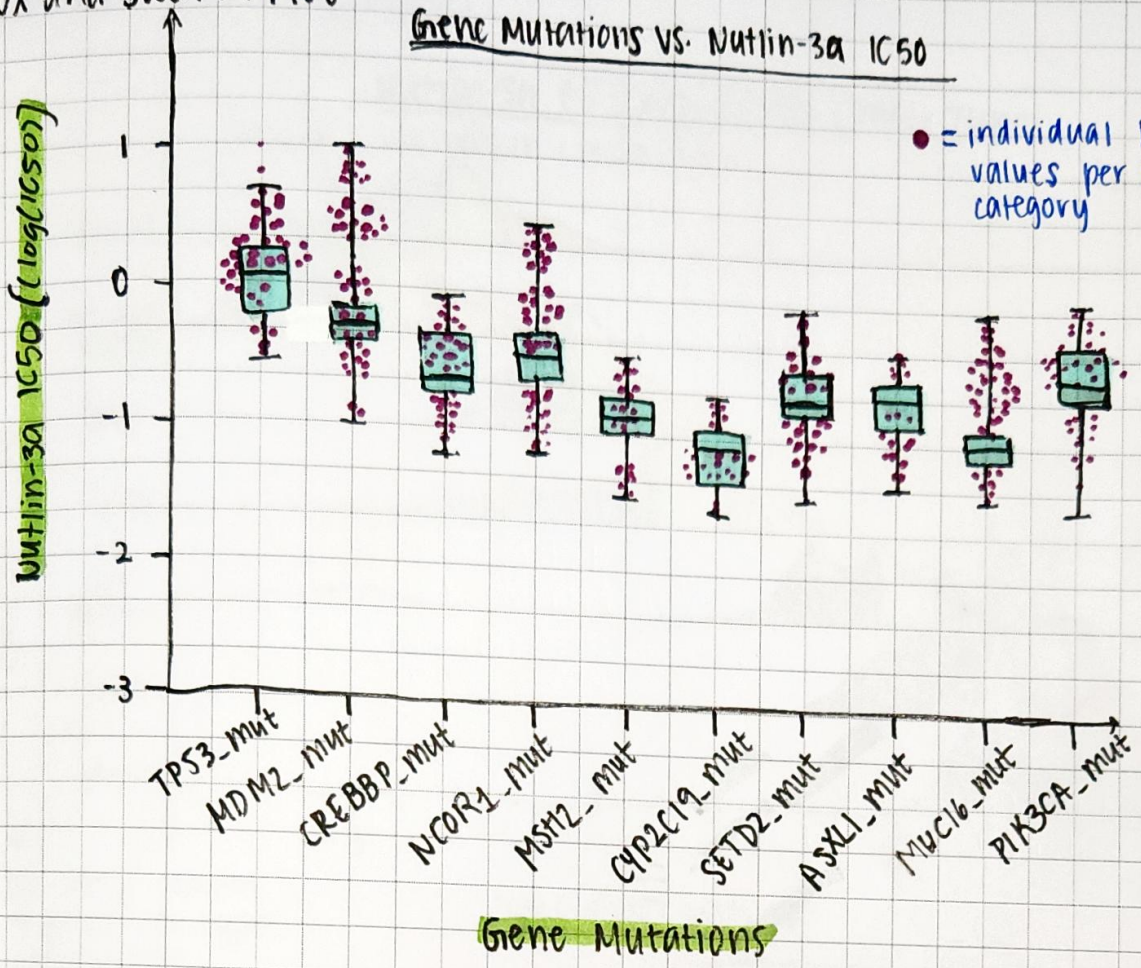
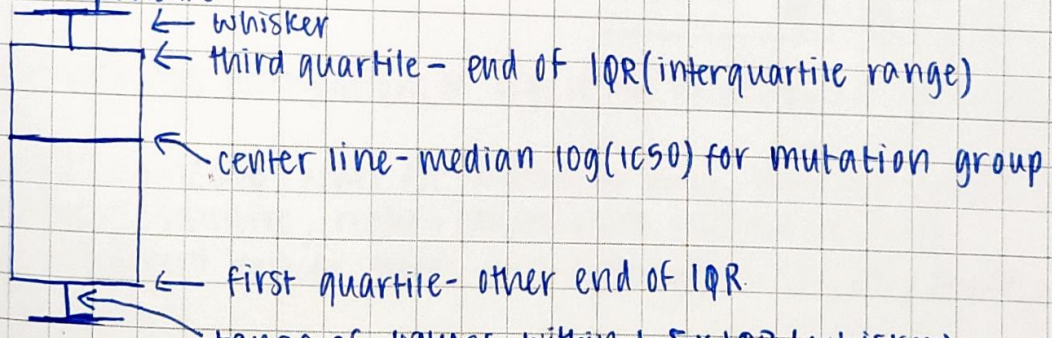
① Box and swarm plot



