



Prediction of Experimental Metadata from Gene Expression

Scientific Advisor: Alexey Sergushichev (ITMO)

Team I:

Andrey Vasilchenko (MSU)

Alexander Illarionov (UToronto)

Anton Karazeev (MIPT)

Maxim Kozin (RNRMU)

Alena Moskalenko (MIPT)

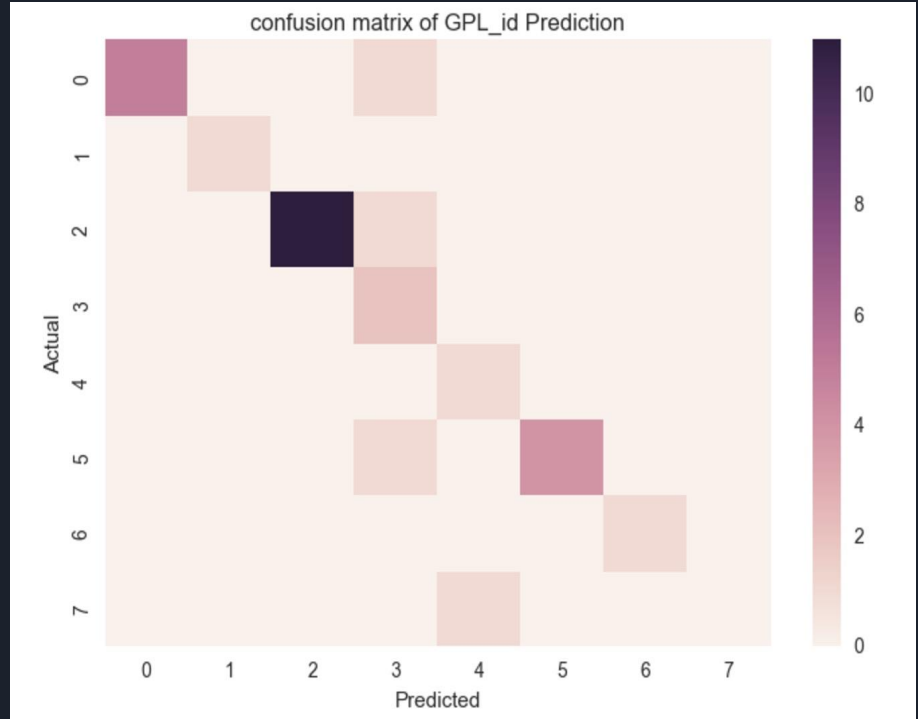
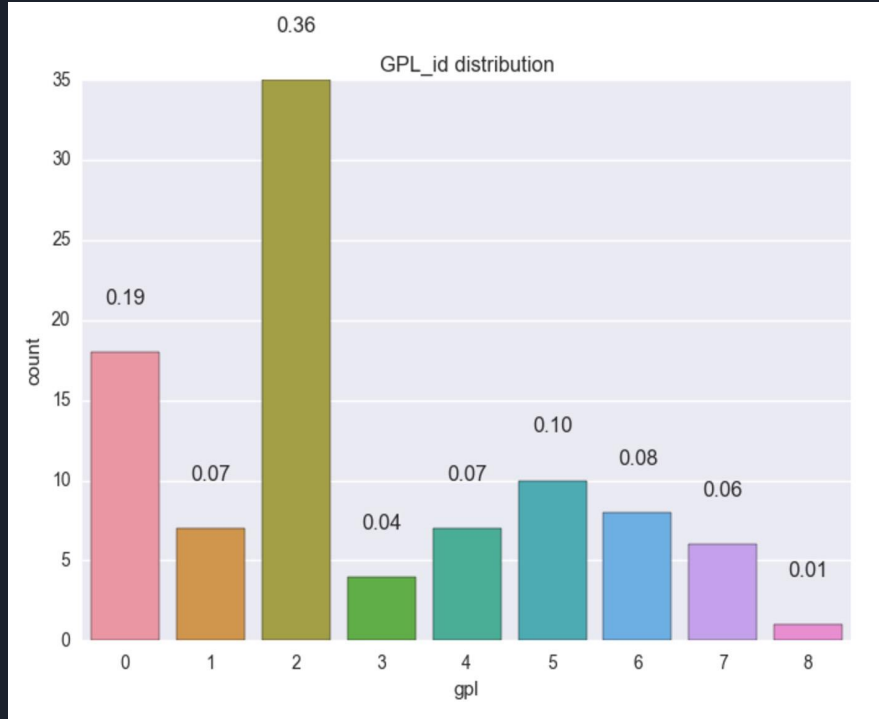
Angelina Mosyagina (KrasSMU)



