**Test 1**

MAAAACTPRAPFPWIRSLSTRS

MAAAACTPRAPFPWIRSLSTRG

MAAAACTPRAPFPWIRSLSTR

PSATTRPQSPPPPASASGSCASRPSLPPPPPHPSESSPSGAPTSS

PSTARRREEGPPPPSPDG--ASSDAEPEPPSGRTESPATAAETAS

PS R + PPP S G AS + P PP +ES + A T+S

findSAPsAndINDELs('MAAAAA---ATTAGAERIPERKVRAVPARSRRAPSSPPPSPP', 'MAAVGSGGYARNDAGEKLPS-VMAGVPAR--RGQSSPPPAPP', 'MAA + A E++P + VPAR R SSPPP+PP','sp|O43379-3|WDR62\_HUMAN Isoform 3 of WD repeat-containing protein 62 OS=Homo sapiens GN=WDR62', 'Dataset\_A\_asmbl\_25116\_ORF30\_Frame\_-1\_136-2', '1')

'MAAAAA---ATTAGAERIPEPKVRAVPARSRRAPSSPPPSPP'

'MAAVGSGGYARNDAGEKLPS-VMAGVPAR--RGQSSPPPAPP'

'MAA + A E++P + VPAR R SSPPP+PP'

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |  |  | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| M | A | A | A | A | A | - | - | - | A | T | T | A | G | A | E | R | I | P | E | R | K | V | R | A | V | P | A | R | S | R | R | A | P | S | S | P | P | P | S | P | P |
| M | A | A | V | G | S | G | G | Y | A | R | N | D | A | G | E | K | L | P | S | - | V | M | A | G | V | P | A | R | - | - | R | G | Q | S | S | P | P | P | A | P | P |
| M | A | A |  |  | + |  |  |  | A |  |  |  |  |  | E | + | + | P |  |  |  | + |  |  | V | P | A | R |  |  | R |  |  | S | S | P | P | P | + | P | P |

**Test 2**

qSeq=”RGEPQQDCCVKTELLGEETPMAADEGSAEKQAGEAHMAADGETNGSCENSDASSHANAAKHTQDSARVNPQDGTNTLTRIAENGVSERDSEAAKQNHVTADDFVQTSVIGSNGYILNKPALQAQPLRTTSTLASSLPGHAAKTLPGGAGKGRTPSAFPQTPAAPPATLGEGSADTEDRKLPAPGADVKVHRARKTMPKSVVGLHAASKDPREVREARDHKEPKEEINKNISDFGRQQLLPPFPSLHQSLPQNQCYMATTKSQTA-------AAVSRKKKRRMGTYSLVPKKKTKVLKQRTVIEMFKSITHSTVGSKGEKDLGASSLHVNGESLEMDSDEDDSEELEEDDGHGAEQAAAFPTEDSRTSKESMSEADRAQKMDGESEEEQESVDTGEEEEGGDESDLSSESSIKKKFLKRKGKTDSPWIKPARKRRRRSRKKPSGALGSESYKSSAGSAEQTAPGDSTGYMEVSLDSLDLRVKGILSS------------QAEGLANGPDV---LETDGLQEVPLCS---CRMETPKSREITTLANN----QCMATES---------------VDHE-----------GNFMECQPESSISHRFHKDCASRVNNASYCPHCGEESSKAKEVTIAKADTTSTVTPVPGQEKGSALEGRADTTTGSAAGPPLSEDDKLQGAASHVPEGFDPTGPAGLGRPTPGLSQGPGKETLESALIALDSEKPKKLRFHPKQLYFSARQGELQKVLLMLVDGIDPNFKMEHQNKRSPLHAAAEAGHVDICHMLVQAGANIDTCSEDQRTPLMEAAENNHLEAVKYLIKAGALVDPKDAEGSTCLHLAAKKGHYEVVQYLLSNGQMDVNCQDDGGWTPMIWATEYKHVDLVKLLLSKGSDINIRDNEENICLHWAAFSGCVDIAEILLAAKCDLHAVNIHGDSPLHIAARENRYDCVVLFLSRDSDVTLKNKEGETPLQCASLNSQVWSALQMSKALQDSAPDRPSPVERIVSRDIARGYERIPIPCVNAVDSEPCPSNYKYVSQNCVTSPMNIDRNITHLQYCVCIDDCSSSNCMCGQLSMRCWYDKDGRLLPEFNMAEPPLIFECNHACSCWRNCRNRVVQNGLRARLQLYRTRDMGWGVRSLQDIPPGTFVCEYVGELISDSEADVREEDSYLFDLDNKDGEVYCIDARFYGNVSRFINHHCEPNLVPVRVFMAHQDLRFPRIAFFSTRLIEAGEQLGFDYGERFWDIKGKLFSCRCGSPKCRHSSAALAQRQASAAQEAQEDGLPDTSSAAAADPL”

