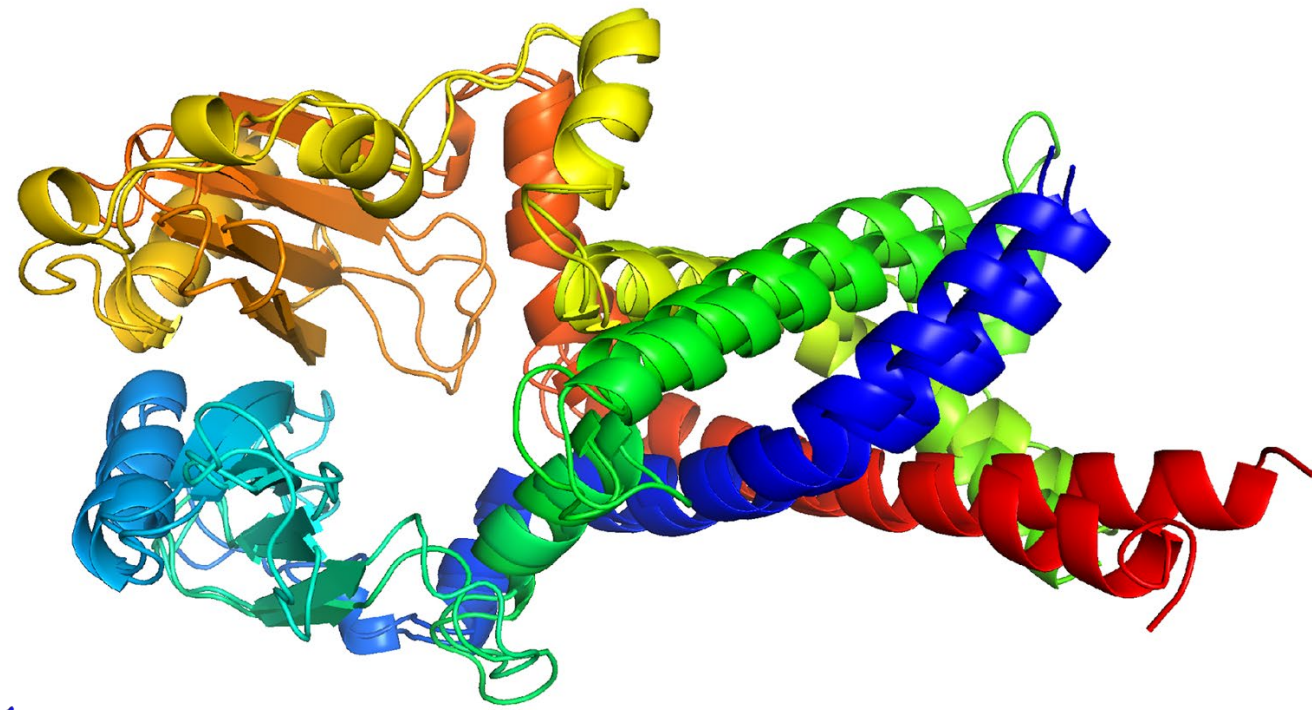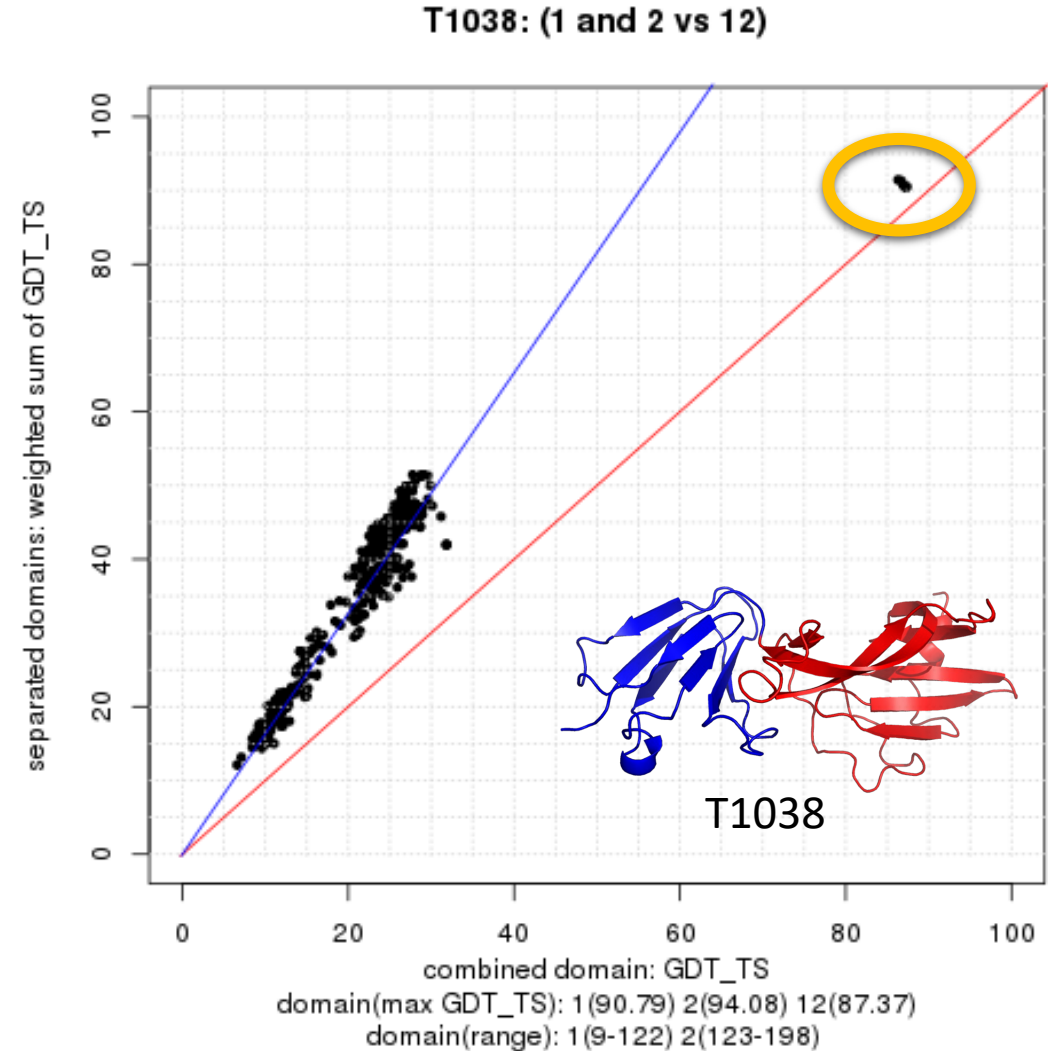**AlphaFold2 (T1094)**

BAKER-Experimental (T1058)

CASP14 : InterDomain Perfomance

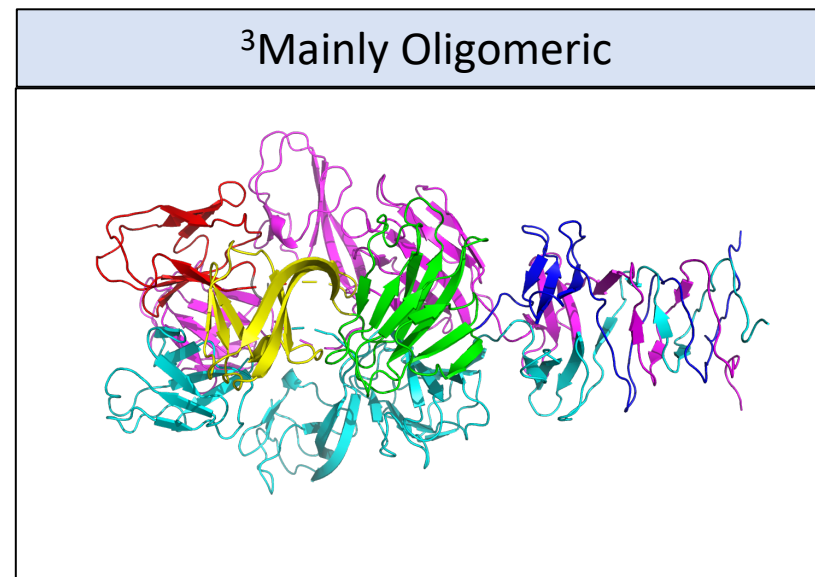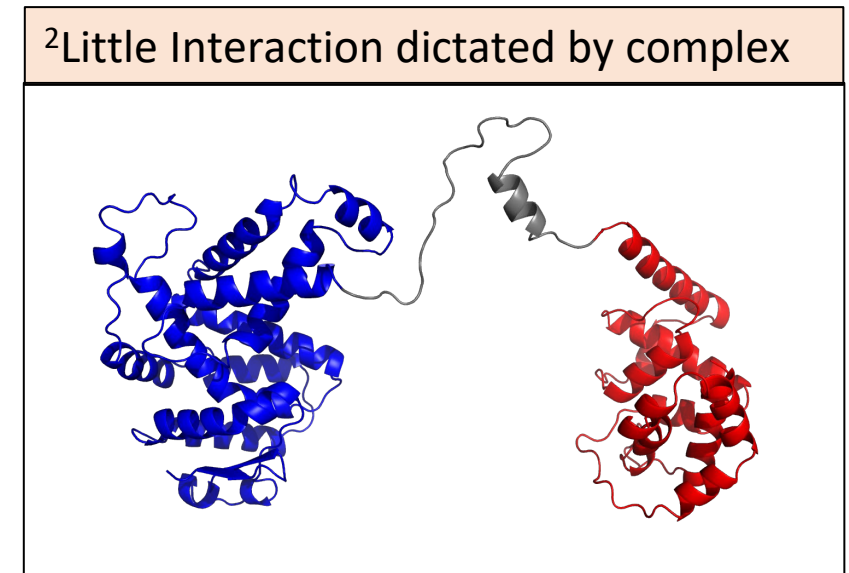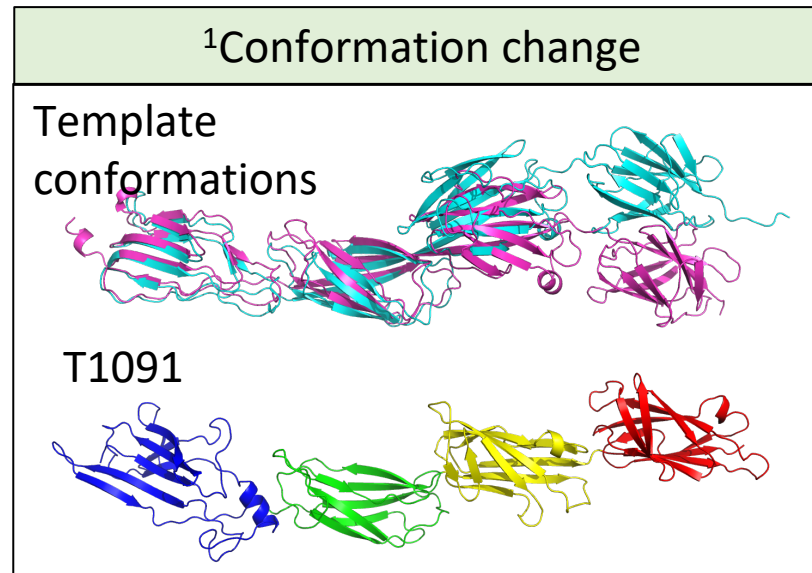R. Dustin Schaeffer, Lisa Kinch, Nick Grishin

