Addendum 1:

This addendum contains two proofs. The first demonstrates that the Fold Enrichment (FE) is directly proportional to the probability of Ssb binding. The second demonstrates that the number of SeRP reads is a function of the elongation rate, and that by using Fold Enrichment, the contribution of the elongation rate is eliminated and the probability of Ssb binding isolated.

Proof 1: Demonstration that the FE is directly proportional to the probability of Ssb binding

The probability of finding Ssb bound to a nascent chain when codon i of transcripts from gene j are being translated at a given instant in time is defined as

$$\boldsymbol{P_{bound}(i,j) = \frac{N_{bound}(i,j)}{N_{Total}(i,j)}}, \quad \textbf{(1)}$$

where $N_{bound}(i,j)$ is the number of ribosomes that have Ssb bound when codon i is being translated, and $N_{Total}(i,j)$ is the total number of ribosomes actively translating codon i. The FE at codon i of gene j is defined as $FE(i,j) = \frac{S(i,j)}{R(i,j)}$, where $S(i,j)$ is the number of reads arising from selective ribosome profiling that map to codon i of gene j, and $R(i,j)$ is the number of reads arising from ribosome profiling that map to the same location. Experiments have shown that the number of RNA-Seq reads is directly proportional to the number of mRNA molecules (Fig. 2C in Pubmed ID: 18516045). We assume this holds for RP and SeRP reads as well, since they are also Next-Gen Sequencing methods. Therefore, $R(i,j)$ is directly proportional to the total number or ribosomes at i and j; that is $R(i,j) \propto N_{Total}(i,j)$, and $R(i,j) = aN_{Total}(i,j)$, where a is a constant of proportionality. This equation can be algebraically rearranged to find

$$\boldsymbol{N_{Total}(i,j) = \frac{R(i,j)}{a}} \quad \textbf{(2)}$$

Likewise, $S(i,j)$ is most likely directly proportional to the number of ribosomes that have Ssb bound at length i on transcript j, and hence $S(i,j) = b\, N_{Bound}(i,j)$ and

$$\boldsymbol{N_{Bound}(i,j) = \frac{S(i,j)}{b}} \quad \textbf{(3)}$$

where b is a constant of proportionality.

Substituting Eqs. 2 and 3 into Eq. 1 results in $P_{bound}(i,j) = \frac{a\,S(i,j)}{b\,R(i,j)}$. Substituting our definition for $F.E.(i,j)$ into this equation yields $P_{bound}(i,j) = \frac{a}{b} F.E.(i,j)$. Therefore,

$$\boldsymbol{P_{bound}(i,j) \propto F.E.(i,j)} \quad \textbf{(4)}$$

This demonstrates that the fold enrichment that is experimentally measured is directly proportional to the probability of an Ssb molecule being bound to the nascent chain.

Proof 2: Demonstration that SeRP reads are a function of the elongation rate, and that the Fold Enrichment metric controls for this effect.

The total number of ribosomes on transcript j, $\boldsymbol{N_{Total}(j) = \sum_{k=1}^{N_C} N_{Total}(k,j)}$, where N_C is the number of codons in the transcript, is equal to

$$\boldsymbol{N_{Total}(j) = k_{int,j} N_{mRNA,j} \tau_{S,j}} \quad \textbf{(5)}$$

where $k_{int,j}$, $N_{mRNA,j}$, $\tau_{S,j}$ are, respectively, the initiation rate, mRNA copy number and average synthesis time of transcript j. The synthesis time is the sum of the codon translation times,

therefore $\tau_{S,j} = \sum_{k=1}^{N_C} \tau_A(k,j)$. $\tau_A(k,j)$ is the average translation time of codon k in transcript j. Substituting this expression into Eq. 5 it can be seen that

$$\boldsymbol{N_{Total}(i,j) = k_{int,j} N_{mRNA,j}\, \tau_A(i,j)} \quad \textbf{(6)}$$

Inserting Eq. 3 into Eq. 1 we get $S(i,j) = bN_{Total}P_{bound}(i,j)$. Subsituting Eq. 6 into this equation we get a key result

$$\boldsymbol{S(i,j) = bk_{int,j} N_{mRNA,j}\, \tau_A(i,j) P_{bound}(i,j)} \quad \textbf{(7)}$$

Thus, we have demonstrated that the number of SeRP reads ($\boldsymbol{S(i,j)}$) is not just a function of the probability of Ssb binding, but also a function of the codon translation time among other factors. Thus, the probability of binding cannot be defined by the SeRP reads alone. We need additional information to define the bound regions.

Likewise, it can be shown using Eq. 2 that $\boldsymbol{R(i,j) = ak_{int,j} N_{mRNA,j}\, \tau_A(i,j)}$. Substituting this expression and Eq. 7 into our definition of FE, we get

$$\boldsymbol{F.E.(i,j)} = \frac{\boldsymbol{S(i,j)}}{\boldsymbol{R(i,j)}}$$

$$= \frac{\boldsymbol{bk_{int,j} N_{mRNA,j}\, \tau_A(i,j) P_{bound}(i,j)}}{\boldsymbol{ak_{int,j} N_{mRNA,j}\, \tau_A(i,j)}}$$

$$= \frac{\boldsymbol{b}}{\boldsymbol{a}} \boldsymbol{P_{bound}(i,j)} \quad \textbf{(8)}$$

These additional factors (including elongation speed) all cancel out and we have thereby isolated the effect of Ssb binding.

$$\boldsymbol{P_{bound}(i,j) \propto F.E.(i,j)} \quad \textbf{(9)}$$

In Eqs. 8 and 9 we have demonstrated that by dividing the SeRP reads by the RP reads we eliminate the effects of elongation times, initiation rates, and mRNA copy number and measure the effect of the probability of Ssb binding.