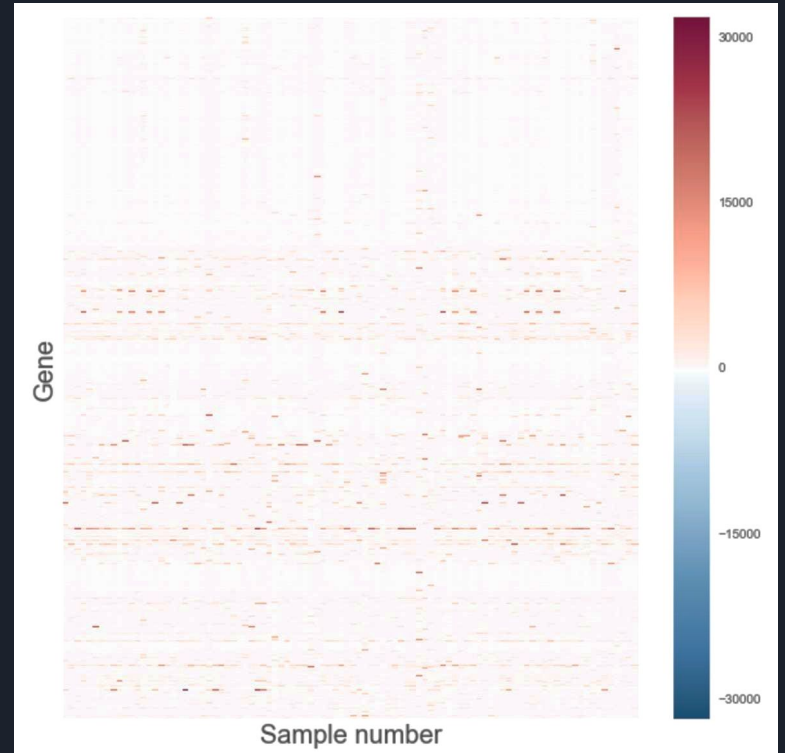
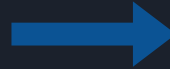
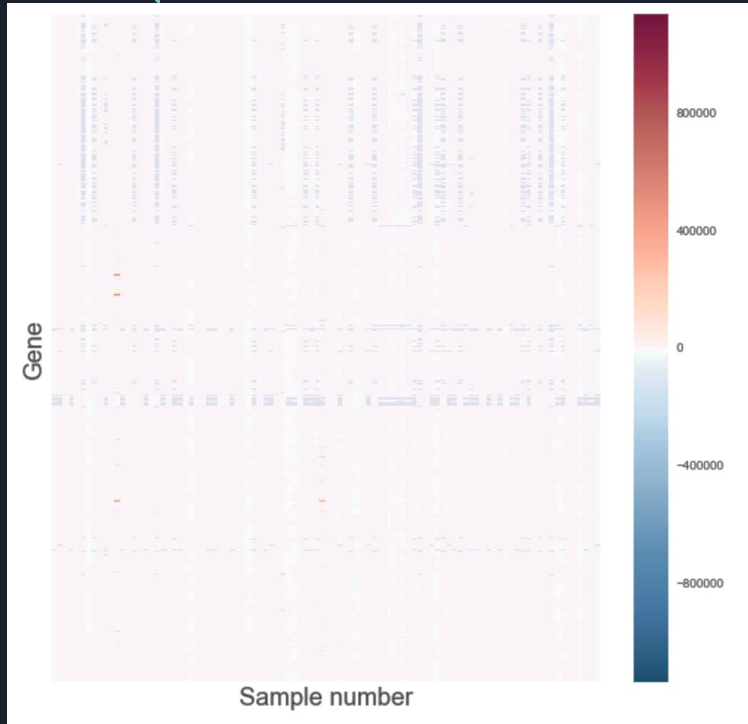
Agenda