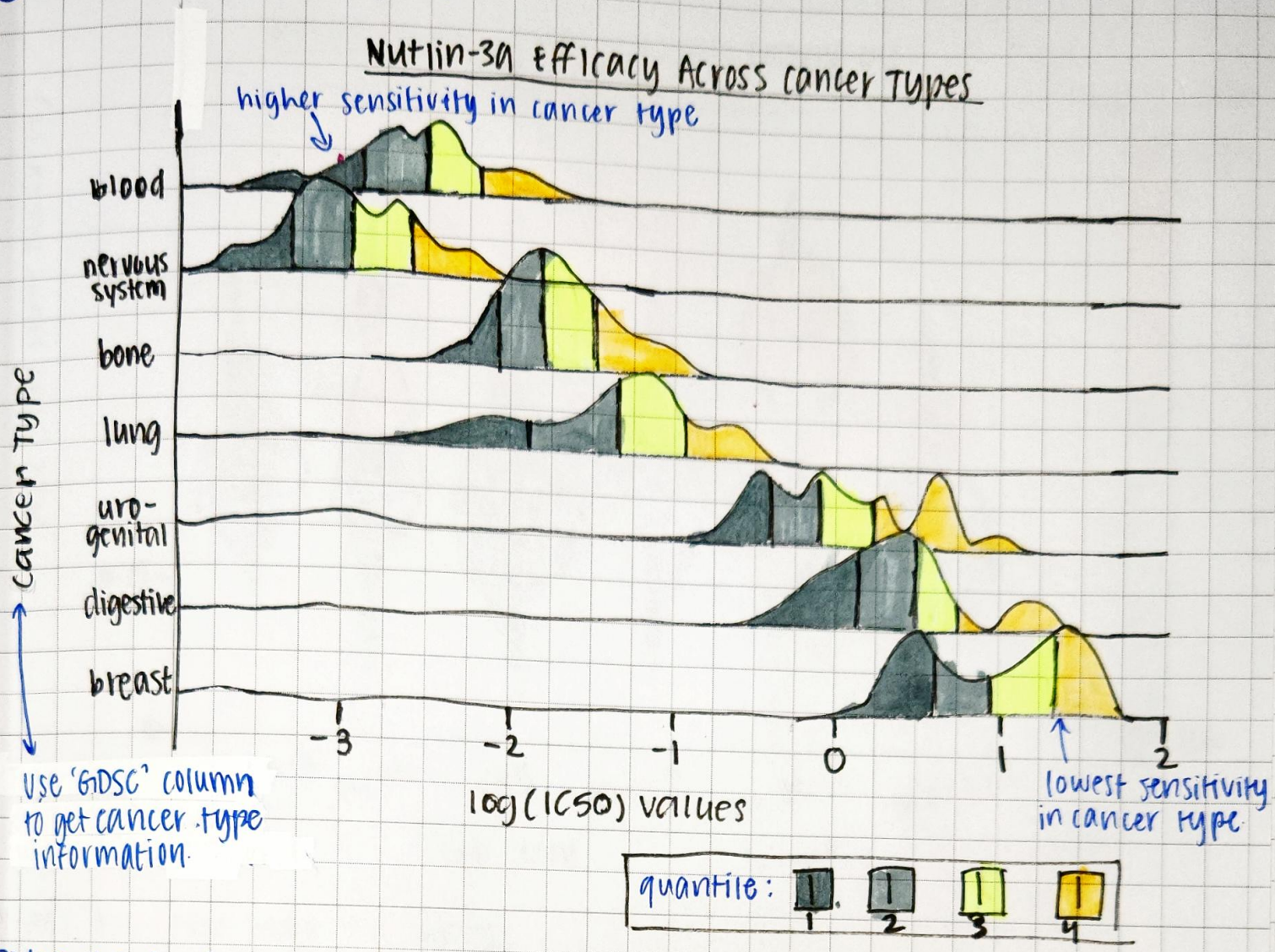
Box Plot components:



Swarm plots overlay individual data points; gives insight into variability & outliers. helps identify which gene mutations are associated with increased drug sensitivity / resistance as IC50 value directly correlates to this.
↳ low IC50 = less of the drug required to have an effect = higher sensitivity

data points can't overlap in swarm plot.