sSeq=”RGEPQQDCCVKTELLGEETPMAADEGSAEKQAGEAHMAADGETNGSCENSDASSHANAAKHTQDSARVNPQDGTNTLTRIAENGVSERDSEAAKQNHVTADDFVQTSVIGSNGYILNKPALQAQPLRTTSTLASSLPGHAAKTLPGGAGKGRTPSAFPQTPAAPPATLGEGSADTEDRKLPAPGADVKVHRARKTMPKSVVGLHAASKDPREVREARDHKEPKEEINKNISDFGRQQLLPPFPSLHQSLPQNQCYMATTKSQTACLPFVLAAAVSRKKKRRMGTYSLVPKKKTKVLKQRTVIEMFKSITHSTVGSKGEKDLGASSLHVNGESLEMDSDEDDSEELEEDDGHGAEQAAAFPTEDSRTSKESMSEADRAQKMDGESEEEQESVDTGEEEEGGDESDLSSESSIKKKFLKRKGKTDSPWIKPARKRRRRSRKKPSGALGSESYKSSAGSAEQTAPGDSTGYMEVSLDSLDLRVKGILSSQAEGLANGPDVLETDGLQEVPLCSCRMETPKSREITTLANNQCMATESVDHELGRCTNSVVKYELMRPSNKAPLLVLCEDHRGRMVKHQCCPGCGYFCTAGNFMECQPESSISHRFHKDCASRVNNASYCPHCGEESSKAKEVTIAKADTTSTVTPVPGQEKGSALEGRADTTTGSAAGPPLSEDDKLQGAASHVPEGFDPTGPAGLGRPTPGLSQGPGKETLESALIALDSEKPKKLRFHPKQLYFSARQGELQKVLLMLVDGIDPNFKMEHQNKRSPLHAAAEAGHVDICHMLVQAGANIDTCSEDQRTPLMEAAENNHLEAVKYLIKAGALVDPKDAEGSTCLHLAAKKGHYEVVQYLLSNGQMDVNCQDDGGWTPMIWATEYKHVDLVKLLLSKGSDINIRDNEENICLHWAAFSGCVDIAEILLAAKCDLHAVNIHGDSPLHIAARENRYDCVVLFLSRDSDVTLKNKEGETPLQCASLNSQVWSALQMSKALQDSAPDRPSPVERIVSRDIARGYERIPIPCVNAVDSEPCPSNYKYVSQNCVTSPMNIDRNITHLQYCVCIDDCSSSNCMCGQLSMRCWYDKDGRLLPEFNMAEPPLIFECNHACSCWRNCRNRVVQNGLRARLQLYRTRDMGWGVRSLQDIPPGTFVCEYVGELISDSEADVREEDSYLFDLDNKDGEVYCIDARFYGNVSRFINHHCEPNLVPVRVFMAHQDLRFPRIAFFSTRLIEAGEQLGFDYGERFWDIKGKLFSCRCGSPKCRHSSAALAQRQASAAQEAQEDGLPDTSSAAAADPL”

