Supplementary Appendix: Probabilities of SNP configurations with $K \geq 3$

If there are up to 2 sequencing errors, the possible SNP configurations and their probabilities are given in METHODS. If there are 3 sequencing errors on the site, the possible configurations and their probabilities are

$$
\begin{aligned}
Pr\left(i, j-3, 1, 1, 1\right) &= \binom{j}{3}\left(1-u\right)\left(1-2u\right)\left(1-3u\right)P_3 \\
Pr\left(i, j-3, 2, 1\right) &= \binom{j}{2}\binom{j-2}{1}u\left(1-u\right)\left(1-2u\right)P_3 \\
Pr\left(i, j-3, 3\right) &= \binom{j}{3}u^2\left(1-u\right)P_3 \\
Pr\left(i+1, j-3, 1, 1\right) &= \binom{j}{1}\binom{j-1}{2}u\left(1-u\right)\left(1-2u\right)P_3 \\
Pr\left(i+1, j-3, 2\right) &= \binom{j}{1}\binom{j-1}{2}u^2\left(1-u\right)P_3 \\
Pr\left(i+2, j-3, 1\right) &= \binom{j}{2}\binom{j-2}{1}u^2\left(1-u\right)P_3 \\
Pr\left(i+3, j-3\right) &= \binom{j}{3}u^3P_3
\end{aligned}
$$

$$
\begin{aligned}
Pr\left(i-1, j-2, 1, 1, 1\right) &= \binom{i}{1}\binom{j}{2}\left(1-u\right)\left(1-2u\right)\left(1-3u\right)P_3 \\
Pr\left(i-1, j-2, 2, 1\right) &= \binom{i}{1}\binom{j}{2}u\left(1-u\right)\left(1-2u\right)P_3 \\
&\quad+\binom{i}{1}\binom{j}{1}\binom{j-1}{1}u\left(1-u\right)\left(1-2u\right)P_3 \\
Pr\left(i-1, j-2, 3\right) &= \binom{i}{1}\binom{j}{2}u^2\left(1-u\right)P_3 \\
Pr\left(i, j-2, 1, 1\right) &= \binom{i}{1}\binom{j}{1}\binom{j-1}{1}u\left(1-u\right)\left(1-2u\right)P_3 \\
Pr\left(i, j-2, 2\right) &= \binom{i}{1}\binom{j}{1}\binom{j-1}{1}u^2\left(1-u\right)P_3 \\
Pr\left(i+1, j-2, 1\right) &= \binom{i}{1}\binom{j}{2}u^2\left(1-u\right)P_3 \\
Pr\left(i-1, j-1, 1, 1\right) &= \binom{i}{1}\binom{j}{2}u\left(1-u\right)\left(1-2u\right)P_3 \\
&\quad+\binom{i}{2}\binom{j}{1}u\left(1-u\right)\left(1-2u\right)P_3 \\
Pr\left(i-1, j-1, 2\right) &= \binom{i}{1}\binom{j}{2}u^2\left(1-u\right)P_3 \\
&\quad+\binom{i}{2}\binom{j}{1}u^2\left(1-u\right)P_3 \\
Pr\left(i, j-1, 1\right) &= \binom{i}{1}\binom{j}{1}\binom{j-1}{1}u^2\left(1-u\right)P_3 \\
Pr\left(i+1, j-1\right) &= \binom{i}{1}\binom{j}{2}u^3P_3
\end{aligned}
$$

where

$$P_3 = \varepsilon^3 (1-\varepsilon)^{i+j-3}.$$

By switching i and j in the above formulas, we can calculate $Pr(i-3,j,1,1,1)$, $Pr(i-3,j,2,1)$, $Pr(i-3,j,3)$, $Pr(i-3,j+1,1,1)$, $Pr(i-3,j+1,2)$, $Pr(i-3,j+2,1)$, $Pr(i-3,j+3)$, $Pr(i-2,j-1,1,1,1)$, $Pr(i-2,j-1,2,1)$, $Pr(i-2,j-1,3)$, $Pr(i-2,j,1,1)$, $Pr(i-2,j,2)$, $Pr(i-2,j+1,1)$, $Pr(i-1,j,1)$ and $Pr(i-1,j+1)$.