② Stacked Ridge Plot (Joyplot)



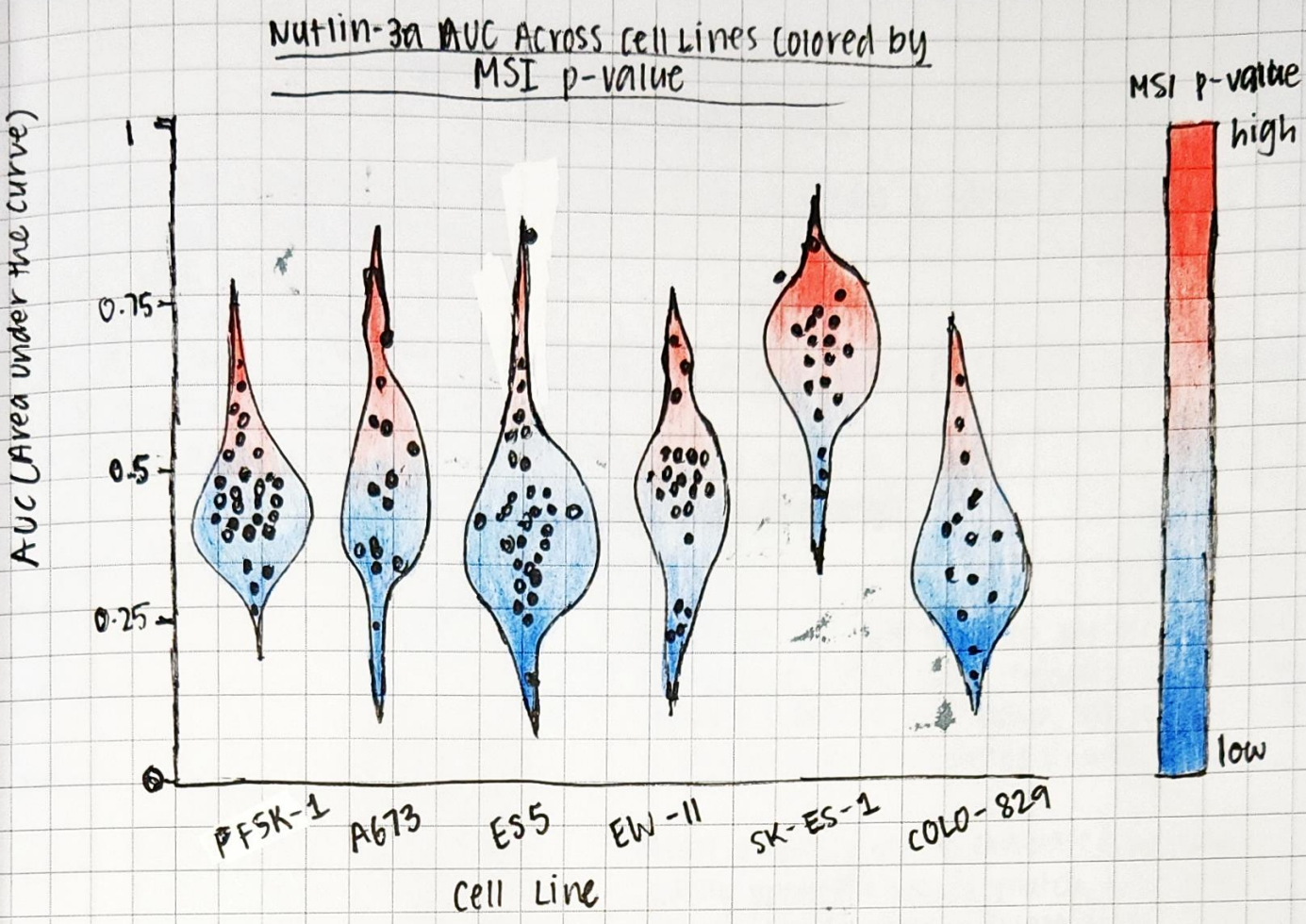
Use 'GDSC' column to get cancer type information.

Ridge plot displays distributions of log(IC50) values across multiple cancer types, in this case:

- x-axis: log(IC50) values for Nutlin-3a drug, showing drug's efficacy (left = cancer more sensitive; right = cancer more resistant to drug)
- y-axis: different cancer types. each type has a smooth density curve.

can compare shape, spread, and central tendency of Nutlin-3a efficacy and see how it varies across different cancer types.