mSeq=”RGEPQQDCCVKTELLGEETPMAADEGSAEKQAGEAHMAADGETNGSCENSDASSHANAAKHTQDSARVNPQDGTNTLTRIAENGVSERDSEAAKQNHVTADDFVQTSVIGSNGYILNKPALQAQPLRTTSTLASSLPGHAAKTLPGGAGKGRTPSAFPQTPAAPPATLGEGSADTEDRKLPAPGADVKVHRARKTMPKSVVGLHAASKDPREVREARDHKEPKEEINKNISDFGRQQLLPPFPSLHQSLPQNQCYMATTKSQTA AAVSRKKKRRMGTYSLVPKKKTKVLKQRTVIEMFKSITHSTVGSKGEKDLGASSLHVNGESLEMDSDEDDSEELEEDDGHGAEQAAAFPTEDSRTSKESMSEADRAQKMDGESEEEQESVDTGEEEEGGDESDLSSESSIKKKFLKRKGKTDSPWIKPARKRRRRSRKKPSGALGSESYKSSAGSAEQTAPGDSTGYMEVSLDSLDLRVKGILSS + +GL P +ET +E+ + C E+ N+ + M + V H+ GNFMECQPESSISHRFHKDCASRVNNASYCPHCGEESSKAKEVTIAKADTTSTVTPVPGQEKGSALEGRADTTTGSAAGPPLSEDDKLQGAASHVPEGFDPTGPAGLGRPTPGLSQGPGKETLESALIALDSEKPKKLRFHPKQLYFSARQGELQKVLLMLVDGIDPNFKMEHQNKRSPLHAAAEAGHVDICHMLVQAGANIDTCSEDQRTPLMEAAENNHLEAVKYLIKAGALVDPKDAEGSTCLHLAAKKGHYEVVQYLLSNGQMDVNCQDDGGWTPMIWATEYKHVDLVKLLLSKGSDINIRDNEENICLHWAAFSGCVDIAEILLAAKCDLHAVNIHGDSPLHIAARENRYDCVVLFLSRDSDVTLKNKEGETPLQCASLNSQVWSALQMSKALQDSAPDRPSPVERIVSRDIARGYERIPIPCVNAVDSEPCPSNYKYVSQNCVTSPMNIDRNITHLQYCVCIDDCSSSNCMCGQLSMRCWYDKDGRLLPEFNMAEPPLIFECNHACSCWRNCRNRVVQNGLRARLQLYRTRDMGWGVRSLQDIPPGTFVCEYVGELISDSEADVREEDSYLFDLDNKDGEVYCIDARFYGNVSRFINHHCEPNLVPVRVFMAHQDLRFPRIAFFSTRLIEAGEQLGFDYGERFWDIKGKLFSCRCGSPKCRHSSAALAQRQASAAQEAQEDGLPDTSSAAAADPL”

findSAPsAndINDELs(qSeq, sSeq, mSeq,' sp|Q9H9B1|EHMT1\_HUMAN', ' Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761', '1')

Before correcting variation count

1 264 0 ACLPFVLA A ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:DEL

1 486 1 SQAEGLANGPDVL S ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:DEL

1 499 3 E Q ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 500 3 T A ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 501 5 D E ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 504 5 QEV ANG ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:ALT

1 508 6 LC DV ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:ALT

1 509 7 CSCR V ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:DEL

1 513 9 M L ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 516 9 PKS DGL ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:ALT

1 519 11 R Q ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 521 12 I V ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 522 12 TTL PLC ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:ALT

1 525 14 A S ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 525 14 ANNQ S ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:DEL

1 530 15 MATESVDH RMETPKSR ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:ALT

1 539 17 L I ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 540 17 GRCT TTLA ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:ALT

