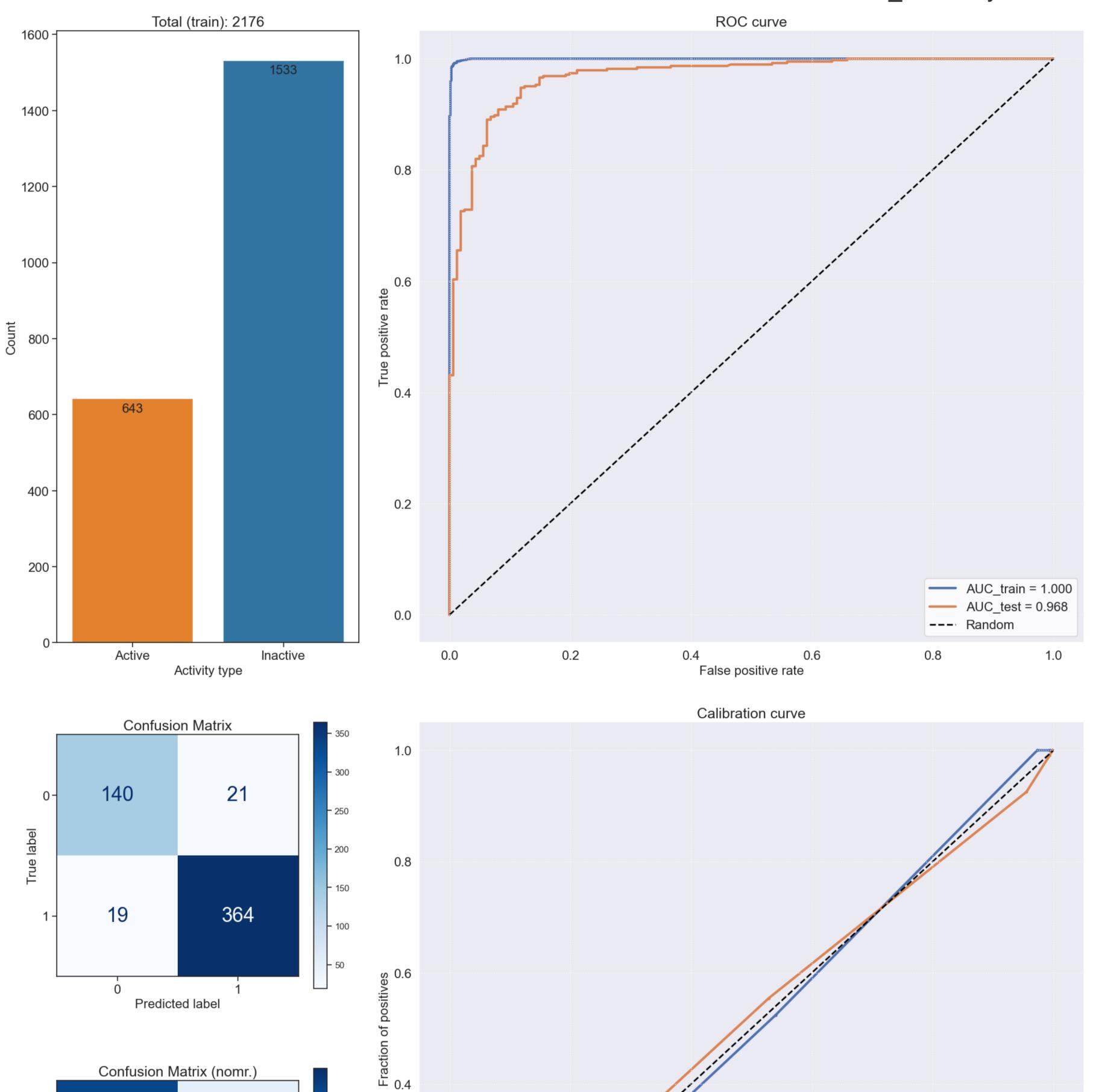
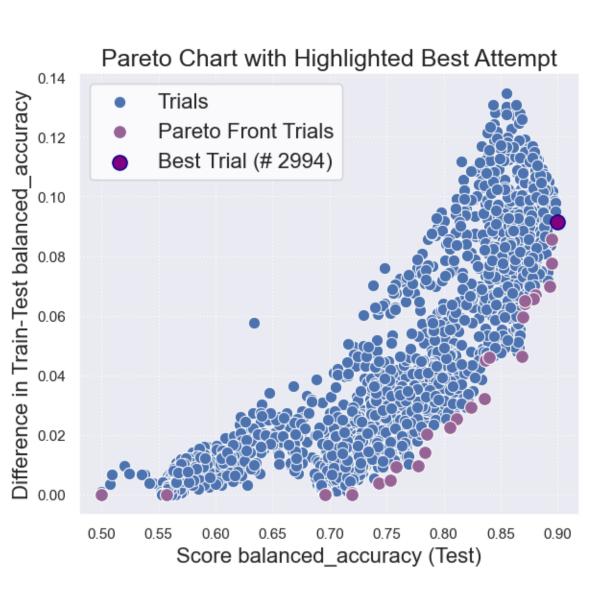
## Protein: Q9UBN7 - metric: balanced\_accuracy