Full-length results suggest the future contains fewer EVUs

- Prediction of domain position in multidomain targets was challenging

- EVUs can belong to multiple assessment categories

- Performance in individual assessment categories suggested full-length predictions worthy of independent assessment
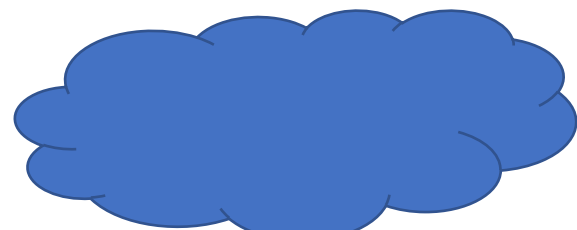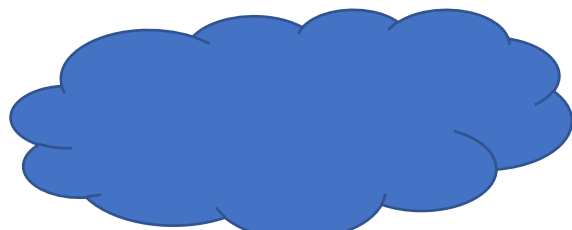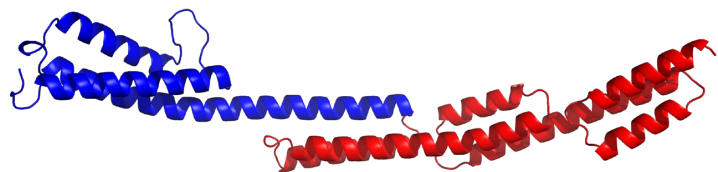

T1038: (1 and 2 vs 12)

combined domain: GDT_TS
domain(max GDT_TS): 1 (90.79) 2(94.08) 12(87.37)
domain(range): 1 (9-122) 2(123-198)

Selection of Domain Interaction Targets

Split Target	Domains	D0
T1024	2	no[1]
T1100	2	no[1]
T1092	2	no[2]
T1096	2	no[2]
T1047s2	3	no[3]
T1050	3	no[1]
T1093	3	no[2]
T1070	4	no[3]
T1091	4	no[1]
T1030	2	yes
T1053	2	yes
T1058	2	yes
T1086	2	yes
T1094	2	yes
T1101	2	yes
T1038	2	yes
T1052	3	yes
T1061	3	yes
T1085	3	yes

[1]Conformation change

Template conformations

T1091

[2]Little Interaction dictated by complex

[3]Mainly Oligomeric

CASP 14 InterDomain targets

T1094

T1101

T1038

T1058

T1053

T1030

T1086

T1052

T1061

T1085
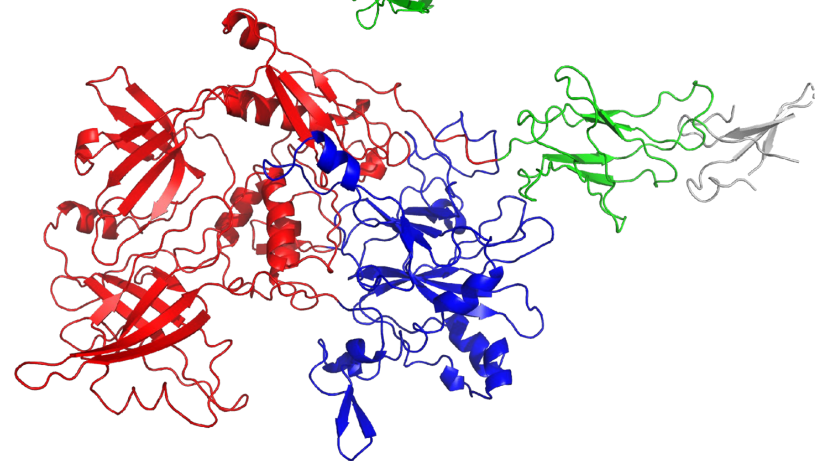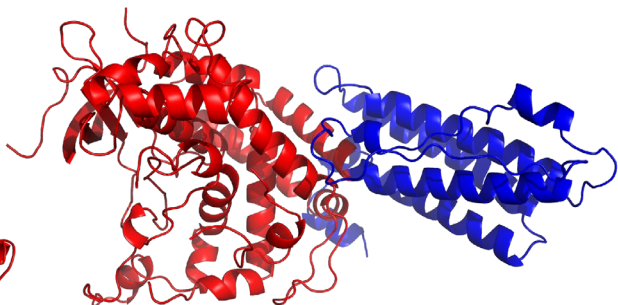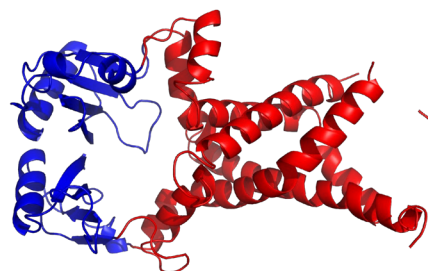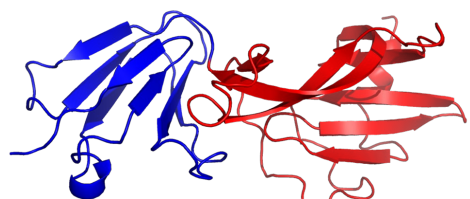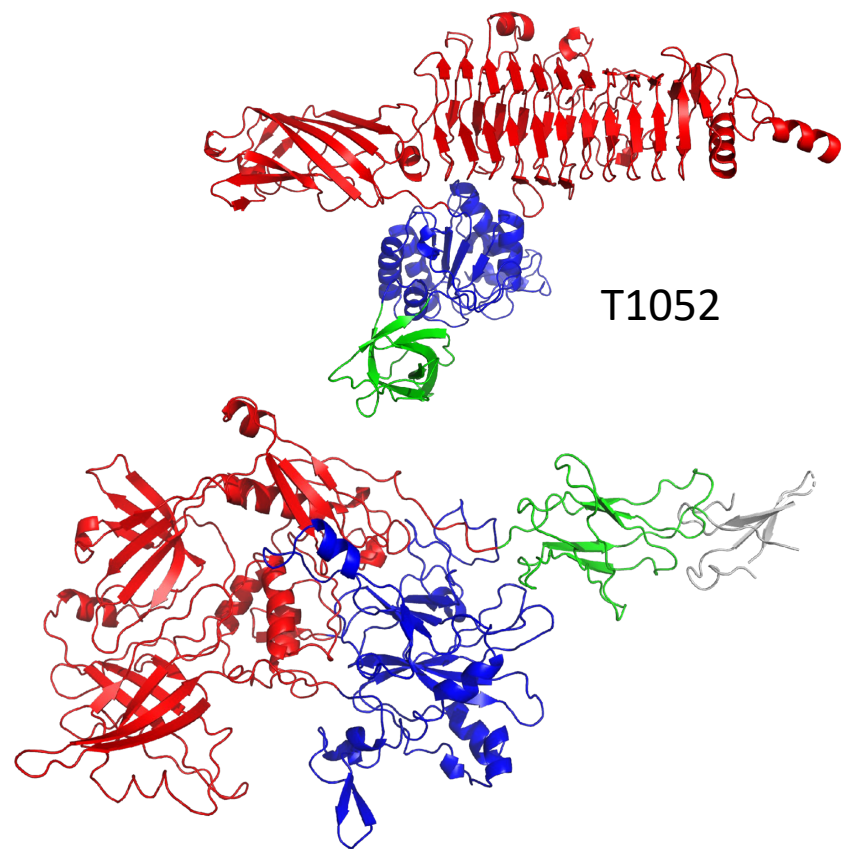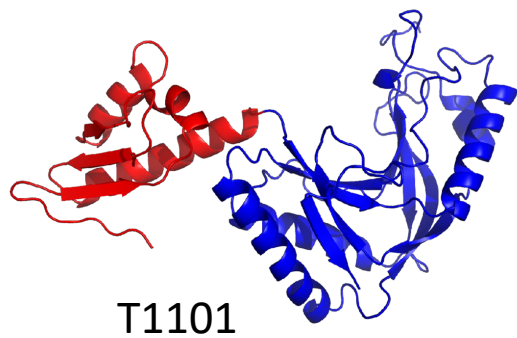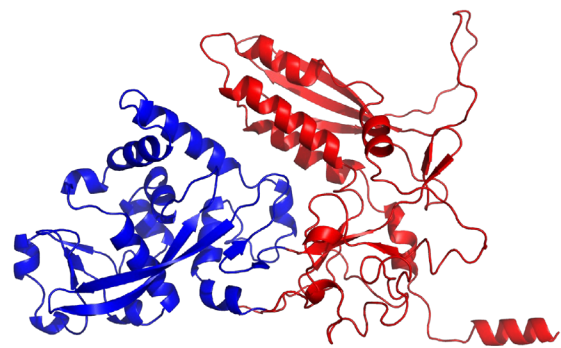
What about T1044?

Very few submitted models / calculated scores
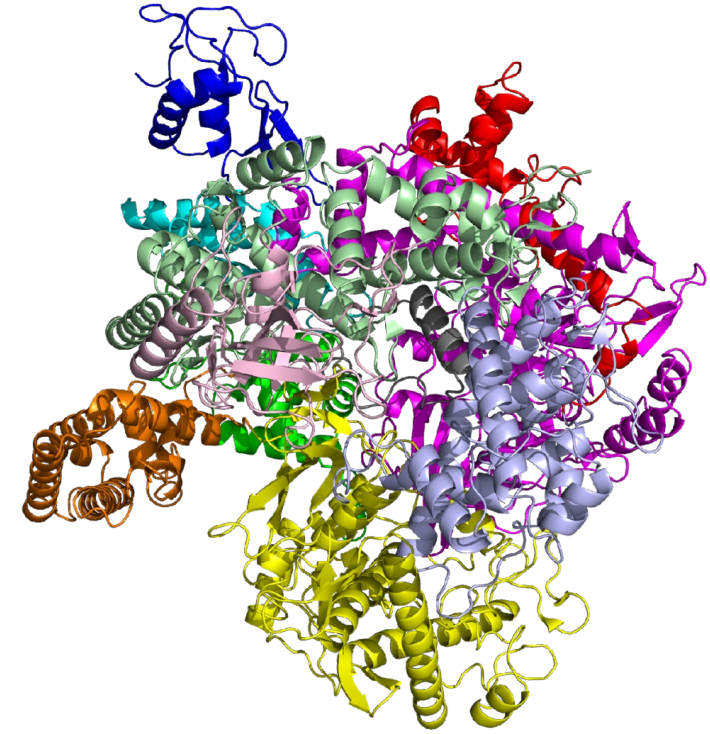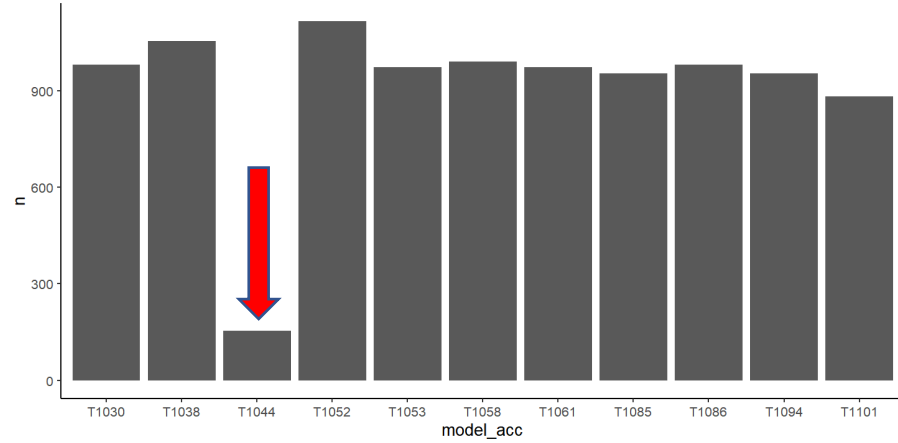
Interdomain scores for submitted models are mostly poor

Interdomain contact statistic for T1044

e.g. F1 score
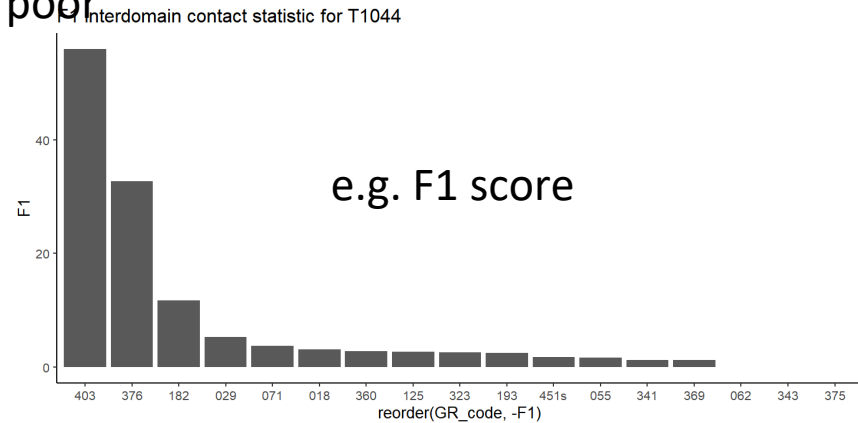
T1044 – 9 EVUs + one previously published region excluded from individual consideration

