# iRecSpot-EF: Effective Sequence Based Features for Recombination Hotspot Prediction

Md Rafsan Jani, Md Toha Khan Mozlish,
Sajid Ahmed, Niger Sultana Tahniat, Dewan Md Farid, and
Swakkhar Shatadba

Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh

# Supporting Information (#1):

Details description of ($l$, $k$, $p$)-mers Composition:

Here, $l$ represents length of sequence, $k$ represents k-gap / total number of gap in a sequence, and $p$ represents positions of gap or where gap occurs. Gap represents don't care or any element of X. Nucleotides, $X = \{A, C, G, T\}$

Table 1: Description of ($l$, $k$, $p$)-mers

| Feature Structure | Values of ($l$, $k$, $p$)-mers |
|---|---|
| XX | l=2, k=0, p=None (because k is zero) |
| X _ X | l=3, k=1, p=2 |
| X __ X | l=4, k=2, p=2, 3 |
| X ___ X | l=5, k=3, p=2, 3, 4 |
| X ____ X | l=6, k=4, p=2, 3, 4, 5 |
| X _____ X | l=7, k=5, p=2, 3, 4, 5, 6 |
| .............. | ....................................... |
| XX _ X | l=4, k=1, p=3 |
| XX __ X | l=5, k=2, p=3, 4 |
| XX ___ X | l=6, k=3, p=3, 4, 5 |
| XX ____ X | l=7, k=4, p=3, 4, 5, 6 |
| XX _____ X | l=8, k=5, p=3, 4, 5, 6, 7 |
| XX _____ X | l=9, k=6, p=3, 4, 5, 6, 7, 8 |
| .............. | ....................................... |
| X _ XX | l=4, k=1, p=2 |
| X __ XX | l=5, k=2, p=2, 3 |
| X ___ XX | l=6, k=3, p=2, 3, 4 |
| X ____ XX | l=7, k=4, p=2, 3, 4, 5 |
| X _____ XX | l=8, k=5, p=2, 3, 4, 5, 6 |
| .............. | ....................................... |
| XX _ XX | l=5, k=1, p=3 |
| XX __ XX | l=6, k=2, p=3, 4 |
| XX ___ XX | l=7, k=3, p=3, 4, 5 |
| XX ____ XX | l=8, k=4, p=3, 4, 5, 6 |
| XX _____ XX | l=9, k=5, p=3, 4, 5, 6, 7 |
| .............. | ....................................... |
| | Continued on next page |

**Table 1 Continued from previous page**

| Feature Structure | Values of $(l,\ k,\ p)$-mers |
|---|---|
| X _ XXX | l=5, k=1, p=2 |
| X __ XXX | l=6, k=2, p=2, 3 |
| X ___ XXX | l=7, k=3, p=2, 3, 4 |
| X ____ XXX | l=8, k=4, p=2, 3, 4, 5 |
| X _____ XXX | l=9, k=5, p=2, 3, 4, 5, 6 |
| ............... | ........................................ |
| XXX _ X | l=5, k=1, p=4 |
| XXX __ X | l=6, k=2, p=4, 5 |
| XXX ___ X | l=7, k=3, p=4, 5, 6 |
| XXX ____ X | l=8, k=4, p=4, 5, 6, 7 |
| XXX _____ X | l=9, k=5, p=4, 5, 6, 7, 8 |
| ............... | ........................................ |
| XXX _ XX | l=6, k=1, p=4 |
| XXX __ XX | l=7, k=2, p=4, 5 |
| XXX ___ XX | l=8, k=3, p=4, 5, 6 |
| XXX ____ XX | l=9, k=4, p=4, 5, 6, 7 |
| XXX _____ XX | l=10, k=5, p=4, 5, 6, 7, 8 |
| ............... | ........................................ |
| XX _ XXX | l=6, k=1, p=3 |
| XX __ XXX | l=7, k=2, p=3, 4 |
| XX ___ XXX | l=8, k=3, p=3, 4, 5 |
| XX ____ XXX | l=9, k=4, p=3, 4, 5, 6 |
| XX _____ XXX | l=10, k=5, p=3, 4, 5, 6, 7 |
| ............... | ........................................ |

Table 1 shows, feature structure, and values of $(l,\ k,\ p)$-mers.