XGBoost Parameters	Value
alpha	2.914
colsample_bytree	0.9302
gamma	0.0043
lambda	0.4714
learning_rate	0.1296
max_depth	10.0
min_child_weight	2.0
n_estimators	149.0
reg_alpha	0.6297
reg_lambda	0.4995
scale_pos_weight	1.0707
subsample	0.9656

Active/Inactive probability								
						type active (383) inactive (161) decoy (541)		
	<del></del>							
.0 0.2	0.		0.6 ility	0.	8	1.0		
		Inactive zone	Inactive zone  0 0.2 0.4	Inactive zone	Inactive zone Active z  0 0.2 0.4 0.6 0.	Inactive zone Active zone  0 0.2 0.4 0.6 0.8		

0.2

0.0

0.0

0.2

0.2

Confusion Matrix (nomr.)

Predicted label

0.13

0.95

0.87

0.05

0-

True label

Metric	Train	Test
AUC	0.9996	0.9684
accuracy	0.9917	0.9265
balanced accuracy	0.9892	0.91
recall_(sens)	0.9954	0.9504
specificity	0.9829	0.8696
precision	0.9928	0.9455
f1_score	0.9917	0.9263

--- train: 0.01

0.8

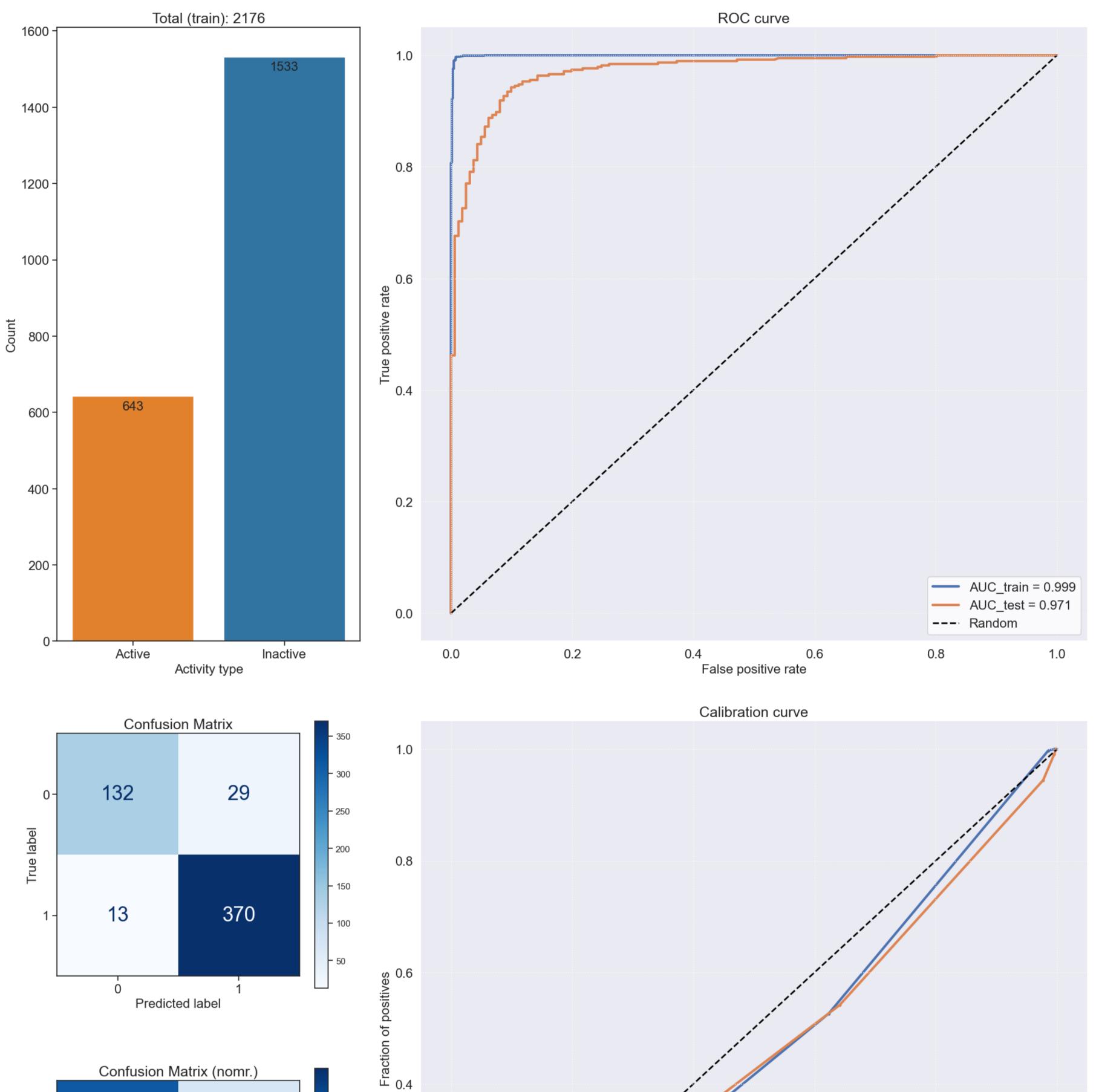
0.4 0.6 Mean predicted probability

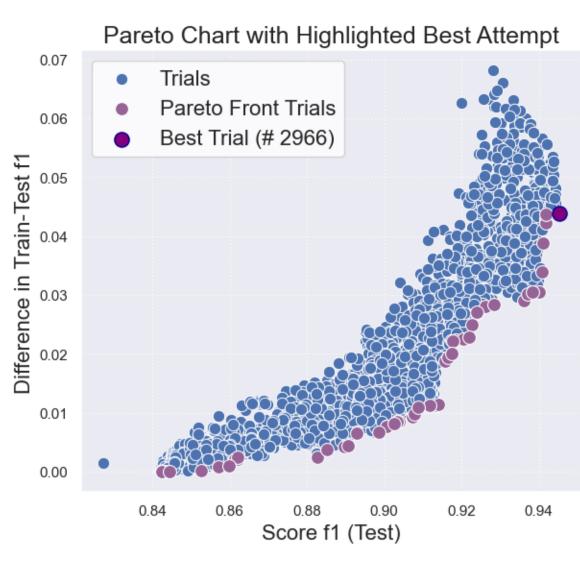
test: 0.0579

--- perfectly\_calibrated

1.0

## Protein: Q9UBN7 - metric: f1





XGBoost Parameters	Value
alpha	1.324
colsample_bytree	0.6052
gamma	0.7014
lambda	4.1559
learning_rate	0.0491
max_depth	9.0
min_child_weight	1.0
n_estimators	845.0
reg_alpha	0.2428
reg_lambda	0.361
scale_pos_weight	2.8884
subsample	0.9041

		Act	tive/Inactive	probability			
60	Ina	active zone			Active zo	inactiv	(383) /e (161)
50						decoy	(541)
40			į				
Percent							
20							
10							
0 —	0.0 0	.2 0.		0.6	0.8		1.0
			Probab	ility			

0.2

0.0

0.0

0.4

0.2

Confusion Matrix (nomr.)

