CASP16 EMA

QMODE3

Alisia Fadini, Randy Read, Gabriel Studer

Choose top 5 models from the MassiveFold dataset

Large sampling of
AlphaFold2

```
PFRMAT QA
TARGET T0999
AUTHOR 1234-5678-9000
METHOD Description of methods used
MODEL 2
QMODE 3
ranked_0_unrelaxed_model_5_ptm_pred_82.pdb
ranked_13_unrelaxed_model_5_ptm_pred_196.pdb
ranked_28_unrelaxed_model_4_ptm_pred_30.pdb
ranked_32_unrelaxed_model_5_ptm_pred_2.pdb
ranked_4_unrelaxed_model_2_ptm_pred_179.pdb
END
```

Choose top 5 models from the MassiveFold dataset

Large sampling of
AlphaFold2

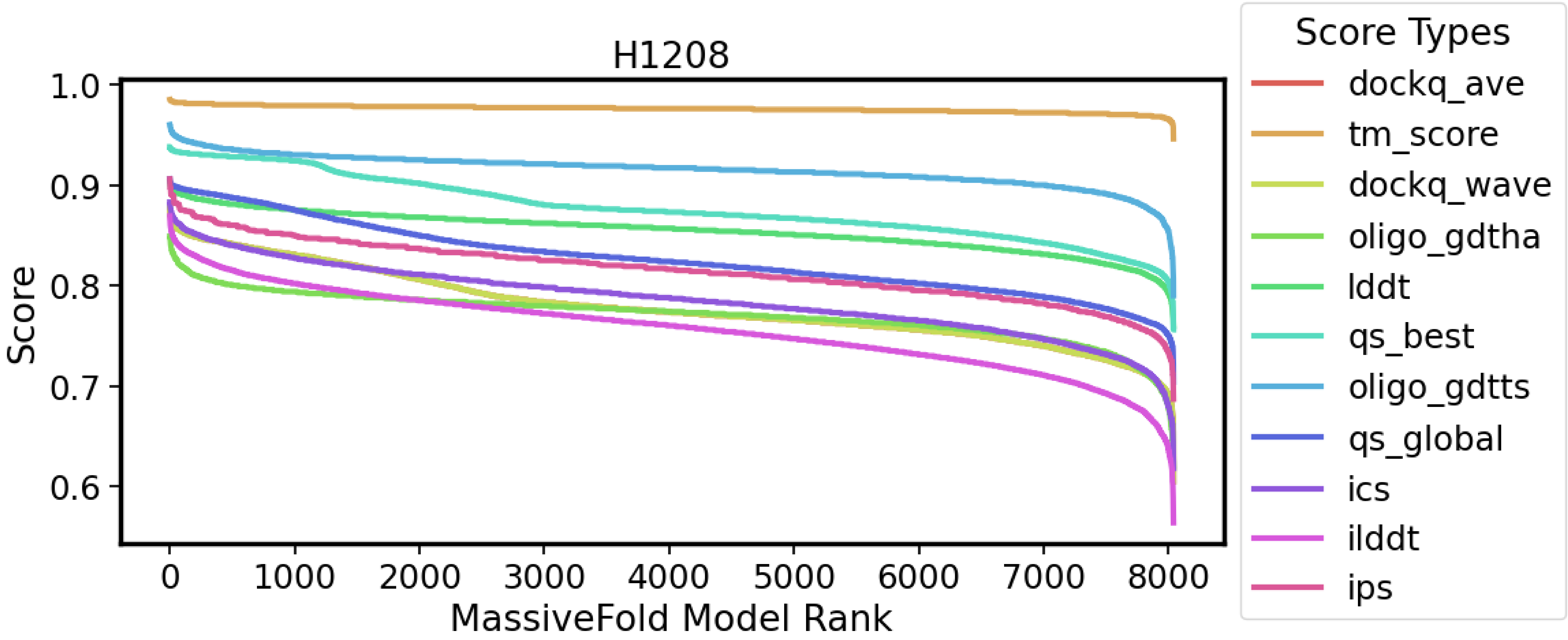
```
PFRMAT QA
TARGET T0999
AUTHOR 1234-5678-9000
METHOD Description of methods used
MODEL 2
QMODE 3
ranked_0_unrelaxed_model_5_ptm_pred_82.pdb
ranked_13_unrelaxed_model_5_ptm_pred_196.pdb
ranked_28_unrelaxed_model_4_ptm_pred_30.pdb
ranked_32_unrelaxed_model_5_ptm_pred_2.pdb
ranked_4_unrelaxed_model_2_ptm_pred_179.pdb
END
```

20 groups predicted
more than 60 targets

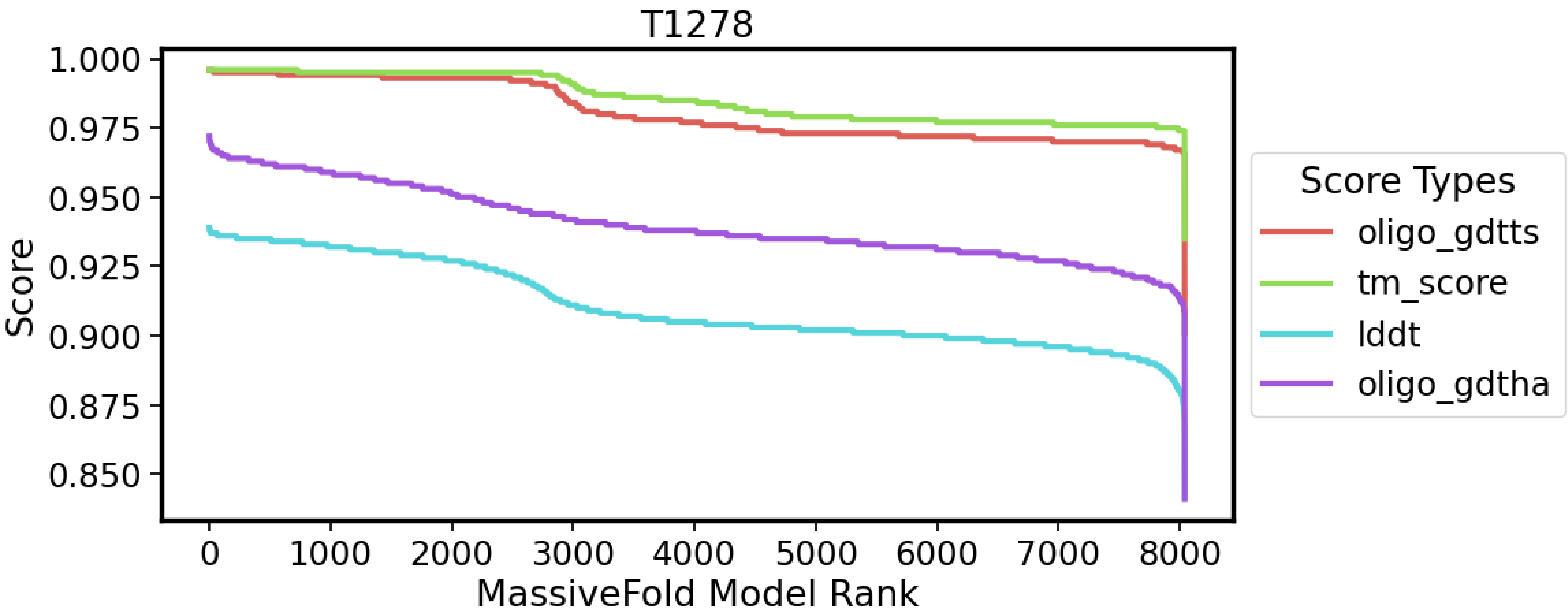
<u>Baselines</u>: iptm,
iptm+ptm, mean plddt
from MassiveFold

OpenStructure (OST) to score each MassiveFold model against experimental target

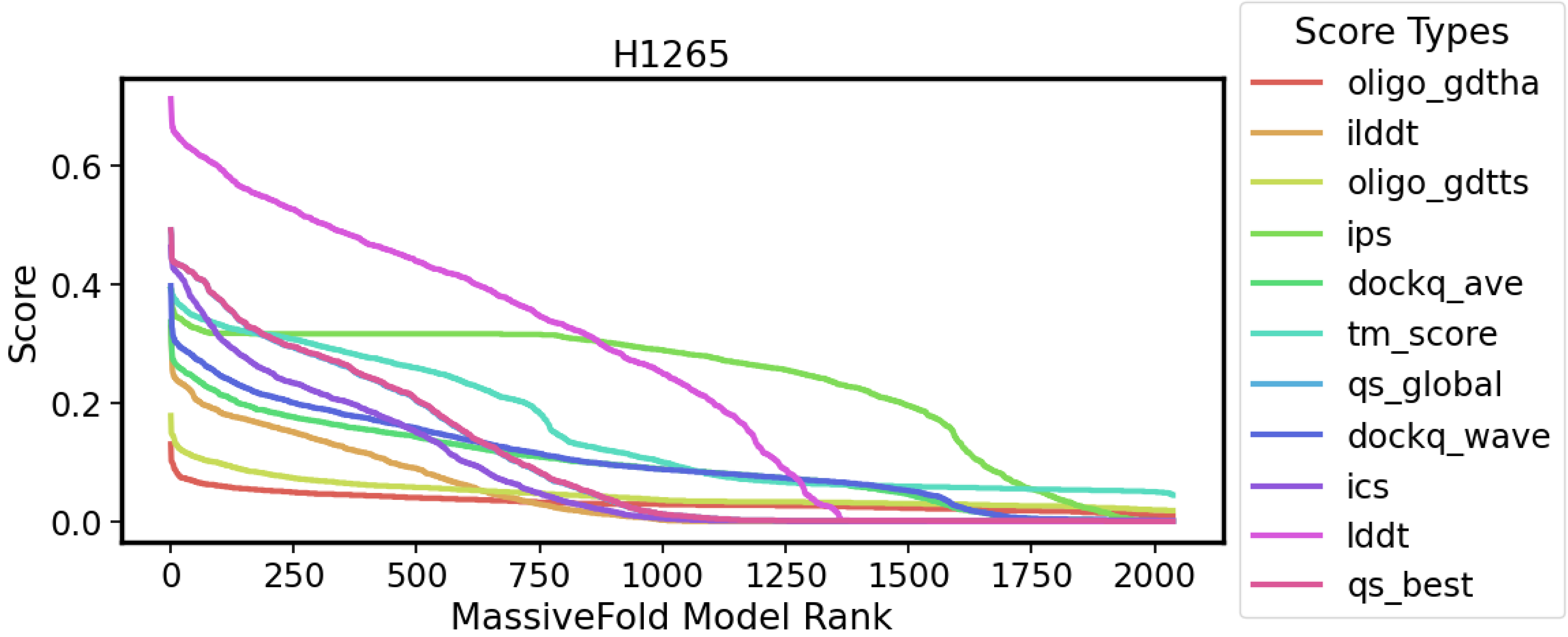
OpenStructure (OST) to score each MassiveFold model against experimental target



OpenStructure (OST) to score each MassiveFold model against experimental target



OpenStructure (OST) to score each MassiveFold model against experimental target



Assessing Group Predictions

Assign a penalty based on predicted ranking and true ranking

$$penalty = \sum_{i=1}^{5} (T_i - P_i)^2$$

T_i, score for i-th model in the true rankings

P_i, score for i-th model in the predicted rankings

Assessing Group Predictions

Assign a penalty based on predicted ranking and true ranking

$$penalty = \sum_{i=1}^{5} (T_i - P_i)^2$$

T_i, score for i-th model in the true rankings

P_i, score for i-th model in the predicted rankings

No further penalty for skipping a target

Assessing Group Predictions

Assign a penalty based on predicted ranking and true ranking

$$penalty = \sum_{i=1}^{5} (T_i - P_i)^2$$

T_i, score for i-th model in the true rankings

P_i, score for i-th model in the predicted rankings

No further penalty for skipping a target

OUTPUT FOR EACH PREDICTOR GROUP: **penalty matrix**

s x t (scores, targets)

