0)	Required:true
	Tissue on which test was performed.		
	Only values of 1,2,3 link to pathology information. Although values 2 and 3 may represent the same tissue type, their pathology information is stored in different database tables. Adjacent adenomas are recorded on the Colorect_Malig table and so are associated with a cancer; Colon polyps are recorded on the Polyps table. As a result, the pathology information for each will differ somewhat. There may be cases where TEST_TISSUE variable has level 2 but the ADJ_ADEN variable of Colorect_Malig table indicates no adjacent adenoma present. In these cases, the TEST_TISSUE value takes precedence and the adjacent adenoma is considered to have been overlooked.		

	Allowable Values	
	1	Cancer
	2	Contiguous (Adjacent) adenoma
	3	Colon polyp
	4	Gastric polyp
	5	Sebaceous adenoma

TUMOR_NO	number (2,0)	Required: true

5

Numbers labeling tumors of the individual.

Numbers are not necessarily sequential.
No adjustment is made when a tumor is deleted from the system. Tumor numbers are never reused. PERSON_NO plus TUMOR_NO uniquely identifies each tumor in the central database. The same tumor number from different tables (including tables outside of the biospecimen module) must always refer to the same physical tumor.

Allowable Values	
-9	NA/Out of scope: Tissue is not cancer or contiguous adenoma.
-15	Information Unknown

POLYP_NO	number (2,0)	Required: true

6

Sequential number over range of 1 to 99 to distinguish a polyp removed on a particular date. The tuple PERSON_ID, DATE_COLLECTED, and POLYP_NO uniquely identify a physical polyp globally within the CFR database.

The colon pathology form has three sets of identical fields for recording information on polyps. The POLYPS table includes corresponding data elements, i.e. POLYP_SITEx, POLYP_TYPEx, POLYP_SIZEx, POLYP_GRADEx, where x is an integer between 1 and 3, inclusive. The physical polyp referenced by x and this data element must be identical.
Polyps are described according to the following rules:
1. The largest polyp is reported first, followed by next largest
2. If size is not indicated, adenomas are reported before hyperplastic polyps
3. Polyps with most severe grade are reported
Not all polyps present in available tissue have an assigned value, as some are not described.

Allowable Values	
-9	NA/Out of scope. Tissue is not a polyp.
-1	Polyp has IHC/MSI result but center is currently unable to locate polyp pathology information. Center review to obtain this information is currently underway.

BLOCK_SPEC_CID	string (40)	Required: true

7

Unique local identifier used at a center to uniquely identify a block tissue specimen.

Allowable Values	
-15	Information Unknown

NUC_ACID_CID_TUMOR	string (40)	Required: true

8

Center's unique identifier for the tumor DNA used in this MSI test. Each DNA extraction has a unique identifier.

Unique Identifier, If this field in not negative, there will be a matching record on the NUC_ACID table that indicates whether remaining tumor DNA from the extraction used for this test is available for distribution to requestors.
If multiple MSI results are provided for the same block and tumor then separate identifiers must be specified.

Allowable Values	
- 15	DNA identifying information lost/unknown.
- 14	Center does not keep normal tissue DNA identifying information.
- 13	Quantity of DNA or tissue not sufficient.

NUC_ACID_CID_NORMAL	string (40)	Required: true

| 9 | Center's unique identifier for the normal DNA used in this MSI test. Each DNA extraction has a unique identifier.

Normal DNA may have been extracted from paraffin embedded normal tissue, blood, mouthwash/buccal cells or fresh frozen normal tissue. Paraffin embedded normal tissue is not necessarily associated with the paraffin block from which the tumor DNA is taken. If this field in not negative, there will be a matching record on the NUC_ACID table that indicates whether remaining normal DNA from the extraction used for this test is available for distribution to requestors. |

Allowable Values	
- 15	DNA identifying information lost/unknown.
- 14	Center does not keep normal tissue DNA identifying information.
- 13	Quantity of DNA or tissue not sufficient.

10	**MSI_STATUS_DNA**	number (1,0)	Required: true

MSI status of DNA extraction.

For each set of scores with a minimum of 4 of 10 markers (excluding D2S123) that are either stable or unstable (Options 1 and 2 only NOT 7 or 8), the percent unstable among stable and unstable markers is calculated and compared to cutpoints to determine if the results are stable (0% unstable), low (1-29% unstable) or high (30-100% unstable). If only 3 markers are scored but they are all unstable then the status is scored as high.

Allowable Values	
-7	NA/Out of scope - Less than 4 marker scores stable or unstable
1	Stable - 0% unstable
2	Low - More than 0% but less than 30% unstable
3	High - At least 30% unstable

Module: derived

Module Contents

1	**PERSON_ID**	string (12)	Required: true
	Number that Uniquely Identifies an Individual.		

