InSiGHT Variant Interpretation Committee: Mismatch Repair Gene Variant Classification Criteria

Rules for Variant Classification:

Rules describing the 5 class system for classification of MMR gene variants were devised and documented by Amanda Spurdle and Bryony Thompson in September 2009 for standardised classification of variants in the Colon Cancer Family Registry database, and revised by Amanda Spurdle, Bryony Thompson, Sean Tavtigian and Marc Greenblatt during April 2011, in collaboration with the InSiGHT Variant Interpretation Committee (VIC), modulated by input from committee members during ongoing VIC meetings. They are based on the following:

- The 5 class system described for quantitative assessment of variant pathogenicity in Plon et al.[1], using a multifactorial likelihood model[2-5] as applied to MMR gene variants[6,7];
- The 5 class system for interpretation of splicing variants and aberrations by Spurdle et al.[8];
- The classification of sequence changes according to standard clinical practice – that is, description of variants generally considered pathogenic (clinically relevant in a genetic counselling setting such that germline variant status is used to inform patient and family management) or non-pathogenic (significant evidence against being a dominant high-risk pathogenic mutation); and
- The documentation of non-quantitative methods that have been used to classify variants in the literature.

For a given class, a variant is required to satisfy all the criteria listed for at least one bullet-point that falls within that class. The symbol "✓" represents an "AND" statement. The footnote # describes the rationale for suggested sample numbers, and the rationale for use of indicative MSI or IHC information. The interpretation of functional assays is assisted by a flowchart developed for this purpose (Figure 1).

These criteria provide a baseline for standardized clinical classification of MMR gene sequence variation that may be linked to patient and family management in the genetic counselling arena according to published guidelines [1]. Use of the InSiGHT database, and associated interpretation relating to pathogenicity, is subject to User discretion and responsibility. Whereas InSiGHT has developed processes to assign pathogenicity utilizing high-level expertise available through its membership, such assignments are subject to change with the availability of new information and interpretation processes. Information submitted to the InSiGHT database is available for individual enquiry for clinical use, but any collective use of the data for research or other purposes transgresses agreements InSiGHT has with the valuable community of researchers, clinicians and diagnostic laboratories that generously support the database through their submissions. All users of the database are encouraged to submit their own variants to support the InSiGHT international collaboration to share gene variant data relating to gastrointestinal tumours. Submissions should be directed to the InSiGHT curator John-Paul Plazzer (johnpaul@variome.org).

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.

[b] As per CMMRD consortium guidelines[11].

[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].

[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]

[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).

[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

Important Notes

PMS2 NGS results need confirmation by other orthogonal assays as well as functional assessment (e.g. Long-Range or cDNA), if variants are located in the pseudogene regions (exons 11-15)

Gene-specific penetrance estimates are available at http://lscarisk.org/

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.

[b] As per CMMRD consortium guidelines[11].

[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].

[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]

[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).

[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

Class 5 – Pathogenic

- Variants with probability of pathogenicity >0.99 (not based on co-segregation in one family, and without conflicting tumour data) using a multifactorial likelihood model
- Variants in the initiation codon of MLH1
- Coding sequence variation resulting in a stop codon i.e. a nonsense or frameshift alteration that is not in the last exon of MSH2, MSH6 or PMS2, and is not after codon 743 in MLH1[a]. For variants in the last exon of MSH2, MSH6 or PMS2 or after codon 743 in MLH1, use other criteria to classify.
- Variants where mRNA assays using **RNA derived from patient constitutional** biological samples indicate that the variant allele results in a splicing aberration (with evidence that the variant allele produces no full-length/reference transcript) leading to premature stop codon or in-frame deletion disrupting a functional domain or protein conformation. **Splicing aberration must be confirmed in a minigene assay or an additional RNA assay from an independent laboratory if it is not a predicted splice site mutation.**
- De novo where both maternity and paternity is confirmed in a case with MMR deficient (MSI/immunoloss consistent with affected gene, with no MLH1 methylation) Lynch spectrum tumour
- Large genomic deletions
- Large genomic duplications shown by laboratory studies (which define the breakpoints of the duplication) to result in a frameshift before the last splice junction
- Variants demonstrating all of the following characteristics with no conflicting results (combined evidence achieves estimated LR >100:1, posterior >0.99 with prior 0.5):