- **Collection and cleansing of REMs data for 3 strains of mice**
- **Data normalization:**
 - Gaussian, Quantile and β -actin comparative
- **Design of ML models for sex, tissue and strain identification**
 - Keras
 - XGBoost
 - mlxtend (Random Forests & Extra Trees Ensembling)

Prediction of the Platform



Quantile Normalization





**Accuracy of GPL_id Prediction without
Quantile Normalization: 82.8%**

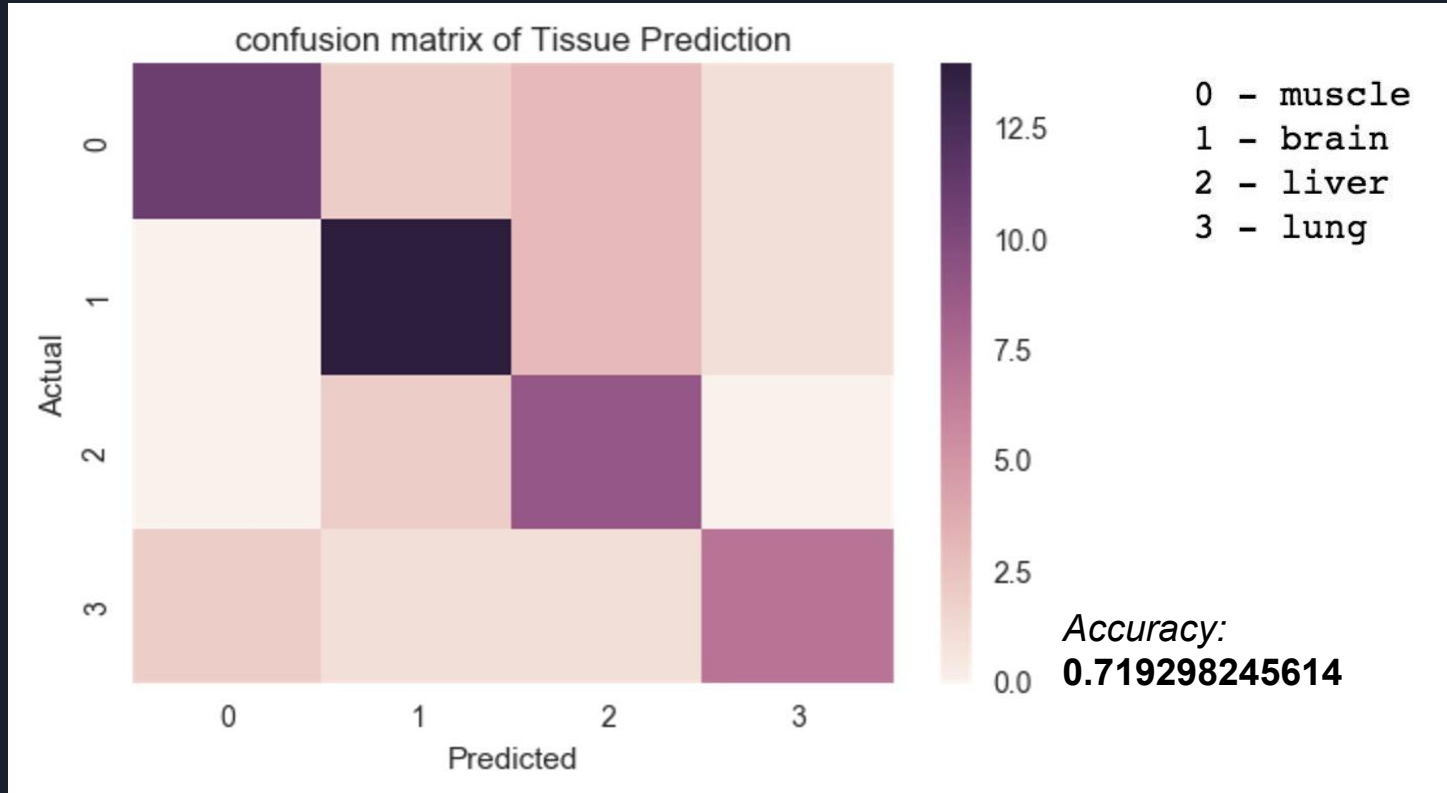


**Accuracy of GPL_id Prediction with
Quantile Normalization: 89.7%**

