---------- autoQTL output ----------

Total number of samples in the dataset: 1523
Total number of features in the dataset: 28
The entire dataset is split into 80-20. The 80% is used in the evolution and the 20% is used as a hold out to assess model performances.
The 80% data is further splitted equally into training and testing datasets.
Training data: 1st split of 80% data.
Testing data: 2nd split of 80% data.
Holdout data: 20% spilt of entire data.
autoQTL parameters:
population size = 100
offspring size = None
generations = 5
mutation rate = 0.9
crossover rate = 0.1
random state = 42

****

Evolution History:

	Best Test R^2 score	Best difference score
Gen 1	0.19339	2.86593
Gen 2	0.20062	8.36525
Gen 3	0.20287	8.36525
Gen 4	0.20287	8.36525
Gen 5	0.20287	8.36525
Gen 1	0.18281	2.49122
Gen 1	0.19339	2.86593
Gen 2	0.20062	8.36525
Gen 1	0.19339	2.86593
Gen 2	0.20062	8.36525
Gen 3	0.20287	8.36525
Gen 4	0.20287	8.36525
Gen 5	0.20287	8.36525

****

Multiple Linear Regression:

Train R^2:	0.24117
Test R^2:	0.15773
Holdout R^2:	0.21584
Entire Dataset R^2:	0.22317

****

Final Pareto Front Statistics:

18 pipelines on the final Pareto front

	Number	Percent
LinearRegression	6	33%
MachineLearning	12	67%
RandomForest	12	67%

```
   DecisionTree       0      0%
--------------------------------
 AdditiveEncoder      13     72%
  3-LevelEncoder       2     11%
  2-LevelEncoder       3     17%
```

Range of Test R^2: (-0.00020, 0.20287)
Range of Difference Score: (1.20518, 8.36525)

****

Final Pareto Front:

Pipeline #1:
Test R^2: 0.2028696692420967 | Difference Score: 1.205184923684204

1. RandomForestRegressor(max_features=0.5, min_samples_leaf=3,
   min_samples_split=4, random_state=42)

No feature selectors in the pipeline
Train R^2: 0.6768771973985885

Pipeline #2:
Test R^2: 0.20062139118458966 | Difference Score: 1.282997370997124

1. RandomForestRegressor(max_features=0.6000000000000001,
   min_samples_leaf=2, min_samples_split=14,
   random_state=42)

No feature selectors in the pipeline
Train R^2: 0.5696813544379606

Pipeline #3:
Test R^2: 0.19710661506190275 | Difference Score: 1.3881514150420797

1. RandomForestRegressor(max_features=0.5, min_samples_leaf=3,
   min_samples_split=20,
   random_state=42)

No feature selectors in the pipeline
Train R^2: 0.46641671401120544

Pipeline #4:
Test R^2: 0.1948336962769427 | Difference Score: 1.3980914839004042

1. RandomForestRegressor(max_features=0.3, min_samples_leaf=5,
   min_samples_split=14,
   random_state=42)

No feature selectors in the pipeline
Train R^2: 0.4565661873682112

Pipeline #5:
Test R^2: 0.1878988496956575 | Difference Score: 1.5145684880053256

1. RandomForestRegressor(max_features=0.2, min_samples_leaf=7,
                         min_samples_split=14,
                         random_state=42)

No feature selectors in the pipeline
Train R^2: 0.37793854842502017

Pipeline #6:
Test R^2: 0.18412895106238658 | Difference Score: 1.5388684929168948

1. RandomForestRegressor(bootstrap=False, max_features=0.2,
                         min_samples_leaf=12,
                         min_samples_split=14,
                         random_state=42)

No feature selectors in the pipeline
Train R^2: 0.36244645172327716

Pipeline #7:
Test R^2: 0.1807212576977093 | Difference Score: 1.5429934207498468

1. RandomForestRegressor(max_features=0.5, min_samples_leaf=11,
                         min_samples_split=19,
                         random_state=42)