✓ Variant-specific abrogated function in protein or mRNA based lab assays (see MMR functional assay supplementary material, Figure 1)
  - OR - co-occurrence *in trans* with a known pathogenic sequence variant in the same gene in a patient with clinical features consistent with Constitutional Mismatch Repair Deficiency (CMMRD) and documented MMR deficiency in normal cells [b]
  - OR - presence of the variant on different haplotypes across families indicating that reported LS clinical features are not due to an undiscovered sequence change *in cis* with the variant

✓ Evidence for co-segregation with disease where pedigree information provided allows calculation of likelihood ratio (LR) of ≥10:1 [c] or >5 informative meioses
  - OR - at least one revised Amsterdam criteria[9] family with ≥4 affected carriers
  - OR - across ≥2 revised Amsterdam criteria families reported to show segregation with disease (no further information provided)
  - OR - ≥2 families with ≥3 affected non-proband carriers in each family

✓ ≥2 independent MSI-H tumours using a standard panel of 5-10 markers [d] **and/or** loss of MMR protein expression consistent with the variant location (may include tumour information from proband)

✓ evidence that variant is not an undescribed polymorphism at allele frequency >1% (minor allele frequency [MAF] >0.01) in appropriate population control reference groups [e]

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.

[b] As per CMMRD consortium guidelines[11].

[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].

[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]

[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).

[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

Class 4 – Likely pathogenic

- Variants with probability of pathogenicity between 0.95-0.99 (or >0.99 if co-segregation in one family and presence of conflicting tumour data) using a multifactorial likelihood model
- Variants in the initiation codon of MSH6 or PMS2
- Variants with monoallelic expression - complete loss of expression of the variant allele (full-length transcript should be analysed with and without NMD block)
- De novo with maternity and/or paternity unconfirmed in a case with MMR deficient (MSI/immunoloss consistent with affected gene, with no MLH1 methylation) Lynch spectrum tumour
- Variants at IVS±1 or IVS±2, or G>non-G at last base of exon if first 6 bases of the intron are not GTRRGT, that are untested for splicing aberrations *in vitro*
- Variants demonstrating (combined evidence achieves LR >20:1, posterior 0.95-0.99 with prior 0.5):

✓ variant-specific abrogated function in protein or mRNA based lab assays (see MMR functional assay supplementary material)
  - OR - co-occurrence *in trans* with a known pathogenic sequence variant in the same gene in a patient with clinical features consistent with CMMRD and documented MMR deficiency in normal cells [b]
  - OR - presence of the variant on different haplotypes across families indicating that reported LS clinical features are not due to an undiscovered sequence change *in cis* with the variant

✓ **plus one of the following:**
  - co-segregation with disease assessed by available pedigree information allows calculation of LR of >5:1 [c] or >3 informative meioses
  - OR - at least one revised Amsterdam criteria family with ≥3 affected carriers
  - OR - ≥2 families with ≥2 affected non-proband carriers
  - OR - ≥2 independent tumours with MSI using a standard panel of 5-10 markers [d] **and/or** loss of MMR protein expression consistent with the variant location (may include tumour information from proband)

- A missense substitution that encodes the same amino acid change with a different underlying nucleotide change as a previously established Class 5 pathogenic missense variant with normal RNA result*, and is absent from appropriate population control reference groups [e]
  * Otherwise, if the previously established Class 5 pathogenic missense variant truly is a splice defect, the new missense variant also has to be investigated on a functional level for RNA splicing.

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.
[b] As per CMMRD consortium guidelines[11].
[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].
[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]
[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).
[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

Class 3 – Uncertain

- Variants with probability of pathogenicity between 0.05-0.949 using a multifactorial likelihood model
- Variants in the initiation codon of MSH2
- Variants that have insufficient evidence (molecular or otherwise) to classify, which may include large genomic duplications not yet shown by laboratory studies to result in a frameshift before the last splice junction, missense alterations, small in-frame insertions/deletions, silent variants, intronic variants, promoter and regulatory region variants

Conflicting lines of evidence (e.g. IHC/MSI does not fit, segregation does not fit, etc)

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.
[b] As per CMMRD consortium guidelines[11].
[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].
[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]
[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).
[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

Class 2 – Likely not pathogenic/little clinical significance

- Variants with probability of pathogenicity between 0.001-0.049 using a multifactorial likelihood model
- Synonymous substitutions and intronic variants with no associated mRNA aberration (either splicing or allelic imbalance) as determined by laboratory assays conducted with nonsense-mediated decay inhibition. Whenever abnormal transcripts are identified at similar levels in controls they will be considered naturally occurring isoforms and not mRNA aberrations
- Variants reported to occur in a specific ethnic group at allele frequency ≥1%, (MAF ≥ 0.01, tested in ≥ 160 individuals) and that have not yet been excluded as known founder pathogenic sequence variants ("founder mutations")
- 5' UTR or promoter region variants which have bi-allelic expression, shown by cDNA experiment that both alleles are expressed ~ 50/50 including one with the variant or at least one heterozygous exonic variant in the transcript (full-length transcript should be analysed with and without NMD block)
- Variants demonstrating:

✓ variant-specific proficient function in protein and mRNA based lab assays (see MMR functional assay supplementary material)
- OR - co-occurrence *in trans* with a known pathogenic sequence variant in the same gene in a patient with colorectal cancer after age 45 (or other LS cancer above the median age of onset for that cancer in LS [f]), and who has no previous or current evidence of clinical manifestations of CMMRD

✓ **plus one of the following:**
- present in control reference groups at allele frequency 0.01-1% (MAF 0.0001-0.01, tested in ≥ 160 individuals)
- OR - lack of co-segregation with disease consistent with a dominant high-risk pathogenic sequence variant in pedigrees (LR≤ 0.01 [c])
- OR - estimated risk with upper bound 95% confidence limit <4, as determined by large well-designed case-control studies that consider size, geography/ethnicity and quality control measures
- OR - ≥3 colorectal tumours with MSS **and/or** no loss of MMR protein expression **and/or** loss of MMR protein(s) in LS spectrum tumours, that is inconsistent with the gene demonstrating genetic variation

- Variants demonstrating any **three** of the following:
- present in control reference groups at allele frequency 0.01-1% (MAF 0.0001-0.01, tested in ≥ 160 individuals)
- OR - lack of co-segregation with disease consistent with a dominant high-risk pathogenic sequence variant in pedigrees (LR≤ 0.01 [c])
- OR - estimated risk with upper bound 95% confidence limit <4, as determined by large well-designed case-control studies that consider size, geography/ethnicity and quality control measures

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.
[b] As per CMMRD consortium guidelines[11].
[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].
[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]
[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).
[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

- OR - ≥3 colorectal tumours with MSS **and/or** no loss of MMR protein expression **and/or** loss of MMR protein(s) in LS spectrum tumours, that is inconsistent with the gene demonstrating genetic variation

- An exonic variant, that encodes the same amino acid change as a previously established Class 1 missense variant with a different underlying nucleotide change, and for which there is no evidence of mRNA aberration (splicing or allelic imbalance) as determined using *in vitro* laboratory assays conducted with nonsense-mediated decay inhibition

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.

[b] As per CMMRD consortium guidelines[11].

[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].

[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]

[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).

[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

Class 1 – Not pathogenic/no clinical significance

- Variants with probability of pathogenicity <0.001 using a multifactorial likelihood model
- Variants reported to occur in control reference groups at allele frequency ≥1% (MAF ≥ 0.01, tested in ≥ 160 individuals) and excluded as founder pathogenic sequence variants
- 5' UTR or promoter region variants demonstrating:

✓ Bi-allelic expression, shown by cDNA experiment that both alleles are expressed ~ 50/50 including one with the variant or at least one heterozygous exonic variant in the transcript (full-length transcript should be analysed with and without NMD block)

✓ **Meets Class 1 Frequency criteria:**
present in control reference groups at allele frequency ≥ 0.01% (MAF ≥ 0.0001, tested in ≥ 160 individuals)

- OR - lack of co-segregation with disease consistent with a dominant high-risk pathogenic sequence variant in pedigrees (LR≤ 0.01 [c])

- Variants demonstrating:

✓ Variant-specific proficient function in protein and mRNA based lab assays (see MMR functional assay supplementary material)
- OR - co-occurrence *in trans* with a known pathogenic sequence variant in the same gene in a patient with colorectal cancer after age 45 (or other LS cancer above the median age of onset for that cancer in LS [f]), and who has no previous or current evidence of clinical manifestations of CMMRD

✓ **Plus any two of the following:**
- present in control reference groups at allele frequency 0.01-1% (MAF 0.0001-0.01, tested in ≥ 160 individuals)
- **OR -** lack of co-segregation with disease consistent with a dominant high-risk pathogenic sequence variant in pedigrees (LR≤ 0.01 [c])
- **OR -** estimated risk with upper bound 95% confidence limit <4, as determined by large well-designed case-control studies that consider size, geography/ethnicity and quality control measures
- **OR -** ≥3 colorectal tumours with MSS **and/or** no loss of MMR protein expression **and/or** loss of MMR protein(s) in LS spectrum tumours, that is inconsistent with the gene demonstrating genetic variation