T1044 was excluded from the PCA+heatmap interdomain analysis due to lack of data

BAKER-Experimental outperforms on T1044

T1044

Baker-Experimental, (403_1)

1281 contacts, **F1 56.0**
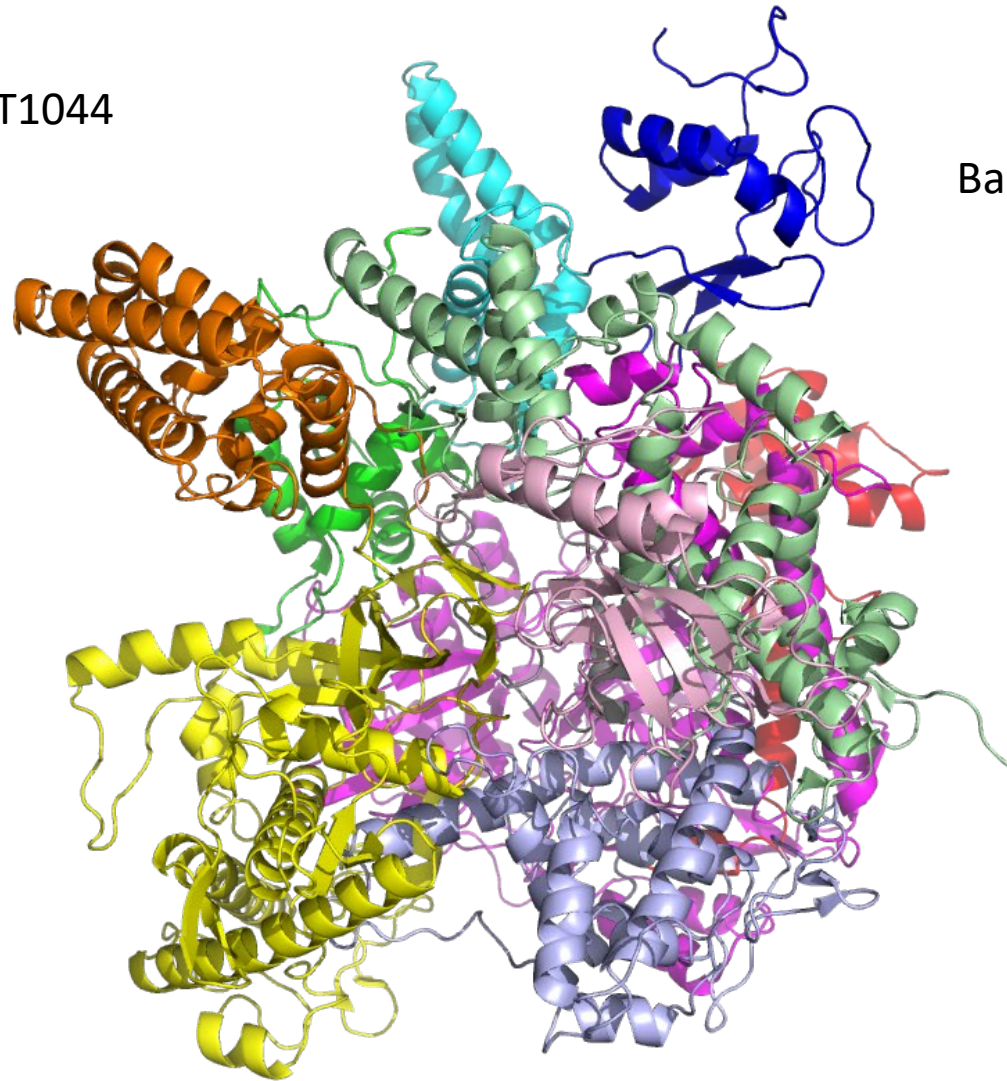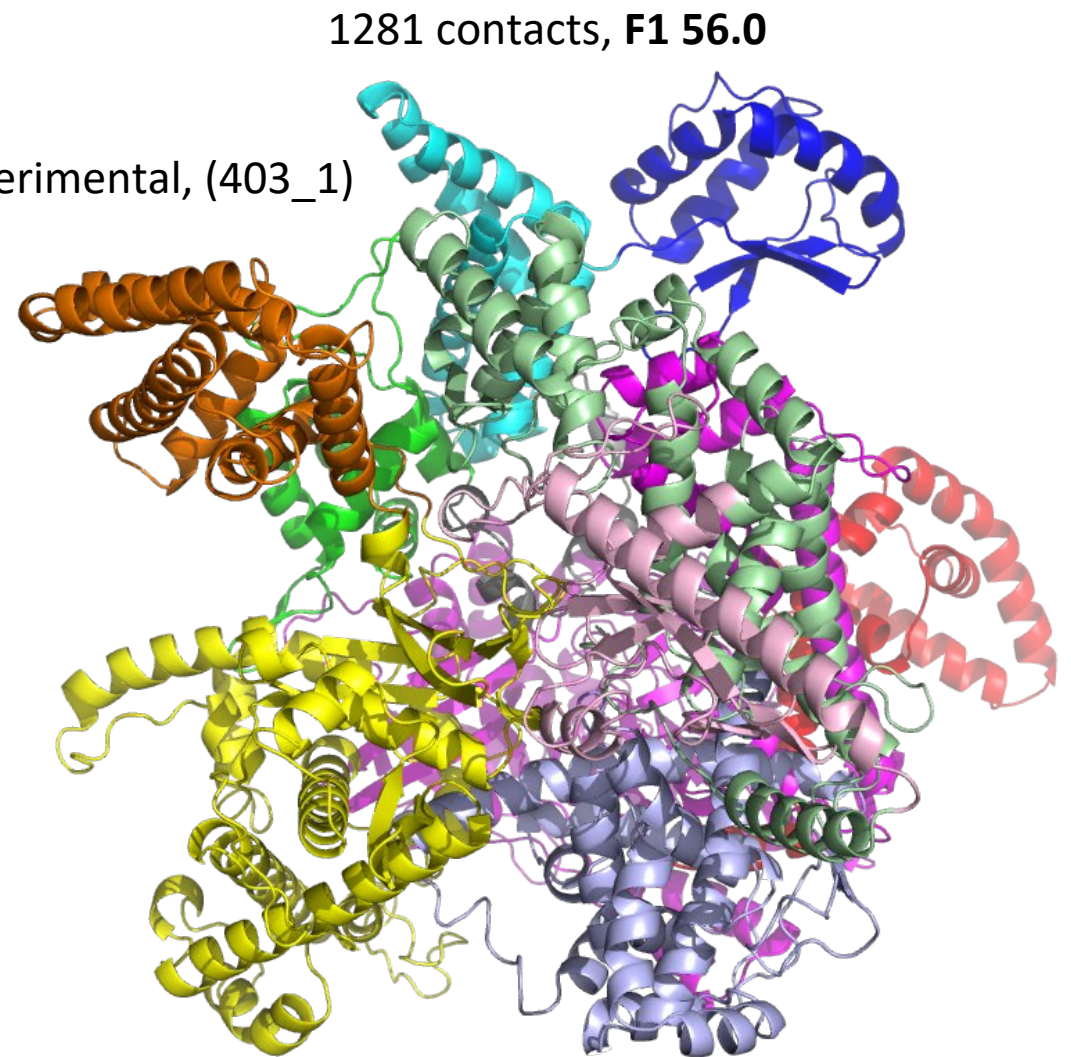
T1044 Morph Movie
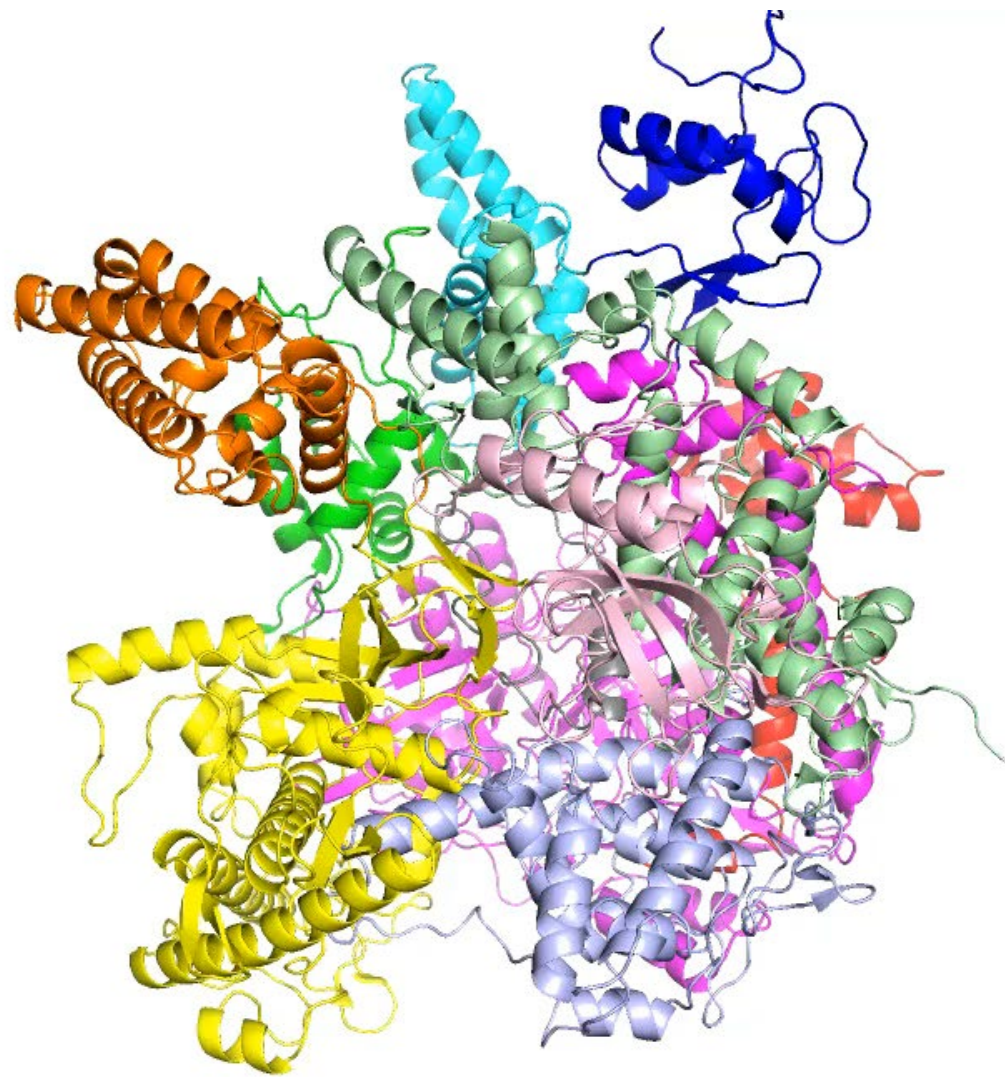
CASP14 interdomain scores repurposed from assembly analysis

Iface-check

Precision - % of correct interdomain contacts over total model interdomain contacts

Recall - % of correct interdomain contacts over total native interdomain contacts

Jacc. Coefficient – Shows the similarity of model and target interfaces given the residues participating in interdomain interfaces in the model

F1 – Harmonic mean of the precision and recall

PMID: 29071742

QS

QS (Contact Agreement Score) – Fraction of correctly modeled interface contacts over the maximum of either **correct (target)** or **predicted (model)** interface contacts

Global.RMSD – RMSD over all domains based on the lowest RMSD domain matching

Iface.RMSD – RMSD of the superposition based on the alignment of interface residues