No feature selectors in the pipeline
Train R^2: 0.3571395861966534

Pipeline #8:
Test R^2: 0.18052560374288362 | Difference Score: 1.6133917874316024

1. RandomForestRegressor(max_features=0.2, min_samples_leaf=9,
                         min_samples_split=16,
                         random_state=42)

No feature selectors in the pipeline
Train R^2: 0.32811006430057

Pipeline #9:
Test R^2: 0.16996860540052283 | Difference Score: 1.703458452860967

1. RandomForestRegressor(bootstrap=False, max_features=0.2,
                         min_samples_leaf=18,
                         min_samples_split=16,

random_state=42)

No feature selectors in the pipeline
Train R²: 0.28872959763520845

Pipeline #10:
Test R²: 0.1630167231246592 | Difference Score: 1.724905876315829

1. RandomForestRegressor(bootstrap=False,
   max_features=0.15000000000000002,
   min_samples_leaf=18, min_samples_split=5,
   random_state=42)

No feature selectors in the pipeline
Train R²: 0.27598028899796534

Pipeline #11:
Test R²: 0.16093319189862598 | Difference Score: 1.7434128950982368

1. RandomForestRegressor(max_features=0.25, min_samples_leaf=15,
   min_samples_split=19,
   random_state=42)

No feature selectors in the pipeline
Train R²: 0.2691759835260087

Pipeline #12:
Test R²: 0.1590162055173121 | Difference Score: 1.7687998550953483

1. RandomForestRegressor(max_features=0.25, min_samples_leaf=16,
   min_samples_split=7, random_state=42)

No feature selectors in the pipeline
Train R²: 0.2611772263770299

Pipeline #13:
Test R²: 0.15783279332220157 | Difference Score: 1.8923379181861386

1. SelectPercentile(percentile=90),
2. LinearRegression()

Number of features selected by selectpercentile after 1 level of feature selection in the pipeline: 25
Feature names of the selected features by selectpercentile after 1 level of feature selection in the pipeline: ['mCV24380249_T', 'rs13476111_G', 'rs6181202_C', 'gnf05.013.094_T', 'rs4226028_A', 'rs13479428_A', 'rs13479907_A', 'rs3662744_A', 'rs3685576_A', 'rs6163251_G', 'rs13482283_G', 'rs8266245_C', 'rs13482691_G', 'rs13475982_G', 'rs3718090_G', 'rs13478123_G', 'rs13478864_A', 'gnf08.122.415_G', 'rs3675233_A', 'rs13479859_A', 'rs13480409_G', 'rs3653651_G', 'rs3670360_G', 'rs13482284_G', 'UT_14_62.125092_A']
Train R²: 0.23581674403958175

Pipeline #14:
Test R^2: 0.1477919361739567 | Difference Score: 2.301834178847108

1. RecessiveEncoder(),
2. VarianceThreshold(threshold=0.15),
3. LinearRegression()

Number of features selected by variancethreshold after 1 level of feature selection in the pipeline: 23
Feature names of the selected features by variancethreshold after 1 level of feature selection in the pipeline: ['rs13476111_G', 'rs6181202_C', 'gnf05.013.094_T', 'rs4226028_A', 'rs13479428_A', 'rs13479907_A', 'rs3662744_A', 'rs3685576_A', 'rs6163251_G', 'rs13482283_G', 'rs8266245_C', 'rs13482644_A', 'rs13475982_G', 'rs3718090_G', 'rs13478864_A', 'rs3714598_G', 'rs3675233_A', 'rs13479859_A', 'rs13480409_G', 'rs3653651_G', 'rs13482200_C', 'rs13482284_G', 'UT_14_62.125092_A']
Train R^2: 0.18341275203668794

Pipeline #15:
Test R^2: 0.12786483053947206 | Difference Score: 2.331023180951703