Assessing Group Predictions

Assign a penalty based on predicted ranking and true ranking

$$penalty = \sum_{i=1}^{5} (T_i - P_i)^2$$

T_i , score for i-th model in the true rankings

P_i , score for i-th model in the predicted rankings

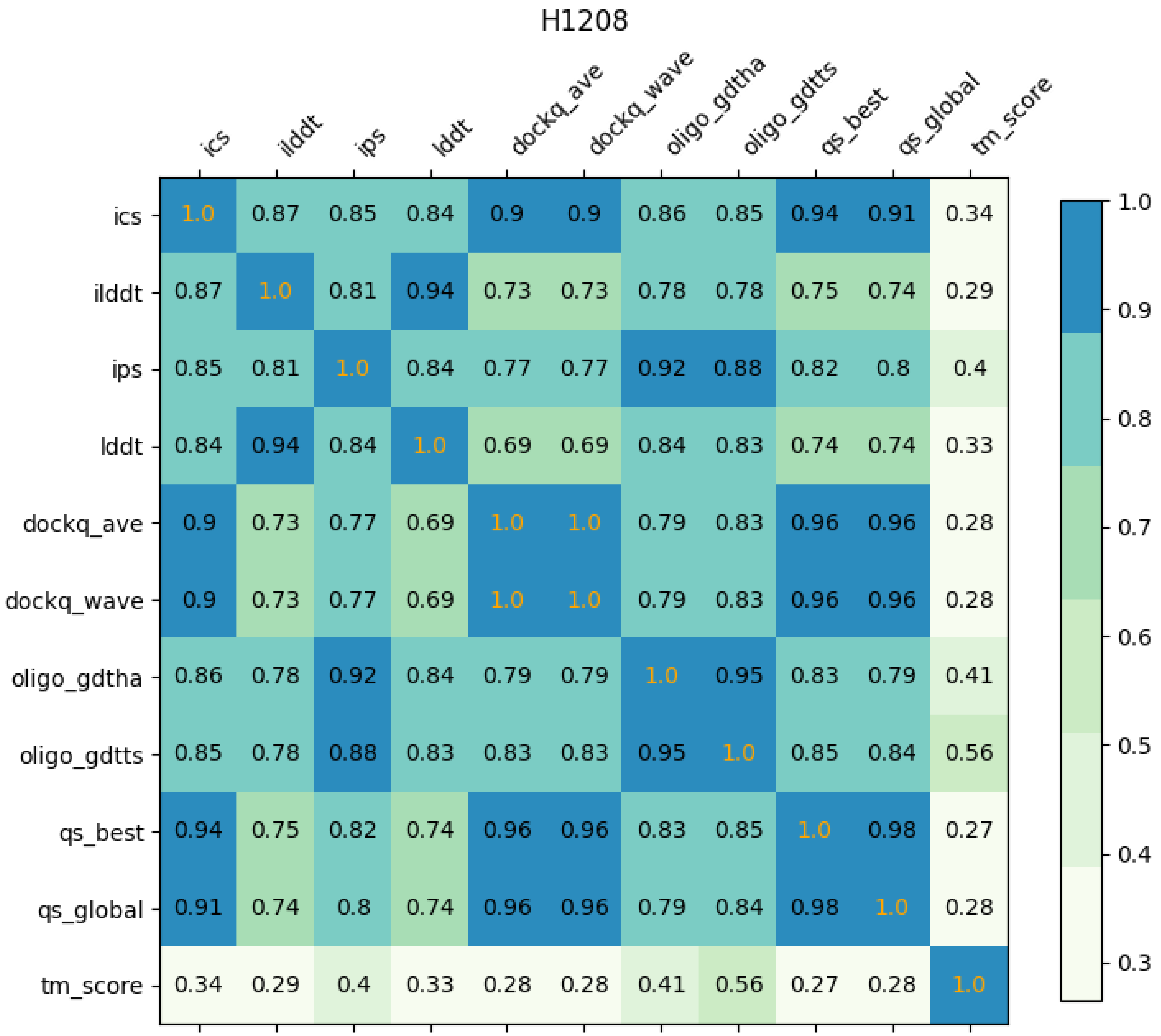
No further penalty for skipping a target

OUTPUT FOR EACH PREDICTOR GROUP: **penalty matrix**

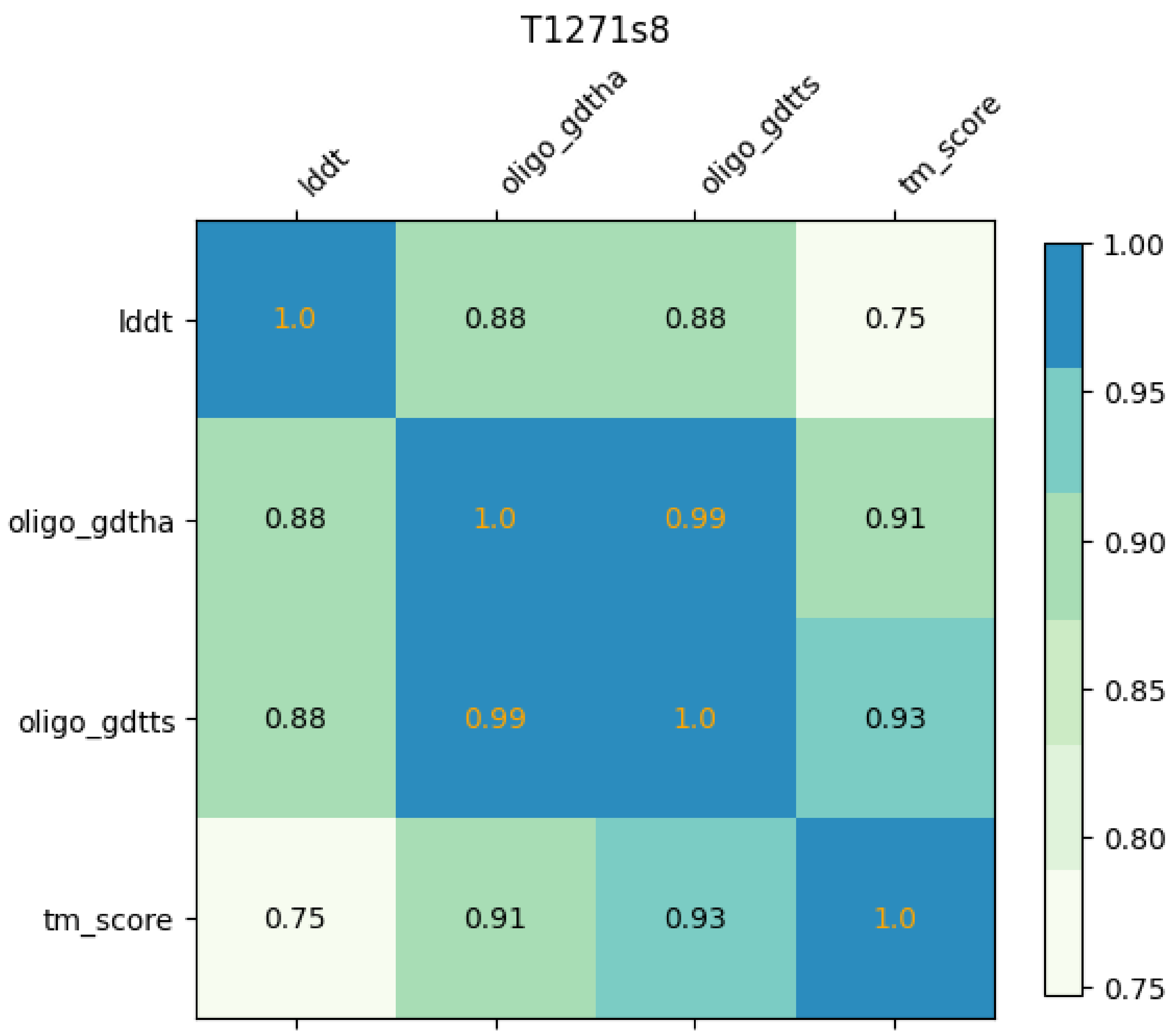
s x t (scores, targets)

3 categories: monomers, homo-oligomers, hetero-oligomers

Correlation Matrix for Scores in MassiveFold Set



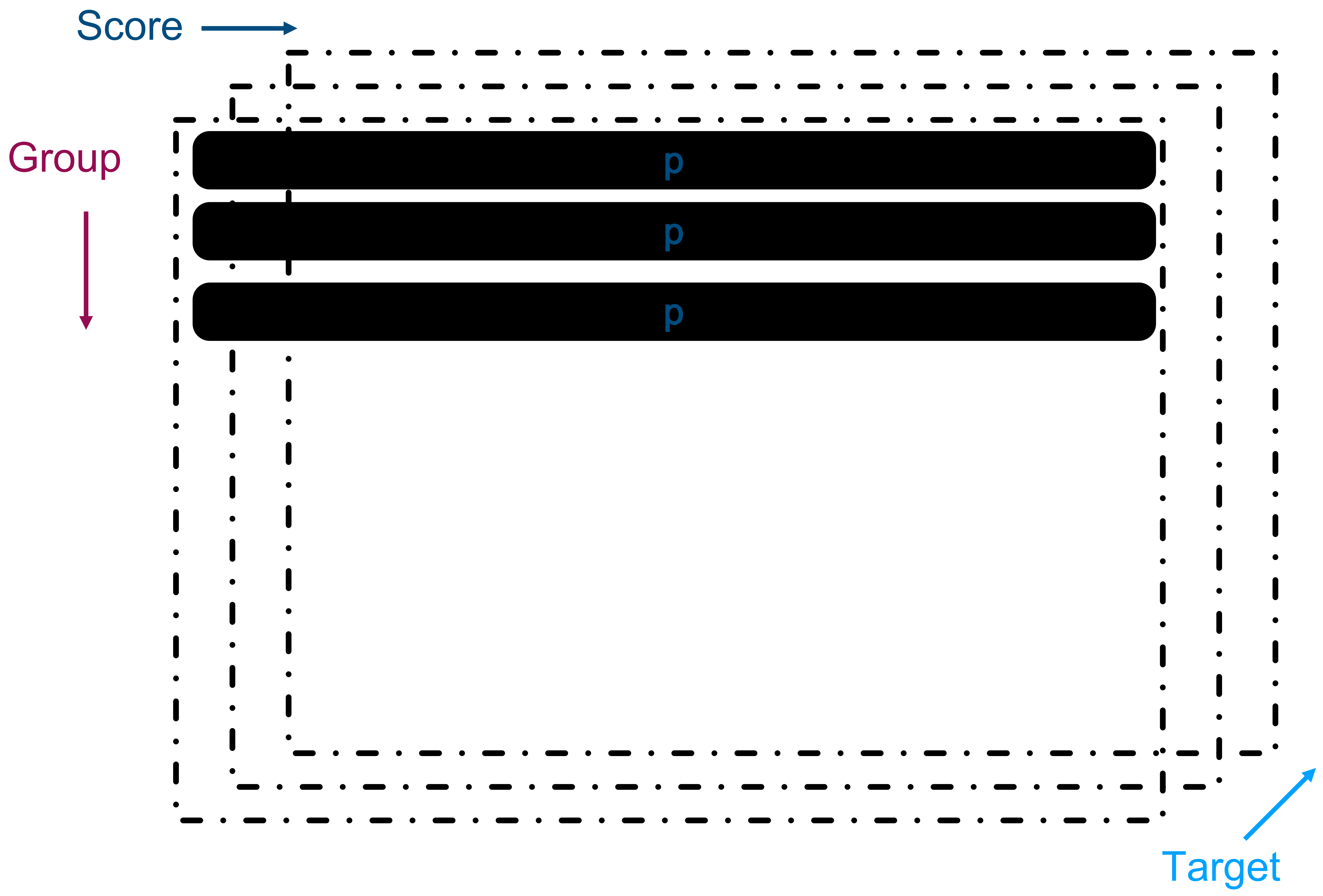
H* target example



T*s* target example

Predictor Group Rankings

weighted_penalty weights the penalties for each score type

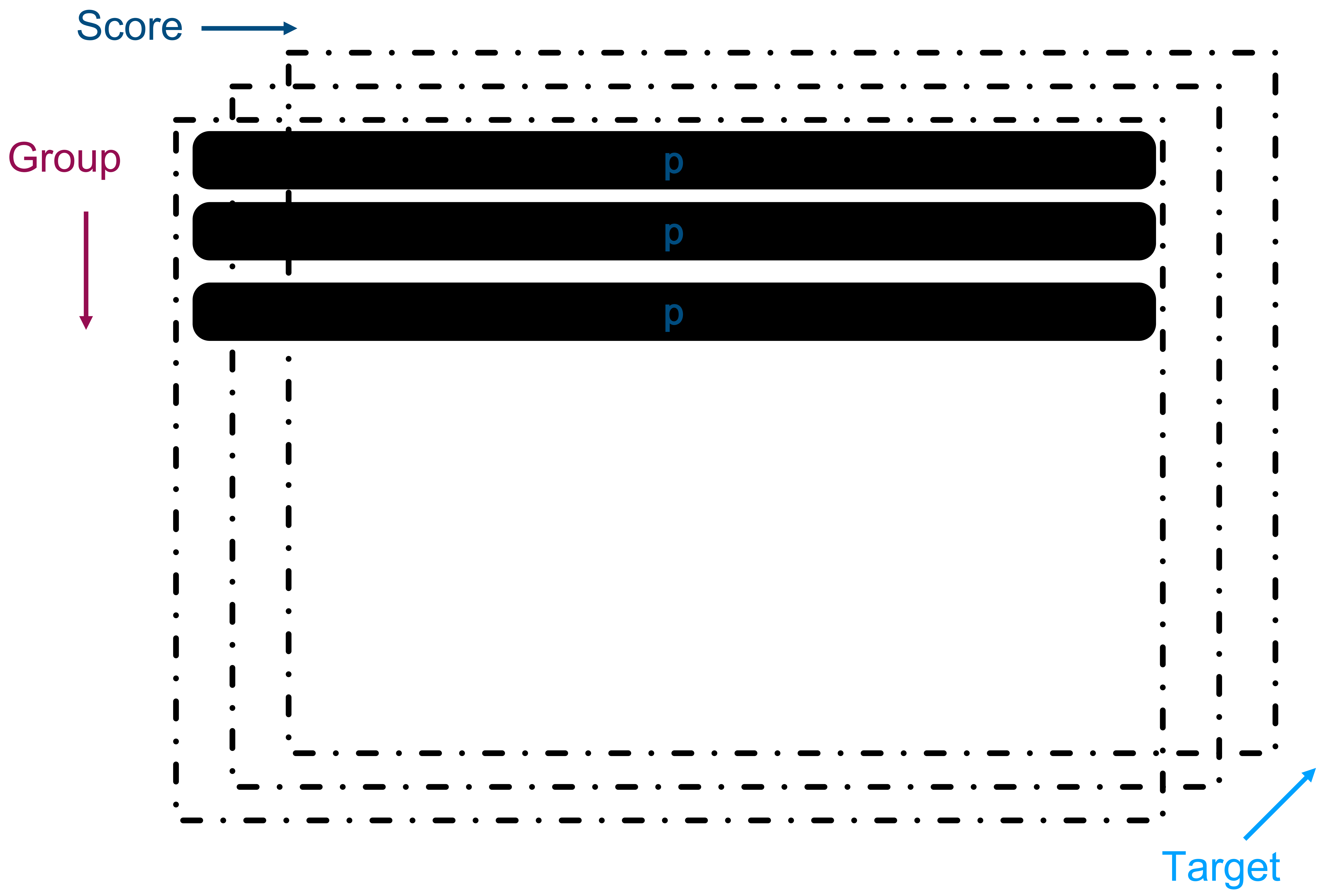


weighted_penalty (w_p)

$$\sqrt{\mathbf{p}\Sigma^{-1}\mathbf{p}^{T}}$$

Predictor Group Rankings

weighted_penalty weights the penalties for each score type



weighted_penalty (w_p)

