**Supplemental Table S4: Examples of Complex, Potentially Clinically Significant Sequence Variants.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Gene** | **Nucleotide Description** | **Protein Description** | **Reason for complexity** |
| *EGFR* | NM\_005228.3:c.2236\_2253delGAATTAAGAGAAGCAACAinsTTG | p.Glu746\_Thr751delinsLeu | Horizontally complex variant, SNV and a deletion flank a larger deletion (Figure 3) |
| *KRAS* | NM\_004985.3:c.35G>A; NM\_004985.3:c.35G>T | p.Gly12Asp; p.G12Val | Vertically complex variant (Figure 4) |
| *FLT3* | Internal tandem duplication (ITD) | Gain of function | Duplication with or without insertion of a variable sequence (3 to 400bp) into the *FLT3* gene. Such inframe mutations confer a gain of function rather than loss of function. Both the size of the alteration and variability in flanking sequence create challenges. |
| *BRCA2* | NM\_000059.3:c.1813dupA | p.Ile605Asnfs | In short homopolymer repeat. Left/right shuffling may create different representations. Some NGS methods produce artifacts at such sites, and mutations may be vertically heterogeneous. |
| *BRCA2* | NM\_000059.3:c.1310\_1313delAAGA | p.Lys437Ilefs | In short quad-nucleotide repeat. Left/right shuffling may create different representations. |
| *BRCA2* | NM\_000059.3:c.5007\_5008insAluY | Protein disruption | Large insertion of a transposable element (Alu). These inserted sequences are typically variable. |
| *MLH1* | NM\_000249.3:c.1852\_1854delAAG | p.Lys618del | In short tri-nucleotide repeat. Left/right shuffling may create different representations. |
| *MLH1* | NM\_000249.3:c.232\_243delGAAAGGTTCACTinsATGTAAGG | p.Glu78Metfs\*2 | Complex call - different variant calling algorithms may represent this alteration differently, including as multiple independent variants. |
| *MSH2* | NM\_000251.2:c.942+3A>T | Splice affecting | In long homopolymer repeat. All current NGS methods produce artifacts at such sites, and mutations will be vertically heterogeneous. |
| *MSH6* | NM\_000179.2:c.2056\_2060delGGTTGinsCTTCTACCTCAAAAA | p.Gly686Leufs | Complex call - different variant calling algorithms may represent this alteration differently, including as multiple independent variants. |
| *PMS2* | NM\_000535.5:c.2243\_2246delAGAA also  NM\_000535.5:c.2253T>C | NP\_000068.1:p.Ala68Leu also  NP\_478102.2:p.Arg82Leu | SNV near indel, both are in pseudogene associated region of *PMS2*, highly homologous to *PMS2CL* |
| *CDKN2A* | NM\_000077.4:c.9\_32dupGGCGGCGGGGAGCAGCATGGAGCC | p.Pro11\_Ser12insAlaAlaGlySerSerMetGluPro | Mutation is 3rd copy of repeat in an endogenous tandem duplication; Also this region is 80% GC and low complexity. |
| *CDKN2A* | NM\_000077.4:c.202\_203delCCinsTT and  NM\_058195.3:c.245\_246delCCinsTT | p.Ala68Leu and  p.Ala68Leu respectively | These two variants are the same and affect the two different gene products (p14ARF and P16INK4a) of the *CDKN2A* locus |
| *TP53* | NM\_000546.5:c.215\_221delCCGTGGCinsGTG | p.Pro72Argfs\*50 | At end of homopolymer repeat region. Complex call - different variant calling algorithms may represent this alteration differently. |
| *MSH2* | NM\_000251.2:inv exons 1-7 (“Boland inversion”) | Loss of protein | Breakpoints in non-coding regions, requires specialized bioinformatics. |

This table is not comprehensive and does not represent all of known complex, potentially clinically significant sequence variants. Each of these either requires specialized methods to reliably detect the alteration, or presents complexity in converting NGS variant calls into standardized nomenclature (ie, HGVS), or both. For the large structural variants, conventional descriptions rather than HGVS are shown, as is common practice currently, although HGVS is becoming increasingly used for copy number alterations. Some of these variants are larger than 21bp and thus are not specifically within scope of this guideline, but if such variants included in a test’s analytic range then their detection and reporting should be specifically validated.