1. UnderDominanceEncoder(),
2. LinearRegression()

No feature selectors in the pipeline
Train R^2: 0.16173470839034398

Pipeline #16:
Test R^2: 0.1248379739592771 | Difference Score: 2.955732655182495

1. VarianceThreshold(threshold=0.2),
2. UnderDominanceEncoder(),
3. LinearRegression()

Number of features selected by variancethreshold after 1 level of feature selection in the pipeline: 26
Feature names of the selected features by variancethreshold after 1 level of feature selection in the pipeline: ['mCV24380249_T', 'rs13476111_G', 'rs6181202_C', 'gnf05.013.094_T', 'rs4226028_A', 'rs13479428_A', 'rs13479907_A', 'rs3662744_A', 'rs3685576_A', 'rs6163251_G', 'rs13482283_G', 'rs8266245_C', 'rs13482644_A', 'rs13482691_G', 'rs13475982_G', 'rs3718090_G', 'rs13478864_A', 'rs3714598_G', 'gnf08.122.415_G', 'rs3675233_A', 'rs13479859_A', 'rs13480409_G', 'rs3653651_G', 'rs13482200_C', 'rs13482284_G', 'UT_14_62.125092_A']
Train R^2: 0.13794002844198505

Pipeline #17:
Test R^2: 0.11295847931917402 | Difference Score: 3.6647028190838418

1. HeterosisEncoder(),
2. VarianceThreshold(threshold=0.15),

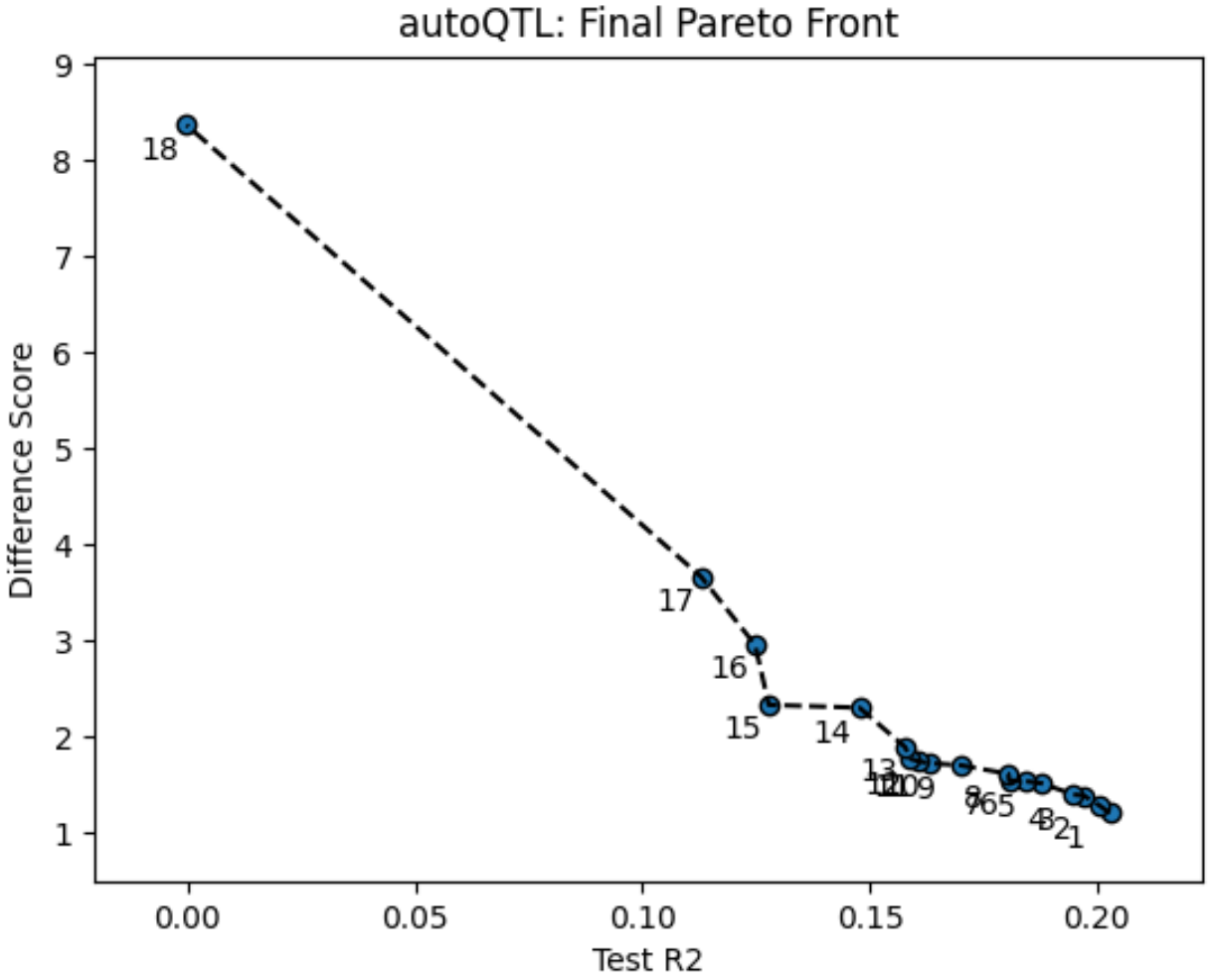
3. LinearRegression()