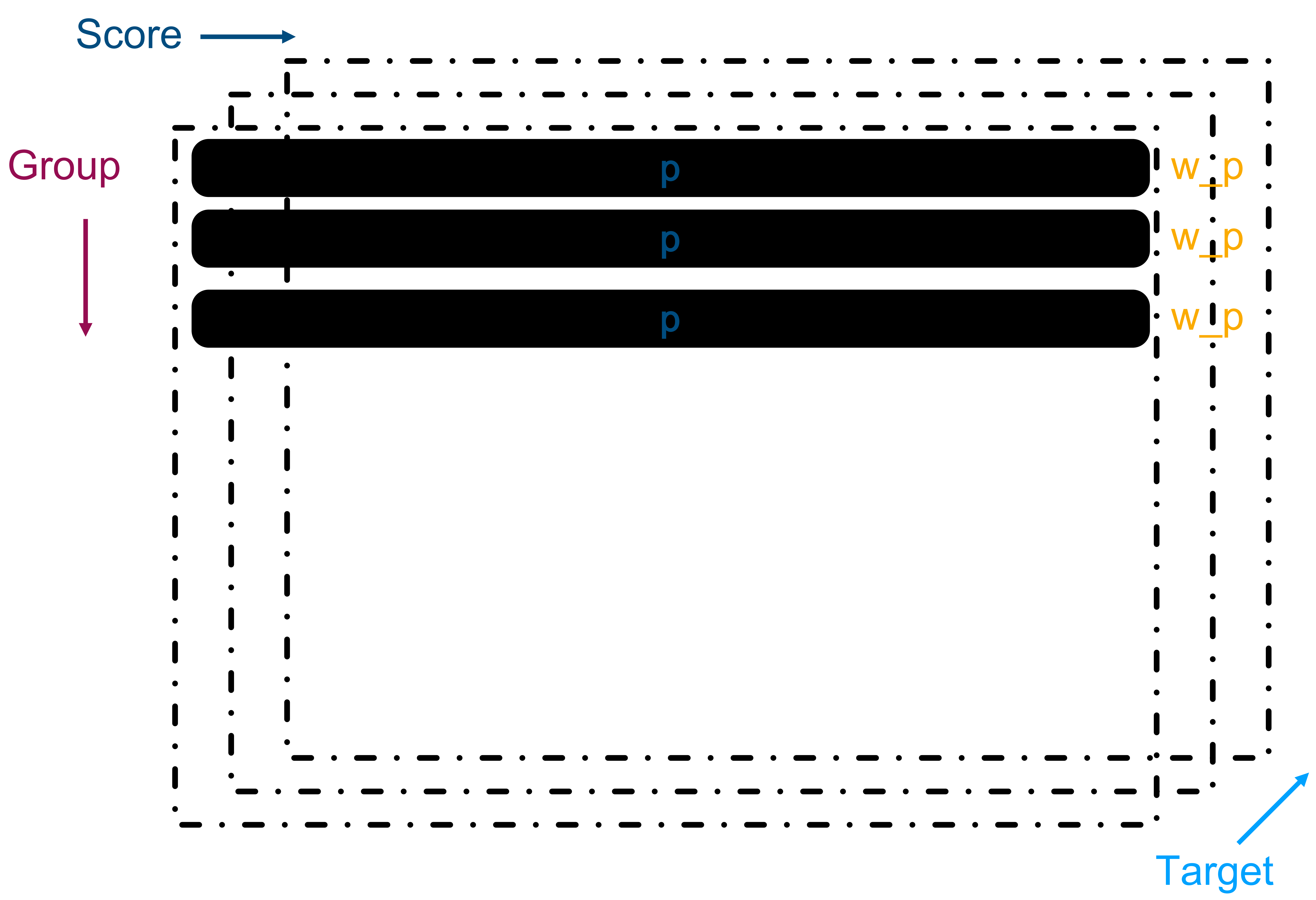
$$\sqrt{\mathbf{p}\Sigma^{-1}\mathbf{p}^T}$$

Σ^{-1} is the covariance matrix of scores for MF models in that category

(also tested the covariance matrix of scores for MF dataset of that target)

Predictor Group Rankings

weighted_penalty weights the penalties for each score type



weighted_penalty (w_p)

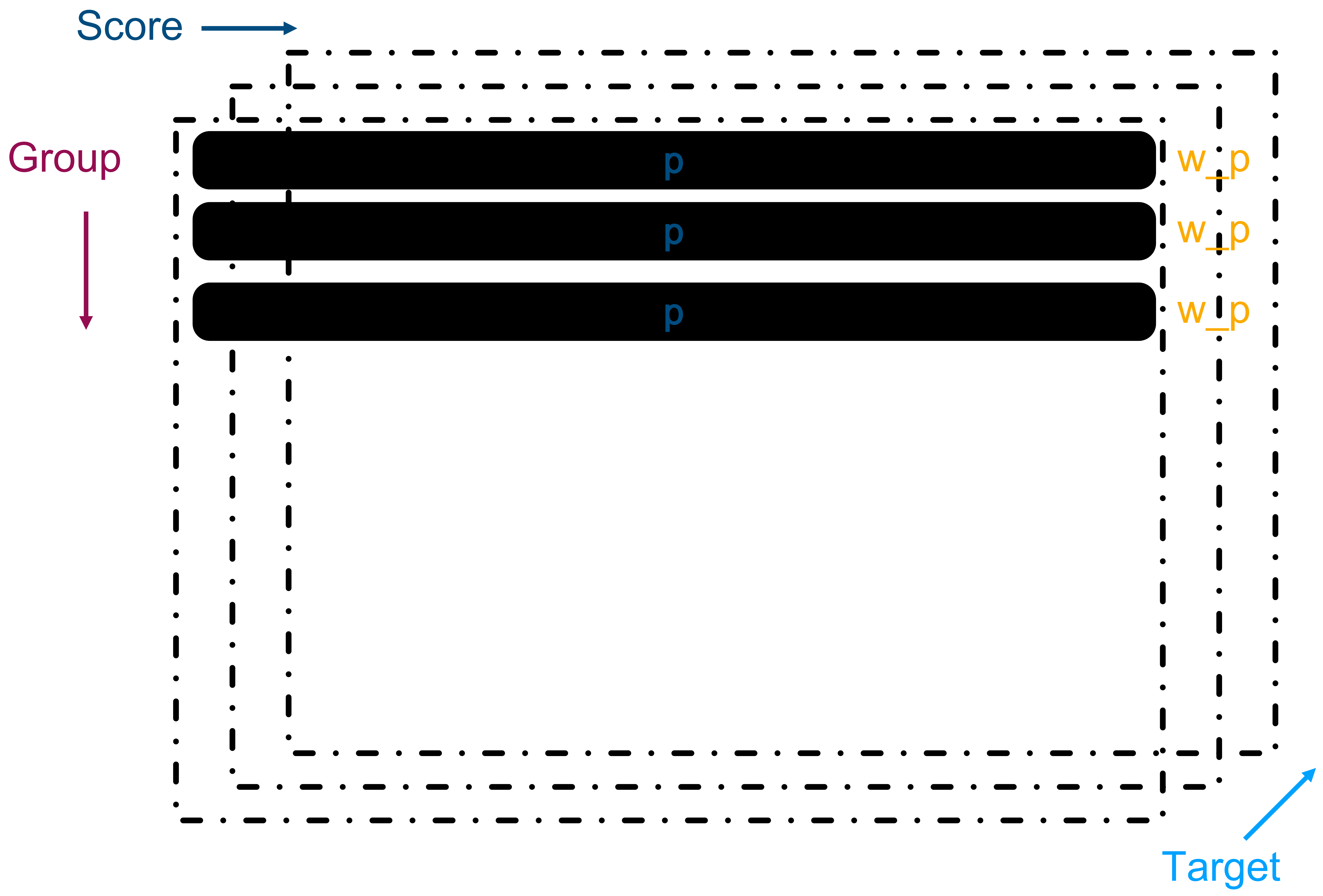
$$\sqrt{\mathbf{p}\Sigma^{-1}\mathbf{p}^T}$$

Σ^{-1} is the covariance matrix of scores for MF models in that category

(also tested the covariance matrix of scores for MF dataset of that target)

Predictor Group Rankings

weighted_penalty weights the penalties for each score type



weighted_penalty (w_p)

$$\sqrt{\mathbf{p}\Sigma^{-1}\mathbf{p}^T}$$

Σ^{-1} is the covariance matrix of scores for MF models in that category

(also tested the covariance matrix of scores for MF dataset of that target)

For each group:

Mean of weighted_penalty over all targets

Predictor Group

Rankings

weighted_penalty, $\Sigma_{\cdot}^{-1}$ outlier rejection

Remove outlier “yarn ball” predictions

Mahalanobis Distance

measure of distance that accounts for correlations between variables and their variances

Useful:
identifying outliers
working with multivariate distributions

Mahalanobis Distance

measure of distance that accounts for correlations between variables and their variances

Useful:
identifying outliers
working with multivariate distributions

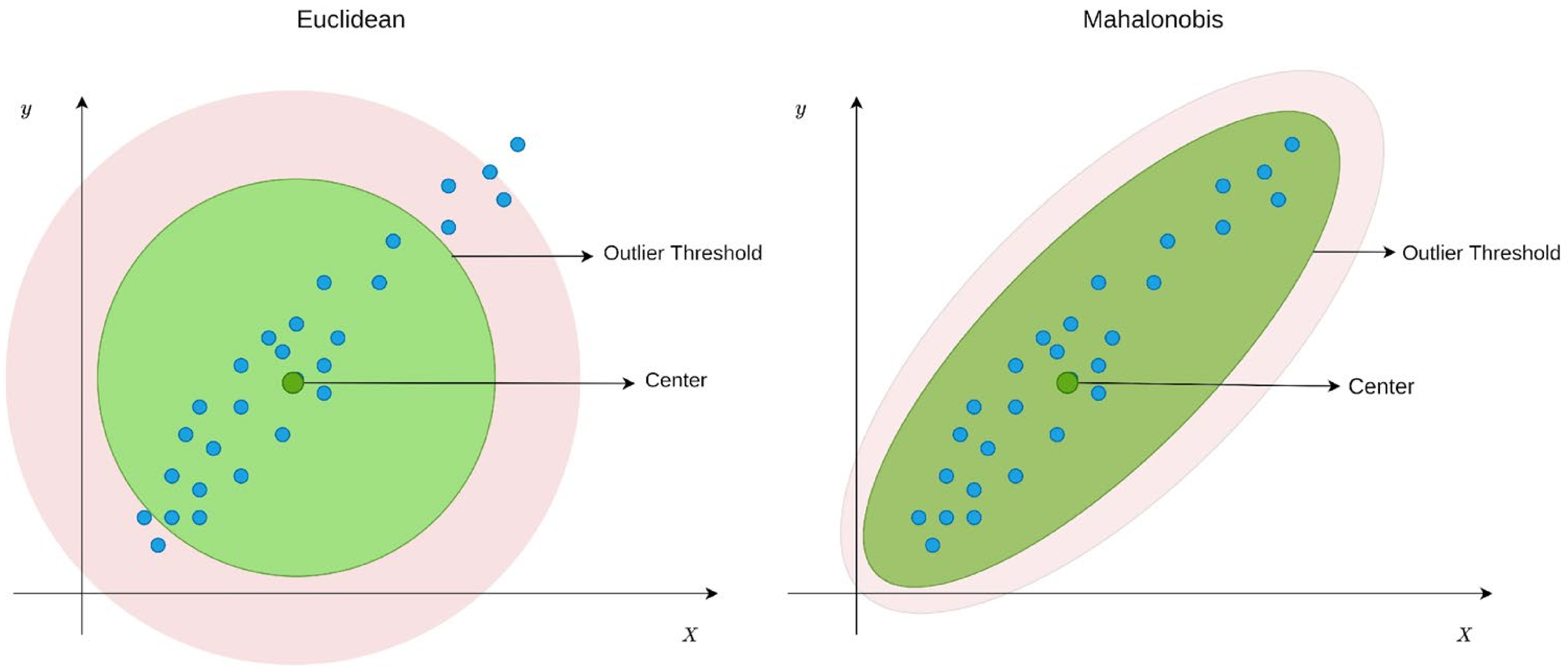
$$D_M(\mathbf{x}) = \sqrt{(\mathbf{x}-\mu)^\top \Sigma^{-1} (\mathbf{x}-\mu)}$$

$D_M(\mathbf{x})$: Mahalanobis distance of point $\mathbf{x}$

$\mathbf{x}$: Data point (vector)