2	**COLLECTION_CID**	string (40)	Required: true
	The label assigned by the center to all tissue associated with a single surgical event - resection or biopsy/endoscopy procedure. This data element is typically used to allow aggregation of all materials, such as blocks, from a single surgery.		

3	**TEST_TISSUE**	number (1,0)	Required: true
	Type of tissue tested. Only values of 1,2,3 link to pathology information. Although values 2 and 3 may represent the same tissue type, their pathology information is stored in different database tables. Adjacent adenomas are recorded on the Colorect_Malig table and so are associated with a cancer; Colon polyps are recorded on the Polyps table. As a result, the pathology information for each will differ somewhat. There may be cases where TEST_TISSUE variable has level 2 but the ADJ_ADEN variable of Colorect_Malig table indicates no adjacent adenoma present. In these cases, the TEST_TISSUE value takes precedence and the adjacent adenoma is considered to have been overlooked.		

Allowable Values	
1	Cancer
2	Contiguous (Adjacent) adenoma
3	Colon Polyp
4	Gastric polyp
5	Sebaceous adenoma

4	**TUMOR_NO**	number (2,0)	Required: true
	Numbers labeling tumors of the individual. Numbers are not necessarily sequential. No adjustment is made when a tumor is deleted from the system. Tumor numbers are never reused. PERSON_NO plus TUMOR_NO uniquely identifies each tumor in the central database. The same tumor number from different tables (including tables outside of the biospecimen module) must always refer to the same physical tumor.		

Allowable Values	
1 to 99 or -9, -15	Range
-9	NA/Out of scope: Tissue does not currently have a TUMOR_NO and is not cancer or contiguous adenoma
-15	Information Unknown

POLYP_NO	number (2,0)	Required: true

5	Sequential number over range of 1 to 99 to distinguish a polyp removed on a particular date. The tuple PERSON_ID, DATE_COLLECTED, and POLYP_NO uniquely identify a physical polyp globally within the CFR database. The colon pathology form has three sets of identical fields for recording information on polyps. The POLYPS table includes corresponding data elements, i.e. POLYP_SITEx, POLYP_TYPEx, POLYP_SIZEx, POLYP_GRADEx, where x is an integer between 1 and 3, inclusive. The physical polyp referenced by x and this data element must be identical. Polyps are described according to the following rules: 1. The largest polyp is reported first, followed by next largest 2. If size is not indicated, adenomas are reported before hyperplastic polyps 3. Polyps with most severe grade are reported Not all polyps present in available tissue have an assigned value, as some are not described.

Allowable Values	
1 to 99 or -9, -1	Range
-9	NA/Out of scope. Tissue is not a polyp.
-1	Polyp has IHC/MSI result but center is currently unable to locate polyp pathology information. Center review to obtain this information is currently underway.

6	**BLOCK_SPEC_CID**	string (40)	Required: true
	Unique local identifier used at a center to uniquely identify a block tissue specimen.		

Allowable Values	
-15	Information Unknown

7	**IHC_STATUS_TISSUE**	number(1,0)	Required: true
	IHC status of tissue from tested block. Expression is absent if any of IHC_MLH1, IHC_MSH2, IHC_MSH6, IHC_PMS2 equal either 0 or 4. Expression is normal if at least one of them equals 1 or 2 and none equal 0 or 4.		

Allowable Values	
-8	No data transmitted from centers
-7	NA/Out of scope: None of IHC_MLH1, IHC_MSH2, IHC_MSH6, IHC_PMS2 recorded values 0,1,2,4.
1	Normal MMR protein expression
2	Absence of MMR protein expression

8	**MSI_STATUS_TISSUE**	number(1,0)	Required: true
	MSI status of tissue from tested block. The maximum of MSI_STATUS_DNA over all DNA extractions from a given tissue-block combination.		

Allowable Values	
-8	No data transmitted from centers
-7	NA/Out of scope - Less than 4 marker scores stable or unstable
1	Stable - 0% unstable
2	Low - More than 0% but less than 30% unstable
3	High - At least 30% unstable

MSIIHC_STATUS_TISSUE	number(1,0)	Required:true

9	MMR status considering all MSI DNA extractions and IHC results on that tissue/block combination. MSIIHC_STATUS_TISSUE has a value for each tissue that has an MSI and/or IHC result within a given block. MSIIHC_STATUS_TISSUE is defective 2 if MSI_STATUS_TISSUE is high and IHC_STATUS_TISSUE is absent. MSIIHC_STATUS_TISSUE is defective 1 if it does not qualify as defective 2 and MSI_STATUS_TISSUE is high or IHC_STATUS_TISSUE is absent. MSIIHC_STATUS_TISSUE is proficient 2 if MSI_STATUS_TISSUE is stable or low and IHC_STATUS_TISSUE is normal. MSIIHC_STATUS_TISSUE is proficient 1 if it does not qualify as proficient 2 or as defective, and MSI_STATUS_TISSUE is stable or low or IHC_STATUS_TISSUE is normal. MSIIHC_STATUS_TISSUE otherwise equals -7, corresponding to both MSI_STATUS_TISSUE and IHC_STATUS_TISSUE equaling -7 or one them equaling -7 with no record available to merge to on the other table.