PMID: 28874689

Chose 3 prediction Center contact scores for overall interdomain ranking

Interdomain Top Performance Similar to Domain Category

Rank Sum(>0) for top20 InterDomain groups
weights F1/JC/Qsb = 1

- Analysis on model 1
- Top Group: 427
- Top Server: 209
- Top5: 427,403,420,473,339
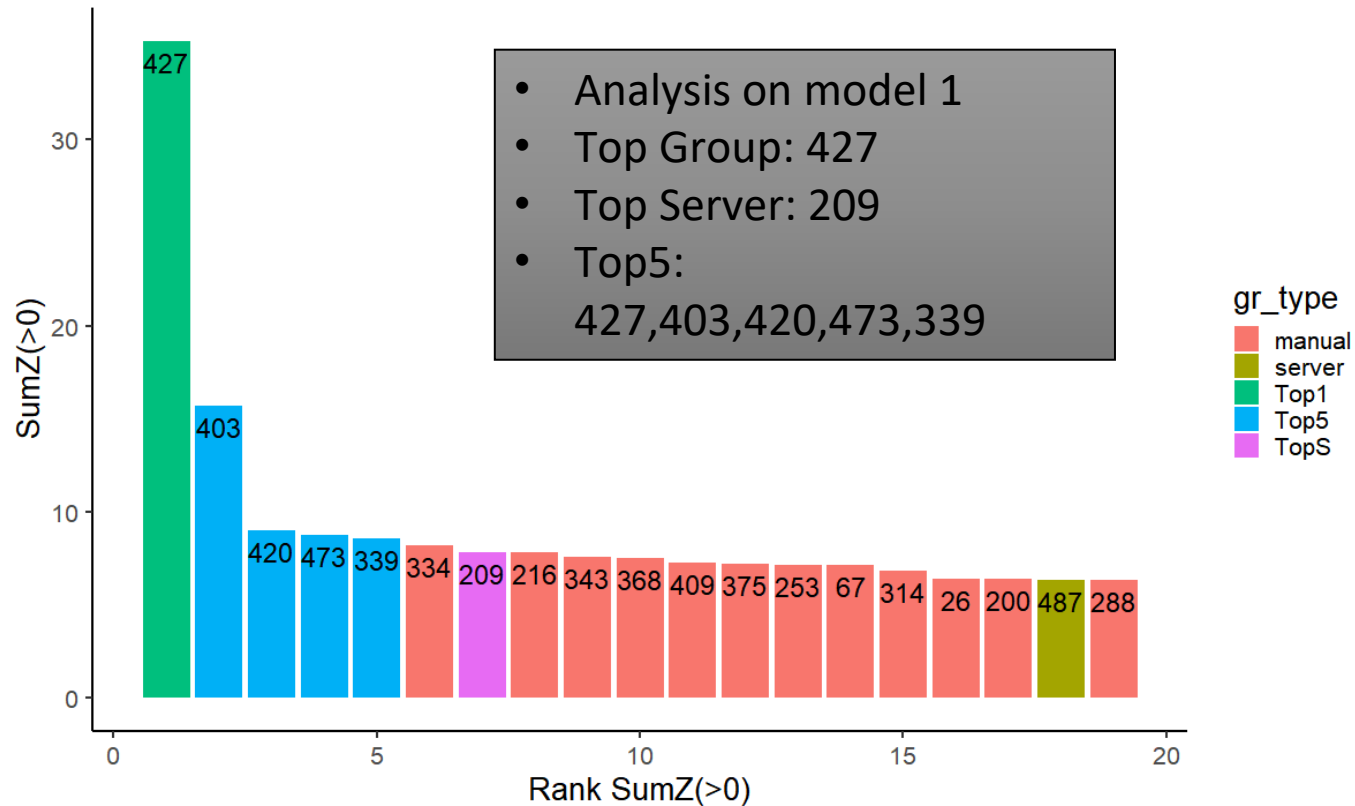
Gr. #	Group	SumZ(>0)
427	AlphaFold2	35.30
403	BAKER-experimental	15.71
420	MultiCom	8.98
473	BAKER	8.75
339	ProQ3D	8.54
334	FEIG-R3	8.17
209	BAKER-ROSETTASERVER	7.84

What is the sensitivity of these rankings to different parameters?

Comparison of ranking schemes show ranks of top groups insensitive to chosen scores/weights/sum, some sensitivity to model selection...

RSumZ(>0) vs RSumZ(>-2) for 1st models, chosen weights

RSumZ(>0) for 1st vs. Best models, chosen weights

427	AlphaFold2
403	BAKER-experimental
420	MultiCom
473	BAKER
339	ProQ3D
209	BAKER-ROSETTASERVER
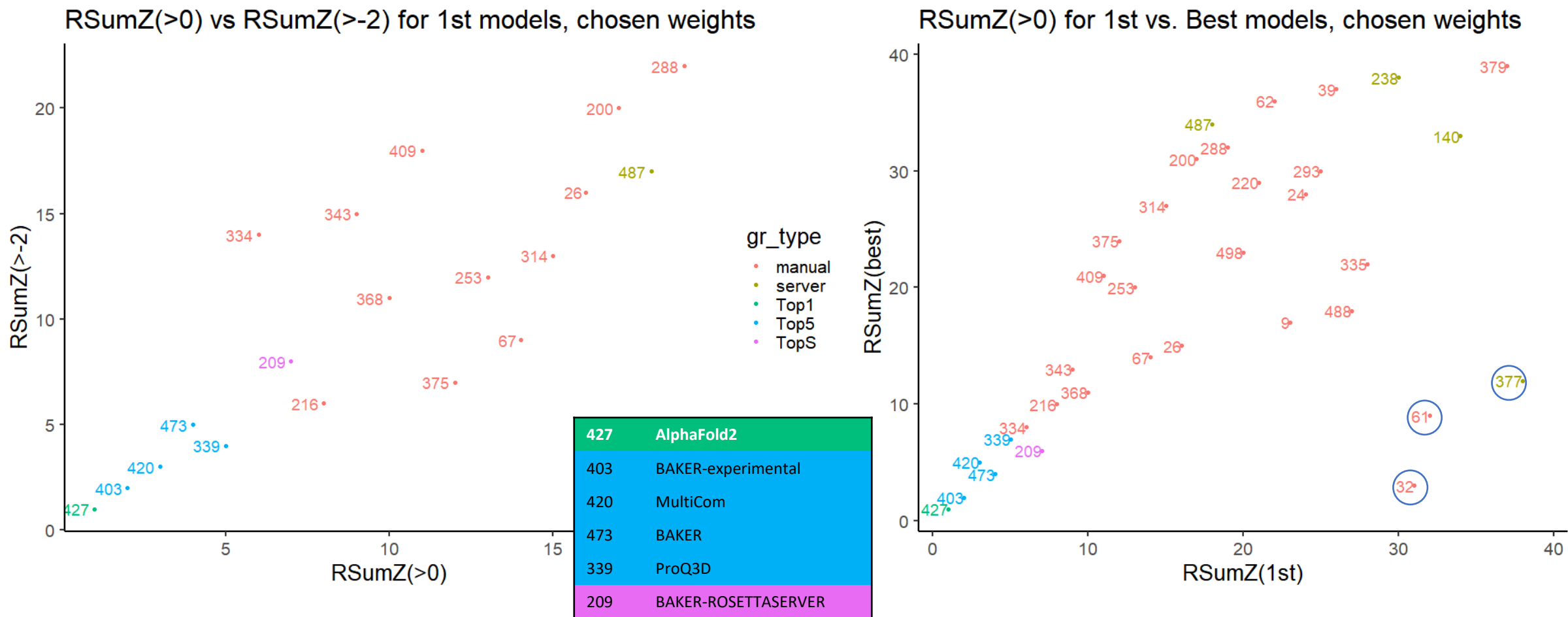
What methods did you use?

- PCA
  - pcaMethods implementation of NIPALS PCA
- Heatmap clustering
  - pheatmap() R implementation
- Repurposed interchain assembly scores for interdomain analysis
  - QS
  - Iface-check

RStudio
Version 1.1.463 – © 2009-2018 RStudio, Inc.

Interdomain Scores – Filter Missing Values

- Selected model: 1st

- Scores: Jacc.Coeff., F1, Qsb100

- Manually scaled to 0-100

- **99/135 groups were considered**

Number of non-NA scores per Group for InterDomain Targets
10 targets * 3 scores

9-10 targets submitted

count(Non-NA scores)

Group

Selected groups compared by heatmap
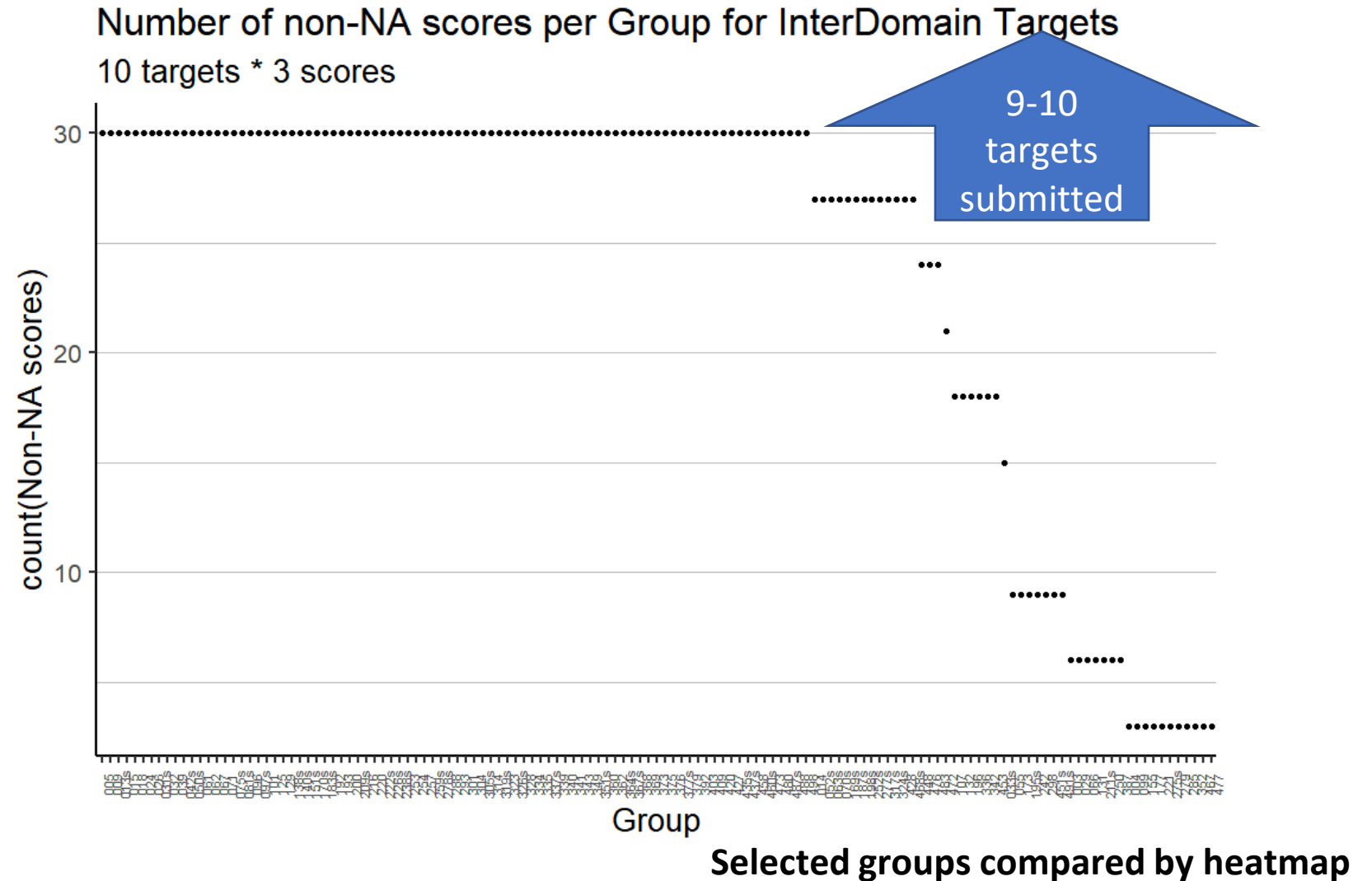
Contact Z-scores for Interdomain Targets for Selected Groups