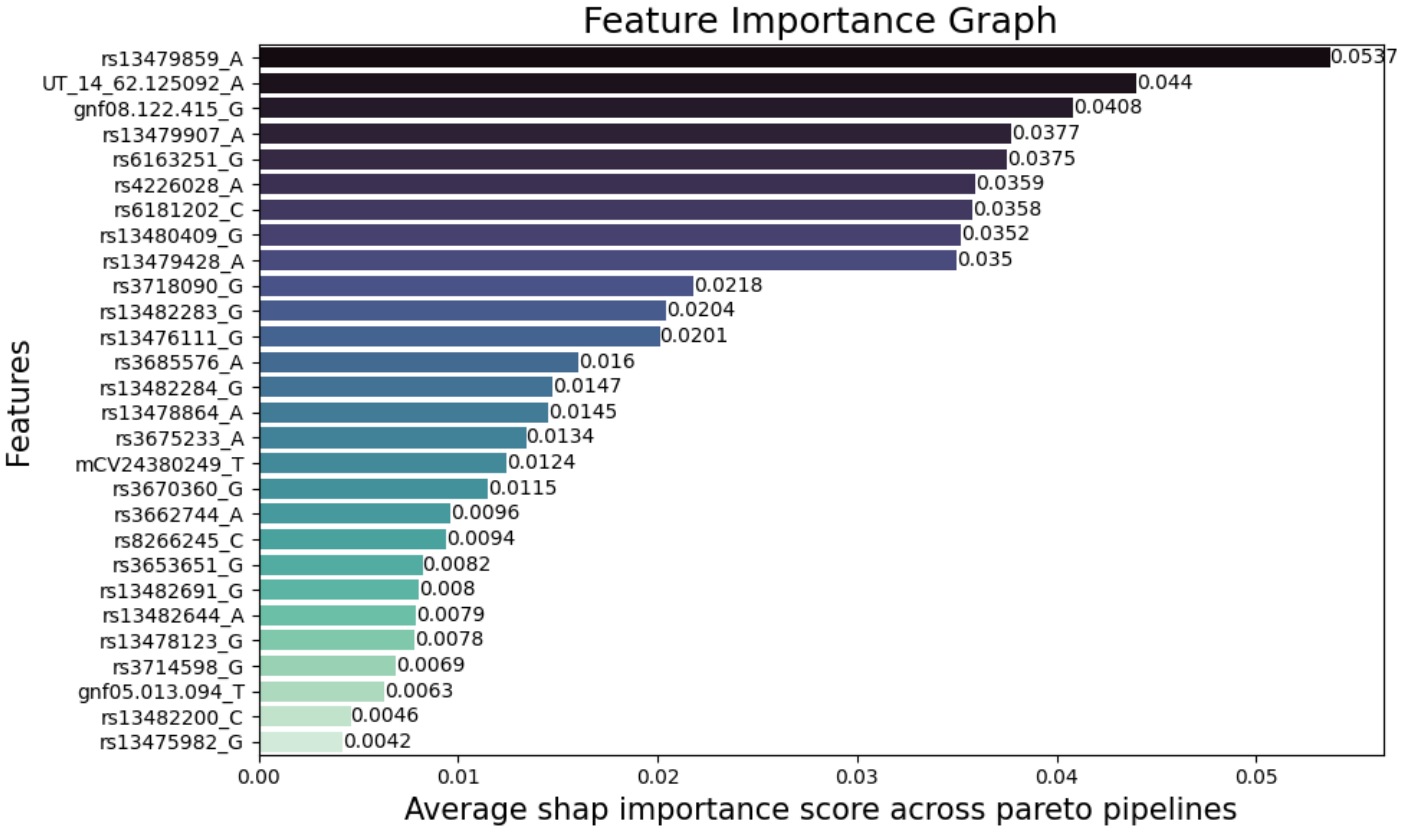
Number of features selected by variancethreshold after 1 level of feature selection in the pipeline: 26
Feature names of the selected features by variancethreshold after 1 level of feature selection in the pipeline: ['mCV24380249_T', 'rs13476111_G', 'rs6181202_C', 'gnf05.013.094_T', 'rs4226028_A', 'rs13479428_A', 'rs13479907_A', 'rs3662744_A', 'rs3685576_A', 'rs6163251_G', 'rs13482283_G', 'rs8266245_C', 'rs13482644_A', 'rs13482691_G', 'rs13475982_G', 'rs3718090_G', 'rs13478864_A', 'rs3714598_G', 'gnf08.122.415_G', 'rs3675233_A', 'rs13479859_A', 'rs13480409_G', 'rs3653651_G', 'rs13482200_C', 'rs13482284_G', 'UT_14_62.125092_A']
Train R^2: 0.11850275671285349

