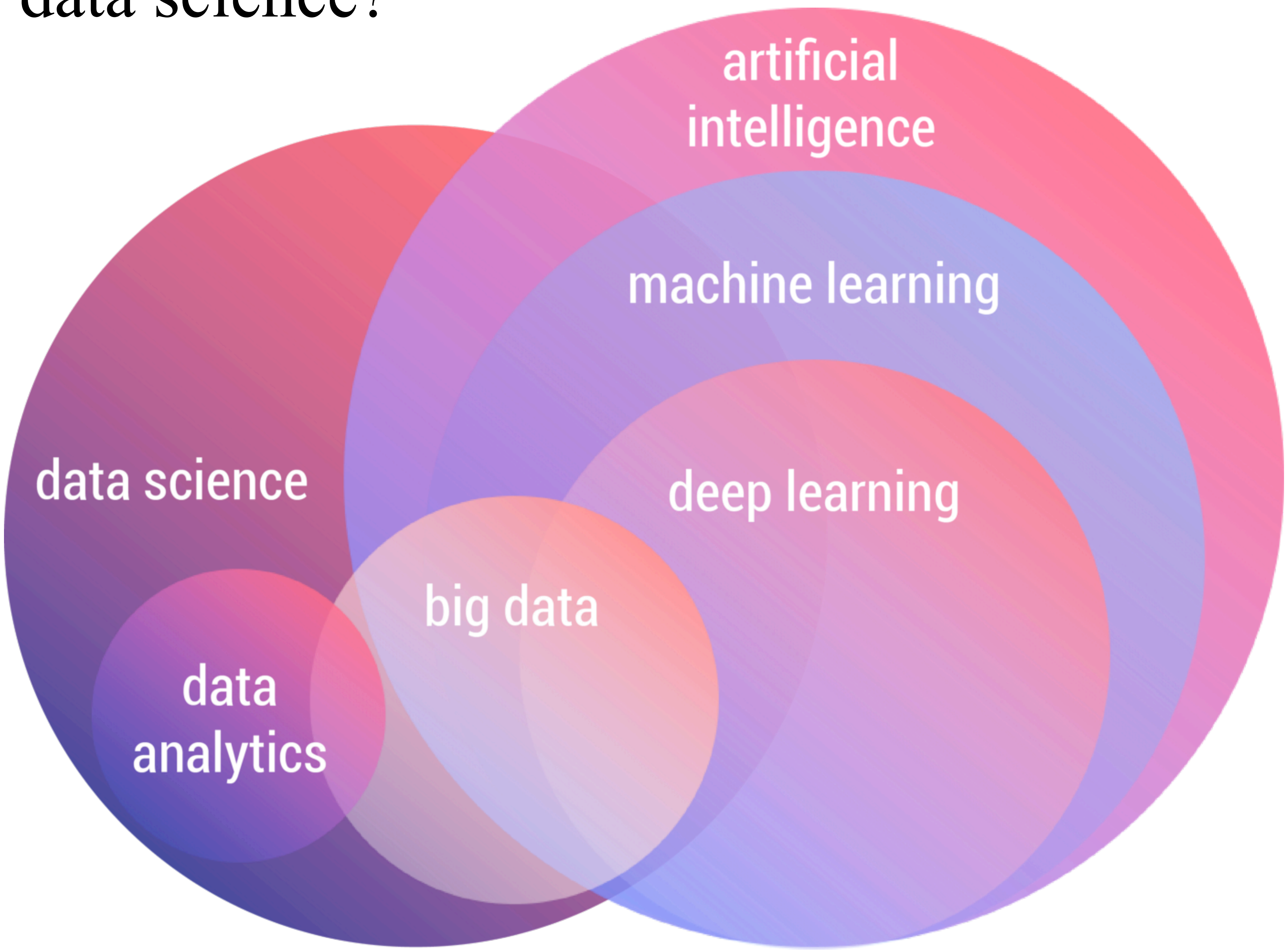
Data Science at the Crossroads of Observational and Computational Galaxy Astronomy

IPMU Colloquium (October 29th, 2021)

Connor Bottrell
(Kavli IPMU | connor.bottrell@ipmu.jp)

Kavli IPMU
The University of Tokyo
UTIAS
Kavli IPMU Institute for the Physics and Mathematics of the Universe
The Kavli Foundation
NSERC CRSNG
WPI World Premier International Research Center Initiative

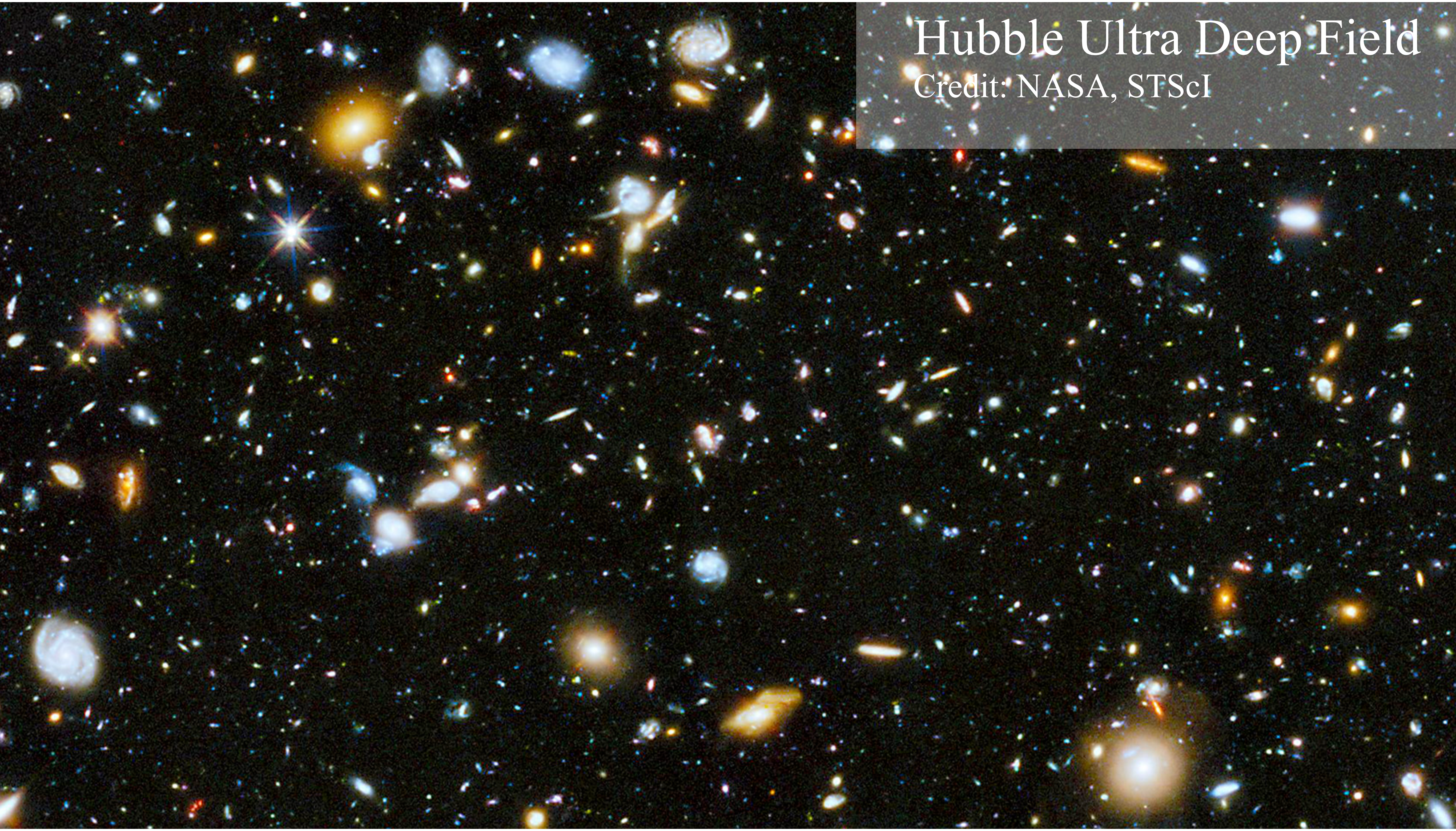
What is data science?



Data science applies *domain-specific knowledge* (e.g. astronomy, business, finance) alongside machine learning and traditional statistics to create *data-driven models* and make real-world predictions.