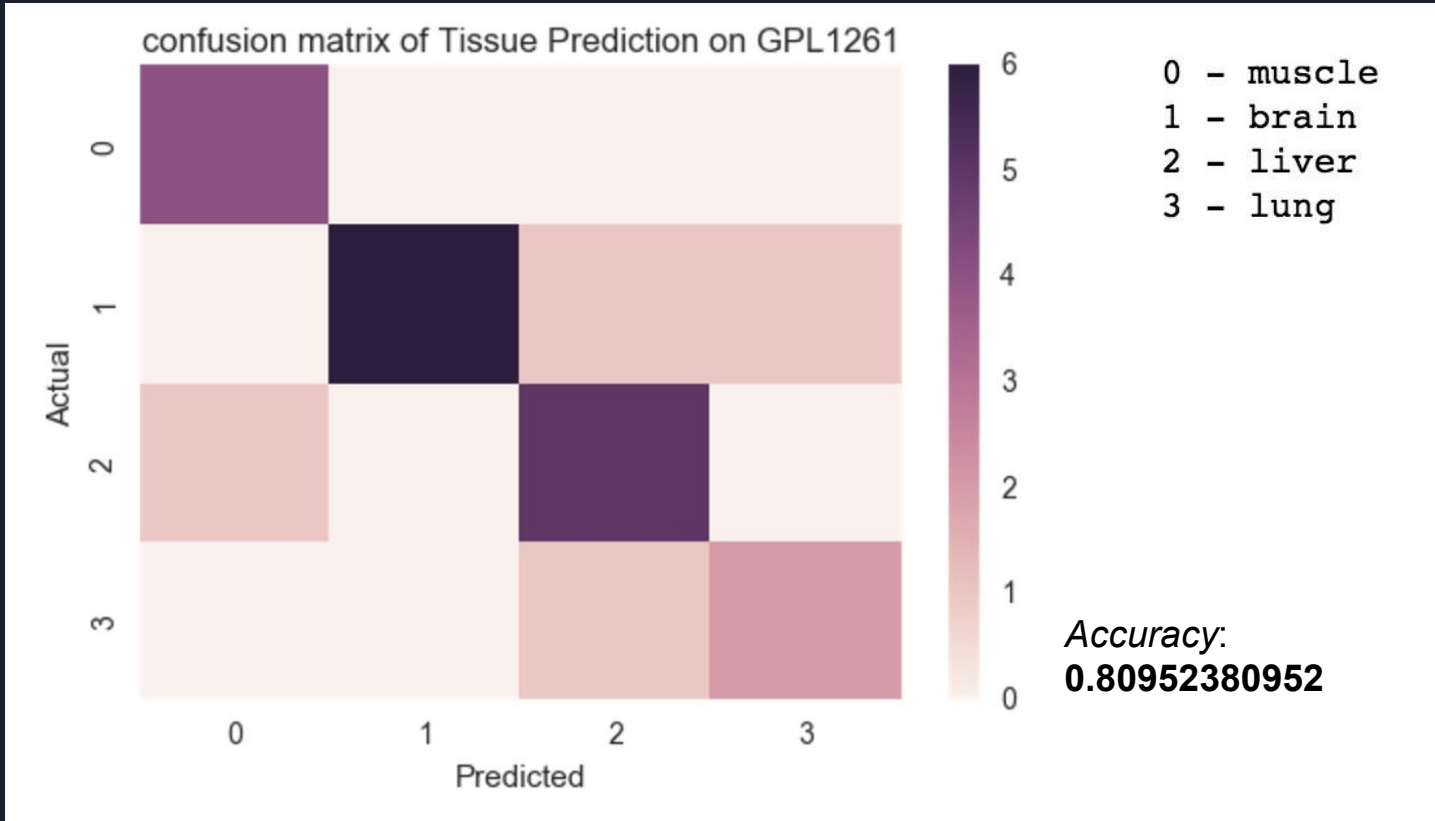
Tissue Prediction



Tissue Prediction



Tissue Prediction on GPL1261



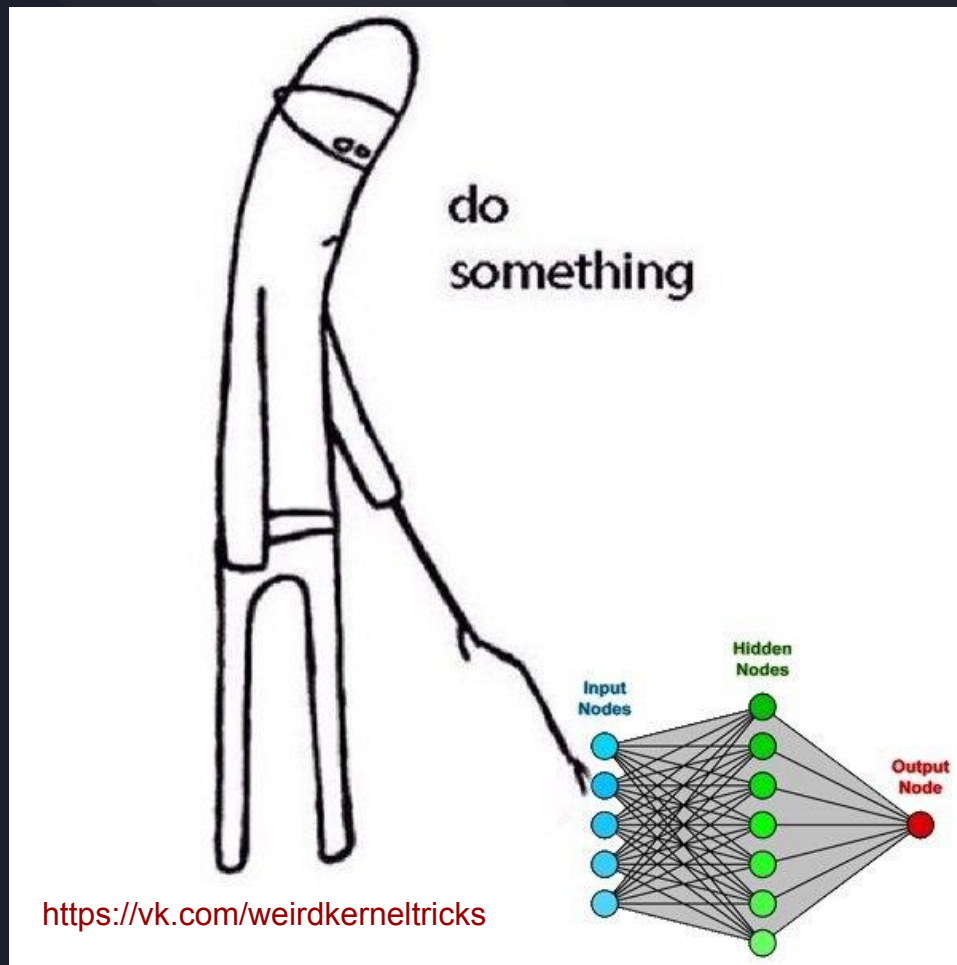


RAN KERAS



ON ALL THE DATA

QQs?



Code -> <https://github.com/akarazeev/MetadataPredictionFromGE>