Allowable Values	
-8	No data transmitted from centers
-7	NA/Out of Scope - Both source variables out of scope
1	Proficient MMR 1
2	Proficient MMR 2
3	Defective MMR 1
4	Defective MMR 2

Module: derived

Module Contents

d_MSI_IHC

1	**PERSON_ID**	string (12)	Required:TRUE
	Number that Uniquely Identifies an Individual.		

2	**IHC_STATUS_PERSON**	number (1,0)	Required:TRUE

IHC status of person tested.

Considers IHC results only for colon cancer tumor types. Excludes all other malignancies, adenomas, metastases and in situ carcinomas. Does not restrict to qualifying tumors.
If at least one test is abnormal, then IHC_STATUS_PERSON is abnormal. If there are no abnormal results and at least one normal result, then IHC_STATUS_PERSON is normal.
Colon cancer is defined by ICDO3 site codes C180, C181, C182, C183, C184, C185, C186, C187, C188, C189, C199, C209, C218, C260, and a malignant tumor (BEHAV=3).
On 1/4/2008 the SITE codes used in deriving this variable were expanded to include C218 and C260.

Allowable Values	
-8	No data transmitted from centers
-7	NA/Out of scope: Individual has IHC record, but no markers having values 0,1,2, or 4.
1	Normal MMR protein expression
2	Absence of MMR protein expression

3	**MSI_STATUS_PERSON**	number (1,0)	Required:TRUE

MSI status of person tested.

Considers MSI results only for colon cancer tumor types. Excludes all other malignancies, adenomas, metastases and in situ carcinomas. Does not restrict to qualifying tumors.
If at least one MSI_STATUS_TISSUE result is high, then MSI_STATUS_PERSON is high. If no MSI_STATUS_TISSUE result is high but at least one is low, then MSI_STATUS_PERSON is low. If no MSI_STATUS_TISSUE results are high or low, but at least one is stable, then MSI_STATUS_PERSON is stable.
Colon cancer is defined by ICDO3 site codes C180, C181, C182, C183, C184, C185, C186, C187, C188, C189, C199, C209, C218, C260, and a malignant tumor (BEHAV=3).
On 1/4/2008 the SITE codes used in deriving this variable were expanded to include C218 and C260.

Allowable Values	
-8	No data transmitted from centers
-7	NA/Out of scope - An MSI record, but no Stable, Low, or High tissue-level result recorded.
1	Stable - No High or Low tissue-level results, but at least one Stable.
2	Low - No High tissue-level results, but at least one Low.
3	High - At least one High tissue-level result

	MSIIHC_STATUS_PERSON	number (1,0)	Required:TRUE

MMR status for person tested.

Considers MSI and IHC results only for colon cancer tumor types. Excludes all other malignancies, adenomas, metastases and in situ carcinomas. Does not restrict to qualifying tumors.
MSIIHC_STATUS_PERSON has a value for each each tissue that has an MSI and/or IHC result within a given block.

4	MSIIHC_STATUS_PERSON is defective 2 if MSI_STATUS_PERSON is High and IHC_STATUS_PERSON is Abnormal. MSIIHC_STATUS_PERSON is defective 1 if it does not qualify as defective 2 and MSI_STATUS_PERSON is High or IHC_STATUS_PERSON is Abnormal. MSIIHC_STATUS_PERSON is proficient 2 if MSI_STATUS_PERSON is Stable or Low and IHC_STATUS_PERSON is Normal. MSIIHC_STATUS_PERSON is proficient 1 if it does not qualify as proficient 2 or as defective, and MSI_STATUS_PERSON is Stable or Low or IHC_STATUS_PERSON is Normal. MSIIHC_STATUS_PERSON otherwise equals -7, corresponding to both MSI_STATUS_PERSON and IHC_STATUS_PERSON equaling -7 or one them equaling -7 with no record available to merge to on the other table.

Allowable Values	
-8	No data transmitted from centers
-7	NA/Out of Scope - Both source variables out of scope
1	Proficient MMR 1
2	Proficient MMR 2
3	Defective MMR 1
4	Defective MMR 2

Module: derived

Module Contents

1	**PERSON_ID**	string (12)	Required:FALSE
	Identifier for person that is globally unique within the C-CFR central database. FK. Consists of concatenation of CENTER_NO (2 digit; digits 1-2) + (10 digit local unique individual id; digits 3-12). The 10 digit individual number component should be right justified, zero filled. Ex: Local unique individual id 98765 from New York Breast Center should be: 020000098765 or 20000098765. Ex: Individual Id 98765 from Australia Colon Center should be: 130000098765.		