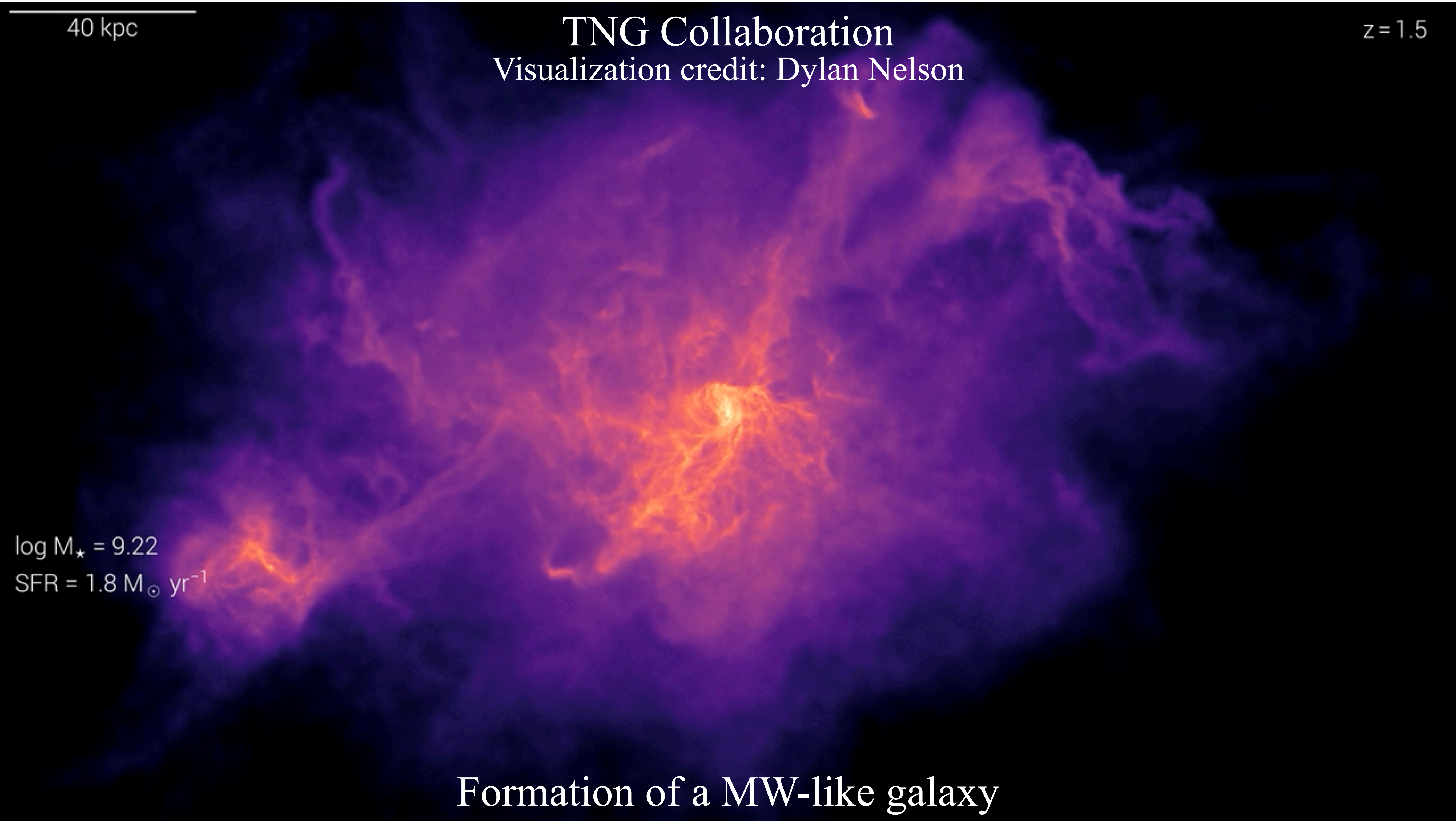
This process usually requires large datasets to create *generalized* models.

Data science applications in astronomy?



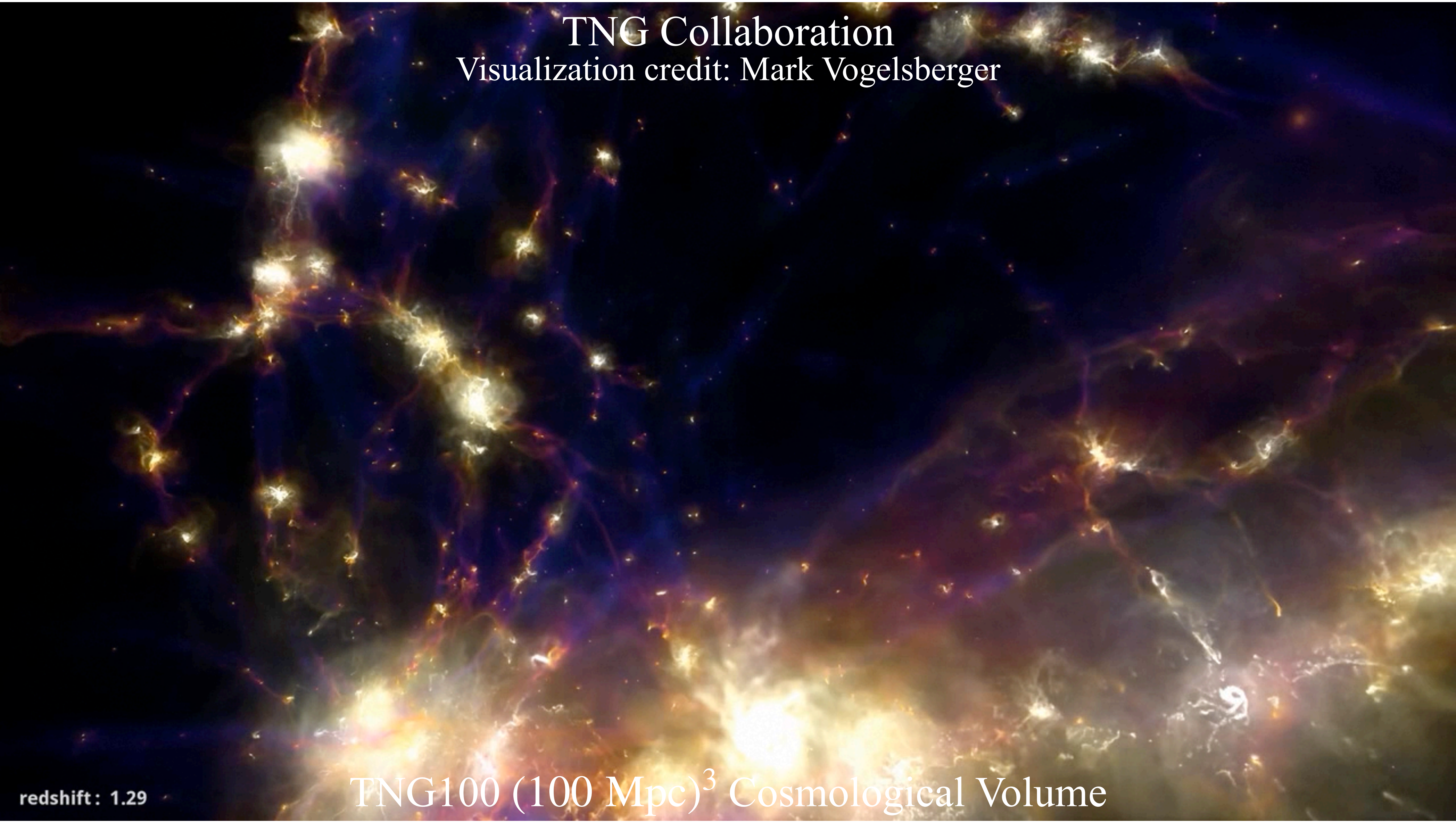
Very broad. But *one* is to create connections between the features we can *observe* and the inaccessible/meta information can only be *inferred*.

Computational astronomy facilitates connections between *observables* and *non-observables*



In a numerical hydrodynamical simulation, the co-evolution of dark and baryonic matter are explicitly tracked (i.e. full access to merger histories).

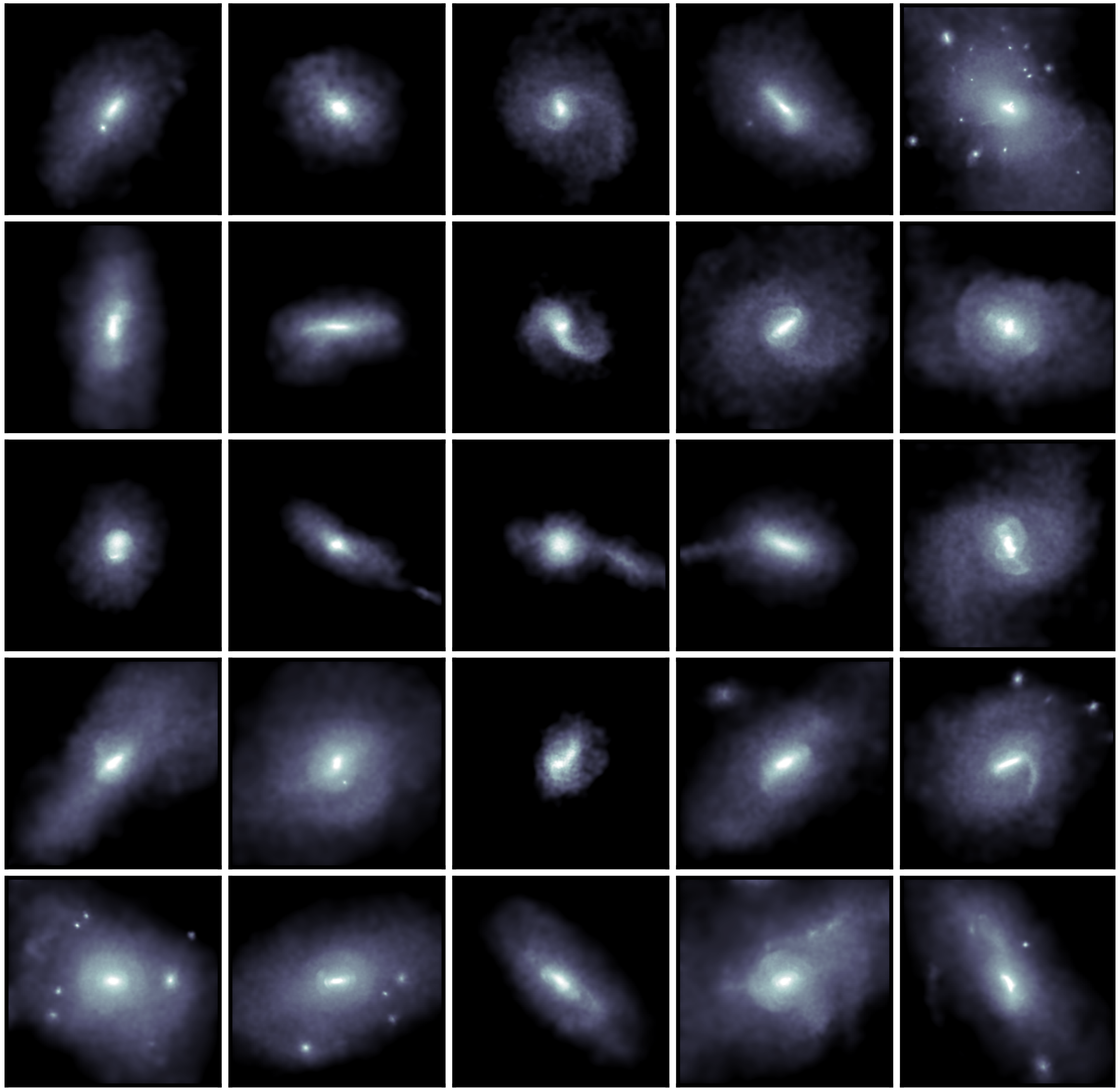
What about generalizability? One galaxy is *not* enough.