Predicted label

0.18

0.97

0.82

0.034

0-

True label

Metric	Train	Test
AUC	0.9995	0.9715
accuracy	0.9844	0.9228
balanced accuracy	0.974	0.893
recall_(sens)	0.9993	0.9661
specificity	0.9487	0.8199
precision	0.9789	0.9273
f1_score	0.9843	0.9216

--- train: 0.013

0.8

test: 0.0575

--- perfectly\_calibrated

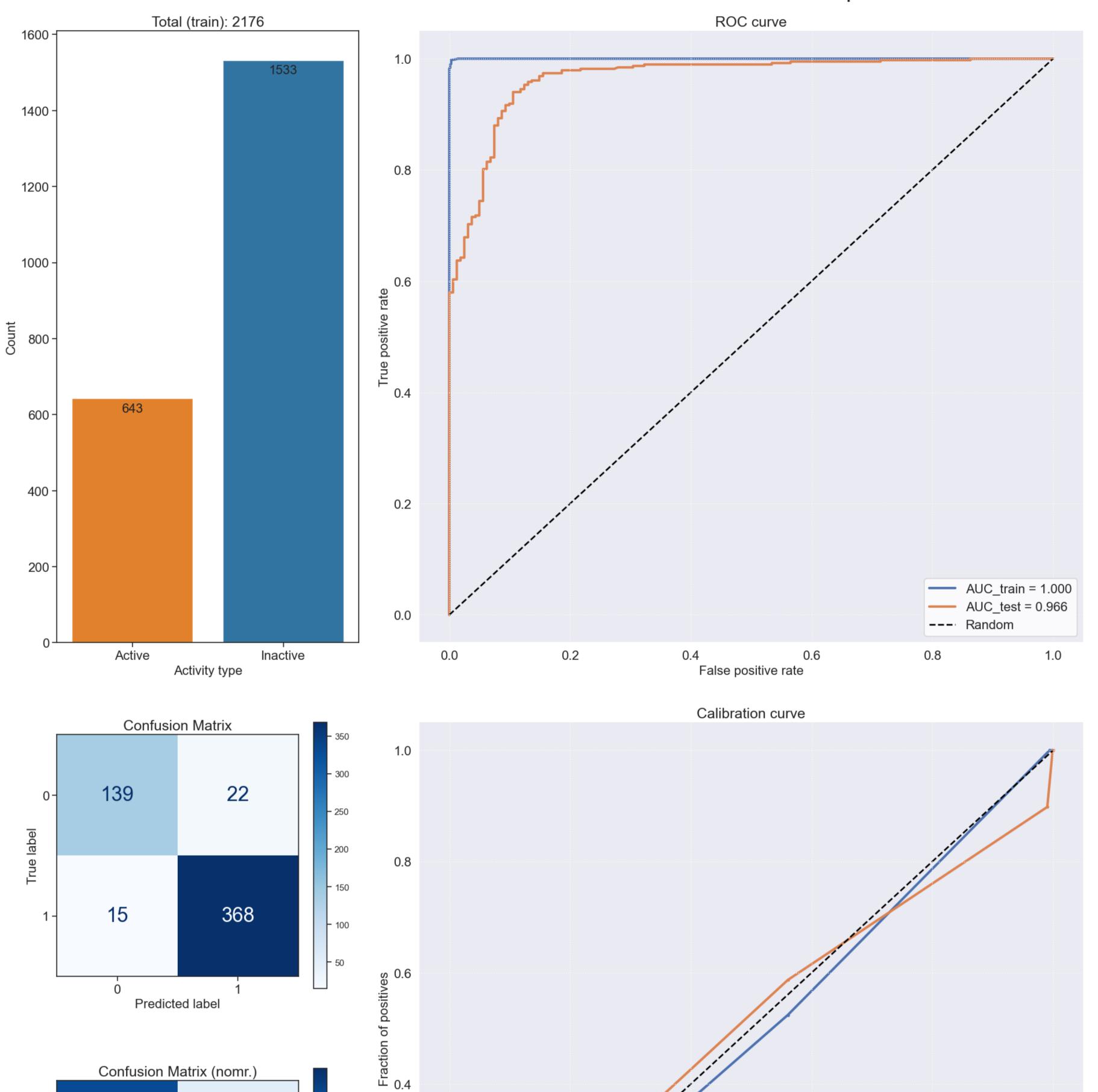
1.0

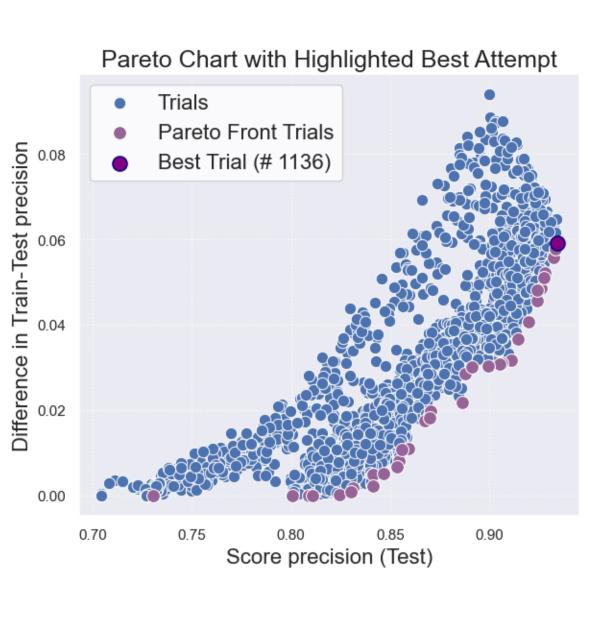
	Active/Inactive	probability				
60	Inactive zone	Active zone	type active (383)	Metric	Train	
50			inactive (161) decoy (541)	AUC	0.9995	
				accuracy	0.9844	
40				balanced accuracy	0.974	
Percent 				recall_(sens)	0.9993	
20				specificity	0.9487	
10				precision	0.9789	
0				f1_score	0.9843	
0.	l l	0.6 0.8	1.0			

0.2

0.4 0.6 Mean predicted probability

## Protein: Q9UBN7 - metric: precision





Value
1.7412
0.9218
0.2111
4.6234
0.4241
10.0
3.0
242.0
0.0051
0.0771
2.2102
0.8909

			Active/Inactive	e probability		
70	<b>I</b>	nactive zone			Active zone	active (383) inactive (161)
60						decoy (541)
50						
Percent						
30						
20						
0-						
	0.0	0.2	0.4 Probal	0.6 bility	0.8	1.0

0.2

0.0

0.0

0.2

0.4

0.2

Confusion Matrix (nomr.)

Predicted label

0.14

0.96

0.86

0.039

0-

True label

Metric	Train	Test
AUC	0.9999	0.9658
accuracy	0.9959	0.932
balanced accuracy	0.993	0.9121
recall_(sens)	1.0	0.9608
specificity	0.986	0.8634
precision	0.9942	0.9436
f1_score	0.9959	0.9315