2	**C_MLH1**	number (1,0)	Required:FALSE
	Presence of deleterious mutation or unclassified variant in MLH1 gene		

Allowable Values	
-8	No data transmitted from centers
-9	Failed or equivocal MLPA result or failed sequencing result
-1	One of the tests (Genomic or MLPA) has been run, and no deleterious variant or unclassified variant detected
1	Both of the tests (Genomics and MLPA) have been run and no deleterious variant or unclassified variant detected
2	One or more deleterious variants detected (Genomic or MLPA) (If 1 should be homozygous)
-3	One or more unclassifed variants detected (Genomic)
3	One or more unclassified variants detected (Genomic or MLPA)
4	One or more deleterious AND one or more unclassifed variants detected (Genomic or MLPA)

3	**C_MSH2**	number (1,0)	Required:FALSE
	Presence of deleterious mutation or unclassified variant in MSH2 gene		

Allowable Values	
-8	No data transmitted from centers
-9	Failed or equivocal MLPA result or failed sequencing result
-1	One of the tests (Genomic or MLPA) has been run, and no deleterious variant or unclassified variant detected
1	Both of the tests (Genomics and MLPA) have been run and no deleterious variant or unclassified variant detected
2	One or more deleterious variants detected (Genomic or MLPA) (If 1 should be homozygous)
-3	One or more unclassifed variants detected (Genomic)
3	One or more unclassified variants detected (Genomic or MLPA)
4	One or more deleterious AND one or more unclassifed variants detected (Genomic or MLPA)

4	**C_MSH6**	number (1,0)	Required:FALSE
	Presence of deleterious mutation or unclassified variant in MSH6 gene		

Allowable Values	
-8	No data transmitted from centers
-9	Failed or equivocal MLPA result or failed sequencing result
-1	One of the tests (Genomic or MLPA) has been run, and no deleterious variant or unclassified variant detected
1	Both of the tests (Genomics and MLPA) have been run and no deleterious variant or unclassified variant detected
2	One or more deleterious variants detected (Genomic or MLPA) (If 1 should be homozygous)
-3	One or more unclassifed variants detected (Genomic)
3	One or more unclassified variants detected (Genomic or MLPA)
4	One or more deleterious AND one or more unclassifed variants detected (Genomic or MLPA)

5	**C_PMS2**	number (1,0)	Required:FALSE
	Presence of deleterious mutation or unclassified variant in PMS2 gene		

Allowable Values	
-8	No data transmitted from centers
-9	Failed or equivocal MLPA result or failed sequencing result
-1	One of the tests (Genomic or MLPA) has been run, and no deleterious variant or unclassified variant detected
1	Both of the tests (Genomics and MLPA) have been run and no deleterious variant or unclassified variant detected
2	One or more deleterious variants detected (Genomic or MLPA) (If 1 should be homozygous)
-3	One or more unclassifed variants detected (Genomic)
3	One or more unclassified variants detected (Genomic or MLPA)
4	One or more deleterious AND one or more unclassifed variants detected (Genomic or MLPA)

6	**C_MUTYH**	number (1,0)	Required:FALSE
	Presence of deleterious mutation or unclassified variant in MUTYH gene		

Allowable Values	
-8	No data transmitted from centers
-9	Failed or equivocal MLPA result or failed sequencing result
-1	One of the tests (Genomic or MLPA) has been run, and no deleterious variant or unclassified variant detected
1	Both of the tests (Genomics and MLPA) have been run and no deleterious variant or unclassified variant detected
2	One or more deleterious variants detected (Genomic or MLPA) (If 1 should be homozygous)
-3	One or more unclassifed variants detected (Genomic)
3	One or more unclassified variants detected (Genomic or MLPA)
4	One or more deleterious AND one or more unclassifed variants detected (Genomic or MLPA)

7	**C_EPCAM**	number (1,0)	Required:FALSE
	Presence of deleterious mutation or unclassified variant in EPCAM gene		

Allowable Values	

-8	No data transmitted from centers
-9	Failed or equivocal MLPA result or failed sequencing result
-1	One of the tests (Genomic or MLPA) has been run, and no deleterious variant or unclassified variant detected
1	Both of the tests (Genomics and MLPA) have been run and no deleterious variant or unclassified variant detected
2	One or more deleterious variants detected (Genomic or MLPA) (If 1 should be homozygous)
-3	One or more unclassifed variants detected (Genomic)
3	One or more unclassified variants detected (Genomic or MLPA)
4	One or more deleterious AND one or more unclassifed variants detected (Genomic or MLPA)