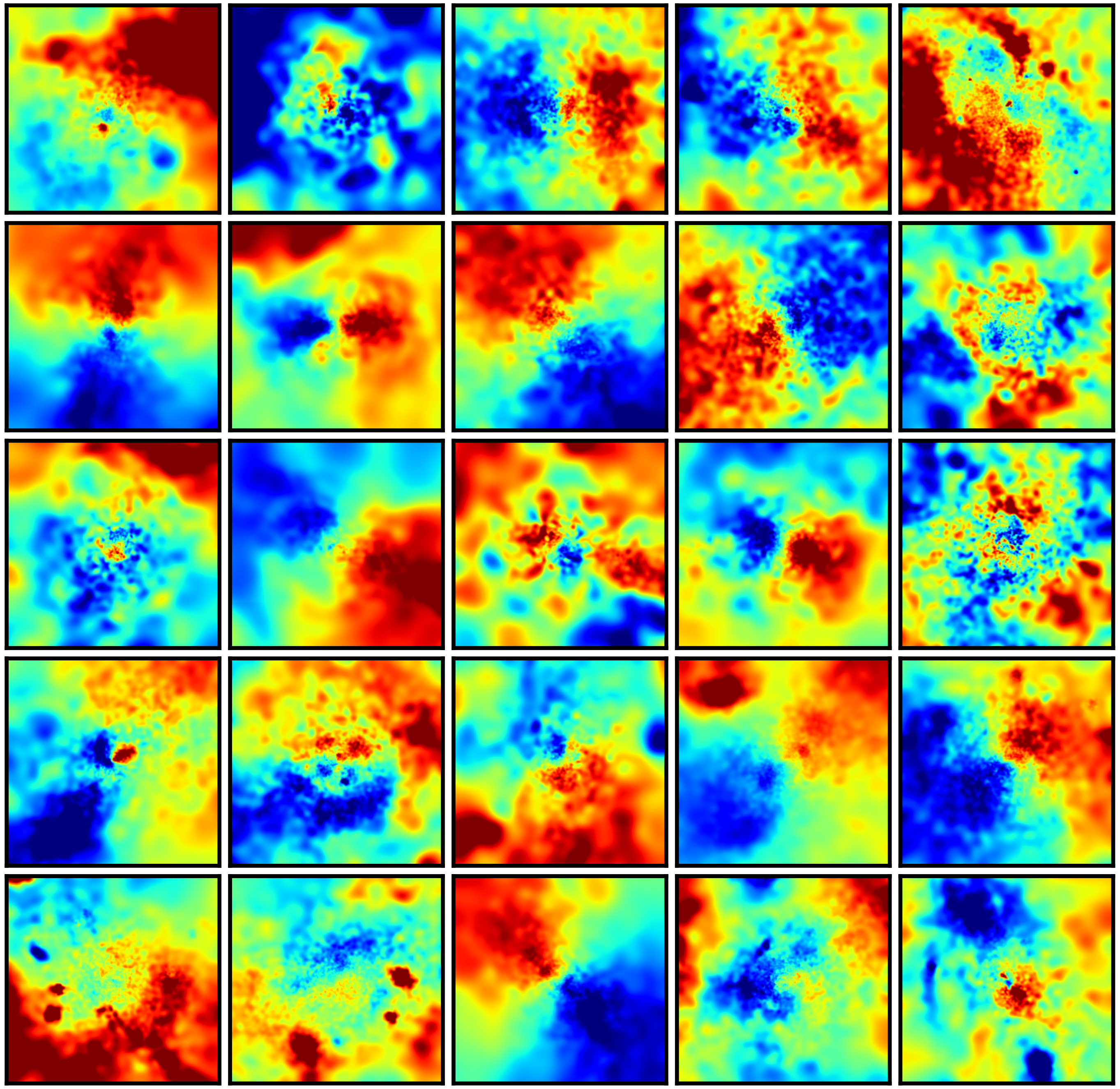
To make predictions for new data, a data-driven model must learn from *training* sets which are representative of the *application* sets. Many galaxies are needed.

How accurately can observables be mapped to merger status?

2,332 post-coalescence merger remnants from TNG100



Stellar Mass Maps



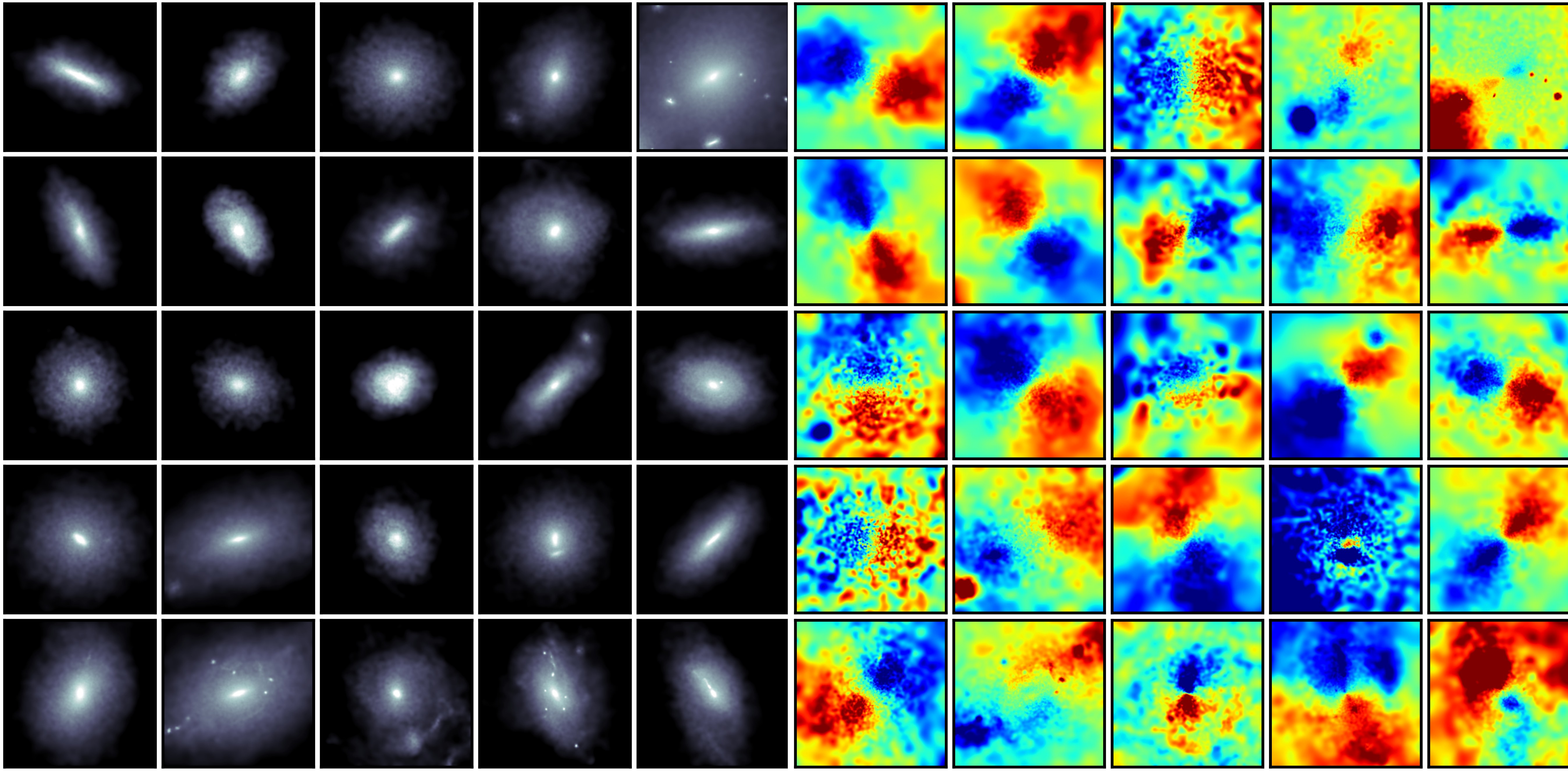
Stellar Velocity Maps

Idealized synthetic observations (no noise, no blurring, no projection effects)

Bottrell et al. 2021a (submitted)

How accurately can observables be mapped to merger status?