1 545 19 S N ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 545 19 SVVKY N ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:DEL

1 550 21 E Q ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 551 21 L C ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 553 22 RPS ATE ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:ALT

1 556 24 N S ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 556 24 NKAPLLVLCEDHRGRM S ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:DEL

1 573 25 K D ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

1 575 27 Q E ProtId= sp|Q9H9B1|EHMT1\_HUMAN;ORFId: Dataset\_A\_asmbl\_55684\_ORF44\_Frame\_3\_33-3761;Type:SAP

**Test 3**

qSeq=”MAADGERSPLLSEPIDGGAGGNGLVGPGGSGAGPGGGLTPSAPPYGAGKHAPPQAFPPFPEGHPAVLPGEDPPPYSPLTSPDSGSAPMITCRVCQSLINVEGKMHQHVVKCGVCNEATPIKNAPPGKKYVRCPCNCLLICKVTSQRIACPRPYCKRIINLGPVHPGPLSPEPQPMGVRVICGHCKNTFLWTEFTDRTLARCPHCRKVSSIGRRYPRKRCICCFLLGLLLAVTATGLAFGTWKHARRYGGIYAAWAFVILLAVLCLGRALYWACMKVSHPVQNFS”

sSeq=”MAADGERSPLLSEPIDGGAGGNGLVGPGGSGAGPGGGLTPSAPPYGA-------AFPPFPEGHPAVLPGEDPPPYSPLTSPDSGSAPMITCRVCQSLINVEGKMHQHVVKCGVCNEATPIKNAPPGKKYVRCPCNCLLICKVTSQRIACPRPYCKRIINLGPVHPGPLSPEPQPMGVRVICGHCKNTFLWTEFTDRTLARCPHCRKVSSIGRRYPRKRCICCFLLGLLLAVTATGLAFGTWKHARRYGGIYAAWAFVILLAVLCLGRALYWACMKVSHPVQNFS”

mSeq=”MAADGERSPLLSEPIDGGAGGNGLVGPGGSGAGPGGGLTPSAPPYGA AFPPFPEGHPAVLPGEDPPPYSPLTSPDSGSAPMITCRVCQSLINVEGKMHQHVVKCGVCNEATPIKNAPPGKKYVRCPCNCLLICKVTSQRIACPRPYCKRIINLGPVHPGPLSPEPQPMGVRVICGHCKNTFLWTEFTDRTLARCPHCRKVSSIGRRYPRKRCICCFLLGLLLAVTATGLAFGTWKHARRYGGIYAAWAFVILLAVLCLGRALYWACMKVSHPVQNFS”

findSAPsAndINDELs(qSeq, sSeq, mSeq,'sp|Q86T03|TM55B\_HUMAN’, 'Dataset\_A\_asmbl\_17297\_ORF19\_Frame\_2\_98-949', '1')

**Test 4**

qSeq=”MAATQKMDM-ASHKMCWLMVLRNS-QF-VFLRK-EERMVMHRLQKMRPPAFHKTQELYRRNPCKDQQRWFTSWRKNILEQQPASSNSWLIRTANCLGNTPCFLPQTTECLEFMCLSRTVPRTLWHGLKIMPTHCSSAALWTGRRTAILPWGSWLRVLGRLVKLSLKQKEY”

sSeq=”VDGVKKLSVCVSEKDNEIEAQRSSWAYPVSHRKSEERMVMHRLQKMRPPAFHKTQELYRRNPCKDQQRWFTSWRKNILEQQPASSNSWLIRTANCLGNTPCFLPQTTECLEFMCLSRTVPRTLWHGLKIMPTHCSSAALWTGRRTAILPWGSWLRVLGRLVKLSLKQKEY”

mSeq=”+ +K+ + S K + R+S + V RK EERMVMHRLQKMRPPAFHKTQELYRRNPCKDQQRWFTSWRKNILEQQPASSNSWLIRTANCLGNTPCFLPQTTECLEFMCLSRTVPRTLWHGLKIMPTHCSSAALWTGRRTAILPWGSWLRVLGRLVKLSLKQKEY”

findSAPsAndINDELs(qSeq, sSeq, mSeq,’sp|Q8IYB7-5|DI3L2\_HUMAN’, ‘Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141’, '1')