③ Violin Plot with strips on top:



y-axis: AUC (Area under the curve)
- measures overall drug efficacy of the Nutlin-3a drug per cell line. Lower AUC = greater sensitivity (more effective). Higher AUC = greater resistance of cancer to drug (less effective). Ranges from 0 to 1.

x-axis: Cell Lines - each violin represents a unique cancer cell line. they are categorical. Violin width reflects density of AUC values w/ multiple measures per cell line

Color Fill = MSI_pval (MSI p-value) represents the statistical significance of microsatellite instability (genetic instability caused by DNA mismatch repair)
- blue = low MSI p-value → MSI-high → more genomic instability (p val < 0.05)
- red = high MSI p-value → MSI-stable → less genomic instability (p val > 0.05)

visually tracks relationship between MSI & AUC.

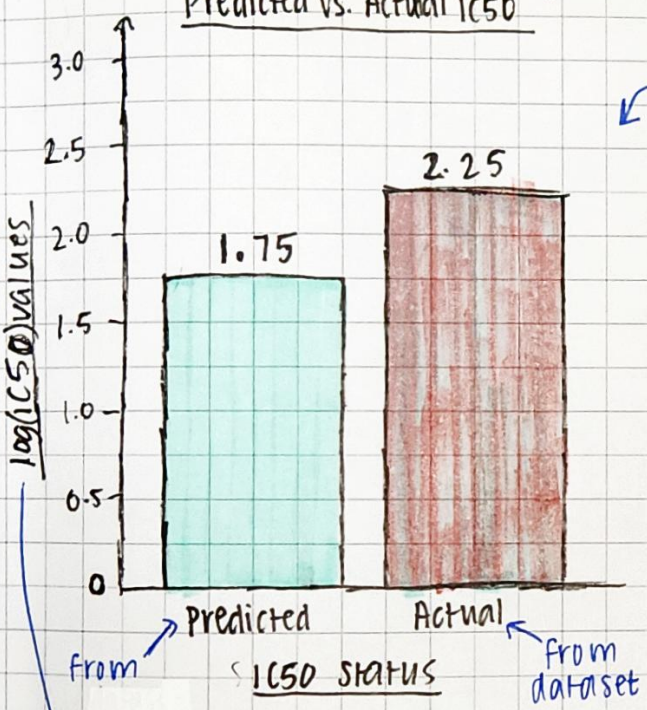
Individual points for better specificity (with the black color). data points overlap allowed.

④ Machine Learning Model

Predicting Nutlin-3a IC50 from Biological Features

Input values (features that will be used to train model & identify patterns):
- MSI_pval (MSI p-value)
- Cell Line Name
- GDSC (tissue classification)
- ic50_effect_size (measures difference in drug response (IC50) between mutated & nonmutated samples)
- AUC or area under the curve: overall drug response.

↓ predicts IC50 / efficacy



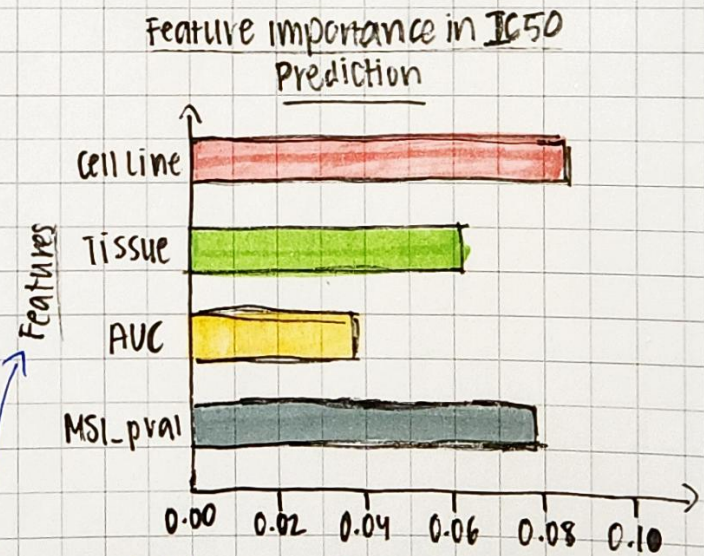
From ~~[illegible]~~

From dataset

plotting log values (for better interpretability) but the original IC50 is measured in μM.

Nutlin-3a efficacy based on various features is predicted. Other drug features can also be inputted (that are not Nutlin-3a) so that IC50 is predicted for other drugs as well

Model accuracy can be evaluated by how closely predicted values match true IC50 values. Larger gaps indicate error.



Feature importance scores show how strongly each feature influences the predicted IC50 values. Higher scores = greater predictive power

different for each kind of model

⑤ Interactive hypothetical prediction simulator