2,332 non-merger *controls* matched in redshift, stellar mass, environment



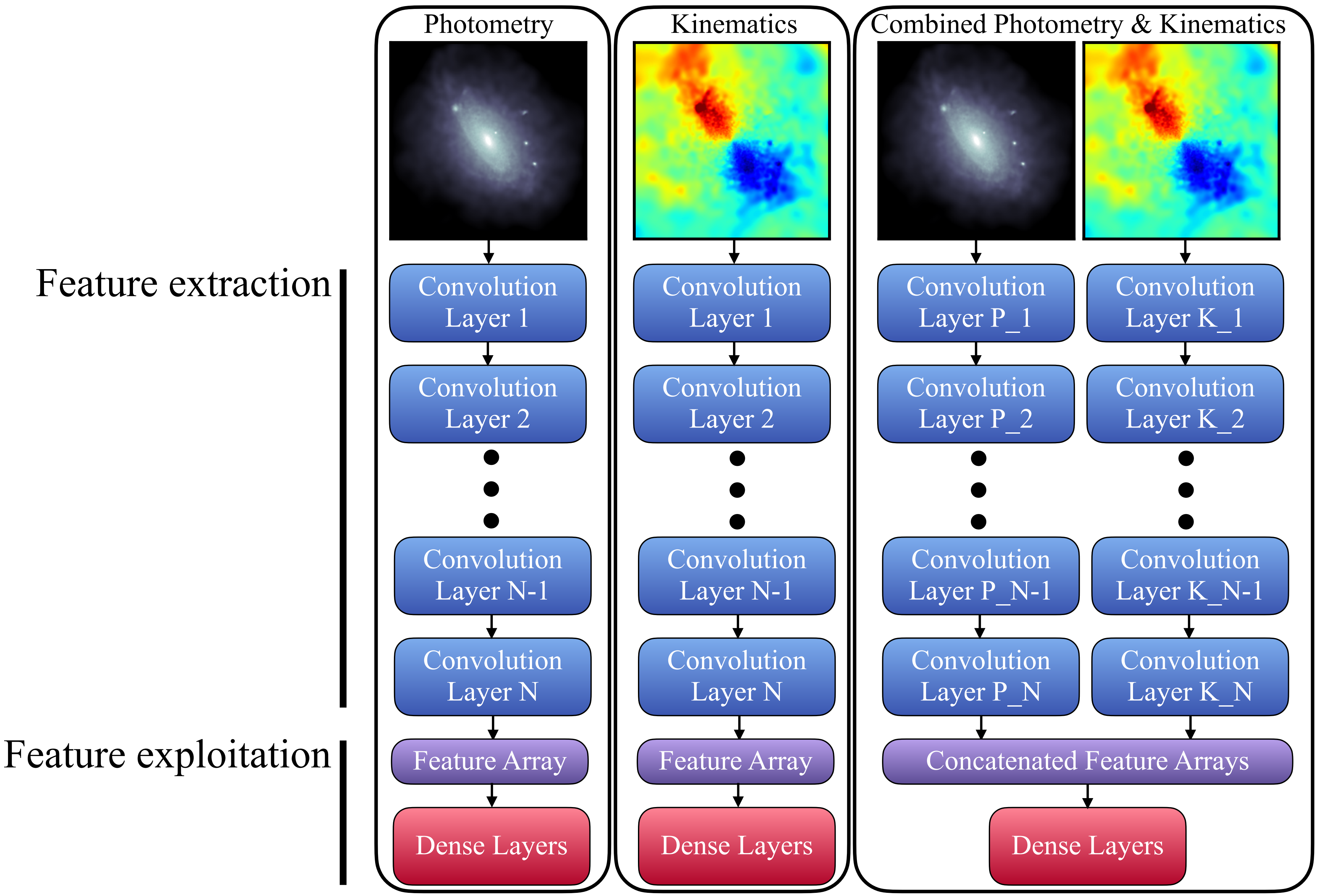
Stellar Mass Maps

Stellar Velocity Maps

Idealized synthetic observations (no noise, no blurring, no projection effects)

Bottrell et al. 2021a (submitted)

Merger Remnant Classification Models

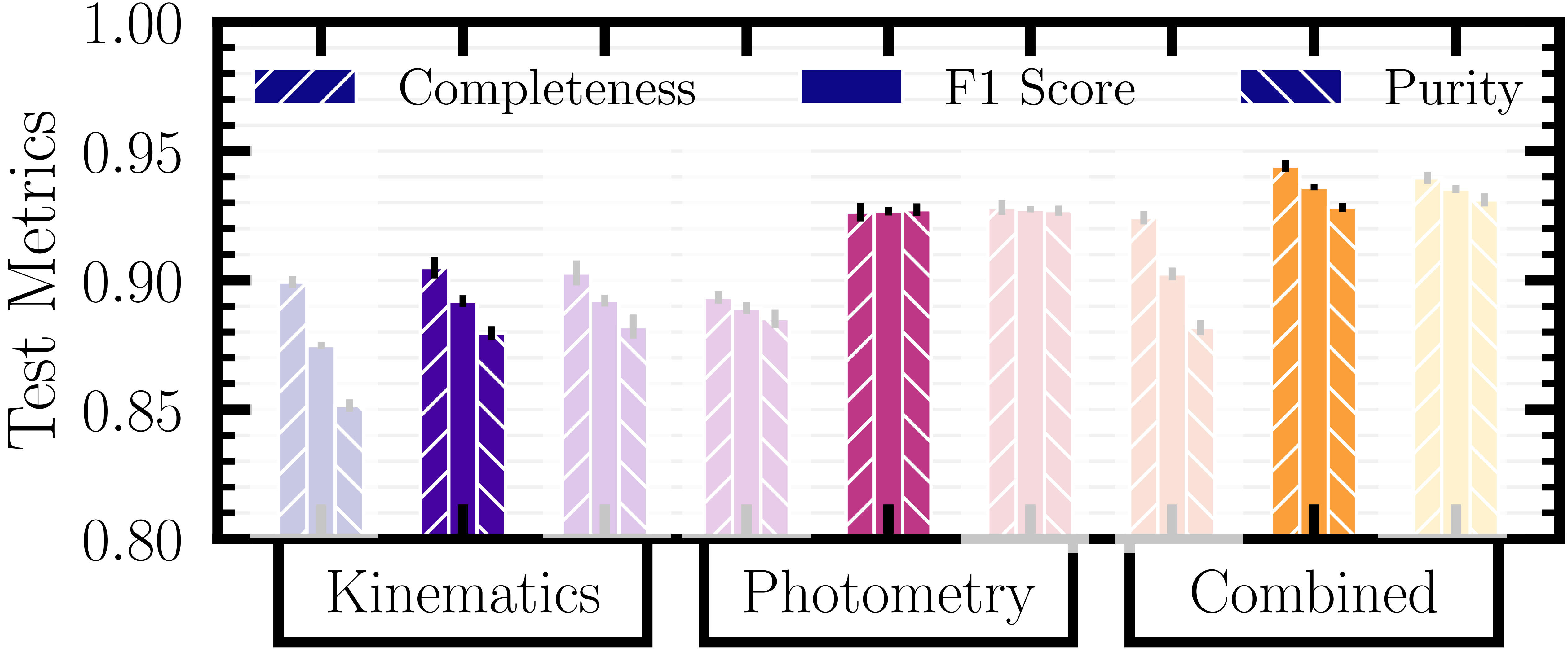


Convolutional neural networks (CNNs, LeCun+1998) are a class of deep learning models that are useful on data that exhibits topological structure (e.g. images).

Specific Architectures:
AlexNet: Krizhevsky+2012; **ResNet-V2**: He+2015,2016

Relative and combined utilities of visual appearance and stellar dynamics in identifying post-mergers

Bottrell et al. 2021a (submitted)

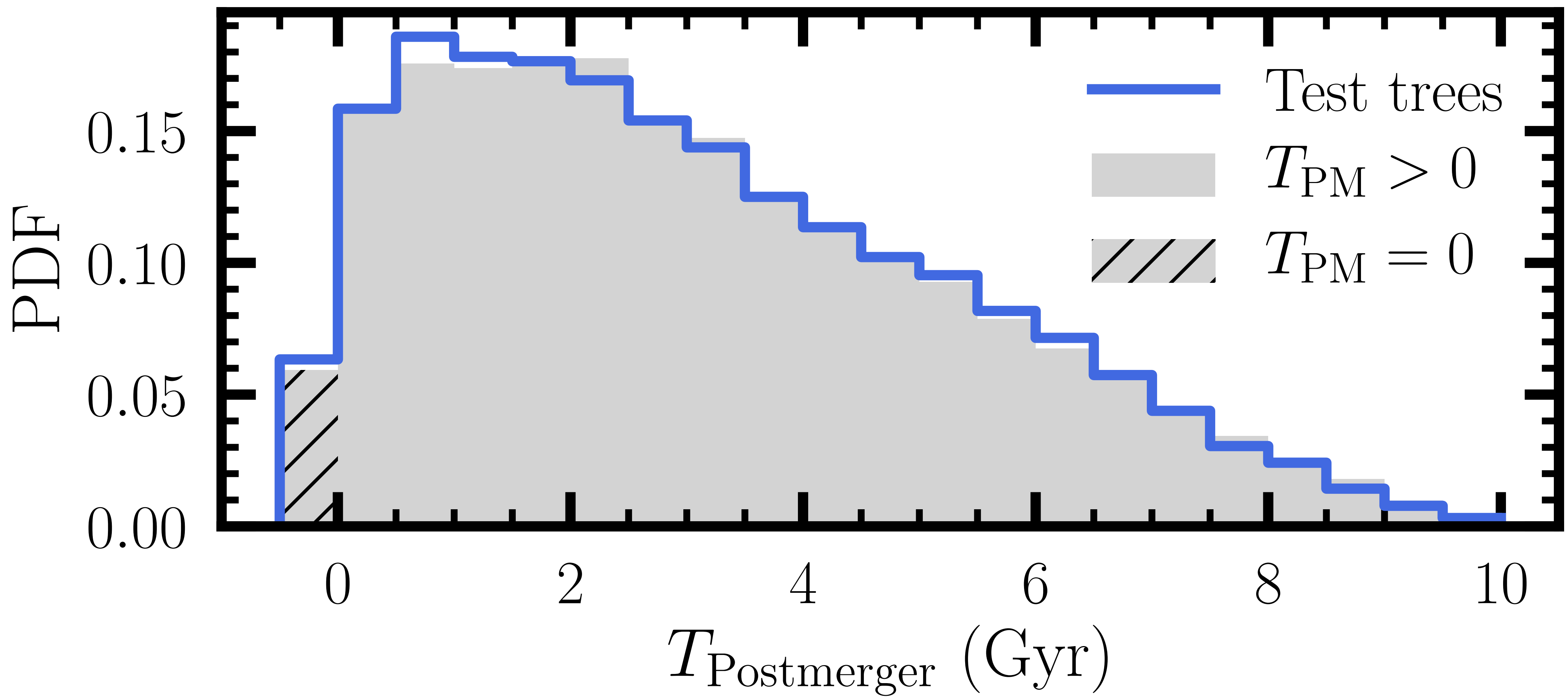


Images enable ***93% completeness and 93% purity***. Stellar kinematics are less useful, at ***90% completeness and 88% purity***. The model combining both data sets has a 2% boost in completeness but no advantage in purity.