Combining all probabilities of identical configurations, we have all possible SNP configurations up to 3 sequencing errors ($K=3$). Their probabilities are

$$\begin{aligned}
Pr(i,j) &= P_0 + iju^2 P_2 \\
Pr(i-1,j+1) &= iuP_1 + \binom{j}{1}\binom{i}{2}u^3 P_3 \\
Pr(i+1,j-1) &= juP_1 + \binom{i}{1}\binom{j}{2}u^3 P_3 \\
Pr(i-1,j,1) &= i(1-u)P_1 + iju(1-u)P_2 + \binom{j}{1}\binom{i}{1}\binom{i-1}{1}u^2(1-u)P_3 \\
Pr(i,j-1,1) &= j(1-u)P_1 + iju(1-u)P_2 + \binom{i}{1}\binom{j}{1}\binom{j-1}{1}u^2(1-u)P_3 \\
Pr(i-1,j-1,2) &= iju(1-u)P_2 + \binom{i}{1}\binom{j}{2}u^2(1-u)P_3 + \binom{i}{2}\binom{j}{1}u^2(1-u)P_3 \\
Pr(i-1,j-1,1,1) &= i(1-u)(1-2u)P_2 + \binom{i}{1}\binom{j}{2}u(1-u)(1-2u)P_3 + \binom{i}{2}\binom{j}{1}u(1-u)(1-2u)P_3 \\
Pr(i-2,j+2) &= \binom{i}{2}u^2 P_2 \\
Pr(i+2,j-2) &= \binom{j}{2}u^2 P_2 \\
Pr(i-2,j+1,1) &= \binom{i}{1}\binom{i-1}{1}u(1-u)P_2 + \binom{j}{1}\binom{i}{2}u^2(1-u)P_3 \\
Pr(i+1,j-2,1) &= \binom{j}{1}\binom{j-1}{1}u(1-u)P_2 + \binom{i}{1}\binom{j}{2}u^2(1-u)P_3 \\
Pr(i-2,j,2) &= \binom{i}{2}u(1-u)P_2 + \binom{j}{1}\binom{i}{1}\binom{i-1}{1}u^2(1-u)P_3 \\
Pr(i,j-2,2) &= \binom{j}{2}u(1-u)P_2 + \binom{i}{1}\binom{j}{1}\binom{j-1}{1}u^2(1-u)P_3 \\
Pr(i-2,j,1,1) &= \binom{i}{2}(1-u)(1-2u)P_2 + \binom{j}{1}\binom{i}{1}\binom{i-1}{1}u(1-u)(1-2u)P_3 \\
Pr(i,j-2,1,1) &= \binom{j}{2}(1-u)(1-2u)P_2 + \binom{i}{1}\binom{j}{1}\binom{j-1}{1}u(1-u)(1-2u)P_3
\end{aligned}$$

$$\begin{aligned}
Pr(i,j-3,1,1,1) &= \binom{j}{3}(1-u)(1-2u)(1-3u)P_3 \\
Pr(i-3,j,1,1,1) &= \binom{i}{3}(1-u)(1-2u)(1-3u)P_3 \\
Pr(i,j-3,2,1) &= \binom{j}{2}\binom{j-2}{1}u(1-u)(1-2u)P_3 \\
Pr(i-3,j,2,1) &= \binom{i}{2}\binom{i-2}{1}u(1-u)(1-2u)P_3 \\
Pr(i,j-3,3) &= \binom{j}{3}u^2(1-u)P_3
\end{aligned}$$