μ: Mean vector of the distribution

Σ: Covariance matrix of the distribution



Plot by Sergen Cansiz, published in Towards Data Science

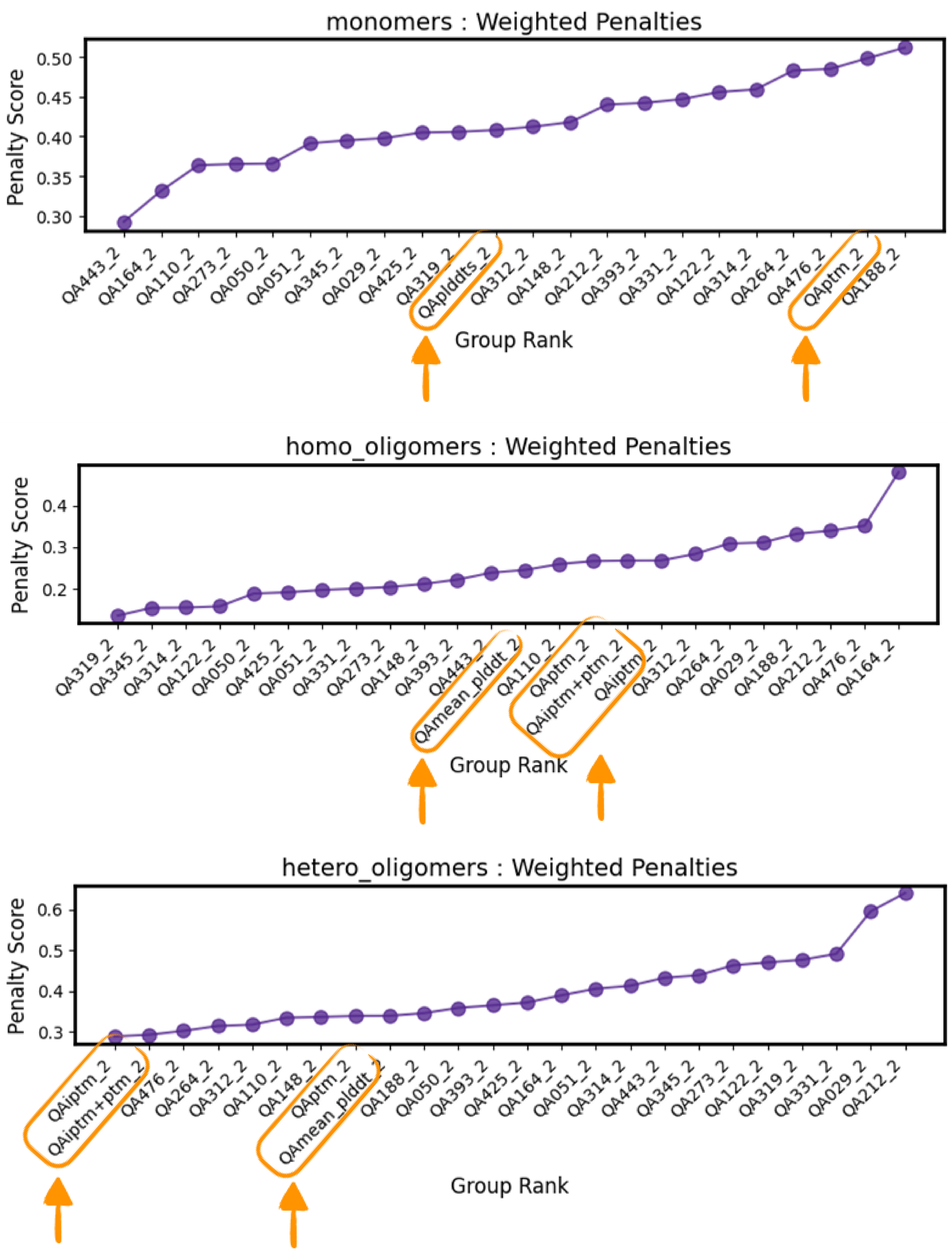
Predictor Group
Rankings
weighted_penalty, Σ^{-1} outlier rejection

Calculated Mahalanobis distance for each model from distribution of all models

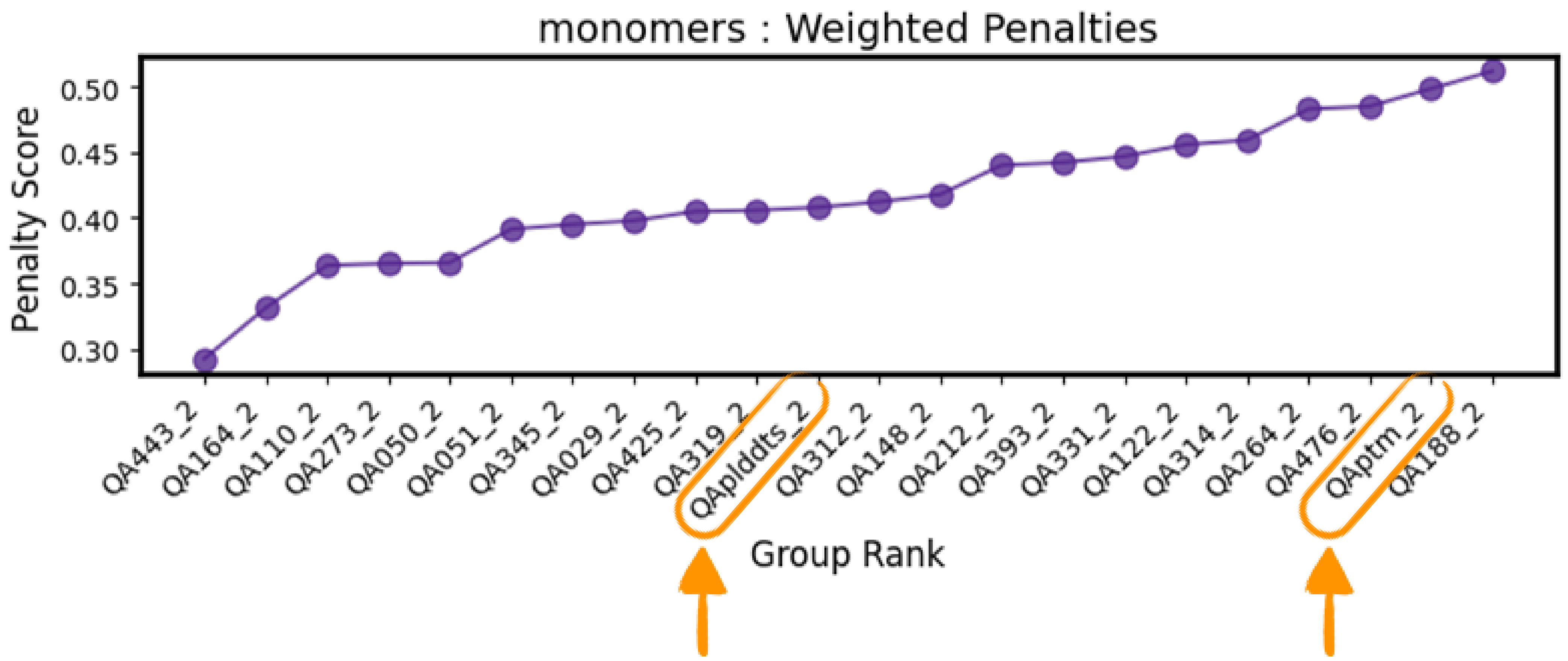
Reject models with Mahalanobis distance larger than 3 standard deviations for Σ^{-1} computation