Images (and stellar kinematics to a lesser extent) can be used to accurately predict recent merger activity (see also **Bickley, Bottrell et al. 2021**)

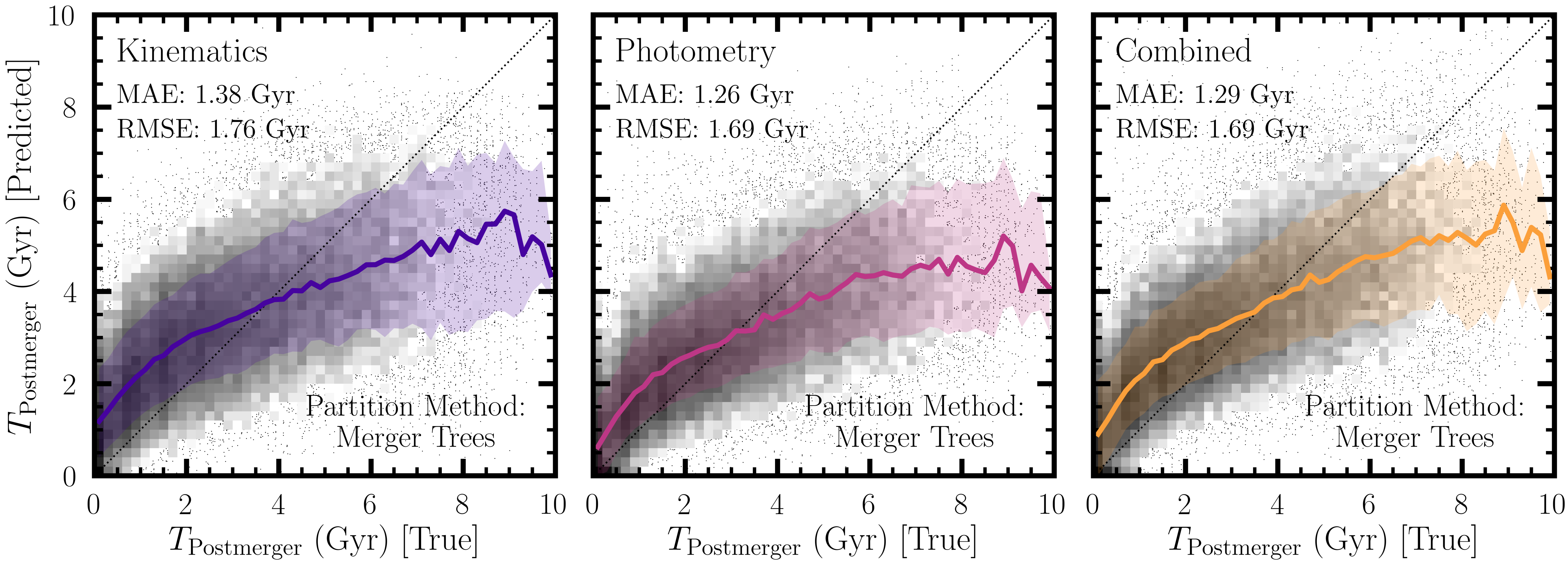
How far back can the line be drawn?

Not every galaxy is a recent merger. But most galaxies have merged in the past. A regression model is better equipped to use this information.



80K galaxies out to $z = 2$. The target quantity is now the *time since the most recent merger*, $T_{\mathrm{Postmerger}}$. Careful partitioning of the data to avoid interpolation along merger histories.

How far back can the line be drawn?

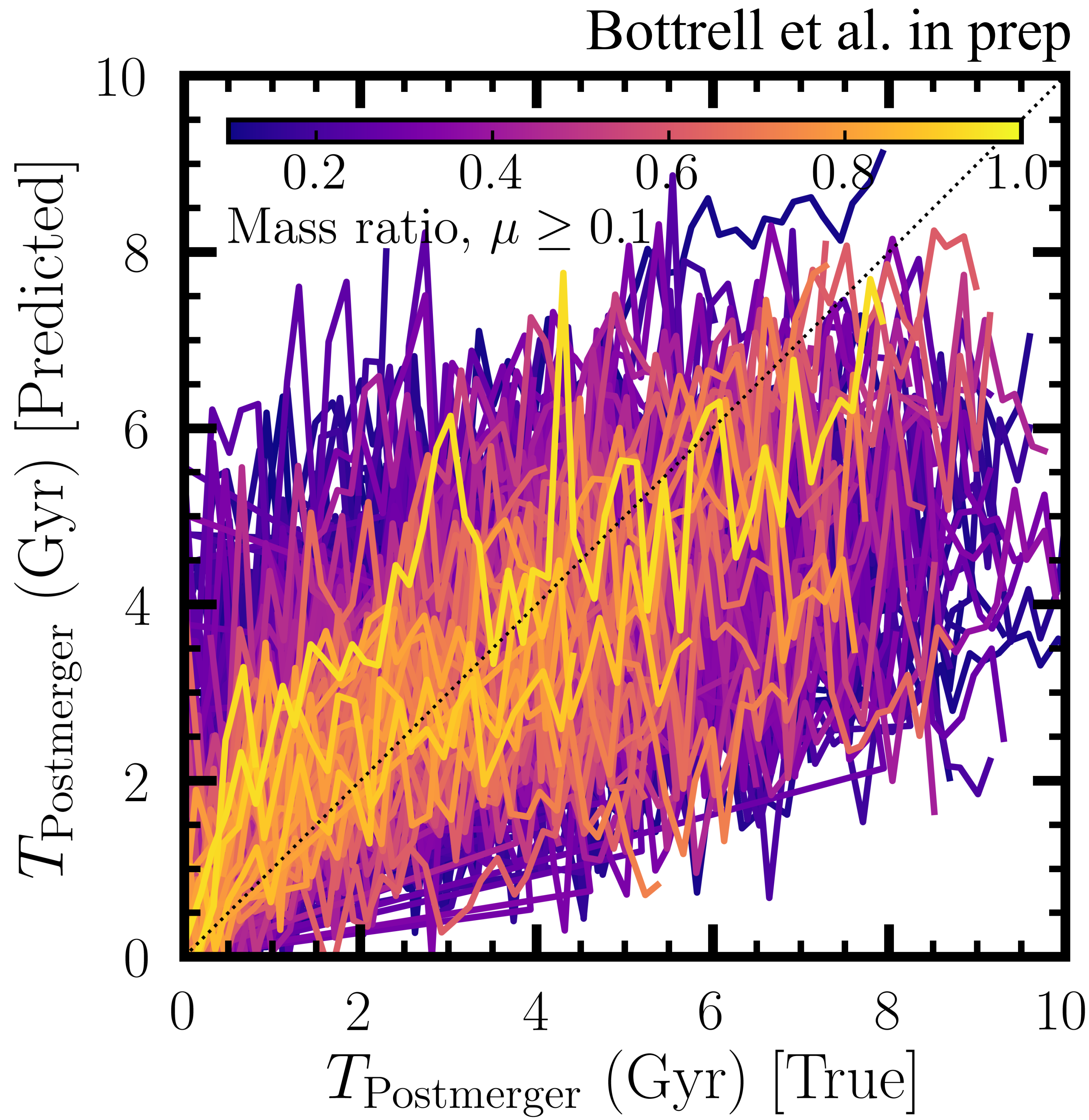


An average galaxy's current morphology and stellar dynamics are strongly connected to its most recent merger event — even many Gigayears after the merger. The predictions then stagnate at larger time-since-merger values.

What makes this predictive power possible?
What is driving the scatter?

