Mathematical model for melting curves and heteroduplex content

In Figure 1, we show high-resolution melting curves of the DNA amplicon from three genotypes of a SNP for which the homozygous mutation is indistinguishable from that of the wild type, due to nearest-neighbor thermodynamic symmetry of the mutation. (This says that the bases immediately surrounding the mutation are identical when the strands are interchanged, e.g., TCA/AGT $\leftrightarrow$ TGA/ACT.) Our goal is to add the right proportion of wild-type DNA to each genotype before PCR, so that after amplification, melting, and reannealing, the mixture with the homozygous mutant sample will develop heteroduplex content but the mixture with the wild-type sample will not, making the curves distinguishable. The mixture with the heterozygous sample will have its heteroduplex content reduced from its natural 50 % value, so we must be careful that its curve remains distinct from the heteroduplex-enhanced homozygous sample.

Therefore, in this section, we develop a model for the melting curve of a mixture of genotypes in terms of the melting curves of the consituent duplex types and the mixture proportions. In the case above that the homozygous mutant and wild-type curves are indistinguishable, we show that the difference among curves of different genotypes depends solely on the heteroduplex content of the mixture. In addition, the temperature of maximum separation does not change with heteroduplex content. This is in spite of the maybe surprising fact that the two heteroduplex species have different thermodynamic melting behavior, even with respect to nearest-neighbor approximation.

Finally, we model and optimize the separation of heteroduplex contents of the three genotypes in terms of mixture proportion.

We begin by asking what kind of duplexes are present for melting in what proportions, and what cumulative melting curve will result when we mix wild-type and homozygous mutant DNA in proportion x of wild-type and $1 - x$ of homozygous mutant and strands, melt them, then rapidly cool them so that complementary and nearly complementary strands form duplexes independently, i.e., solely in proportion to their concentrations.

This answer will model the results of all spiking experiments. When $x = 1$, it describes a pure wild-type sample, whether unspiked, or spiked in any proportion. When $x = 0$, it describes a pure homozygous mutant sample. For arbitrary $0 < x < 1$, it describes spiking a homozygous mutant sample with wild-type DNA in the given proportions. And finally, if we spike a heterozygous sample with wild-type DNA in a proportion x of wild-type to $1 - x$ of heterozygous sample, then since half of the heterozygous samples strands are of wild-type, and half are of homozygous mutant type, in terms of strand concentrations, it is equivalent to spiking a homozygous mutant sample with wild-type DNA in proportion $x + \frac{1-x}{2}$ of wild-type and $\frac{1-x}{2}$.

In the situation described, there are concentrations x of strands W and $\bar{W}$, the 'forward' and complementary 'reverse' strand of the wild-type DNA. There are also concentrations $1 - x$ of strands M and $\bar{M}$, the 'forward' and complementary 'reverse' strand of the mutant homozygous DNA.

After independent reannealing then, there will be the following concentrations of four duplex species arranged in 'standard binomial order', a.k.a., 'FOIL', $(xW + (1 - x)M)(xW + (1 - x)M)$.

$$D_1 = x^2 W\bar{W}$$

$$D_2 = x(1-x)W\bar{M}$$

$$D_3 = x(1-x)M\bar{W}$$

$$D_4 = (1-x)^2 M\bar{M}$$

Corresponding to each duplex species D_j will be its standardized fluorescence curve for a fixed concentration, $F_j(T)$. The function $F_j(T)$ refers to the curve with background fluorescence removed, and automatically accounts for whatever fluorescence per duplex variation there may be among the different species types, i.e., if duplex flourescences are proportional by unequal factors to concentration, just replace F_j by $c_j F_j$, where c_j describes the relationship between duplex concentration and its contribution to fluorescence.)

As noted above, if the unknown sample is wild-type, then regardless of the spike proportion x, the entire sample is wild-type, and the resulting melting curve is described by

$$W(T, x) = F_1(T).$$

If the unknown sample is a homozyogous SNP, then we are in the situation described above, and assuming fluorescence is additive, the resulting melting curve is described by

$$M(T, x) = x^2 F_1(T) + x(1-x)F_2(T) + x(1-x)F_3(T)(1-x)^2 F_4(T).$$

If the unknown sample is heterozyogous, then by the equivalence noted above, wild-type homoduplexes will be present in proportion $\frac{(1+x)^2}{4}$, homozygous SNP homoduplexes will be present in proportion $\frac{(1-x)^2}{4}$, and each type of heteroduplex will be present in proportion $\frac{(1+x)(1-x)}{4} = \frac{1-x^2}{4}$.

The resulting negative derivative of the melting curve is described by

$$H(T, x) = \frac{(1+x)^2}{4} F_1(T) + \frac{1-x^2}{4} F_2(T) + \frac{1-x^2}{4} F_3(T) + \frac{(1-x)^2}{4} F_4(T).$$

To simplify the subsequent analysis we write

$$W(T,x) = 1F_1(T)$$

$$M(T,x) = m_1(x)F_1(T) + m_{23}(x)(F_2(T) + F_3(T)) + m_4(x)F_4(T)$$

$$H(T,x) = h_1(x)F_1(T) + h_{23}(x)(F_2(T) + F_3(T)) + h_4(x)F_4(T)$$

where $m_1(x) = x^2$, $m_{23}(x) = x(1-x)$, $m_4(x) = (1-x)^2$, and $h_1(x) = \frac{(1+x)^2}{4}$, $h_{23}(x) = \frac{1-x^2}{4}$, $h_4(x) = \frac{(1-x)^2}{4}$.