Pipeline #18:
Test R^2: -0.00020421346263055362 | Difference Score: 8.365249673882376

1. RecessiveEncoder(),
2. HeterosisEncoder(),
3. DominantEncoder(),
4. LinearRegression()

No feature selectors in the pipeline
Train R^2: 0.0


autoQTL: Final Pareto Front
Difference Score
Test R2
0.00
0.05
0.10
0.15
0.20
1
2
3
4
5
6
7
8
9
18
17
16
15
14


Feature Importance Graph
rs13479859_A
UT_14_62.125092_A
gnf08.122.415_G
rs13479907_A
rs6163251_G
rs4226028_A
rs6181202_C
rs13480409_G
rs13479428_A
rs3718090_G
rs13482283_G
rs13476111_G
rs3685576_A
rs13482284_G
rs13478864_A
rs3675233_A
mCV24380249_T
rs3670360_G
rs3662744_A
rs8266245_C
rs3653651_G
rs13482691_G
rs13482644_A
rs13478123_G
rs3714598_G
gnf05.013.094_T
rs13482200_C
rs13475982_G
0.0537
0.044
0.0408
0.0377
0.0375
0.0359
0.0358
0.0352
0.035
0.0218
0.0204
0.0201
0.016
0.0147
0.0145
0.0134
0.0124
0.0115
0.0096
0.0094
0.0082
0.008
0.0079
0.0078
0.0069
0.0063
0.0046
0.0042
Features
0.00
0.01
0.02
0.03
0.04
0.05
Average shap importance score across pareto pipelines