- Variants demonstrating all of the following:

✓ Present in control reference groups at allele frequency 0.01-1% (MAF 0.0001-0.01, tested in ≥ 160 individuals)

✓ Lack of co-segregation with disease consistent with a dominant high-risk pathogenic sequence variant in pedigrees (LR≤ 0.01 [c])

✓ Estimated risk <1.5, with upper bound 95% confidence limit <4, as determined by large well-designed case-control studies that consider size, geography/ethnicity and quality control measures

✓ ≥3 colorectal tumours with MSS **and/or** no loss of MMR protein expression **and/or** loss of MMR protein(s) in LS spectrum tumours, that is inconsistent with the gene demonstrating genetic variation

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.
[b] As per CMMRD consortium guidelines[11].
[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].
[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]
[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).
[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

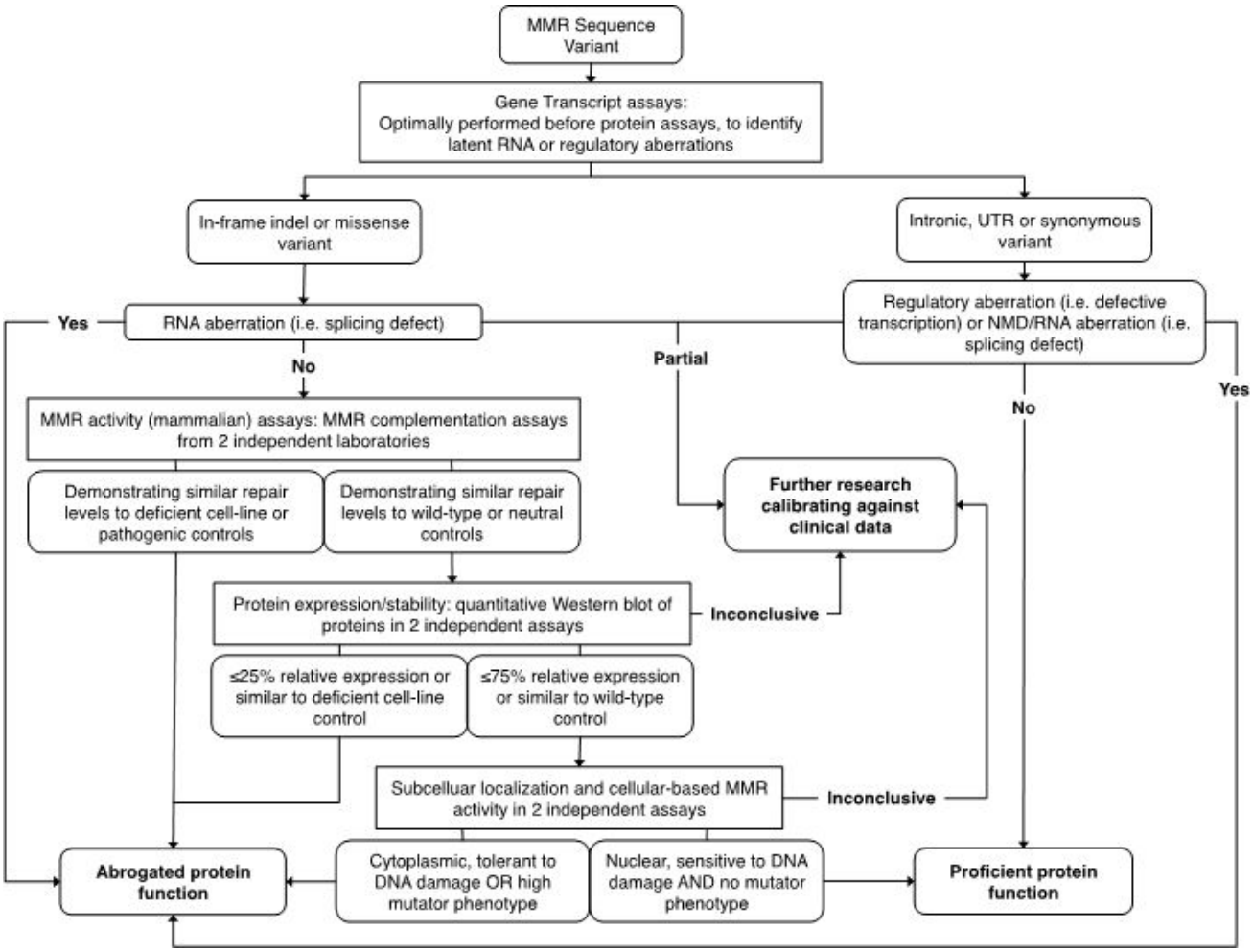


Figure 1. Flowchart used to assist in interpretation of available functional assay data. For variants that had normal/inconclusive/intermediate MMR activity in 2 independent assays, but deficient protein function in 2 independent assays, abrogated function was assigned. NMD – nonsense mediated decay.