Remove outlier "yarn ball" predictions

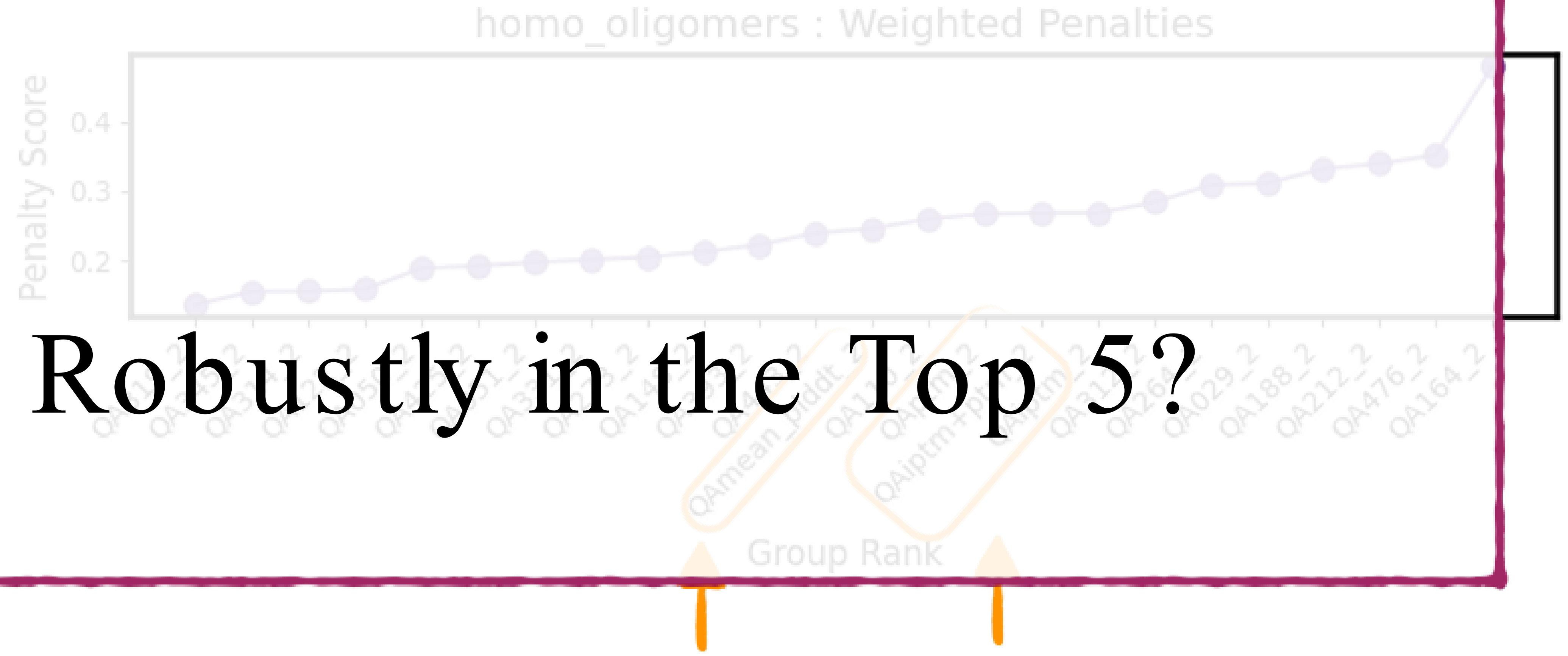
Weighted Penalty-Based Rankings

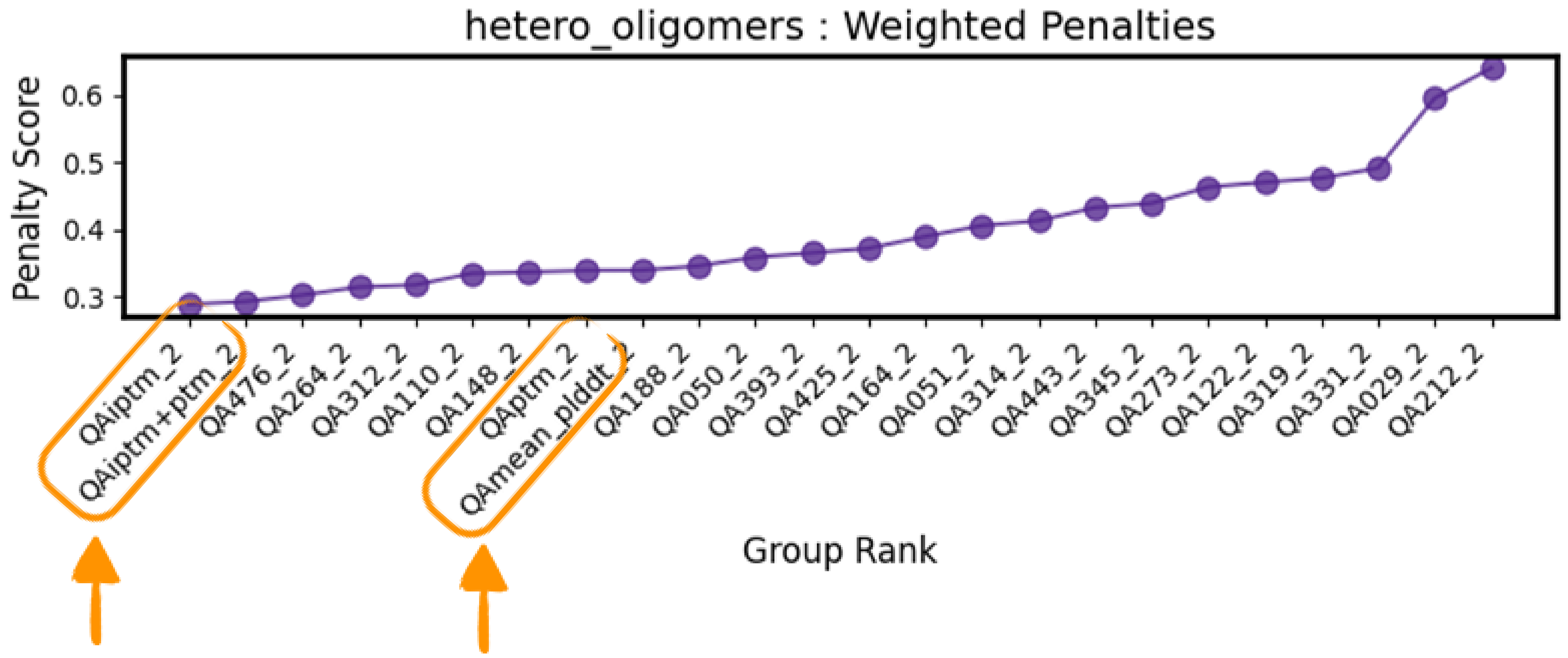


Weighted Penalty-Based Rankings

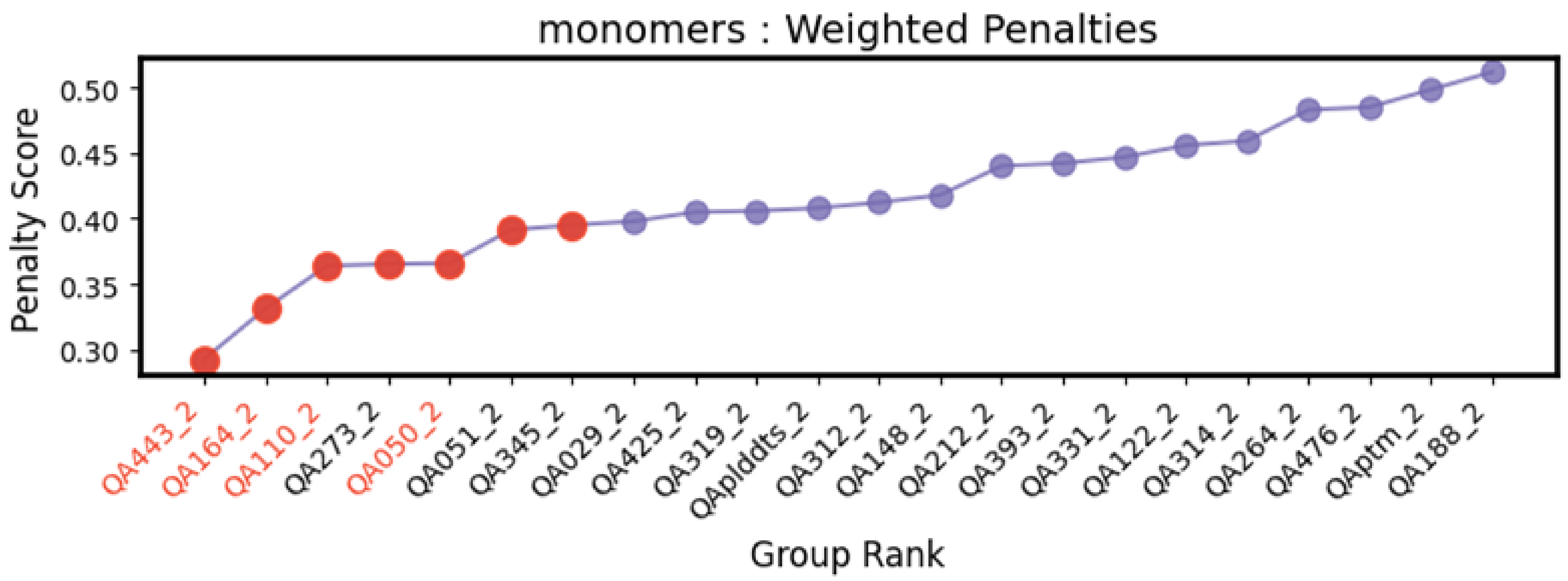


Groups Robustly in the Top 5?





Robust to ranking by Z-score and per-target covariance:



050 David Shortle – Human

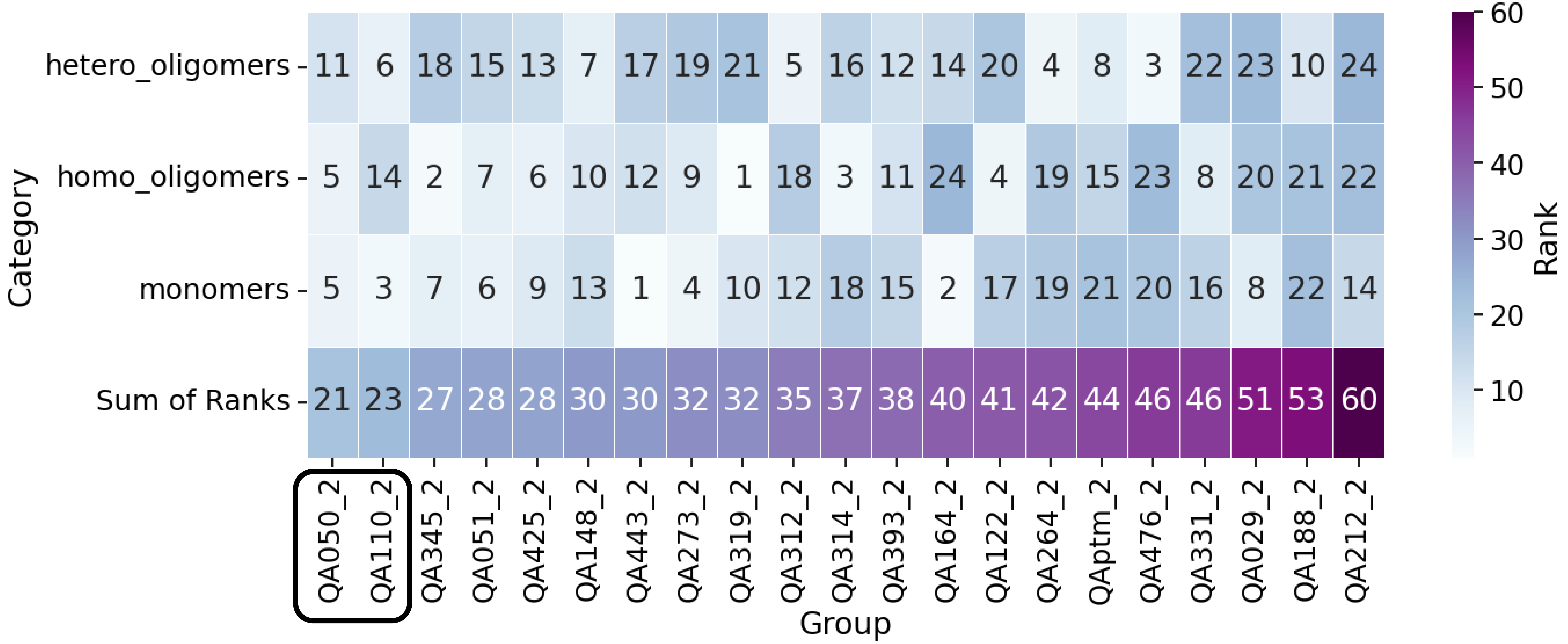
110 Wei Zheng – MIEnsembles-Server

164 Liam McGuffin – Human

Strong for Generality

050 David Shortle – Human

110 Wei Zheng – MIEnsembles-Server



050 David Shortle – Human
110 Wei Zheng – MIEnsembles-Server

050 David Shortle – Human

Method

- Scored models with a collection of statistical parameters and potentials.
- Compared model scores with the scores of 6000 high resolution PDBs.

Characteristics

- High resolution structure set was collection of single chain PDBs
- 3 sets of parameters combined and added with weights for a combination score

110 Wei Zheng – MIEnsembles-Server

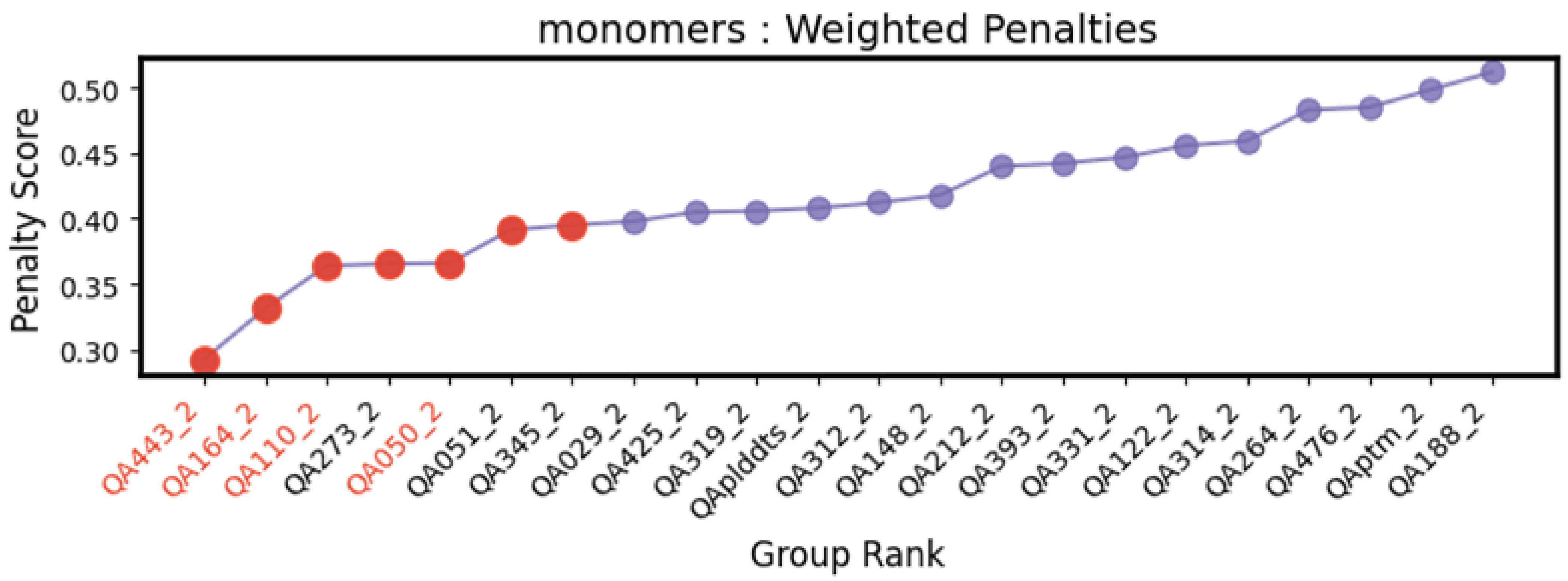
Method

- Use DMFold to construct high-quality MSAs for improved structure predictions.
- QA method integrates DMFold models to assess the quality of MassiveFold models

Characteristics

- Quality of the reference model influences the performance of the QA method
- For complexes: DMFold confidence (0.8ipTM+0.2pTM) is less sensitive in picking correct models than mean plddt for monomers

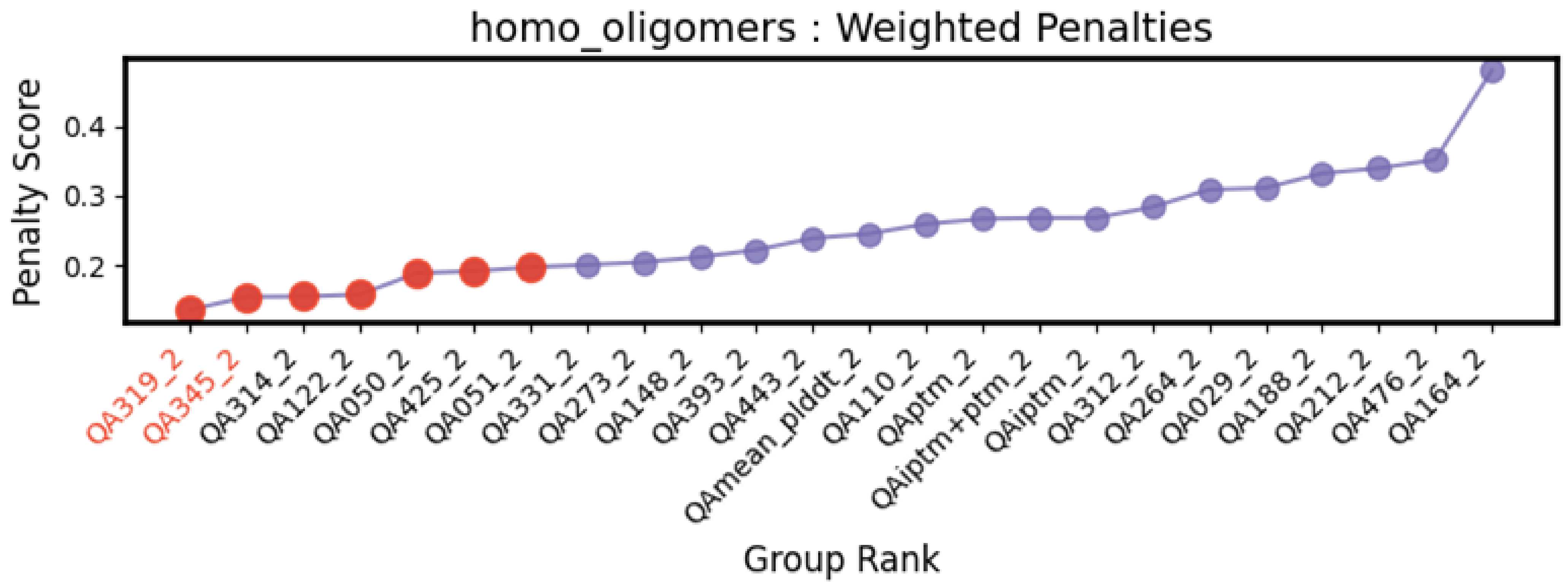
Robust to ranking by Z-score and per-target covariance:



050 David Shortle – Human

110 Wei Zheng – MIEnsembles-Server

164 Liam McGuffin – Human



319 Jianlin Cheng – MULTICOM_LLM

345 Jianlin Cheng – MULTICOM_HUMAN

319 Jianlin Cheng – MULTICOM_LLM
345 Jianlin Cheng – MULTICOM_HUMAN

319 Jianlin Cheng – MULTICOM_LLM

Method

- 200 MassiveFold models selected based on confidence scores
- MULTICOM QAs to select the top 5 models