Branches weighted by perfomance
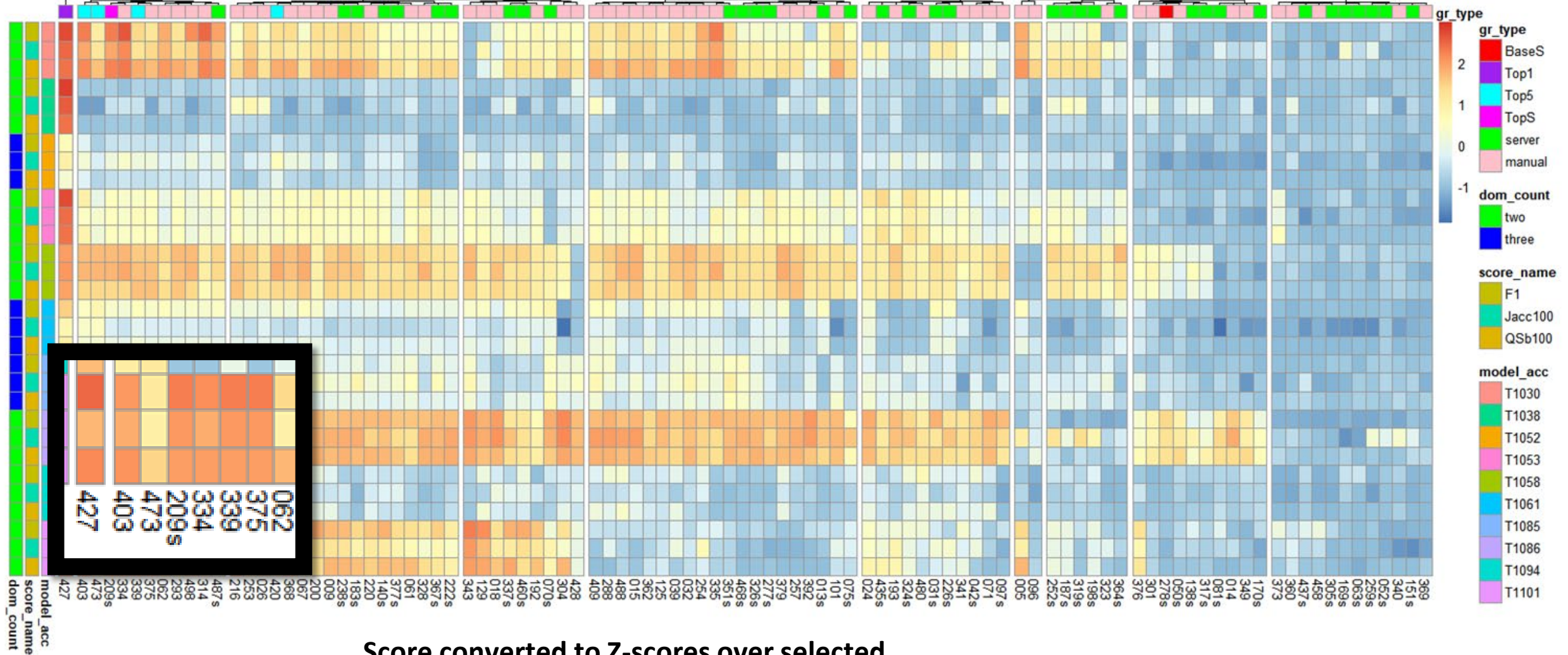
Score converted to Z-scores over selected groups/scores/models. 1.3% missing data imputed

Contact Z-scores for Interdomain Targets for Selected Groups

Scaled data over selected groups/scores/models

Contact Z-scores for Interdomain Targets for Selected Groups
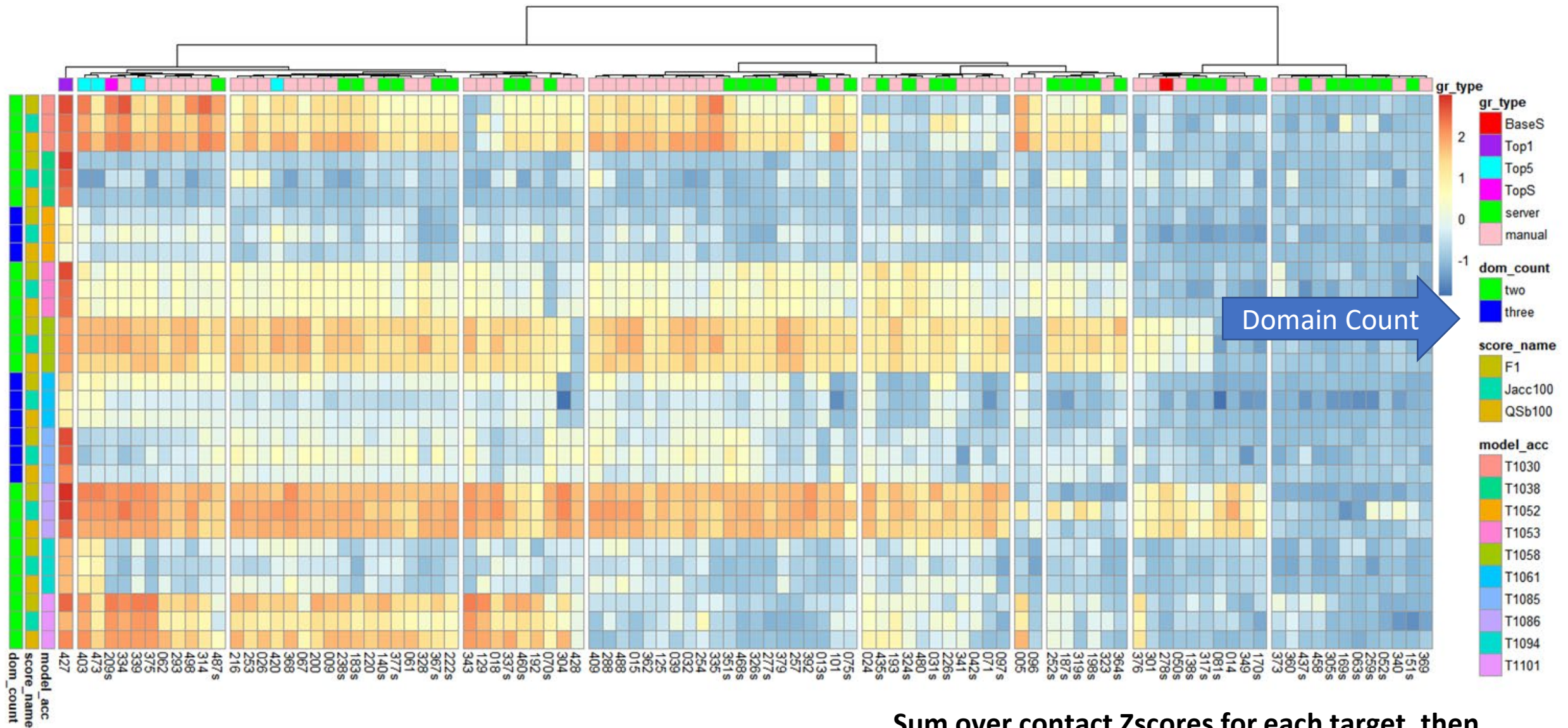
Sum over contact Zscores for each target, then cluster by target

SumZ of selected contact scores cluster groups by well-predicted domain interfaces
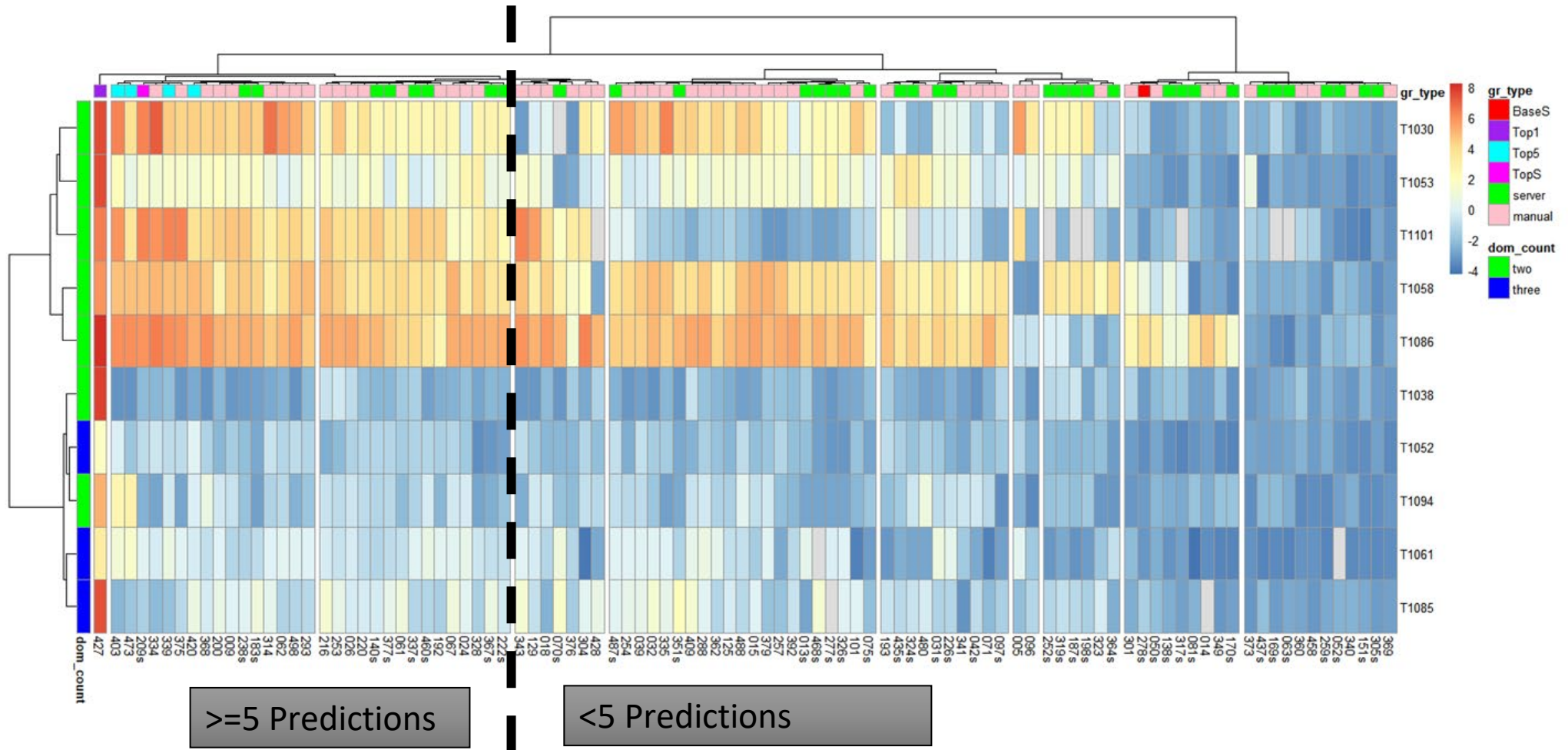
SumZ of selected contact scores shows clustering by target domain count

