**Supplementary Information**

1. **Data transform for one- class SVM.**

For a certain annotation, the “right-tail” case means outliers should have positive deviations, the “left-tail” case means outliers should have negative deviations, and the “both-tail” case means that outliers could have either positive or negative deviations. Reference deviations were then calculated for different cases. For the left-tail and right-tail cases,

$$σ=\sqrt{\frac{\sum\_{i=1}^{n}(x\_{i}-x\_{extreme})^{2}}{n}}$$

σ is the reference deviation, *xi* is annotation value of the *i*th site from the random regions. *n* is the number of sites from the random regions (*n* = 4000). *xextreme* is either the minimum of *xi* for the right-tail case, or the maximum of *xi* for left-tail case. For any observation *x* of the same annotation for any SV, the transform *y* is

$$y=2/(1+cosh(|x-x\_{extreme}|/σ))$$

For both-tail case, we define two reference deviations σright and σleft for either positive or negative deviations from the median *xmed* of *xi,*

$$σ\_{right}=\sqrt{\frac{\sum\_{x\_{i}>x\_{med}}^{}(x\_{i}-x\_{med})^{2}}{\sum\_{x\_{i}>x\_{med}}^{}1}},σ\_{left}=\sqrt{\frac{\sum\_{x\_{i}<x\_{med}}^{}(x\_{i}-x\_{med})^{2}}{\sum\_{x\_{i}<x\_{med}}^{}1}}$$

The transform *y* is

$y=2/(1+cosh(|x-x\_{med}|/σ\_{right}))$, if *x* > *xmed*

$y=2/(1+cosh(|x-x\_{med}|/σ\_{left}))$, if *x* < *xmed*

*y* = 0, if *x* = *xmed*

Therefore outliers indicating potential SVs approach 0 in this transform, which is required by the application of one-class SVM. In the transformed metric space, linear classifiers were trained by the one-class SVM (implemented with package e1071 in the Comprehensive R Archive Network) with SVs from the random regions as the training set. The proportion of SVs in the training set identified as outliers (false positive rate) 1-*p* was approximately controlled by a factor ν in the training algorithm defined by the authors. In short, ν ∈ (0,1) defines the ratio of penalty induced by margin size (e.g. distance from origin point to the class boundary with linear kernel) and penalty induced by number of outliers in the training set in the total penalty function for soft margin case. Higher ν allows more training data points to be on the outliers side to maximize the margin. Classifiers at different ν’s were then applied to predict other SV data sets.