--- train: 0.0041

0.8

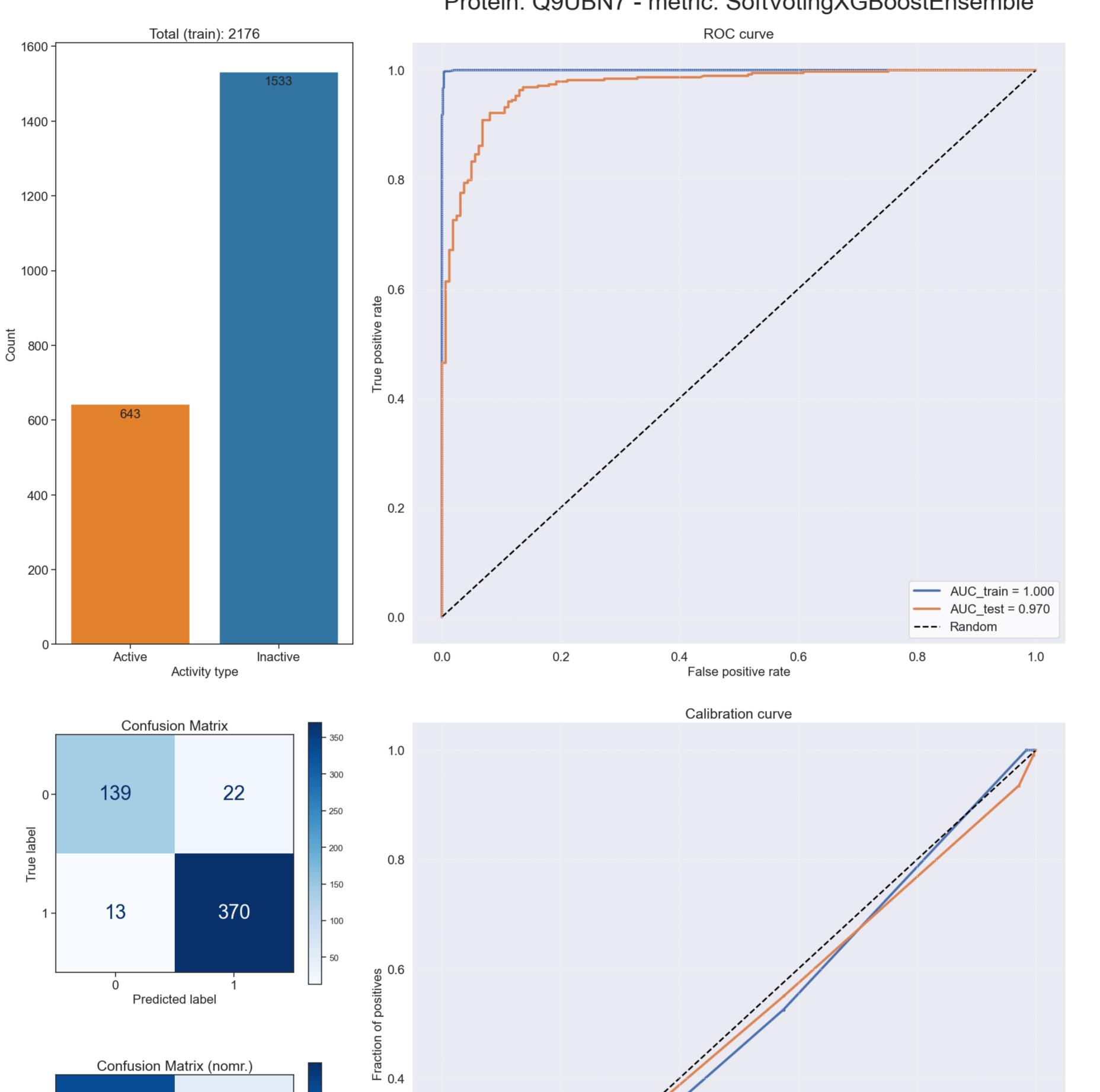
0.4 0.6 Mean predicted probability

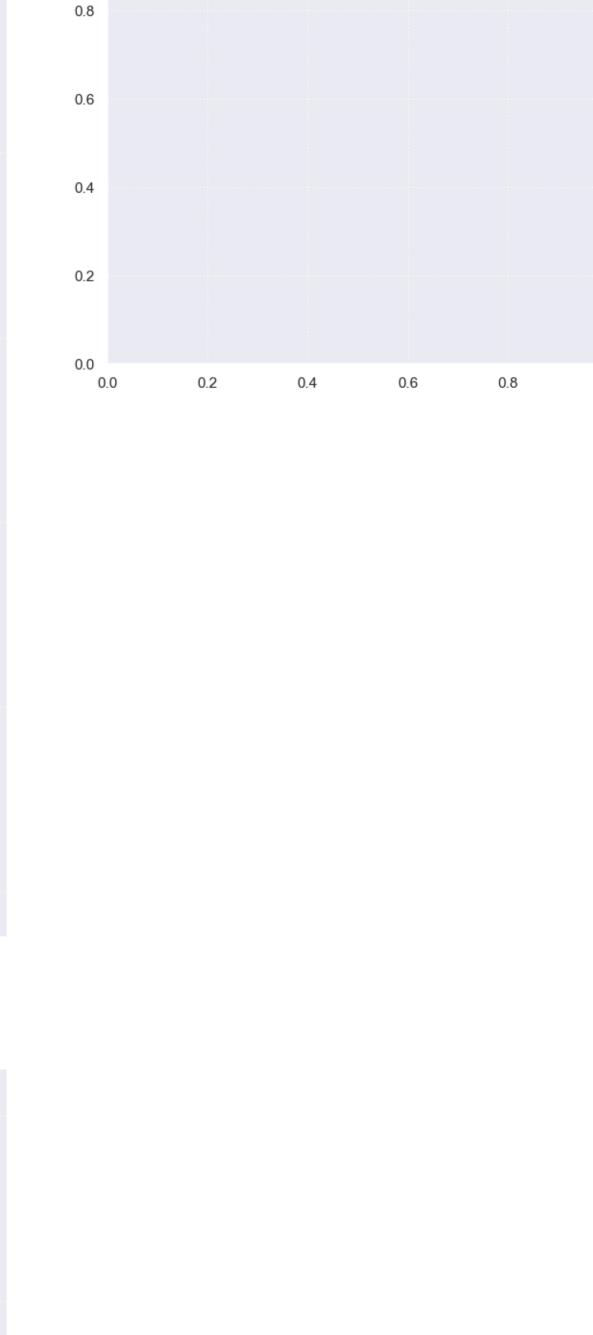
test: 0.057

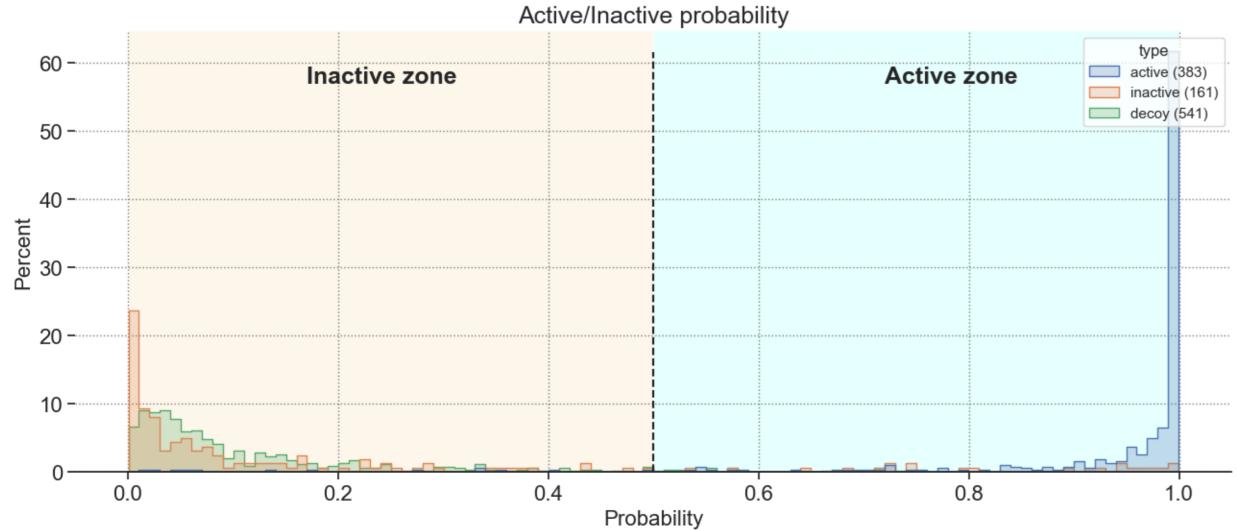
--- perfectly\_calibrated

1.0

## Protein: Q9UBN7 - metric: SoftVotingXGBoostEnsemble







0.86

0.034

Predicted label

0-

True label

0.14

0.97

Metric	Train	Test
AUC	0.9998	0.9702
accuracy	0.994	0.9357
balanced accuracy	0.9903	0.9147
recall_(sens)	0.9993	0.9661
specificity	0.9813	0.8634
precision	0.9922	0.9439
f1_score	0.994	0.9351

train: 0.0079

test: 0.0548

8.0

0.4 0.6 Mean predicted probability

perfectly\_calibrated

1.0

		Active/Inactiv	e probability		
60					type
00	Inactive z	one		Active zone	active (383) inactive (161)
50					decoy (541)
30					
40 -					
40					
ercent					
P 30					
00					
20					
10	<b>7</b> 572_				
				_	
0—	0.0 0.2	0.4	0.6	0.8	1.0
		Proba			

0.2

0.0

0.0

0.2

- 0.4

0.2