Characteristics

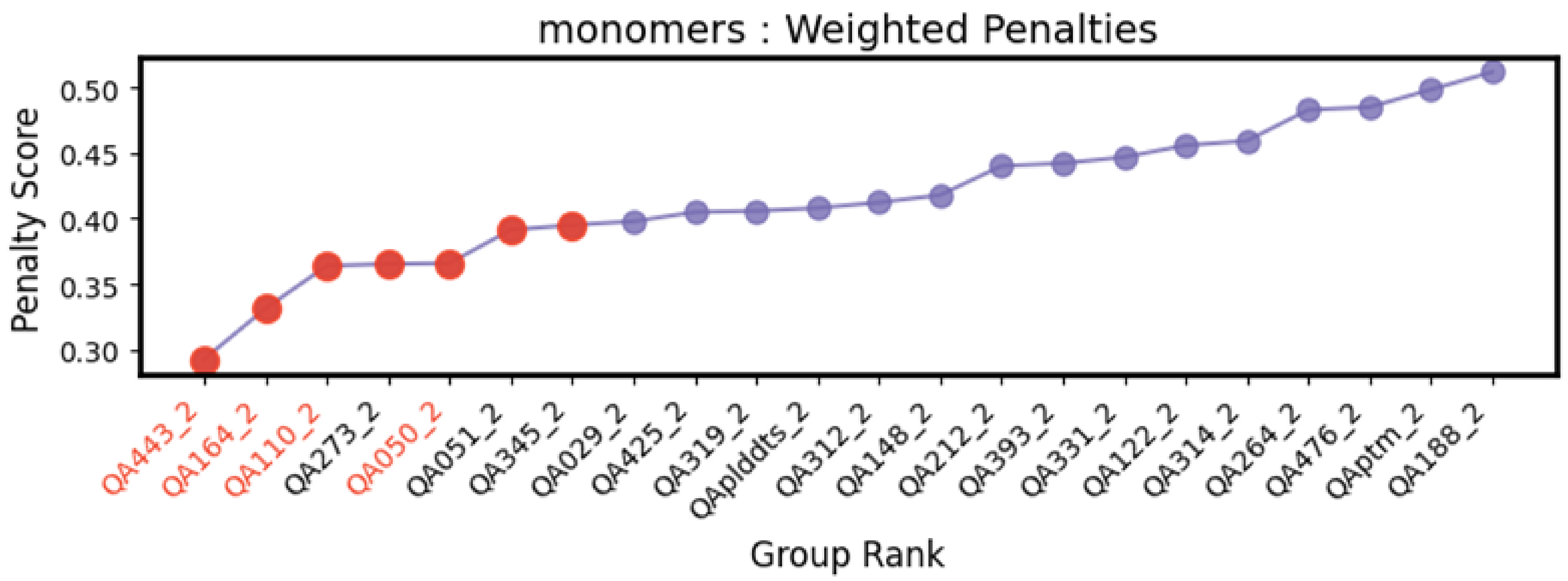
- Average pairwise similarity score to estimate the quality scores

345 Jianlin Cheng – MULTICOM_HUMAN

Method

- Also includes MULTICOM_GATE score
- Deep learning method that combines the pairwise similarity score and the single-model QAs to estimate the quality of the structures
- Model similarity graph

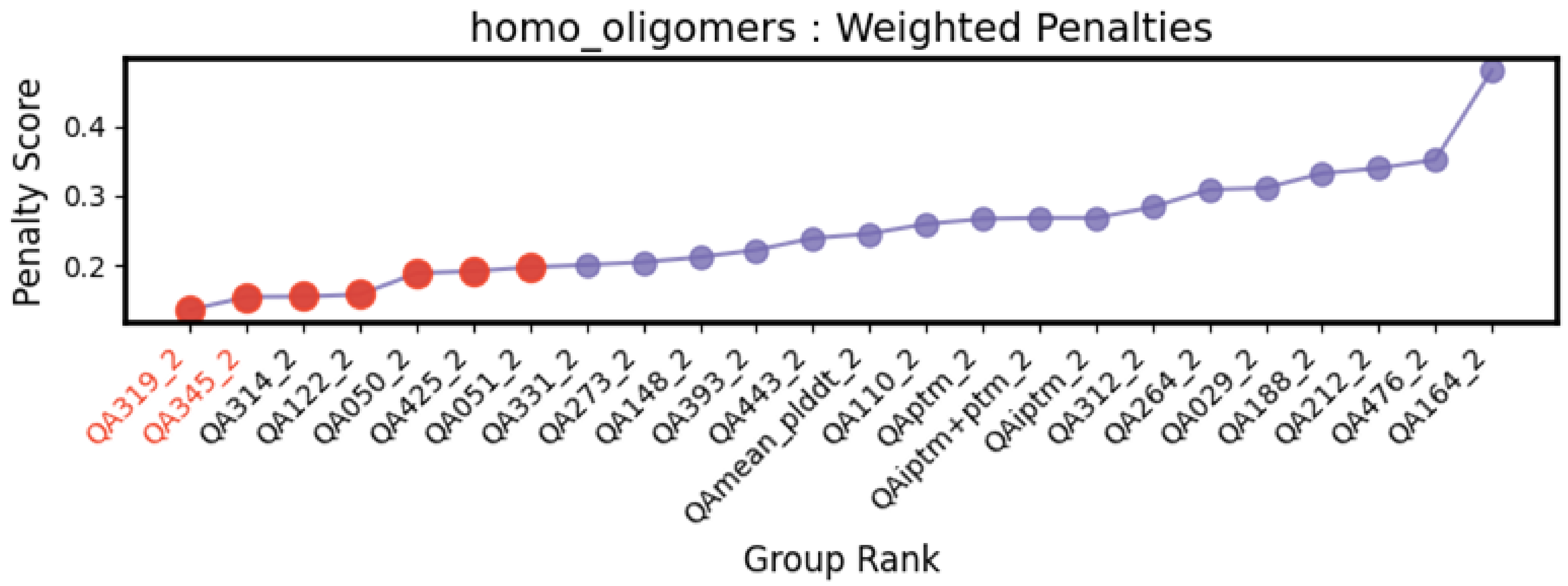
Robust to ranking by Z-score and per-target covariance:



050 David Shortle – Human

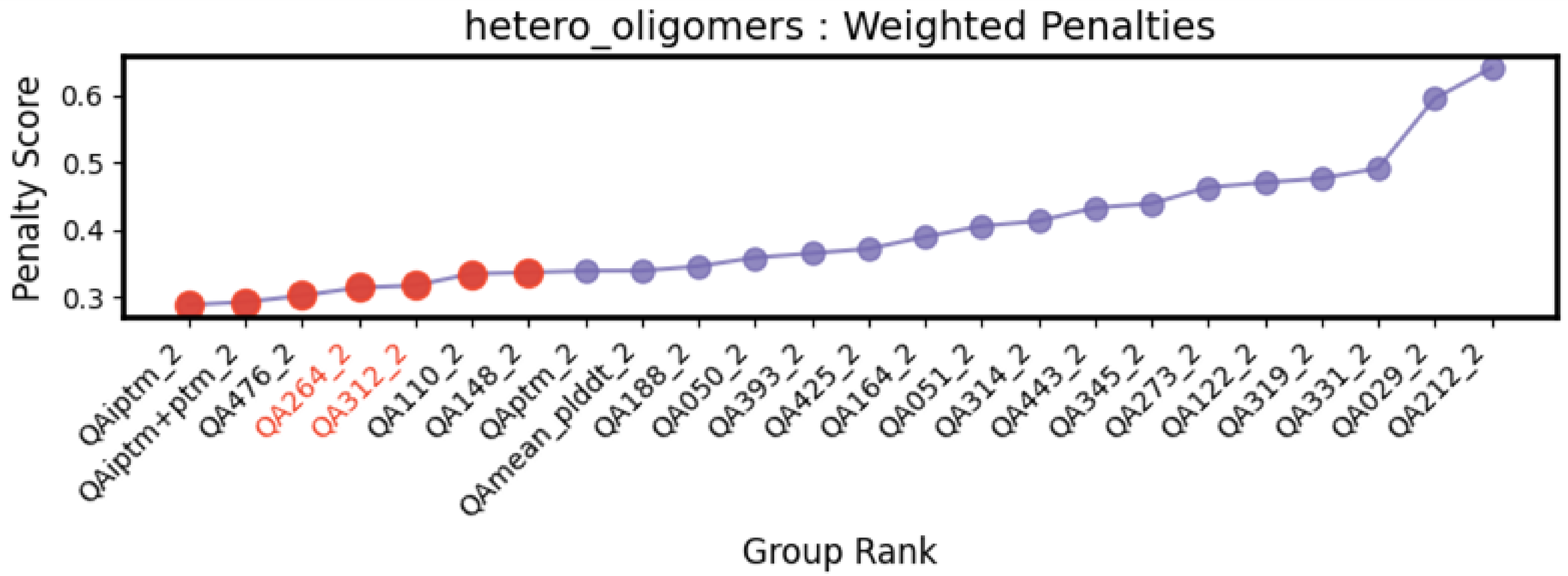
110 Wei Zheng – MIEnsembles-Server

164 Liam McGuffin – Human



319 Jianlin Cheng – MULTICOM_LLM

345 Jianlin Cheng – MULTICOM_HUMAN



264 Guijun Zhang – GuijunLab-Human

312 Guijun Zhang – GuijunLab-Assembly

264 Guijun Zhang – GuijunLab-Human

312 Guijun Zhang – GuijunLab-Assembly

264 Guijun Zhang – GuijunLab-Human

Characteristics

- Did not use the ptm, iptm and plddt of MassiveFold

312 Guijun Zhang – GuijunLab-Assembly

Characteristics

- Used the confidence ranking score from MassiveFold for preliminary screening on 8040 structures.

CASP16 EMA

Final Thoughts

CASP16 EMA

Final Thoughts

Confidence Estimates and Experiment

- Finer-grained (atomic) estimates for confidence add value

Make this an explicit goal in CASP17?

- For practical uses of predicted models, confidence in relative domain orientations is essential

Should CASP17 evaluate PAE matrices? Or define something more general?

CASP16 EMA

Final Thoughts

Confidence Estimates and Experiment

- Finer-grained (atomic) estimates for confidence add value

Make this an explicit goal in CASP17?

- For practical uses of predicted models, confidence in relative domain orientations is essential

Should CASP17 evaluate PAE matrices? Or define something more general?

QMODE1 and QMODE2

- Improvement in score distributions compared to CASP15

CASP16 EMA

Final Thoughts

Confidence Estimates and Experiment

- Finer-grained (atomic) estimates for confidence add value

Make this an explicit goal in CASP17?

- For practical uses of predicted models, confidence in relative domain orientations is essential

Should CASP17 evaluate PAE matrices? Or define something more general?

QMODE1 and QMODE2

- Improvement in score distributions compared to CASP15

QMODE3

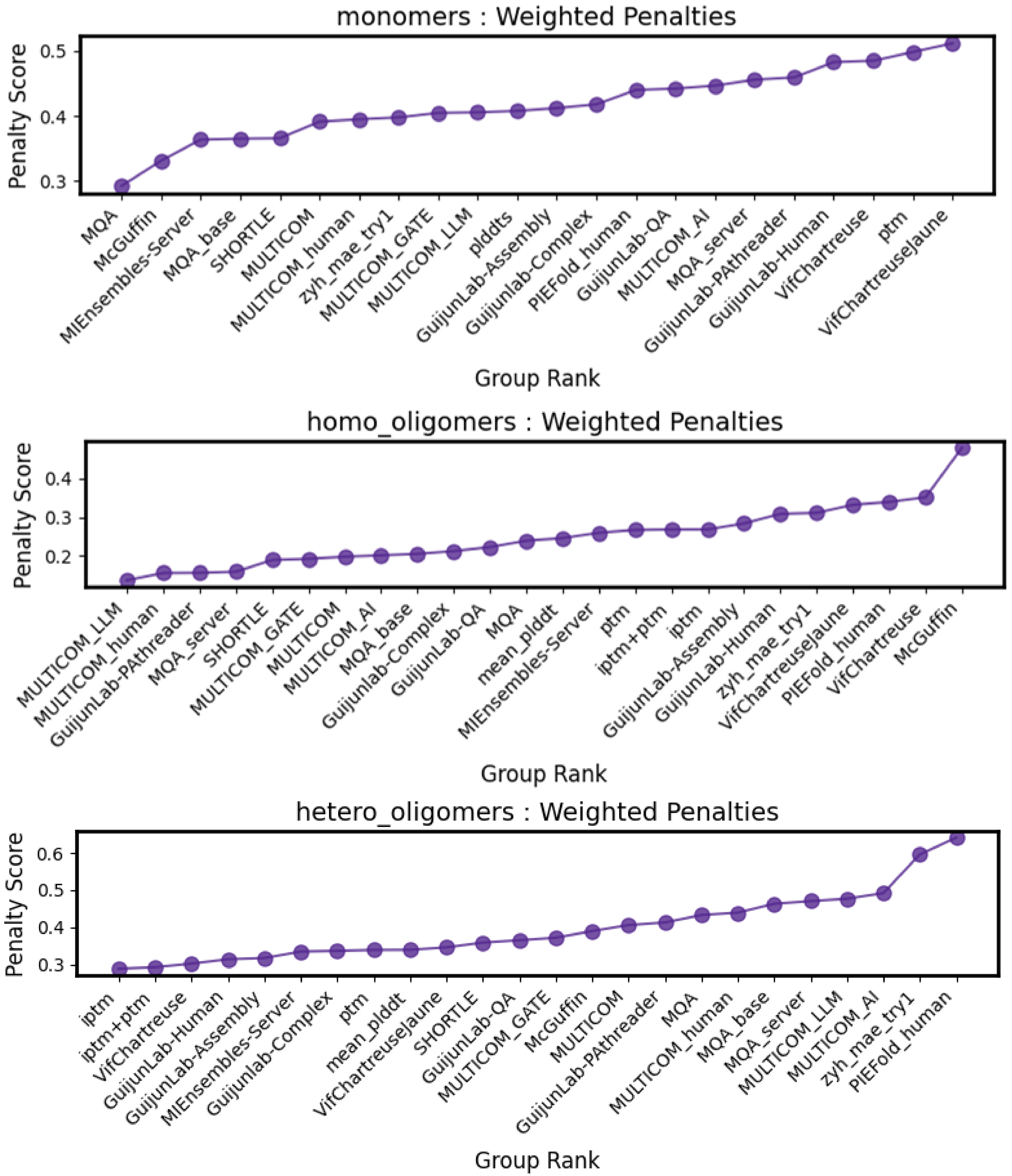
- Methods struggle to outperform AlphaFold2 iptm metric for evaluating hetero-oligomers