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.

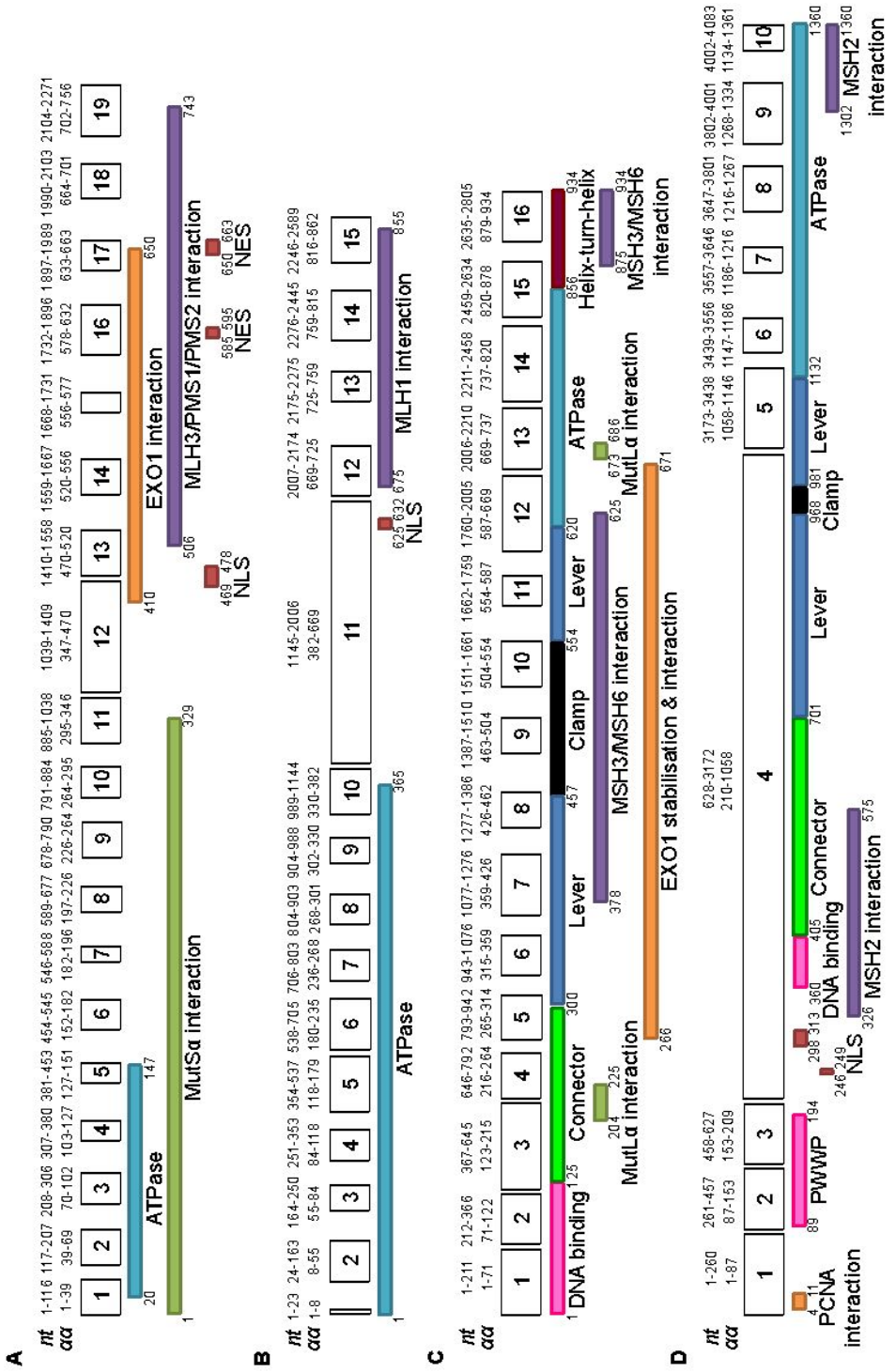
[b] As per CMMRD consortium guidelines[11].

