



Figure 6- figure supplement 1. Density plot for the number of distribution of predictions per genomic region. The genomic data were compiled into 3611 10 kb windows. For machine learning training and testing (related to Figure 6), only 20% of the data could be used for prediction. To generate predictions genome wide, we randomly and independently split the data into training and testing (80:20) and generated predictions. Therefore, each region could have received more than one prediction. The above distribution profile shows that a majority of the regions received three predictions, with a large proportion of the data having received between 2 and 4 predictions. Only 124 regions received no prediction by chance. For each split, we ensured that the population distribution of ~20:1 (core:LS) was maintained in the training and testing data.