Before correction

1 1 1 V M ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 2 2 DGV AAT ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:ALT

1 5 3 K Q ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 7 4 L M ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 8 5 S D ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SAP

1 9 6 V M ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 9 7 VCV M ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:DEL

1 13 8 E H ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SAP

1 15 9 DNE MCW ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:ALT

1 18 10 I L ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 19 11 EAQ MVL ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:ALT

1 23 12 S N ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 24 13 SWA S ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:DEL

1 27 14 Y F ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 27 15 YP F ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:DEL

1 30 16 SH FL ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:ALT

1 33 17 KS K ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:DEL

After correction

1 1 1 V M ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 2 2 DGV AAT ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:ALT

1 5 3 K Q ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 7 4 L M ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 8 5 S D ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SAP

1 9 6 V M ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 9 7 VC M ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:DEL

1 11 8 V A ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:ALT

1 13 9 E H ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SAP

1 15 10 DNE MCW ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:ALT

1 18 11 I L ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 19 12 EAQ MVL ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:ALT

1 23 13 S N ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 24 14 SW S ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:DEL

1 26 15 A Q ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:ALT

1 27 16 Y F ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:SSAP

1 27 17 YP F ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:DEL

1 30 18 SH FL ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:ALT

1 33 19 KS K ProtId=sp|Q8IYB7-5|DI3L2\_HUMAN;ORFId:Dataset\_A\_asmbl\_31213\_ORF20\_Frame\_2\_644-1141;Type:DEL

**Test 5**

qSeq=”MAAAAAAGEARRVLVYGGRGALGSRCVQAFRARNWVTAEVGKLLGEEKVDAILCVAGGWAGGNAKSKSLFKNCDLMWKQSIWTSTISSHLATKHLKEGGLLTLAGAKAALDGTPGL”

sSeq=”MAAAAAAGEARRVLVYGGRGALGSRCVQAFRARNWVTAEVGKLLGEEKVDAILCVAGGWAGGNAKSKSLFKNCDLMWKQSIWTSTISSHLATKHLKEGGLLTLAGAKAALDGTPGM”

mSeq=”MAAAAAAGEARRVLVYGGRGALGSRCVQAFRARNWVTAEVGKLLGEEKVDAILCVAGGWAGGNAKSKSLFKNCDLMWKQSIWTSTISSHLATKHLKEGGLLTLAGAKAALDGTPG+”

findSAPsAndINDELs(qSeq, sSeq, mSeq,’prt1’, ‘orf1’, '1')

Test 6

qSeq=”MAATQKMDM-ASHKMCWLMVLRNS-QF-VFLRK-EERMVMHRLQKMRPPAFHKTQELYRRNPCKDQQRWFTSWRKNILEQQPASSNSWLIRTANCLGNTPCFLPQTTECLEFMCLSRTVPRTLWHGLKIMPTHCSSAALWTGRRTAILPWGSWLRVLGRLVKLSLKQKEY”

sSeq=”VDGVKKLSVCVSEKDNEIEAQRSSWAYPVSHRKSEERMVMHRLQKMRPPAFHKTQELYRRNPCKDQQRWFTSWRKNILEQQPASSNSWLIRTANCLGNTPCFLPQTTECLEFMCLSRTVPRTLWHGLKIMPTHCSSAALWTGRRTAILPWGSWLRVLGRLVKLSLKQKEY”

mSeq=”+ +K+ + S K + R+S + V RK EERMVMHRLQKMRPPAFHKTQELYRRNPCKDQQRWFTSWRKNILEQQPASSNSWLIRTANCLGNTPCFLPQTTECLEFMCLSRTVPRTLWHGLKIMPTHCSSAALWTGRRTAILPWGSWLRVLGRLVKLSLKQKEY”

findSAPsAndINDELs(qSeq, sSeq, mSeq,’prt1’, ‘orf1’, '1')