$$
\begin{aligned}
Pr\left(i-3,j,3\right) &= \binom{i}{3}u^{2}\left(1-u\right)P_{3}\\
Pr\left(i+1,j-3,1,1\right) &= \binom{j}{1}\binom{j-1}{2}u\left(1-u\right)\left(1-2u\right)P_{3}\\
Pr\left(i-3,j+1,1,1\right) &= \binom{i}{1}\binom{i-1}{2}u\left(1-u\right)\left(1-2u\right)P_{3}\\
Pr\left(i+1,j-3,2\right) &= \binom{j}{1}\binom{j-1}{2}u^{2}\left(1-u\right)P_{3}\\
Pr\left(i-3,j+1,2\right) &= \binom{i}{1}\binom{i-1}{2}u^{2}\left(1-u\right)P_{3}\\
Pr\left(i+2,j-3,1\right) &= \binom{j}{2}\binom{j-2}{1}u^{2}\left(1-u\right)P_{3}\\
Pr\left(i-3,j+2,1\right) &= \binom{i}{2}\binom{i-2}{1}u^{2}\left(1-u\right)P_{3}\\
Pr\left(i+3,j-3\right) &= \binom{j}{3}u^{3}P_{3}\\
Pr\left(i-3,j+3\right) &= \binom{i}{3}u^{3}P_{3}\\
Pr\left(i-1,j-2,1,1,1\right) &= \binom{i}{1}\binom{j}{2}\left(1-u\right)\left(1-2u\right)\left(1-3u\right)P_{3}\\
Pr\left(i-2,j-1,1,1,1\right) &= \binom{j}{1}\binom{i}{2}\left(1-u\right)\left(1-2u\right)\left(1-3u\right)P_{3}\\
Pr\left(i-1,j-2,2,1\right) &= \binom{i}{1}\binom{j}{2}u\left(1-u\right)\left(1-2u\right)P_{3}+\binom{i}{1}\binom{j}{1}\binom{j-1}{1}u\left(1-u\right)\left(1-2u\right)P_{3}\\
Pr\left(i-2,j-1,2,1\right) &= \binom{j}{1}\binom{i}{2}u\left(1-u\right)\left(1-2u\right)P_{3}+\binom{j}{1}\binom{i}{1}\binom{i-1}{1}u\left(1-u\right)\left(1-2u\right)P_{3}\\
Pr\left(i-1,j-2,3\right) &= \binom{i}{1}\binom{j}{2}u^{2}\left(1-u\right)P_{3}\\
Pr\left(i-2,j-1,3\right) &= \binom{j}{1}\binom{i}{2}u^{2}\left(1-u\right)P_{3}
\end{aligned}
$$

where

$$
\begin{aligned}
P_{0} &= \left(1-\varepsilon\right)^{i+j}\\
P_{1} &= \varepsilon\left(1-\varepsilon\right)^{i+j-1}\\
P_{2} &= \varepsilon^{2}\left(1-\varepsilon\right)^{i+j-2}.
\end{aligned}
$$

For $K\geq 4$, it is easier to code a program to calculate all possible resulting SNP configurations. An example follows, for a SNP with i copies of the "A" allele and j copies of the "C" allele, one possible event is that

1. There are l_A errors occurring on the "A" allele and l_C errors occurring on the "C" allele;
2. Among the l_A "A" alleles where errors occurred, $l_{A>C}$ of them changed to "C" , $l_{A>G}$ of them changed to "G" and $l_{A>T}=l-l_{A>C}-l_{A>G}$ of them changed to "T";
3. Among the l_C "C" alleles where errors occurred, $l_{C>A}$ of them changed to "A" , $l_{C>G}$ of them changed to "G" and $l_{C>T}=l_C-l_{C>A}-l_{C>G}$ of them changed to "T".

The result are $i-l_A+l_{C>A}$ "A" alleles, $j-l_C+l_{A>C}$ "C" alleles, $l_{A>G}+l_{C>G}$ "G" alleles and $l_{A>T}+l_{C>T}$ "T" alleles with configuration $(i-l_A+l_{C>A},j-l_C+l_{A>C},l_{A>G}+l_{C>G},l_{A>T}+l_{C>T})$. The probability of such event is

$$\frac{i!j!\varepsilon^{l_A+l_C}\left(1-\varepsilon\right)^{i+j-l_A-l_C}}{(i-l_A)!l_{A>C}!l_{A>G}!l_{A>T}!\,(j-l_C)!l_{C>A}!l_{C>G}!l_{C>T}!3^{l_A+l_C}}.$$

Of course this event is only one possible event that can produce such a configuration. To calculate the total probability of any given configuration, one needs to sum the probabilities of all possible events that produce that configuration.