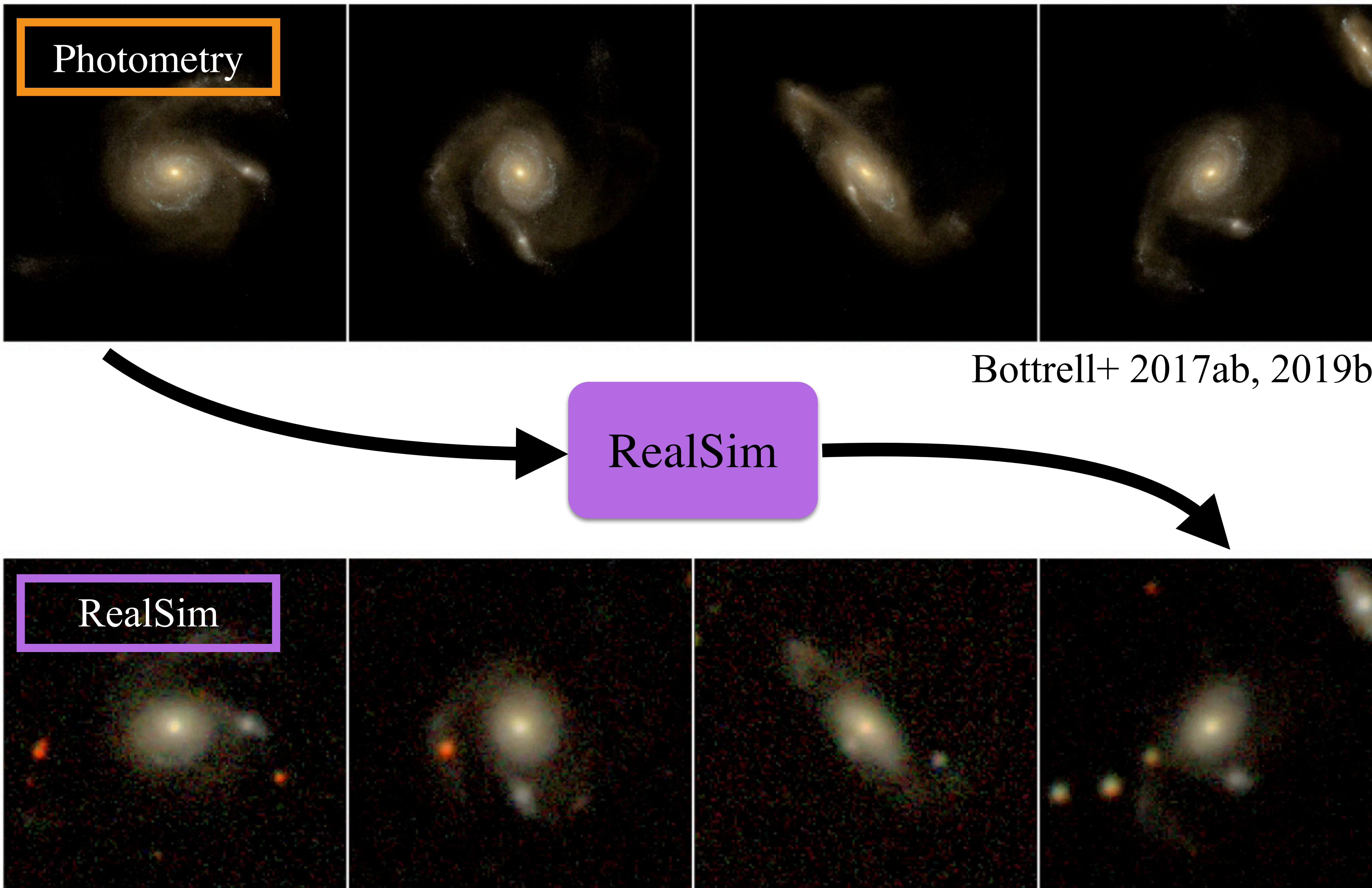
Bottrell et al. in prep

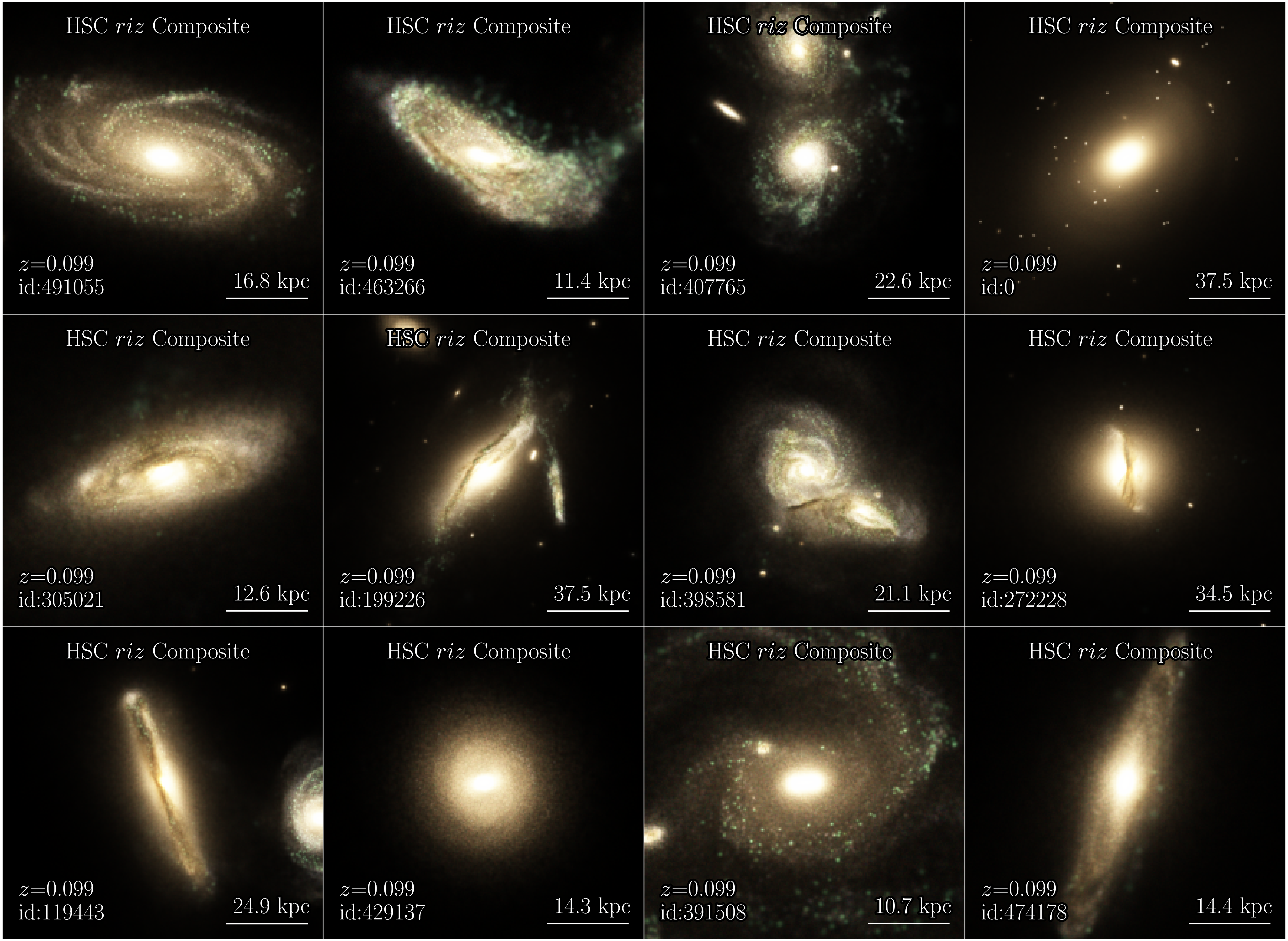
How far back can the line be drawn?



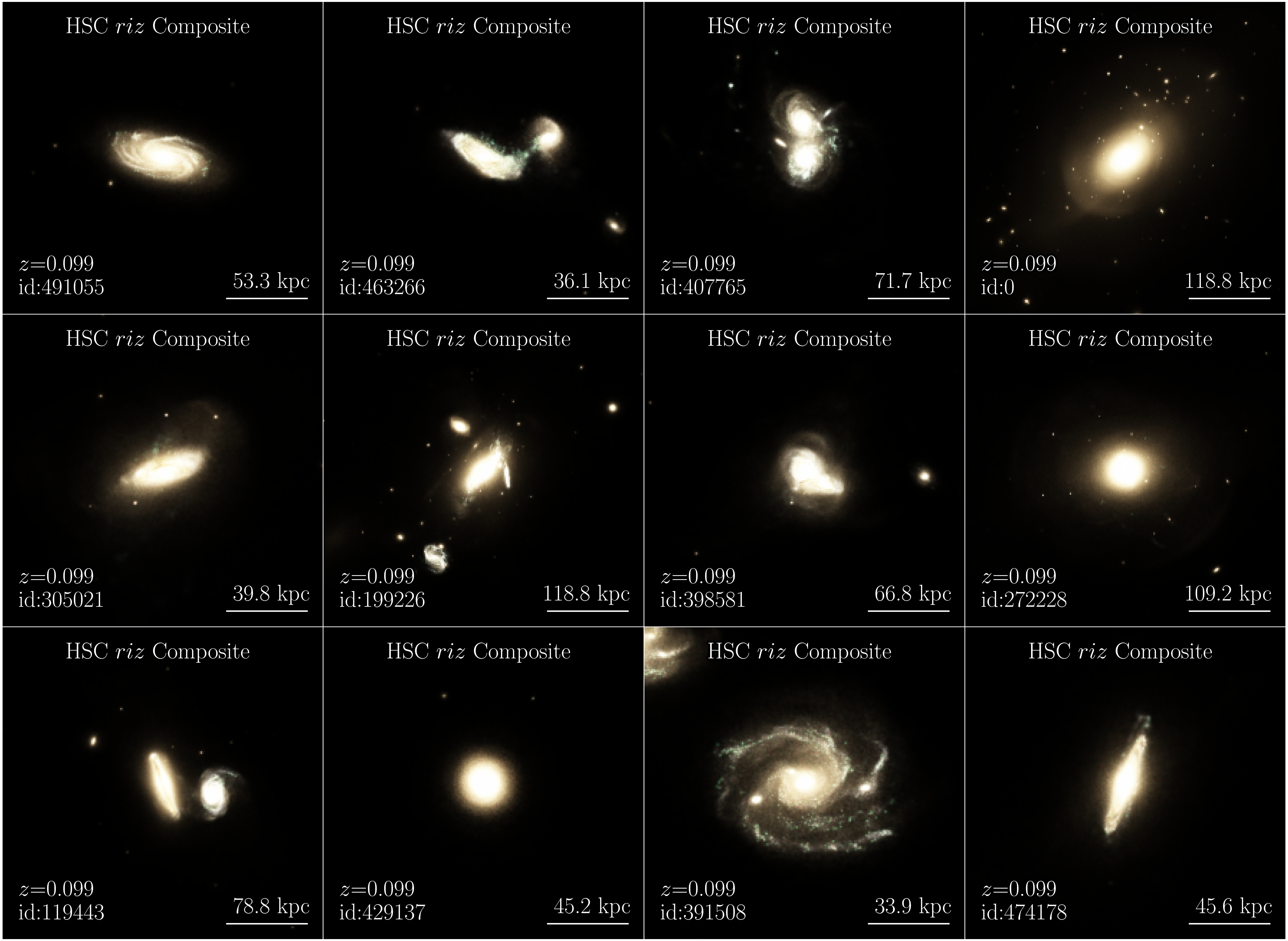
Each line shows the predictions for all descendants of a merger event for galaxies in the test sample. Mergers between similar mass galaxies (yellow/orange) are rarer, but have a tighter scatter and less stagnation.