Note that

$$m_1(x) + 2m_{23}(x) + m_4(x) = 1 = h_1(x) + 2h_{23}(x) + h_4(x).$$

Our goal is to maximize our ability to distinguish these three curves, as measured by the minimum separation between any two of them. The separation will be defined by the maximum absolute value of their difference. So our goal is to find

$$\max_{x\in[0,1]} \min\{\max_T |W(T,x) - M(T,x)|, \max_T |W(T,x) - H(T,x)|, \max_T |M(T,x) - H(T,x)|\}.$$

For this reason we compute

$$W(T,x) - M(T,x) = (1 - m_1(x))F_1(T) - m_{23}(x)(F_2(T) + F_3(T))) - m_4(x)F_4(T).$$

$$W(T,x) - H(T,x) = (1 - h_1(x))F_1(T) - h_{23}(x)(F_2(T) + F_3(T))) - h_4(x)F_4(T).$$

$$H(T,x) - M(T,x) = (h_1(x) - m_1(x))F_1(T) + (h_{23}(x) - m_{23}(x)(F_2(T) + F_3(T))) + (h_4(x) - m_4(x))F_4(T).$$

In the situation where the nearest-neighbor model of the homozygous SNP has the same thermodynamic parameters as the wild-type and the corresponding melting curves and their (negative) derivatives are identical, (experimental curves indistinguishable) i.e., $F_1(T) = F_4(T)$, requiring us to use the spiking protocol, we may simplify these differences considerably. (Even when $F_1(T) = F_4(T)$ according to nearest neighbor theory, the thermodynamic parameters that determine F_2 and $F_3(x)$ are not identical. The differences are reported in Table x, but do not affect the subsequent analysis.) We will use the subscript $_=$ to indicate this case.

Then assuming $F_1 = F_4$ we can combine their of the F_1 and F_4 in the expressions for $W(T,x) - M(T,x)$ and $W(T,x) - H(T,x)$ to obtain

$$W_{=}(T,x) - M_{=}(T,x) = (1 - m_1(x) - m_4(x))F_1(T) - m_{23}(x)(F_2(T) + F_3(T)))$$

and

$$W_{=}(T,x) - H_{=}(T,x) = (1 - h_1(x) - h_4(x))F_1(T) - h_{23}(x)(F_2(T) + F_3(T))).$$

Using

$$1 - (m_1(x) + m_4(x)) = 2m_{23}(x)$$

$$1 - (h_1(x) + h_4(x)) = 2h_{23}(x)$$

then distributing the result back equally between F_1 and F_4 for symmetry ($F_1 = \frac{1}{2}(F_1 + F_4)$), we obtain

$$W_{=}(T,x) - M_{=}(T,x) = m_{23}(x)(F_1(T) + F_2(T) + F_3(T) + F_4(T))$$

$$W_{=}(T,x) - H_{=}(T,x) = h_{23}(x)(F_1(T) + F_2(T) + F_3(T) + F_4(T))$$

and writing the third difference as the difference of these differences,

$$H_{=}(T,x) - M_{=}(T,x) = (m_{23}(x) - h_{23}(x))(F_1(T) + F_2(T) + F_3(T) + F_4(T))$$

The graphs of these three functions are given in Figure 2, annotated with key features derived below.

These expressions have two important consequences. First, they show that the pointwise separation of the curves is proportional solely to the difference in heteroduplex fraction of the mixtures. Second, they uncouple the x (spike proportion) and T (temperature) dependence of the differences among fluorescence curves of different genotypes. This allows us to optimize the spike proportion independently, regardless of the specific nature of individual duplex curves contributing to the superpositions. This is perhaps surprising, since in general, different convex combinations (superpositions with positive coefficients summing to 1) of different functions result in varying shapes and locations of their extrema.

Also note that if one prefers to consider melting 'peaks', i.e., the negative derivative curves, $P_j(T) = -F_j'(T)$, since all expressions are linear, we can obtain the difference of the negative derivative curves of the different genotypes simply by replacing P_j for F_j in

the expressions computed above, and all of the results we will obtain regarding maxima and minima will also hold in this framework.

Symbolically, our ultimate genotype separation problem reduces to

$$\begin{aligned}
&\max_{x\in[0,1]} \min\{\max_T |W_=(T,x) - M_=(T,x)|, \max_T |W_=(T,x) - H_=(T,x)|, \max_T |M_=(T,x) - H_=(T,x)|\} \\
&= \max_{x\in[0,1]} \min\{m_{23}(x), h_{23}(x), |m_{23}(x) - h_{23}(x)|\} \max_T |F_1(T) + F_2(T) + F_3(T) + F_4(T)| \\
&= G \max_{x\in[0,1]} \min\{m(x), h(x), |m(x) - h(x)|\},
\end{aligned}$$

where

$$m(x) = 2m_{23}(x) = 2x(1-x)$$

and

$$h(x) = 2h_{23}(x) = \frac{1-x^2}{2},$$

make $m(x)$ and $h(x)$ the (non-negative) total heteroduplex proportion in the spiked homozygous SNP and heterozygous samples, respectively, and

$$G = \frac{1}{2}|F_1(T) + F_2(T) + F_3(T) + F_4(T)|,$$

For example, when $x = 0$, $h(x) = \frac{1}{2}$, and $m(x) = 0$, $\frac{1}{2}G$ is the separation of the unspiked heterozygous curve from the common unspiked wild-type and homozygous SNP curves. What remains is to solve the spike dependent optimization problem

$$\max_{x\in[0,1]} \min\{m(x), h(x), |m(x) - h(x)|\}.$$

(The rest is the same.)