[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].

[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]

[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).

[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.



[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.

[b] As per CMMRD consortium guidelines[11].

[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].

[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]

[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).

[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

Figure 2. Linear schematic of mismatch repair gene structure and functional domains. The nucleotide (nt) and amino acid (αα) positions are numbered for the exon boundaries of *MLH1* (A), *PMS2* (B), *MSH2* (C), and *MSH6* (D). Position numbering is based on Human Genome Variation Society (HGVS) nomenclature[12]. The proteins include the following functional domains: ATPase domain – MLH1[13-18], PMS2[19], MSH2[20,21], and MSH6[21-23]; MutSα interaction in MLH1[17,24]; EXO1 interaction and/or stabilisation domain – MLH1 and MSH2[25]; nuclear localisation signal (NLS) and/or nuclear export signal (NES) – MLH1[26], PMS2[27], and MSH6[28]; MLH1/MLH3/PMS1/PMS2 interaction domain – MLH1 and PMS2[29,30]; DNA binding domain – MSH2[20,31] and MSH6[32]; connector and lever domains – MSH2[20,31] and MSH6[22]; MutLα interaction domains in MSH2[33]; clamp domain – MSH2[20,31] and MSH6[34]; helix-turn-helix domain in MSH2[20,31]; MSH2/MSH3/MSH6 interaction domains – MSH2 and MSH6[35]; PCNA interaction domain in MSH6[36]; PWWP domain in MSH6 binds double-stranded DNA non-specifically[37].

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.

[b] As per CMMRD consortium guidelines[11].

[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].

[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]

[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).

[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

Footnotes to Mismatch Repair Gene Variant Classification Guidelines:

The upper 95% confidence limit of frequency for an allele observed once in 160 individuals (320 chromosomes) is <0.01, the allele frequency considered sufficient for classification as class 1 (not pathogenic/low clinical significance). Characteristics of a pathogenic variant were selected to achieve a combined likelihood ratio of >100:1 in conjunction with a prior of 0.5 (estimated prior irrespective of *in silico* predictions). Namely, the co-segregation and family history descriptions were selected to estimate segregation odds of 10:1, results from functional assays were assumed to have likelihood ratio 5:1 and MSI/IHC data was conservatively assumed to carry a likelihood ratio of 5:1. Tumour data was based on two tumours for MSI and/**or** IHC to allow for the possibility that pathogenic missense alterations may not all demonstrate loss of MMR protein expression. Where both MSI and IHC information are available, MSI-H results will take precedence over normal immunohistochemical results, since MSI is more specific than MMR immunohistochemistry in colorectal cancers[38,39]. For variants that demonstrate tumour MSS and loss of MMR protein expression, technical or other explanations[38,40-42] for the discrepancy should be investigated. The need for multiple results was implemented to minimize the chance of a "sporadic" MSI-H or negative immunohistochemical result for MLH1 methylated tumours (0.15x0.15 = 0.0225). Independent tumours can include multiple primary tumours from a single individual.

References:

1. Plon, S.E. *et al.* Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* **29**, 1282-91 (2008).
2. Goldgar, D.E. *et al.* Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Genet* **75**, 535-44 (2004).
3. Easton, D.F. *et al.* A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. *Am J Hum Genet* **81**, 873-83 (2007).
4. Goldgar, D.E. *et al.* Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. *Hum Mutat* **29**, 1265-72 (2008).
5. Tavtigian, S.V., Greenblatt, M.S., Lesueur, F. & Byrnes, G.B. In silico analysis of missense substitutions using sequence-alignment based methods. *Hum Mutat* **29**, 1327-36 (2008).
6. Arnold, S. *et al.* Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. *Hum Mutat* **30**, 757-70 (2009).
7. Spurdle, A.B. Clinical relevance of rare germline sequence variants in cancer genes: evolution and application of classification models. *Curr Opin Genet Dev* **20**, 315-23 (2010).
8. Spurdle, A.B., Couch, F.J., Hogervorst, F.B., Radice, P. & Sinilnikova, O.M. Prediction and assessment of splicing alterations: implications for clinical testing. *Hum Mutat* **29**, 1304-13 (2008).

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.

[b] As per CMMRD consortium guidelines[11].

[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].

[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]

[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).