**Use cases: when both alt and ref contains ‘-’**

Case 1:

M=Abdf jhkushd

Q=Abdfgb--hjhkushd

S=Abdfd-fg-jhkushd

Alt = gb—h

Ref = d-fg-

Case 2:

M=Abdf jhkushd

Q=Abdfb--hgjhkushd

S=Abdf-fg-djhkushd

Alt = b—hg

Ref = -fg-d

Case 3:

M=Abdf jhkushd

Q=Abdfg--bhjhkushd

S=Abdfdfg--jhkushd

Alt = g—bh

Ref = dfg--

Case 4:

M=Abdf jhkushd

Q=Abdf--bhjhkushd

S=Abdffg--jhkushd

Alt = —bh

Ref = fg--

**Bug Fixing:**

**Case 1:**

RGEPQQDCCVKTELLGEETPMAADEGSAEKQAGEAHMAADGETNGSCENSDASSHANAAKHTQDSARVNPQDGTNTLTRIAENGVSERDSEAAKQNHVTADDFVQTSVIGSNGYILNKPALQAQPLRTTSTLASSLPGHAAKTLPGGAGKGRTPSAFPQTPAAPPATLGEGSADTEDRKLPAPGADVKVHRARKTMPKSVVGLHAASKDPREVREARDHKEPKEEINKNISDFGRQQLLPPFPSLHQSLPQNQCYMATTKSQTA-------AAVSRKKKRRMGTYSLVPKKKTKVLKQRTVIEMFKSITHSTVGSKGEKDLGASSLHVNGESLEMDSDEDDSEELEEDDGHGAEQAAAFPTEDSRTSKESMSEADRAQKMDGESEEEQESVDTGEEEEGGDESDLSSESSIKKKFLKRKGKTDSPWIKPARKRRRRSRKKPSGALGSESYKSSAGSAEQTAPGDSTGYMEVSLDSLDLRVKGILSS------------QAEGLANGPDV---LETDGLQEVPLCS---CRMETPKSREITTLANN----QCMATES---------------VDHE-----------GNFMECQPESSISHRFHKDCASRVNNASYCPHCGEESSKAKEVTIAKADTTSTVTPVPGQEKGSALEGRADTTTGSAAGPPLSEDDKLQGAASHVPEGFDPTGPAGLGRPTPGLSQGPGKETLESALIALDSEKPKKLRFHPKQLYFSARQGELQKVLLMLVDGIDPNFKMEHQNKRSPLHAAAEAGHVDICHMLVQAGANIDTCSEDQRTPLMEAAENNHLEAVKYLIKAGALVDPKDAEGSTCLHLAAKKGHYEVVQYLLSNGQMDVNCQDDGGWTPMIWATEYKHVDLVKLLLSKGSDINIRDNEENICLHWAAFSGCVDIAEILLAAKCDLHAVNIHGDSPLHIAARENRYDCVVLFLSRDSDVTLKNKEGETPLQCASLNSQVWSALQMSKALQDSAPDRPSPVERIVSRDIARGYERIPIPCVNAVDSEPCPSNYKYVSQNCVTSPMNIDRNITHLQYCVCIDDCSSSNCMCGQLSMRCWYDKDGRLLPEFNMAEPPLIFECNHACSCWRNCRNRVVQNGLRARLQLYRTRDMGWGVRSLQDIPPGTFVCEYVGELISDSEADVREEDSYLFDLDNKDGEVYCIDARFYGNVSRFINHHCEPNLVPVRVFMAHQDLRFPRIAFFSTRLIEAGEQLGFDYGERFWDIKGKLFSCRCGSPKCRHSSAALAQRQASAAQEAQEDGLPDTSSAAAADPL