SumZ of selected contacts scores

Targets in which AlphaFold2, Baker groups outperformed on InterDomain targets
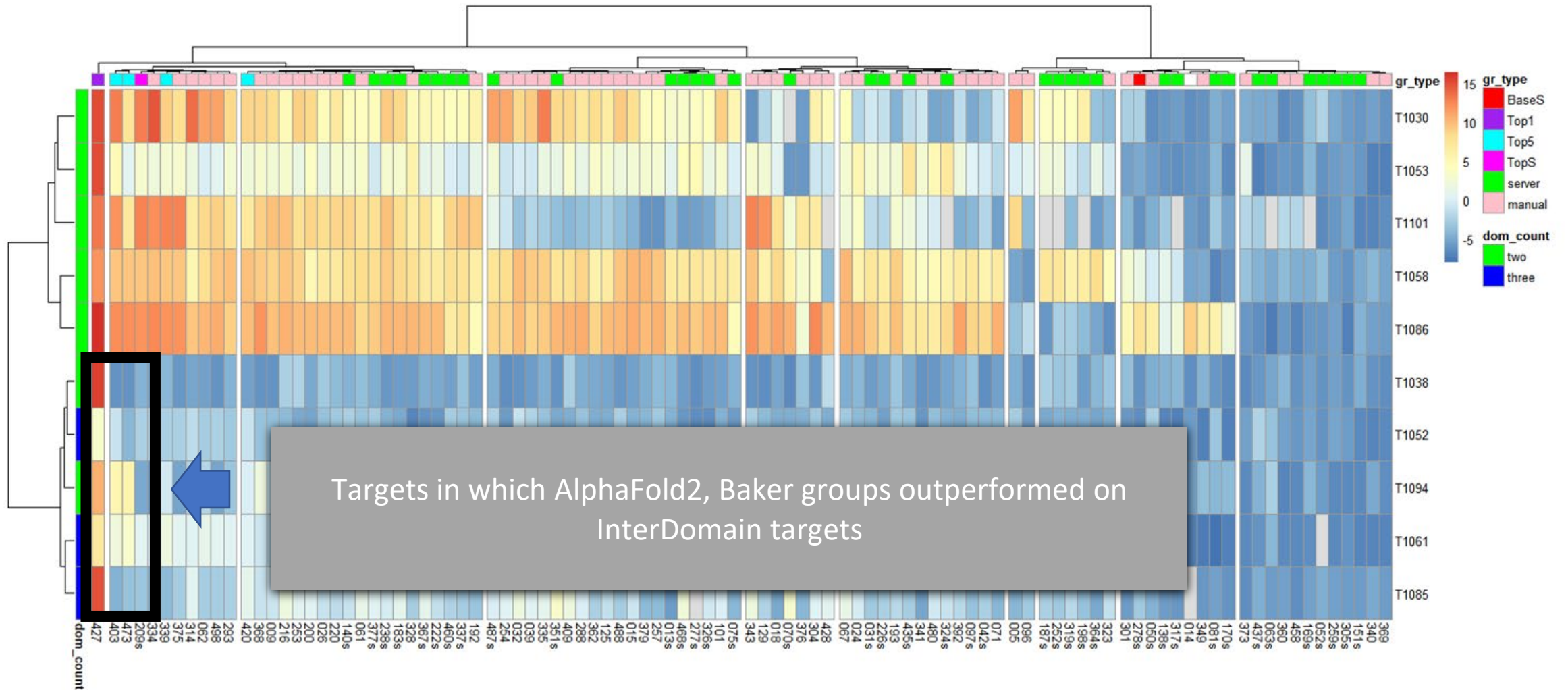
T1094: Two domains that look like 3

AlphaFold (427), F1 = 68.3
SumZ(contact) = 10.5
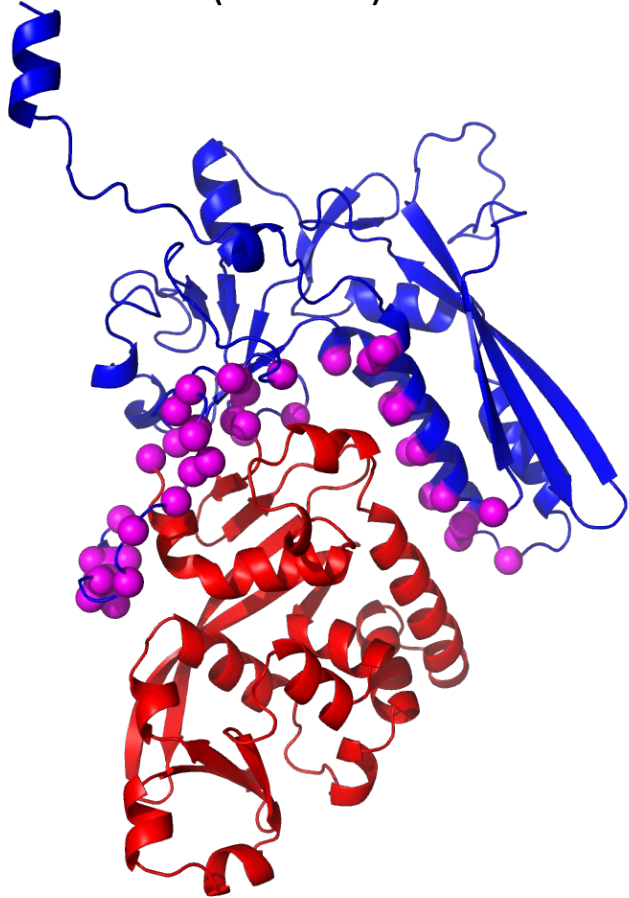
T1094
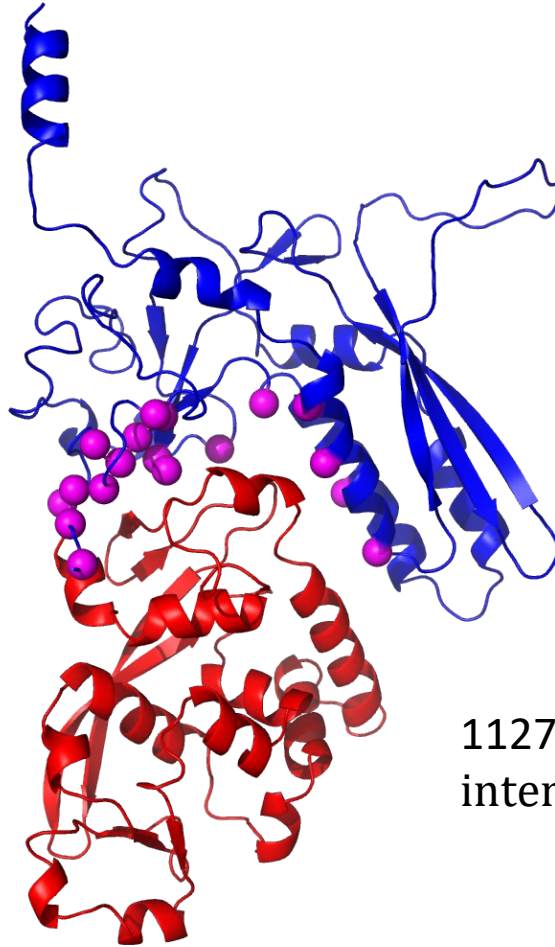
BAKER-Experimental (403), F1= 50
SumZ(contact) = 5.36
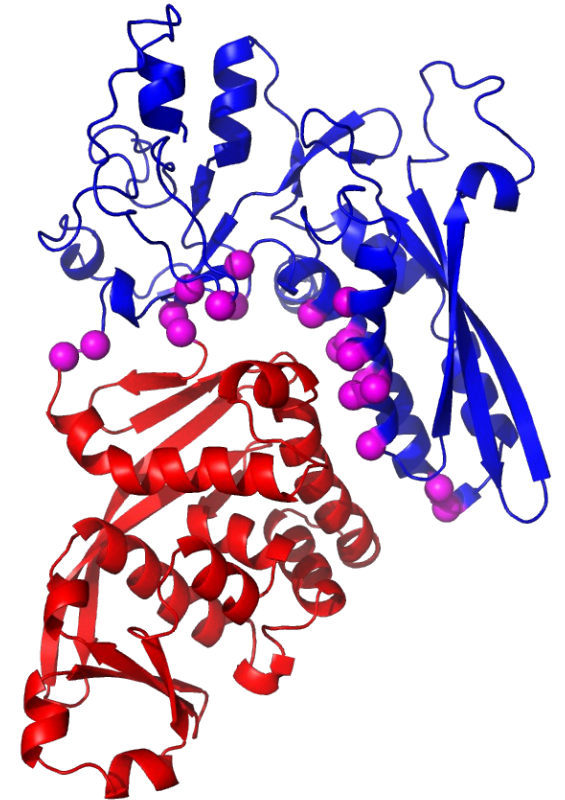
1127.3 Å^2 interface area

PISA PMID: 17681537

Comparison of 2-domain/EVU contact score annotated by interface buried ASA quartile

target	interface_asa (A**2)	quartile
T1101	345.08	first
T1038	733.57	first
T1030	802.13	median
T1086	829.21	median
T1058	990.69	median
T1094	1127.33	third
T0153	1468.53	third
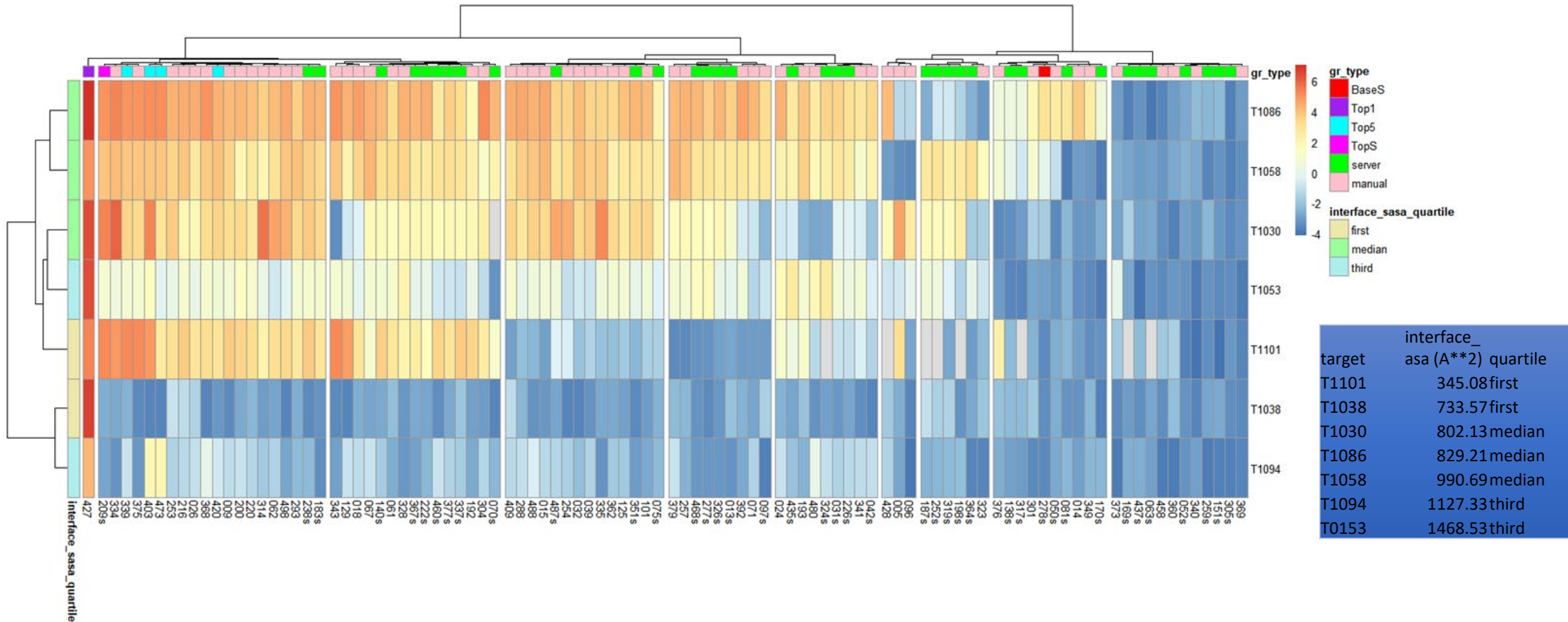
T1038: Interdomain interactions in the presence of multimeric interaction