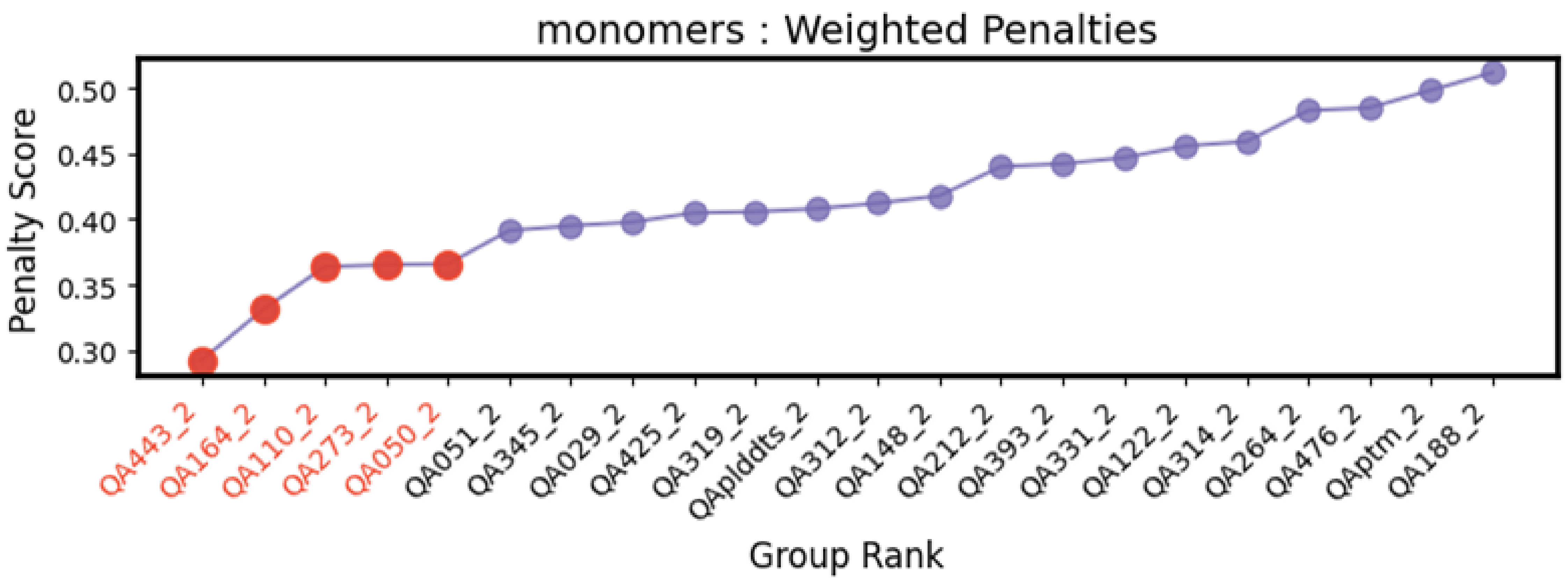
Informative future variant of QMODE3 could request ranking of all available models, instead of top 5

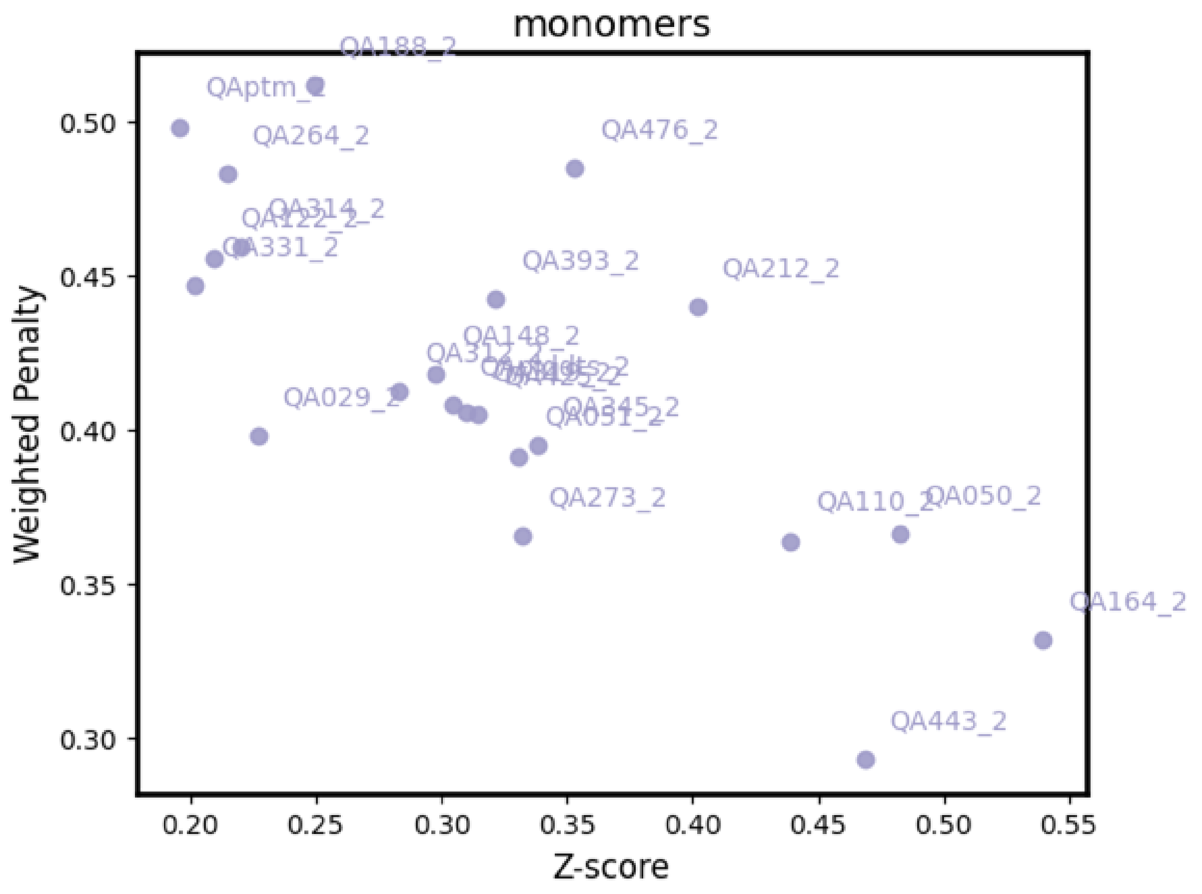
THANK YOU!

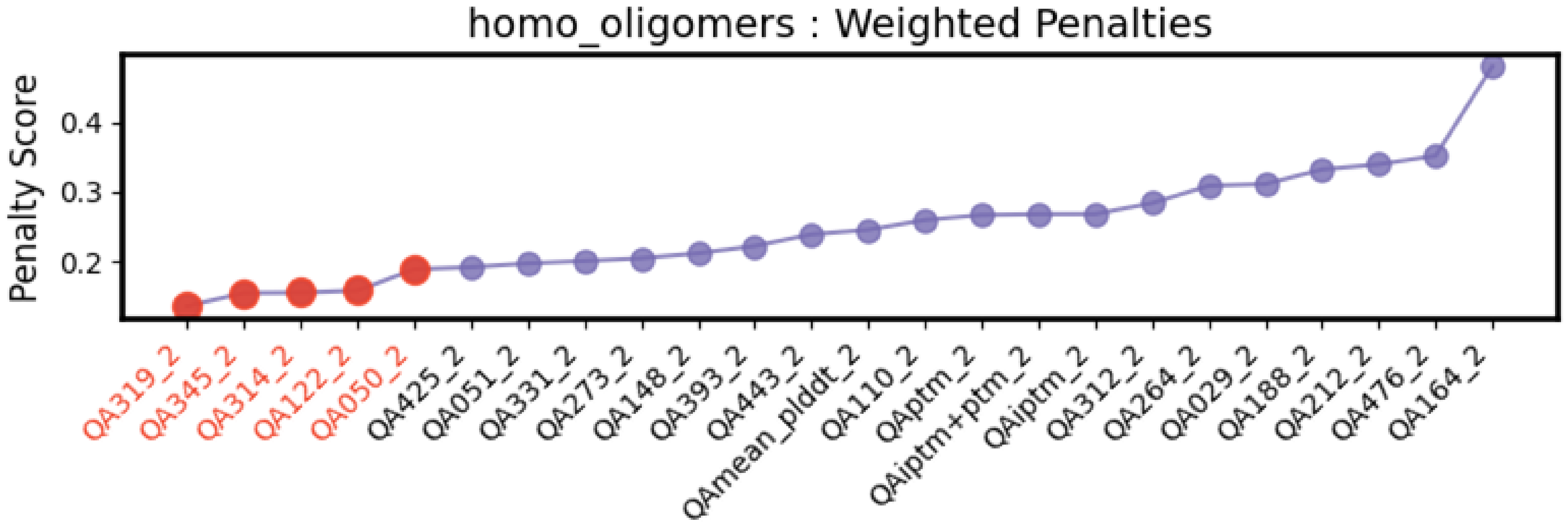
Questions?