RGEPQQDCCVKTELLGEETPMAADEGSAEKQAGEAHMAADGETNGSCENSDASSHANAAKHTQDSARVNPQDGTNTLTRIAENGVSERDSEAAKQNHVTADDFVQTSVIGSNGYILNKPALQAQPLRTTSTLASSLPGHAAKTLPGGAGKGRTPSAFPQTPAAPPATLGEGSADTEDRKLPAPGADVKVHRARKTMPKSVVGLHAASKDPREVREARDHKEPKEEINKNISDFGRQQLLPPFPSLHQSLPQNQCYMATTKSQTACLPFVLAAAVSRKKKRRMGTYSLVPKKKTKVLKQRTVIEMFKSITHSTVGSKGEKDLGASSLHVNGESLEMDSDEDDSEELEEDDGHGAEQAAAFPTEDSRTSKESMSEADRAQKMDGESEEEQESVDTGEEEEGGDESDLSSESSIKKKFLKRKGKTDSPWIKPARKRRRRSRKKPSGALGSESYKSSAGSAEQTAPGDSTGYMEVSLDSLDLRVKGILSSQAEGLANGPDVLETDGLQEVPLCSCRMETPKSREITTLANNQCMATESVDHELGRCTNSVVKYELMRPSNKAPLLVLCEDHRGRMVKHQCCPGCGYFCTAGNFMECQPESSISHRFHKDCASRVNNASYCPHCGEESSKAKEVTIAKADTTSTVTPVPGQEKGSALEGRADTTTGSAAGPPLSEDDKLQGAASHVPEGFDPTGPAGLGRPTPGLSQGPGKETLESALIALDSEKPKKLRFHPKQLYFSARQGELQKVLLMLVDGIDPNFKMEHQNKRSPLHAAAEAGHVDICHMLVQAGANIDTCSEDQRTPLMEAAENNHLEAVKYLIKAGALVDPKDAEGSTCLHLAAKKGHYEVVQYLLSNGQMDVNCQDDGGWTPMIWATEYKHVDLVKLLLSKGSDINIRDNEENICLHWAAFSGCVDIAEILLAAKCDLHAVNIHGDSPLHIAARENRYDCVVLFLSRDSDVTLKNKEGETPLQCASLNSQVWSALQMSKALQDSAPDRPSPVERIVSRDIARGYERIPIPCVNAVDSEPCPSNYKYVSQNCVTSPMNIDRNITHLQYCVCIDDCSSSNCMCGQLSMRCWYDKDGRLLPEFNMAEPPLIFECNHACSCWRNCRNRVVQNGLRARLQLYRTRDMGWGVRSLQDIPPGTFVCEYVGELISDSEADVREEDSYLFDLDNKDGEVYCIDARFYGNVSRFINHHCEPNLVPVRVFMAHQDLRFPRIAFFSTRLIEAGEQLGFDYGERFWDIKGKLFSCRCGSPKCRHSSAALAQRQASAAQEAQEDGLPDTSSAAAADPL