943.9 Å^2 inter-chain interface area

733.5 Å^2 Inter-domain interface area

T1038 w/ dimer partner

AlphaFold2 (427_1), SumZ(contact) = 7.77, GDT_TS = 86.7

BAKER-Experimental (403_1), SumZ(contact) = -3.29, GDT_TS = 26.4
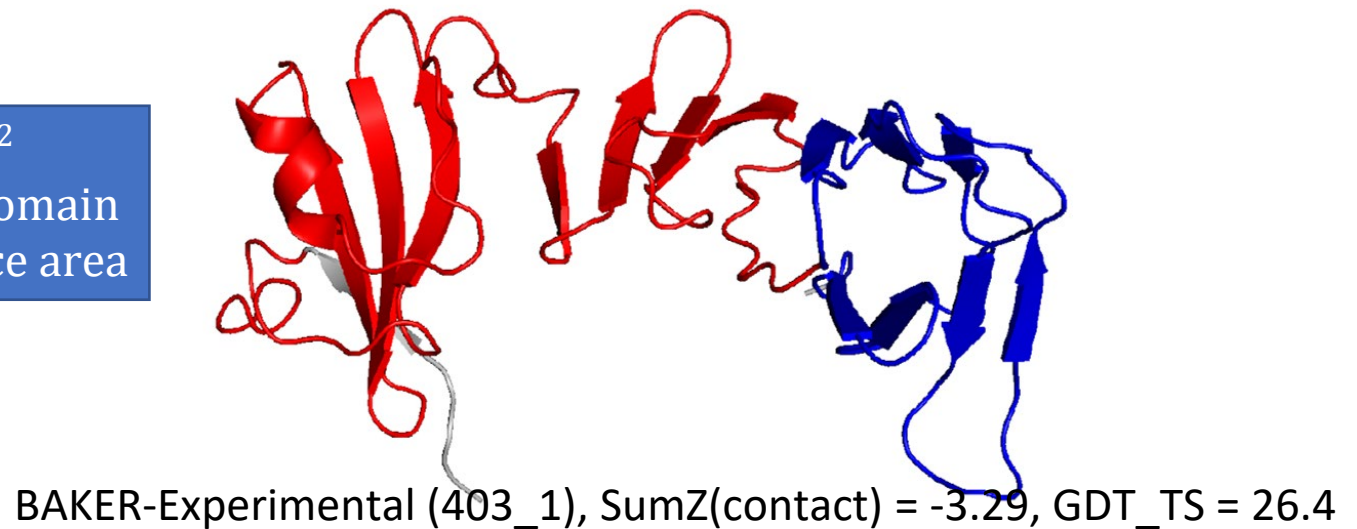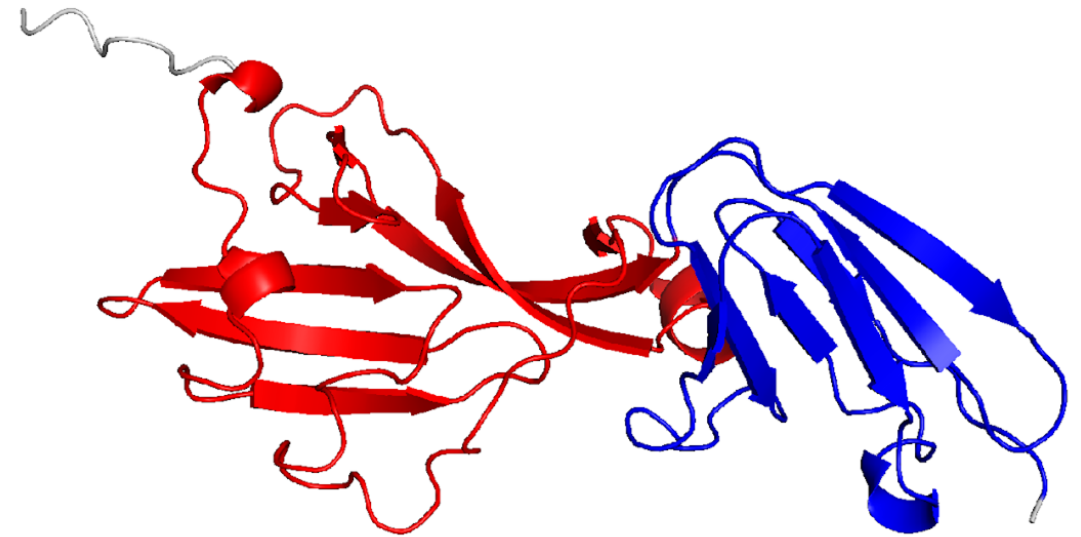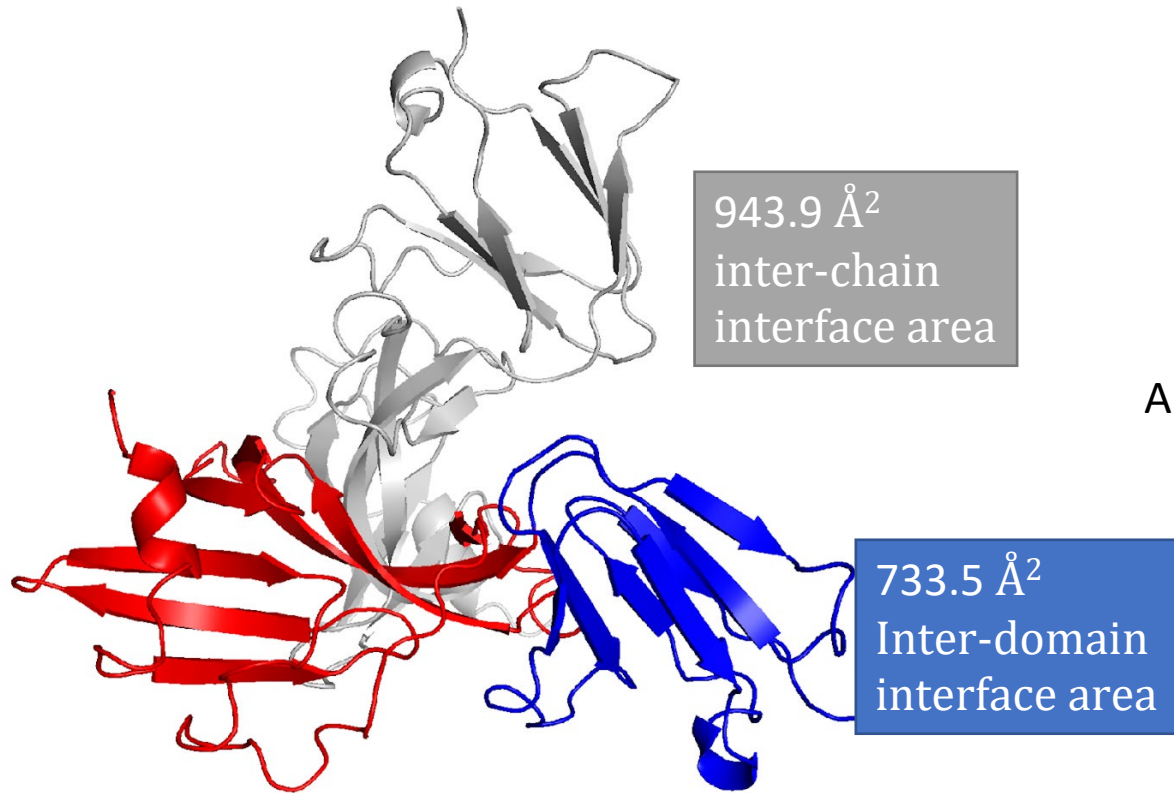
PCA – InterDomain target / scores (w427)

- Scaling = prescaled raw contact scores
- Centered
- NIPALS imputation (1.3 % missing data)
- 30 variables (scores*models)
- 99 samples (groups)
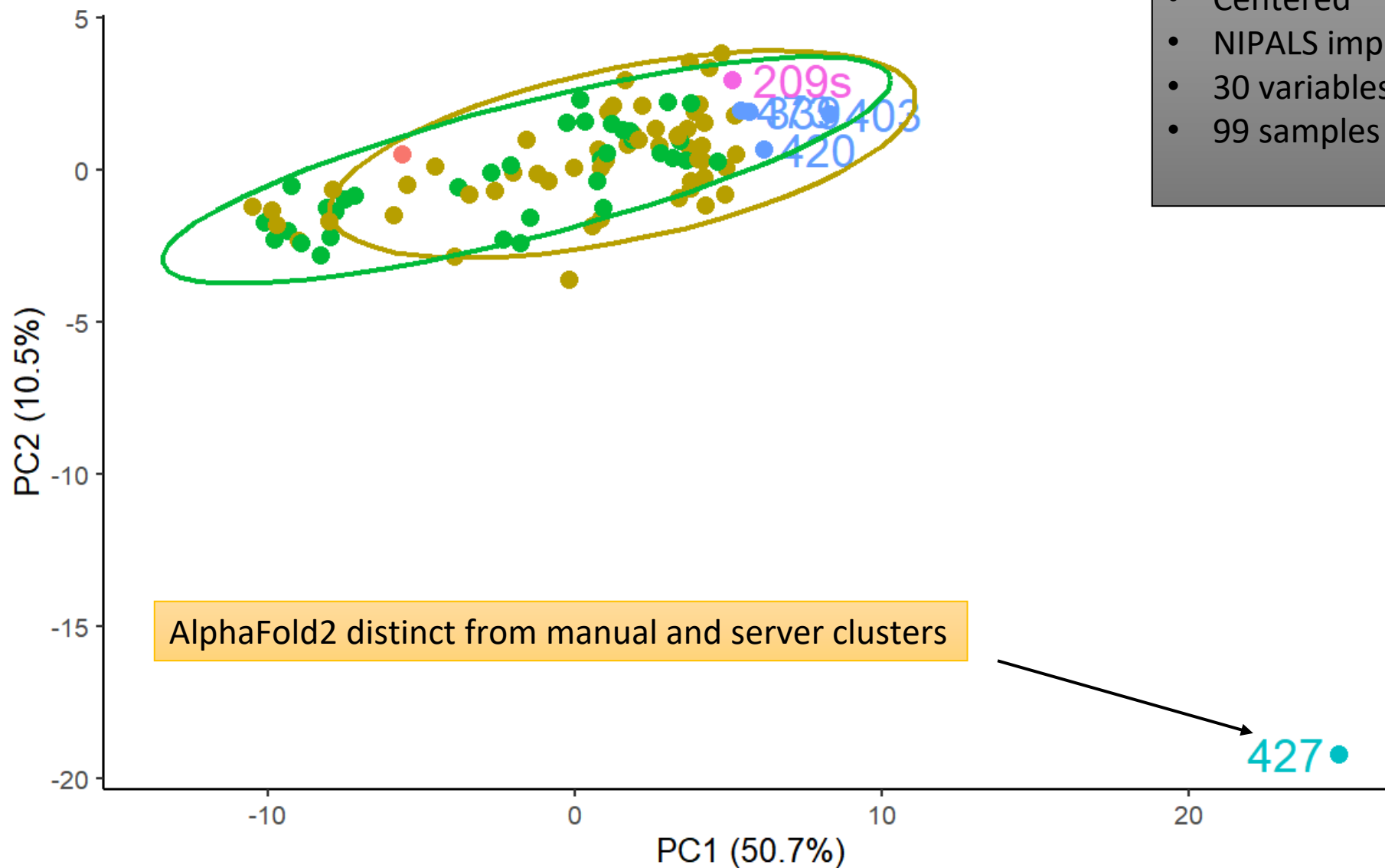
gr_type
- BaseS
- manual
- server
- Top1
- Top5
- TopS

AlphaFold2 distinct from manual and server clusters

PCA – Interdomain targets / contact scores (no427)

JC/F1/Qsb InterDomain targets

Baker also distinct from manual cluster removed from PCA

MultiCom and ProQ3D distinct from server cluster removed from PCA

403

473

339

420

209s

gr_type
- BaseS
- manual
- server
- Top5
- TopS

PC2 (10.5%)

PC1 (46.8%)