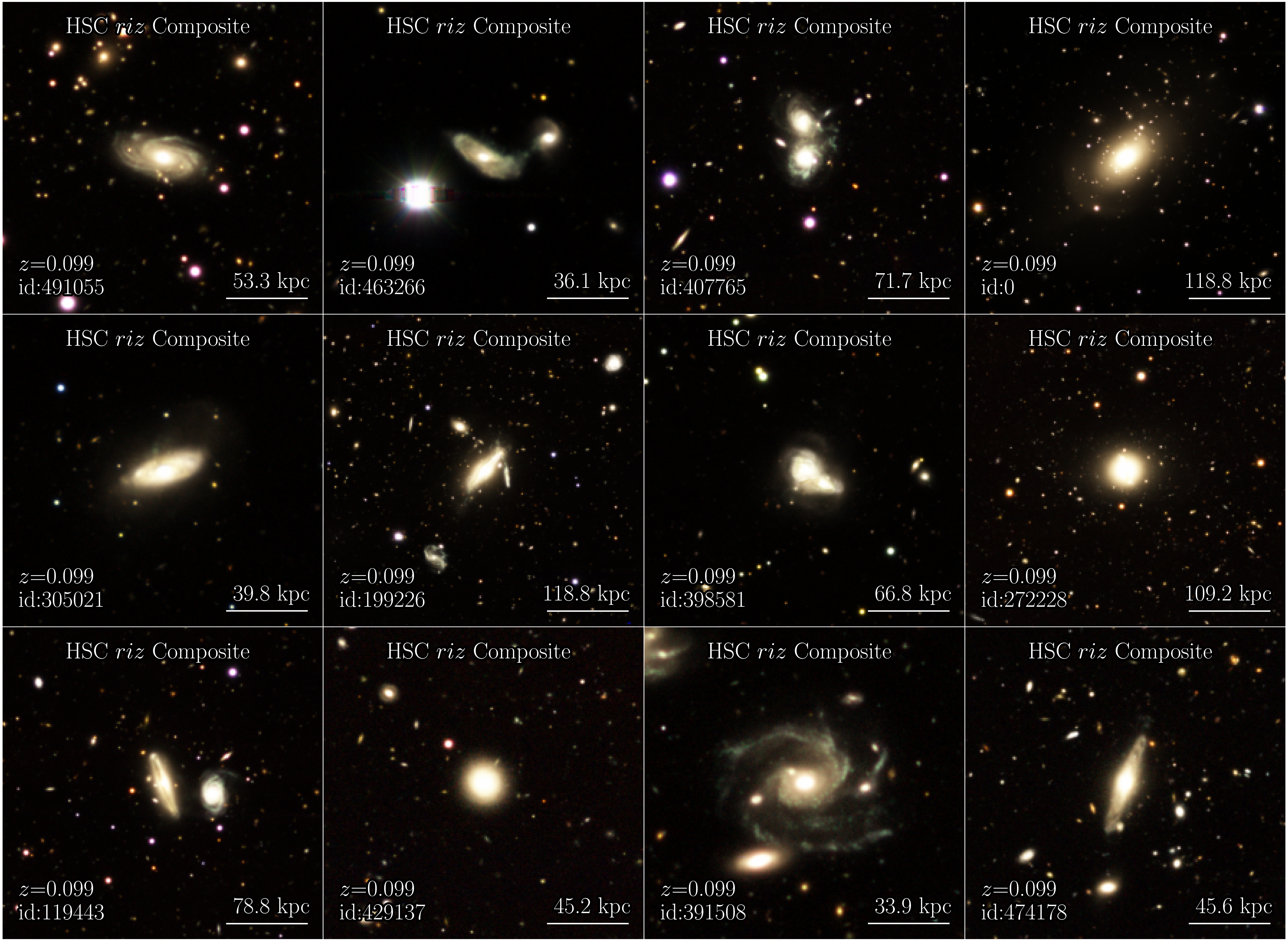
How can the simulation-trained models be applied to observations?