RGEPQQDCCVKTELLGEETPMAADEGSAEKQAGEAHMAADGETNGSCENSDASSHANAAKHTQDSARVNPQDGTNTLTRIAENGVSERDSEAAKQNHVTADDFVQTSVIGSNGYILNKPALQAQPLRTTSTLASSLPGHAAKTLPGGAGKGRTPSAFPQTPAAPPATLGEGSADTEDRKLPAPGADVKVHRARKTMPKSVVGLHAASKDPREVREARDHKEPKEEINKNISDFGRQQLLPPFPSLHQSLPQNQCYMATTKSQTA AAVSRKKKRRMGTYSLVPKKKTKVLKQRTVIEMFKSITHSTVGSKGEKDLGASSLHVNGESLEMDSDEDDSEELEEDDGHGAEQAAAFPTEDSRTSKESMSEADRAQKMDGESEEEQESVDTGEEEEGGDESDLSSESSIKKKFLKRKGKTDSPWIKPARKRRRRSRKKPSGALGSESYKSSAGSAEQTAPGDSTGYMEVSLDSLDLRVKGILSS + +GL P +ET +E+ + C E+ N+ + M + V H+ GNFMECQPESSISHRFHKDCASRVNNASYCPHCGEESSKAKEVTIAKADTTSTVTPVPGQEKGSALEGRADTTTGSAAGPPLSEDDKLQGAASHVPEGFDPTGPAGLGRPTPGLSQGPGKETLESALIALDSEKPKKLRFHPKQLYFSARQGELQKVLLMLVDGIDPNFKMEHQNKRSPLHAAAEAGHVDICHMLVQAGANIDTCSEDQRTPLMEAAENNHLEAVKYLIKAGALVDPKDAEGSTCLHLAAKKGHYEVVQYLLSNGQMDVNCQDDGGWTPMIWATEYKHVDLVKLLLSKGSDINIRDNEENICLHWAAFSGCVDIAEILLAAKCDLHAVNIHGDSPLHIAARENRYDCVVLFLSRDSDVTLKNKEGETPLQCASLNSQVWSALQMSKALQDSAPDRPSPVERIVSRDIARGYERIPIPCVNAVDSEPCPSNYKYVSQNCVTSPMNIDRNITHLQYCVCIDDCSSSNCMCGQLSMRCWYDKDGRLLPEFNMAEPPLIFECNHACSCWRNCRNRVVQNGLRARLQLYRTRDMGWGVRSLQDIPPGTFVCEYVGELISDSEADVREEDSYLFDLDNKDGEVYCIDARFYGNVSRFINHHCEPNLVPVRVFMAHQDLRFPRIAFFSTRLIEAGEQLGFDYGERFWDIKGKLFSCRCGSPKCRHSSAALAQRQASAAQEAQEDGLPDTSSAAAADPL

Case 2:

MHVATVFTDGGPRTLRSLTVSLGPVSKTEGFPKDSRIATTSSSVLLSPSAVESRRNSRVTGNPGDEEFIEPSTENEFGLTSLRWQNDSPTFGEHQLASSSEVQNGSPMSQTETVSRSVAPMRGGEITAHWLLTNSTTSADVTGSSASYPEGVNASVLTQFSDSTVQSGGSHTALGDRSYSESSSTSSSESLNSSAPRGERSTLEDSREPGQALGDSSANAEDRTSGVPSLGTHTLATVTGNGERTLRSVT

MHVATVFTDGGPRTLRSLTVSLGPVSKTEGFPKDSRIATTSSSVLLSPSAVESRRNSRVTGNPGDEEFIEPSTENEFGLTSLRWQNDSPTFGEHQLASSSEVQNGSPMSQTETVSRSVAPMRGGEITAHWLLTNSTTSADVTGSSASYPEGVNASVLTQFSDSTVQSGGSHTALGDRSYSESSSTSSSESLNSSAPRGERSIAGISY--GQVRGTA---IEQRTSSDHTDHTYLSSTFT-KGERALLSIT

MHVATVFTDGGPRTLRSLTVSLGPVSKTEGFPKDSRIATTSSSVLLSPSAVESRRNSRVTGNPGDEEFIEPSTENEFGLTSLRWQNDSPTFGEHQLASSSEVQNGSPMSQTETVSRSVAPMRGGEITAHWLLTNSTTSADVTGSSASYPEGVNASVLTQFSDSTVQSGGSHTALGDRSYSESSSTSSSESLNSSAPRGERS S GQ G + E RTS + T+ +T T GER L S+T

In vcf file no of unique Info column value=4329, but unique ORFs=2297,

2297 passes filters in contigstat file