[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

9. Vasen, H.F., Watson, P., Mecklin, J.P. & Lynch, H.T. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* **116**, 1453-6 (1999).
10. Bakry, D. *et al.* Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: Report from the constitutional mismatch repair deficiency consortium. *Eur J Cancer* (2014).
11. Thompson, B.A. *et al.* Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Hum Mutat* **34**, 255-65 (2013).
12. Taschner, P.E. & den Dunnen, J.T. Describing structural changes by extending HGVS sequence variation nomenclature. *Hum Mutat* **32**, 507-11 (2011).
13. Ban, C., Junop, M. & Yang, W. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell* **97**, 85-97 (1999).
14. Tran, P.T. & Liskay, R.M. Functional studies on the candidate ATPase domains of Saccharomyces cerevisiae MutLalpha. *Mol Cell Biol* **20**, 6390-8 (2000).
15. Raschle, M., Dufner, P., Marra, G. & Jiricny, J. Mutations within the hMLH1 and hPMS2 subunits of the human MutLalpha mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSalpha. *J Biol Chem* **277**, 21810-20 (2002).
16. Kondo, E., Suzuki, H., Horii, A. & Fukushige, S. A yeast two-hybrid assay provides a simple way to evaluate the vast majority of hMLH1 germ-line mutations. *Cancer Res* **63**, 3302-8 (2003).
17. Plotz, G., Raedle, J., Brieger, A., Trojan, J. & Zeuzem, S. N-terminus of hMLH1 confers interaction of hMutLalpha and hMutLbeta with hMutSalpha. *Nucleic Acids Res* **31**, 3217-26 (2003).
18. Hardt, K. *et al.* Missense variants in hMLH1 identified in patients from the German HNPCC consortium and functional studies. *Fam Cancer* (2011).
19. Guarne, A., Junop, M.S. & Yang, W. Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase. *EMBO J* **20**, 5521-31 (2001).
20. Lutzen, A., de Wind, N., Georgijevic, D., Nielsen, F.C. & Rasmussen, L.J. Functional analysis of HNPCC-related missense mutations in MSH2. *Mutat Res* **645**, 44-55 (2008).
21. Iaccarino, I., Marra, G., Palombo, F. & Jiricny, J. hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSalpha. *EMBO J* **17**, 2677-86 (1998).
22. Schopf, B. *et al.* Interplay between mismatch repair and chromatin assembly. *Proc Natl Acad Sci U S A* (2012).
23. Iaccarino, I., Marra, G., Dufner, P. & Jiricny, J. Mutation in the magnesium binding site of hMSH6 disables the hMutSalpha sliding clamp from translocating along DNA. *J Biol Chem* **275**, 2080-6 (2000).
24. Plotz, G. *et al.* Mutations in the MutSalpha interaction interface of MLH1 can abolish DNA mismatch repair. *Nucleic Acids Res* **34**, 6574-86 (2006).
25. Schmutte, C., Sadoff, M.M., Shim, K.S., Acharya, S. & Fishel, R. The interaction of DNA mismatch repair proteins with human exonuclease I. *J Biol Chem* **276**, 33011-8 (2001).

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.

[b] As per CMMRD consortium guidelines[11].

[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].

[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]

[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).