427	AlphaFold2
403	BAKER-experimental
420	MultiCom
473	BAKER
339	ProQ3D
209	BAKER-ROSETTASERVER
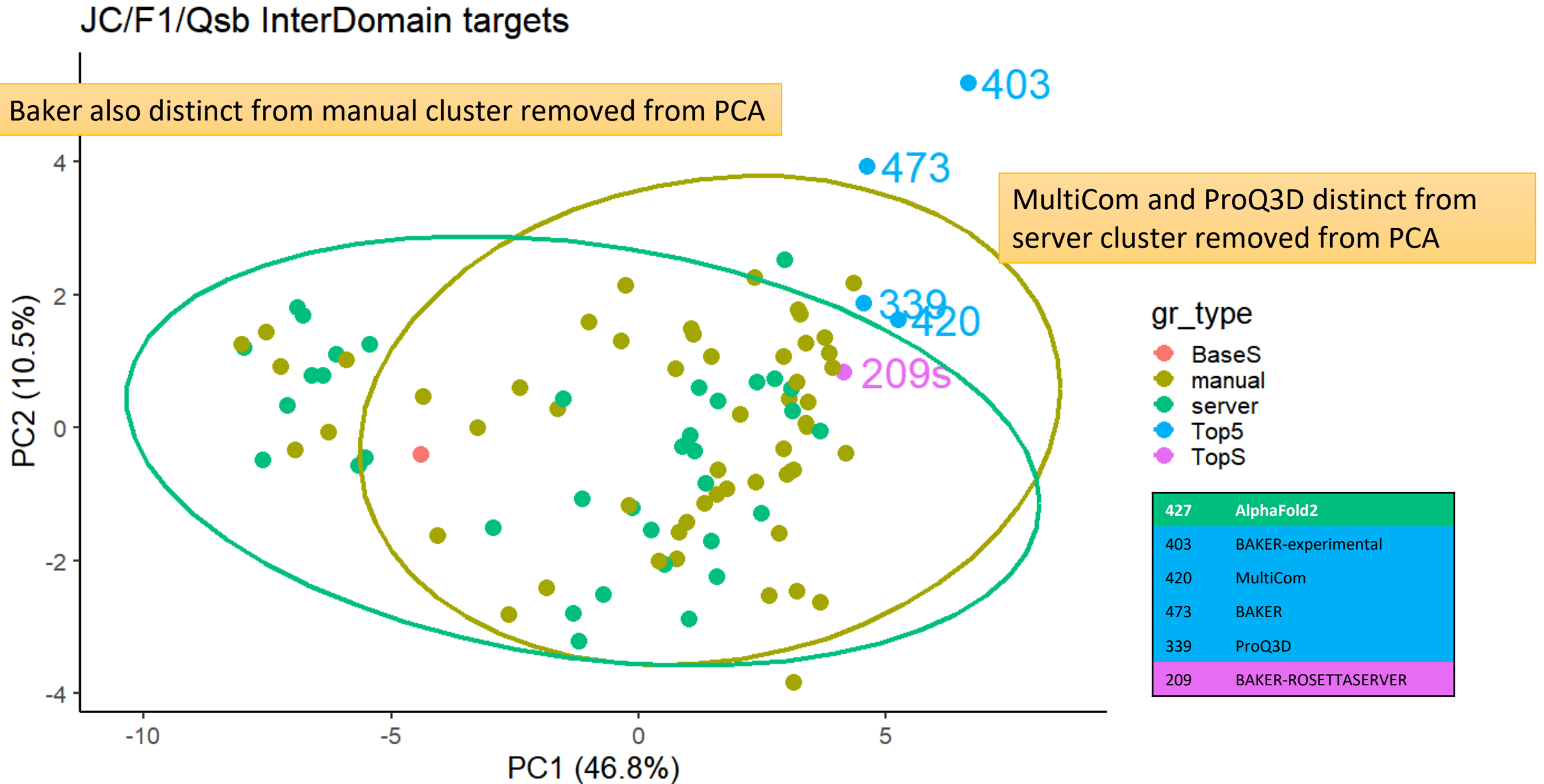
Sum of General Z-scores (GDT/IDDT) on Interdomain targets

Rank SumZ(>0) for top20 InterDomain groups (general scores)

Weights: GDT_TS/IDDT = 1
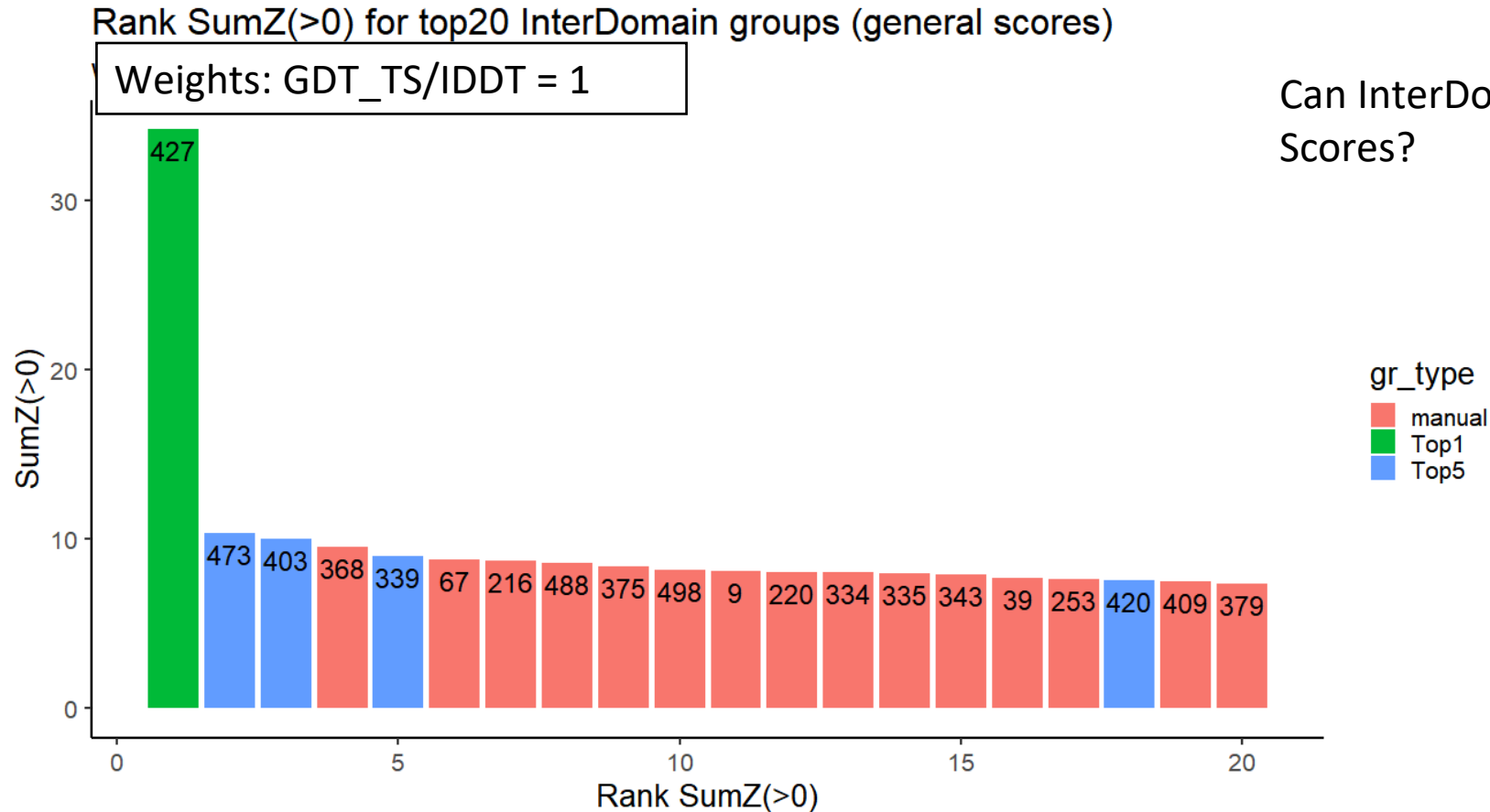
Can InterDomain targets by assessed by structural Scores?

Performance between Interdomain and general targets correlates

Correlation of RSumZ(>0) for Interdomain vs. General targets, structure scores

Group	#	Ri	Rg
AlphaFold2	427	1	1
BAKER-experimental	403	3	3
MULTICOM	420	5	18
BAKER	473	6	2
ProQ3D	339	7	5
FEIG-R3	334	8	13
BAKER-ROSETTASERVER	209	9	26
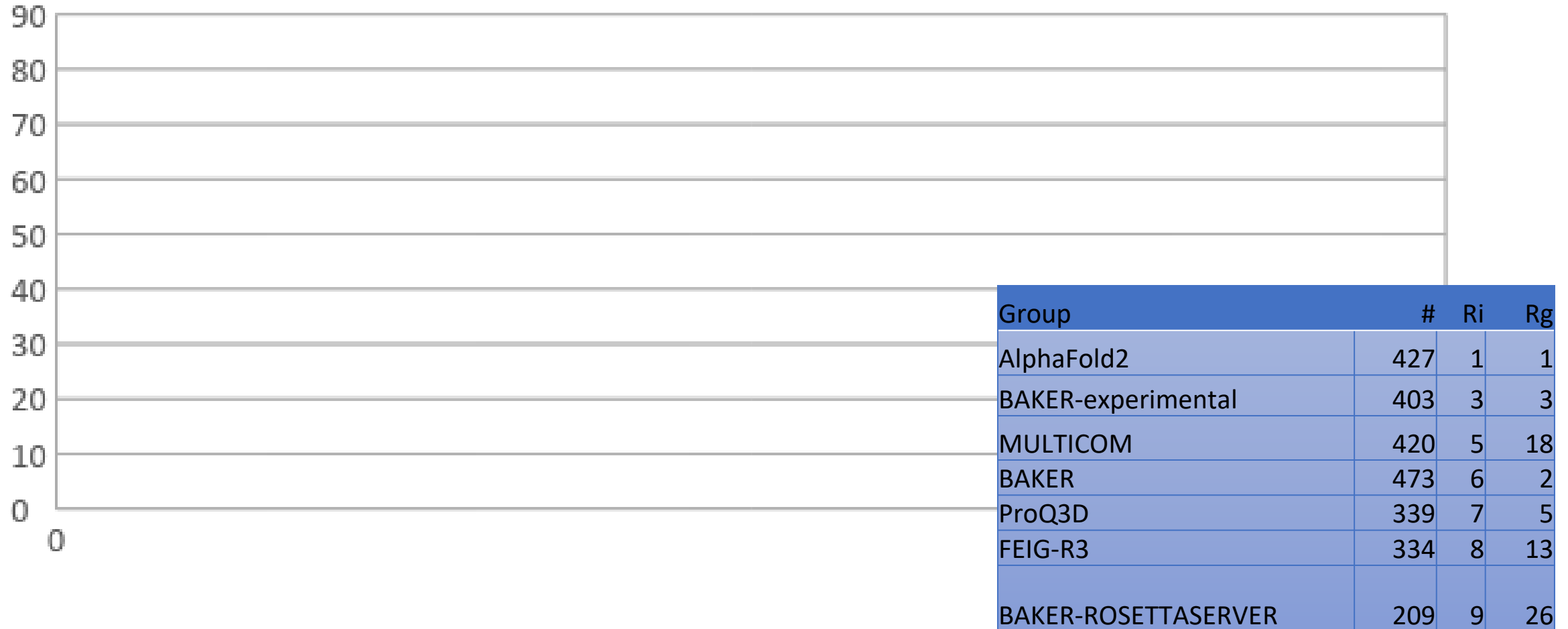
Conclusions

- Groups which perform well on Interdomain targets perform well on general targets
  - AlphaFold clearly top performer, scores well even by GDT
  - Baker clear second, top T1044 prediction
- 2-domain targets are being predicted well and above baseline by many groups
  - Targets with multiple domain interfaces are still not being predicted well

Thank You!

Collaborators
Nick Grishin (UTSW)
Lisa Kinch(UTSW)
Jimin Pei (UTSW)
Andriy Kryshtafovych (Prediction Center)
CASP Assessors
Andrei Lupas (High Accuracy Models)
Alfonso Valencia (Contacts)
Daniel Rigden (Refinement)
Ezgi Karaca (Assembly)
Chaok Seok (Model Accuracy)
Sandor Vajda (Function)
CASP Organizing Committee
John Moult, CASP chair and founder; IBBR, University of Maryland, USA
Krzysztof Fidelis, founder, University of California, Davis, USA
Andriy Kryshtafovych, University of California, Davis, USA
Torsten Schwede, University of Basel, Switzerland
Maya Topf, Birkbeck, University of London, UK

hhmi | Howard Hughes Medical Institute