A Temporal Distributions

The following section details supporting information about each assay's distribution of the feature space and the label space over time. In Fig. 9 the distributions of the observed labels from every assay, are shown for each of the five folds that constitute the three temporal settings illustrated in Fig. 1. Similarly, in Fig. 10 the feature space is shown as t-SNE projections and colored by the five temporal folds. Generally, ADME-T assays exhibit smaller shifts over time compared to target-based assays. This is true in terms of both the feature space and the label space.

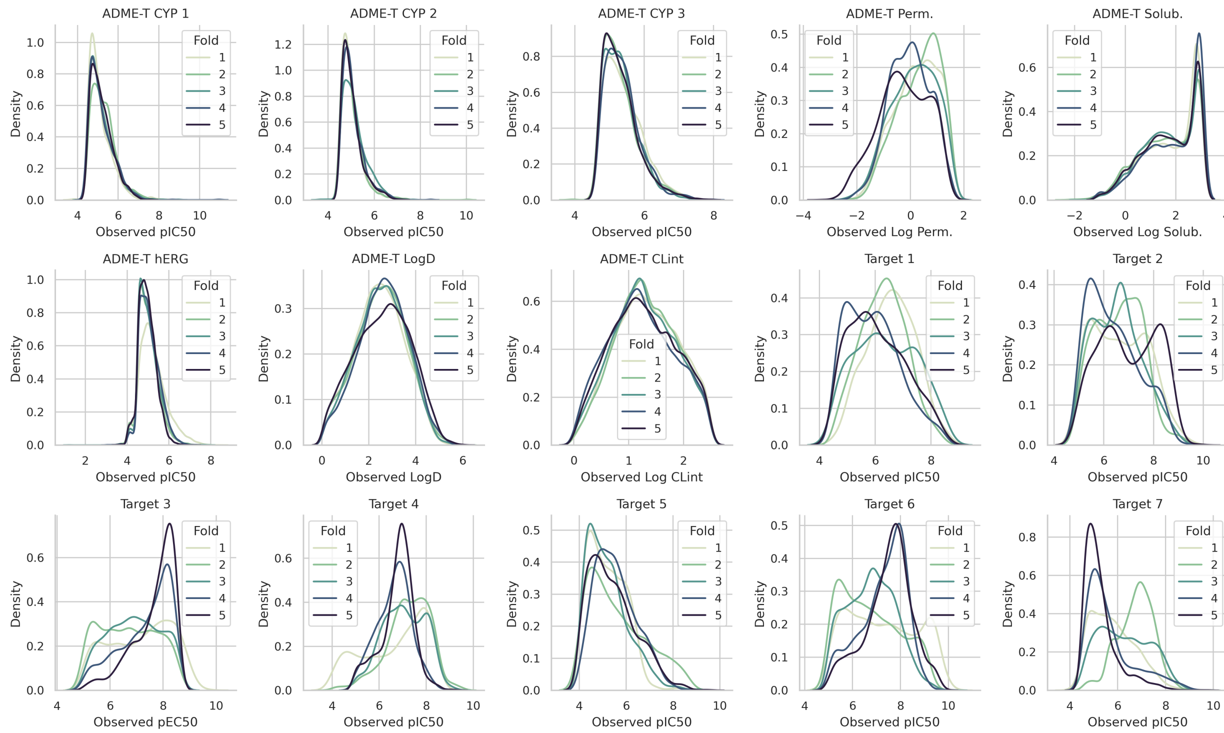

Figure 9: **Temporal Distribution of the Label-space.** Distribution of observed experimental labels across the five temporal folds illustrated per assay.

B Model Selection

For the model selection of all models used in this work, we optimized the hyperparameters described in Table 2 using a grid search according to the loss function of each base model. Each temporal setting was optimized individually. As such, the Random Forest model and the base neural network architecture used in the Ensemble and the MC-Dropout model were all optimized using the validation MSE for observed datasets or the validation CensoredMSE described in Section 2 for censored datasets. For simplicity, we also used the resulting neural network architecture for the Bayes by Backprop model and the evidential deep learning framework. The base neural network architecture for the Gaussian models was optimized using the validation Gaussian NLL.

During the grid search for the neural network architectures, we used a maximum of 100 epochs per experiment. However, a maximum of 500 epochs was used for the final experiments, inspired by the neural architecture search proposed by Jiang et al. [61]. Additionally, the neural networks were trained using the Adam optimizer with a weight decay of 0.0005, the learning rate was reduced when plateauing with a patience of 50 epochs. The hyperparameter named *decreasing dimension* for the neural networks determines if the hidden dimension is decreased by a factor of two for every layer. Thus, for hyperparameters: 4 number of layers, hidden dimension of size 512, and decreasing dimension True, the resulting hidden layers have sizes [512, 256, 128, 64].