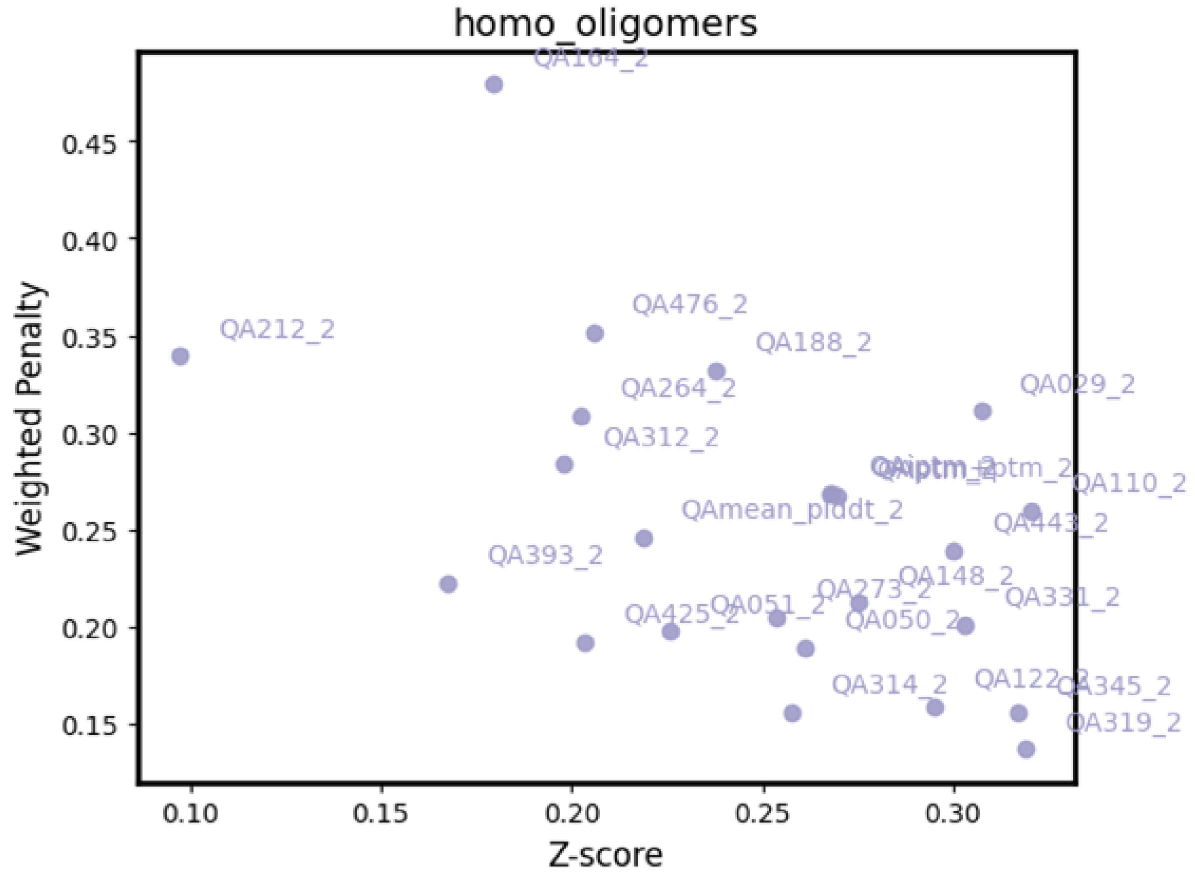
Extra Slides

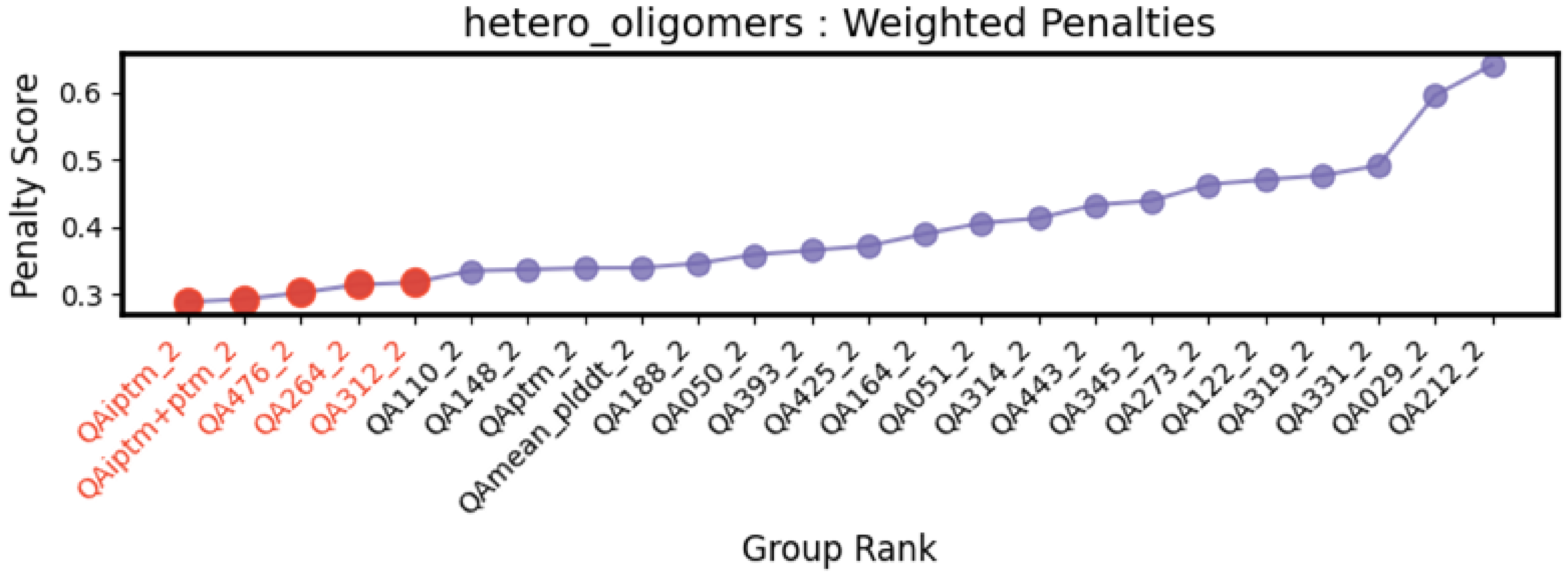

monomers : Weighted Penalties
Penalty Score
0.3
0.4
0.5
MQA
McGuffin
MIEnsembles-Server
MQA_base
SHORTLE
MULTICOM
MULTICOM_human
zyh_mae_try1
MULTICOM_GATE
MULTICOM_LLM
plddts
GuijunLab-Assembly
Guijunlab-Complex
PIEFold_human
GuijunLab-QA
MULTICOM_AI
MQA_server
GuijunLab-PAthreader
GuijunLab-Human
VifChartreuse
ptm
VifChartreuseJaune
Group Rank
homo_oligomers : Weighted Penalties
Penalty Score
0.2
0.3
0.4
MULTICOM_LLM
MULTICOM_human
GuijunLab-PAthreader
MQA_server
SHORTLE
MULTICOM_GATE
MULTICOM
MULTICOM_AI
MQA_base
Guijunlab-Complex
GuijunLab-QA
MQA
mean_plddt
MIEnsembles-Server
ptm
iptm+ptm
iptm
GuijunLab-Assembly
GuijunLab-Human
zyh_mae_try1
VifChartreuseJaune
PIEFold_human
VifChartreuse
McGuffin
Group Rank
hetero_oligomers : Weighted Penalties
Penalty Score
0.3
0.4
0.5
0.6
iptm
iptm+ptm
VifChartreuse
GuijunLab-Human
GuijunLab-Assembly
MIEnsembles-Server
Guijunlab-Complex
ptm
mean_plddt
VifChartreuseJaune
SHORTLE
GuijunLab-QA
MULTICOM_GATE
McGuffin
MULTICOM
GuijunLab-PAthreader
MQA
MULTICOM_human
MQA_base
MQA_server
MULTICOM_LLM
MULTICOM_AI
zyh_mae_try1
PIEFold_human
Group Rank

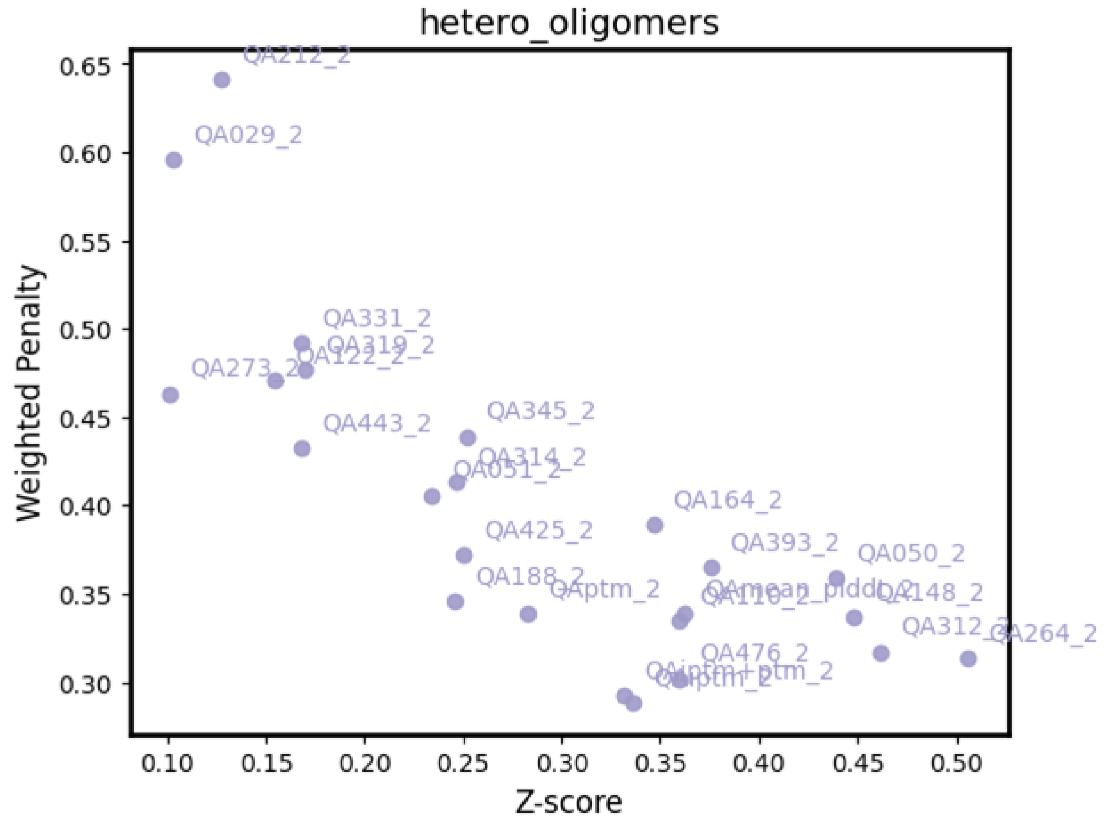

monomers : Weighted Penalties
Penalty Score
Group Rank
QA443_2
QA164_2
QA110_2
QA273_2
QA050_2
QA051_2
QA345_2
QA029_2
QA425_2
QA319_2
QAplddts_2
QA312_2
QA148_2
QA212_2
QA393_2
QA331_2
QA122_2
QA314_2
QA264_2
QA476_2
QAptm_2
QA188_2


monomers
Weighted Penalty
Z-score


homo_oligomers : Weighted Penalties
Penalty Score
QA319_2
QA345_2
QA314_2
QA122_2
QA050_2
QA425_2
QA051_2
QA331_2
QA273_2
QA148_2
QA393_2
QA443_2
QAmean_plddt_2
QA110_2
QAptm_2
QAiptm+ptm_2
QAiptm_2
QA312_2
QA264_2
QA029_2
QA188_2
QA212_2
QA476_2
QA164_2


homo_oligomers
Weighted Penalty
Z-score


hetero_oligomers : Weighted Penalties
Penalty Score
Group Rank
QAiptm_2
QAiptm+ptm_2
QA476_2
QA264_2
QA312_2
QA110_2
QA148_2
QAptm_2
QAmean_plddt_2
QA188_2
QA050_2
QA393_2
QA425_2
QA164_2
QA051_2
QA314_2
QA443_2
QA345_2
QA273_2
QA122_2
QA319_2
QA331_2
QA029_2
QA212_2

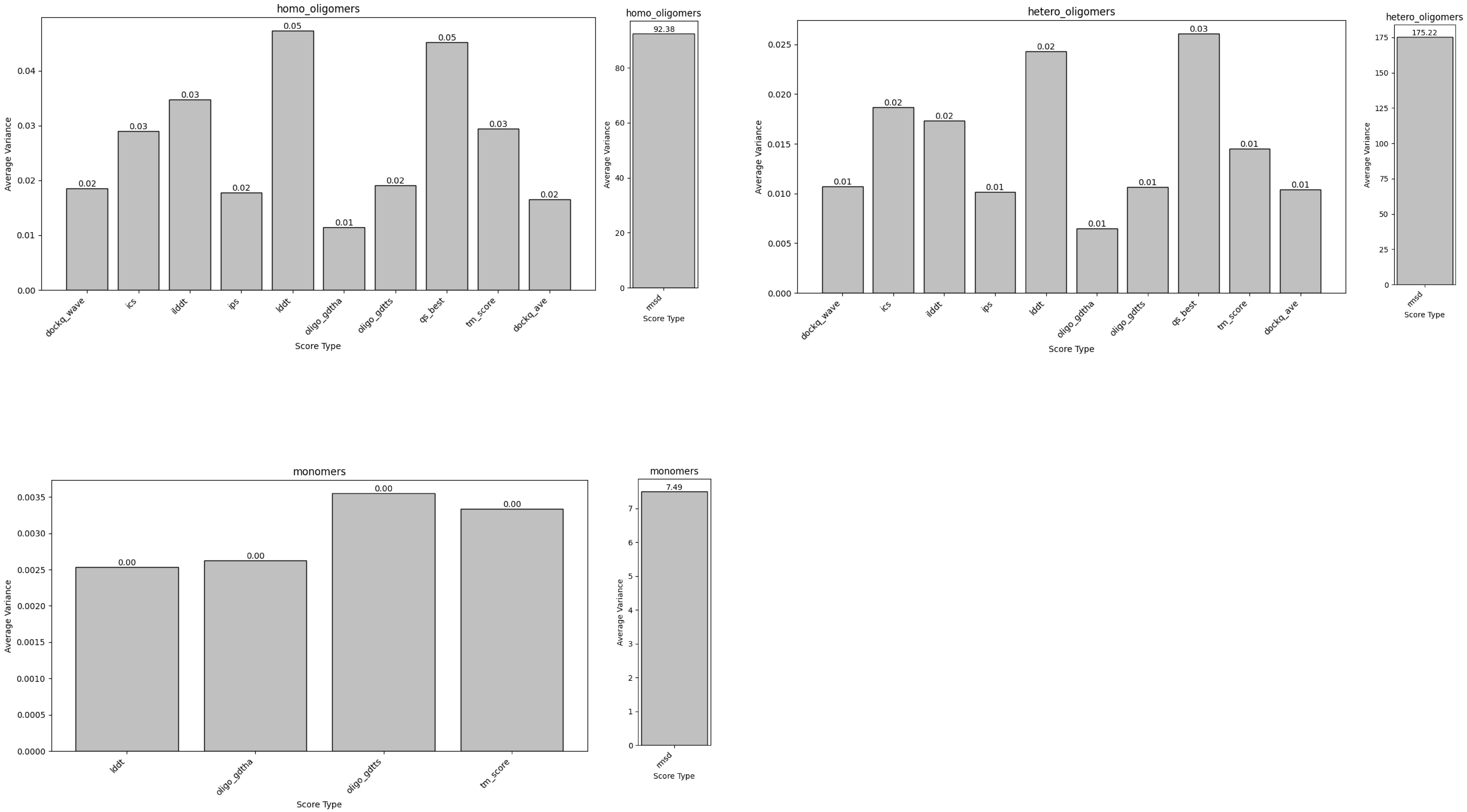

hetero_oligomers
Weighted Penalty
Z-score

Including MF baseline + min(Z)=0 + Z filter at -3

Targets missing:

H1272: 9 components, too big and complicated to be done in MassiveFold

T1247: structure was released early and target was cancelled


homo_oligomers
Average Variance
Score Type
dockq_wave
ics
ilddt
ips
lddt
oligo_gdtha
oligo_gdtts
qs_best
tm_score
dockq_ave
0.02
0.03
0.03
0.02
0.05
0.01
0.02
0.05
0.03
0.02
homo_oligomers
rmsd
92.38
hetero_oligomers
Average Variance
Score Type
dockq_wave
ics
ilddt
ips
lddt
oligo_gdtha
oligo_gdtts
qs_best
tm_score
dockq_ave
0.01
0.02
0.02
0.01
0.02
0.01
0.01
0.03
0.01
0.01
hetero_oligomers
rmsd
175.22
monomers
Average Variance
Score Type
lddt
oligo_gdtha
oligo_gdtts
tm_score
0.00
0.00
0.00
0.00
monomers
rmsd
7.49