[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

26. Brieger, A. *et al.* Characterization of the nuclear import of human MutLalpha. *Mol Carcinog* **43**, 51-8 (2005).
27. Leong, V., Lorenowicz, J., Kozij, N. & Guarne, A. Nuclear import of human MLH1, PMS2, and MutLalpha: redundancy is the key. *Mol Carcinog* **48**, 742-50 (2009).
28. Gassman, N.R. *et al.* Cooperative Nuclear Localization Sequences Lend a Novel Role to the N-Terminal Region of MSH6. *PLoS One* **6**, e17907 (2011).
29. Guerrette, S., Acharya, S. & Fishel, R. The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. *J Biol Chem* **274**, 6336-41 (1999).
30. Kondo, E., Horii, A. & Fukushige, S. The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. *Nucleic Acids Res* **29**, 1695-702 (2001).
31. Lamers, M.H. *et al.* The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature* **407**, 711-7 (2000).
32. Clark, A.B., Deterding, L., Tomer, K.B. & Kunkel, T.A. Multiple functions for the N-terminal region of Msh6. *Nucleic Acids Res* **35**, 4114-23 (2007).
33. Mendillo, M.L. *et al.* A conserved MutS homolog connector domain interface interacts with MutL homologs. *Proc Natl Acad Sci U S A* **106**, 22223-8 (2009).
34. Mukherjee, S. & Feig, M. Conformational change in MSH2-MSH6 upon binding DNA coupled to ATPase activity. *Biophys J* **96**, L63-5 (2009).
35. Guerrette, S., Wilson, T., Gradia, S. & Fishel, R. Interactions of human hMSH2 with hMSH3 and hMSH2 with hMSH6: examination of mutations found in hereditary nonpolyposis colorectal cancer. *Mol Cell Biol* **18**, 6616-23 (1998).
36. Kleczkowska, H.E., Marra, G., Lettieri, T. & Jiricny, J. hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes Dev* **15**, 724-36 (2001).
37. Laguri, C. *et al.* Human mismatch repair protein MSH6 contains a PWWP domain that targets double stranded DNA. *Biochemistry* **47**, 6199-207 (2008).
38. Lucci-Cordisco, E., Boccuto, L., Neri, G. & Genuardi, M. The use of microsatellite instability, immunohistochemistry and other variables in determining the clinical significance of MLH1 and MSH2 unclassified variants in Lynch syndrome. *Cancer Biomark* **2**, 11-27 (2006).
39. Mangold, E. *et al.* Tumours from MSH2 mutation carriers show loss of MSH2 expression but many tumours from MLH1 mutation carriers exhibit weak positive MLH1 staining. *J Pathol* **207**, 385-95 (2005).
40. Pastrello, C. *et al.* Stability of BAT26 in tumours of hereditary nonpolyposis colorectal cancer patients with MSH2 intragenic deletion. *Eur J Hum Genet* **14**, 63-8 (2006).
41. Samowitz, W.S. & Slattery, M.L. Missense mismatch repair gene alterations, microsatellite instability, and hereditary nonpolyposis colorectal cancer. *J Clin Oncol* **20**, 3178; author reply 3178-9 (2002).
42. Xie, J. *et al.* An MLH1 Mutation Links BACH1/FANCJ to Colon Cancer, Signaling, and Insight toward Directed Therapy. *Cancer Prev Res (Phila)* **3**, 1409-1416 (2010).

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.

[b] As per CMMRD consortium guidelines[11].

[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].

[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]

[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).

[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.

43. Thompson, D., Easton, D.F. & Goldgar, D.E. A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am J Hum Genet* **73**, 652-5 (2003).
44. Dowty, J.G. *et al*. Cancer risks for MLH1 and MSH2 mutation carriers. *Hum Mutat* **34**, 490-7 (2013)..
45. Baglietto, L. *et al.* Risks of Lynch syndrome cancers for MSH6 mutation carriers. *J Natl Cancer Inst* **102**, 193-201 (2010).
46. Senter, L. *et al.* The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. *Gastroenterology* **135**, 419-28 (2008).
47. Lindor, N.M. *et al.* Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. *J Clin Oncol* **20**, 1043-8 (2002).
48. Muller, W. *et al.* The reliability of immunohistochemistry as a prescreening method for the diagnosis of hereditary nonpolyposis colorectal cancer (HNPCC)--results of an international collaborative study. *Fam Cancer* **1**, 87-92 (2001).
49. Buhard, O. *et al.* Multipopulation analysis of polymorphisms in five mononucleotide repeats used to determine the microsatellite instability status of human tumors. *J Clin Oncol* **24**, 241-51 (2006).

[a] A known functional protein domain is reported to harbor sequence variants that introduce deleterious changes to protein function (via missense alteration, protein sequence deletion, or protein truncation in the last exon) AND are associated with high risk of cancer. Physical boundaries for functional domains are shown in Figure 2.
[b] As per CMMRD consortium guidelines[11].
[c] Likelihood ratios for segregation can be derived by Bayes factor analysis adapted from the method of Thompson et al[44], as described previously[6]. Penetrance estimates for *MLH1* and *MSH2* variants to be derived from Dowty et al[45] and those for *MSH6* and *PMS2* variants to be derived from Baglietto et al[46] and Senter et al[47].
[d] Standard MSI markers panel: BAT25, BAT26, BAT40, BAT34, D5S346, D17S250, ACTC, D18S55, D10S197, MYCL [45]; D2S123, D18S69 [46]; NR21, NR24, NR27 [47]
[e] Outbred control reference groups currently used for this purpose include datasets from the Exome Aggregation Consortium (exac.broadinstitute.org) and Genome Aggregation Database (gnomad.broadinstitute.org).
[f] Lynch syndrome tumours include: colorectal/colon/rectal, endometrial, ovarian, small bowel/small intestine, renal pelvis, ureter, and stomach/gastric carcinomas, sebaceous skin tumours (adenomas and carcinomas), gliomas.