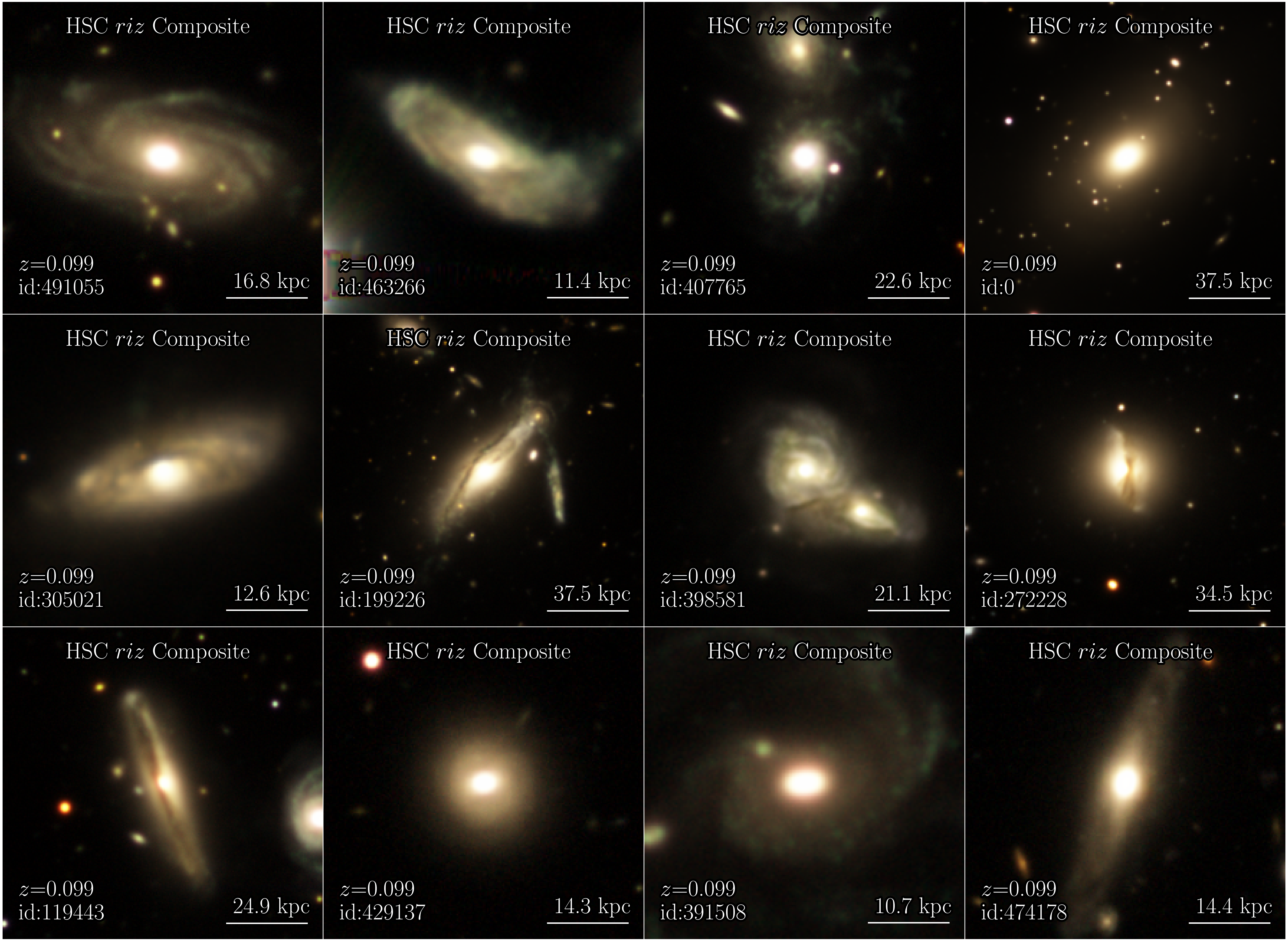


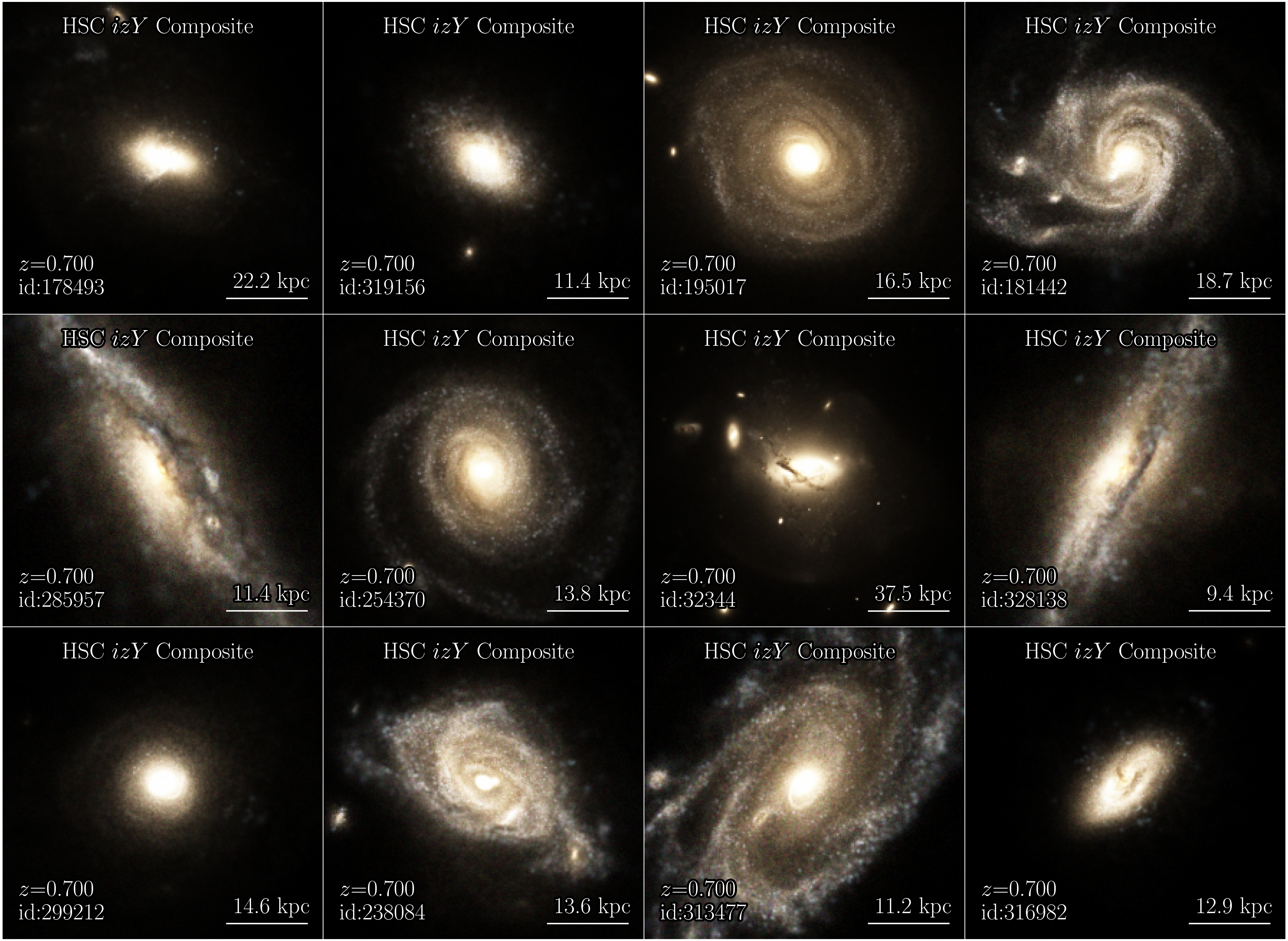


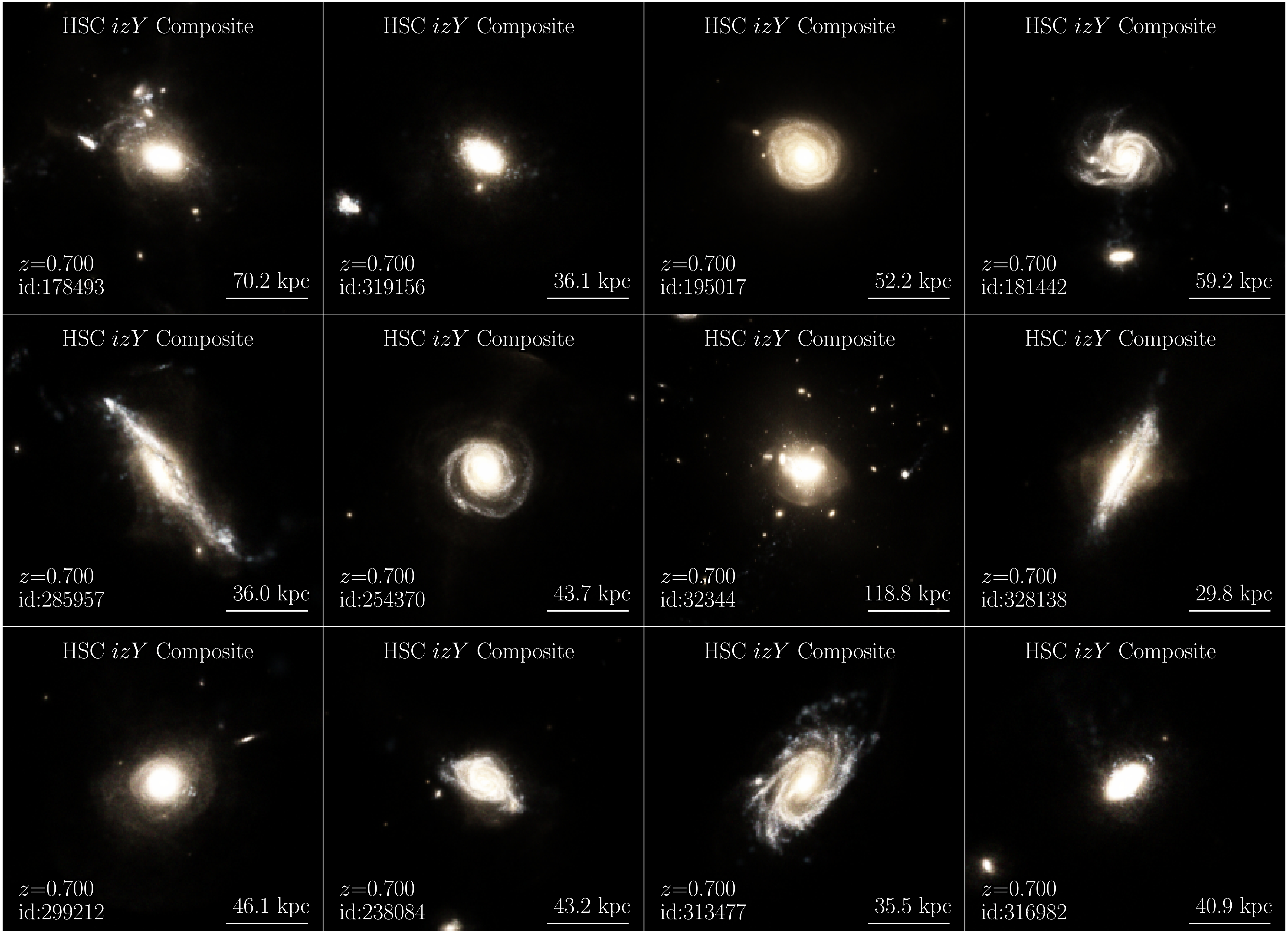
HSC riz Composite
z=0.099
id:491055
16.8 kpc
HSC riz Composite
z=0.099
id:463266
11.4 kpc
HSC riz Composite
z=0.099
id:407765
22.6 kpc
HSC riz Composite
z=0.099
id:0
37.5 kpc
HSC riz Composite
z=0.099
id:305021
12.6 kpc
HSC riz Composite
z=0.099
id:199226
37.5 kpc
HSC riz Composite
z=0.099
id:398581
21.1 kpc
HSC riz Composite
z=0.099
id:272228
34.5 kpc
HSC riz Composite
z=0.099
id:119443
24.9 kpc
HSC riz Composite
z=0.099
id:429137
14.3 kpc
HSC riz Composite
z=0.099
id:391508
10.7 kpc
HSC riz Composite
z=0.099
id:474178
14.4 kpc


HSC riz Composite
z=0.099
id:491055
53.3 kpc
HSC riz Composite
z=0.099
id:463266
36.1 kpc
HSC riz Composite
z=0.099
id:407765
71.7 kpc
HSC riz Composite
z=0.099
id:0
118.8 kpc
HSC riz Composite
z=0.099
id:305021
39.8 kpc
HSC riz Composite
z=0.099
id:199226
118.8 kpc
HSC riz Composite
z=0.099
id:398581
66.8 kpc
HSC riz Composite
z=0.099
id:272228
109.2 kpc
HSC riz Composite
z=0.099
id:119443
78.8 kpc
HSC riz Composite
z=0.099
id:429137
45.2 kpc
HSC riz Composite
z=0.099
id:391508
33.9 kpc
HSC riz Composite
z=0.099
id:474178
45.6 kpc

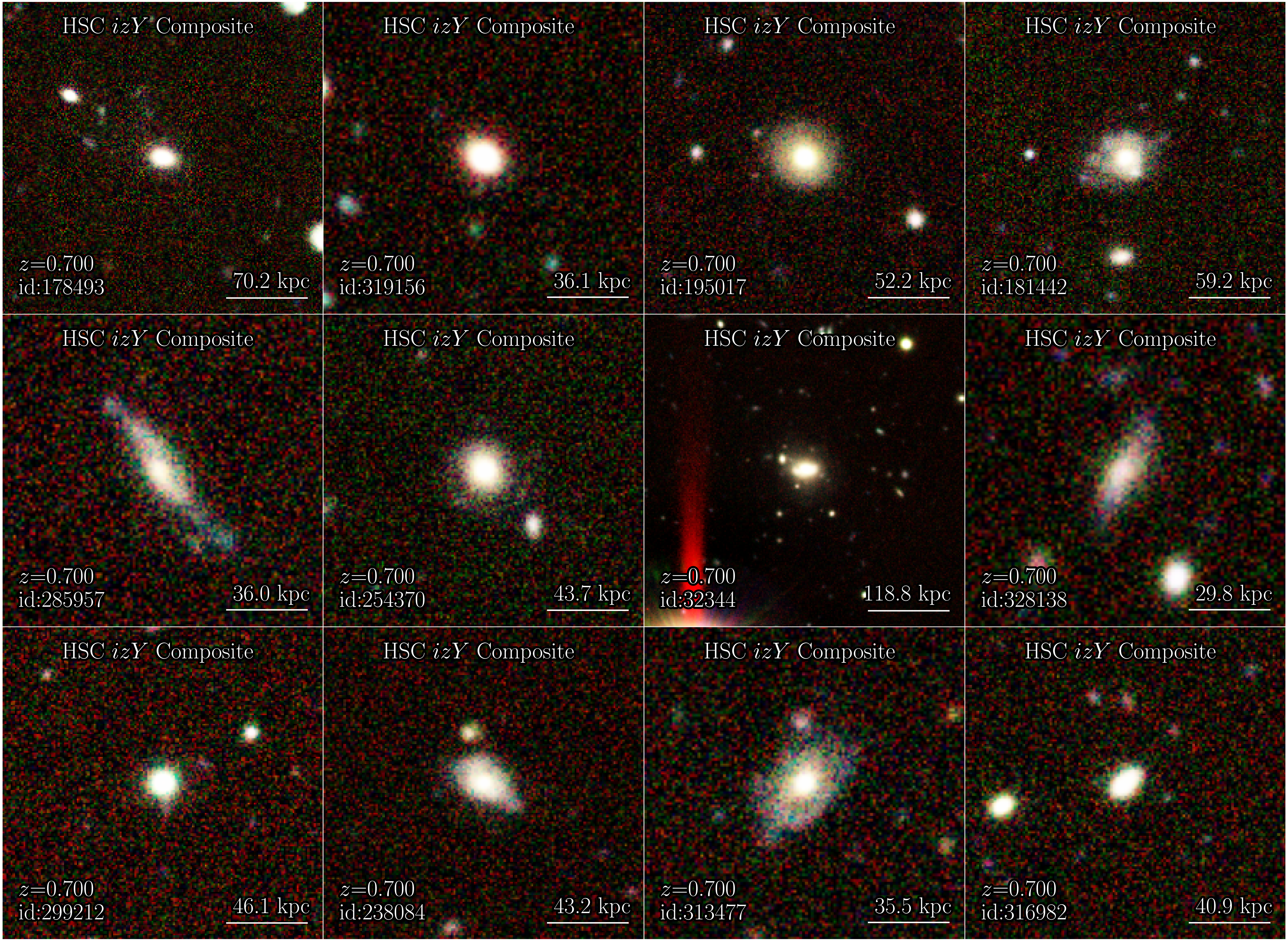

HSC riz Composite
z=0.099
id:491055
53.3 kpc
HSC riz Composite
z=0.099
id:463266
36.1 kpc
HSC riz Composite
z=0.099
id:407765
71.7 kpc
HSC riz Composite
z=0.099
id:0
118.8 kpc
HSC riz Composite
z=0.099
id:305021
39.8 kpc
HSC riz Composite
z=0.099
id:199226
118.8 kpc
HSC riz Composite
z=0.099
id:398581
66.8 kpc
HSC riz Composite
z=0.099
id:272228
109.2 kpc
HSC riz Composite
z=0.099
id:119443
78.8 kpc
HSC riz Composite
z=0.099
id:429137
45.2 kpc
HSC riz Composite
z=0.099
id:391508
33.9 kpc
HSC riz Composite
z=0.099
id:474178
45.6 kpc

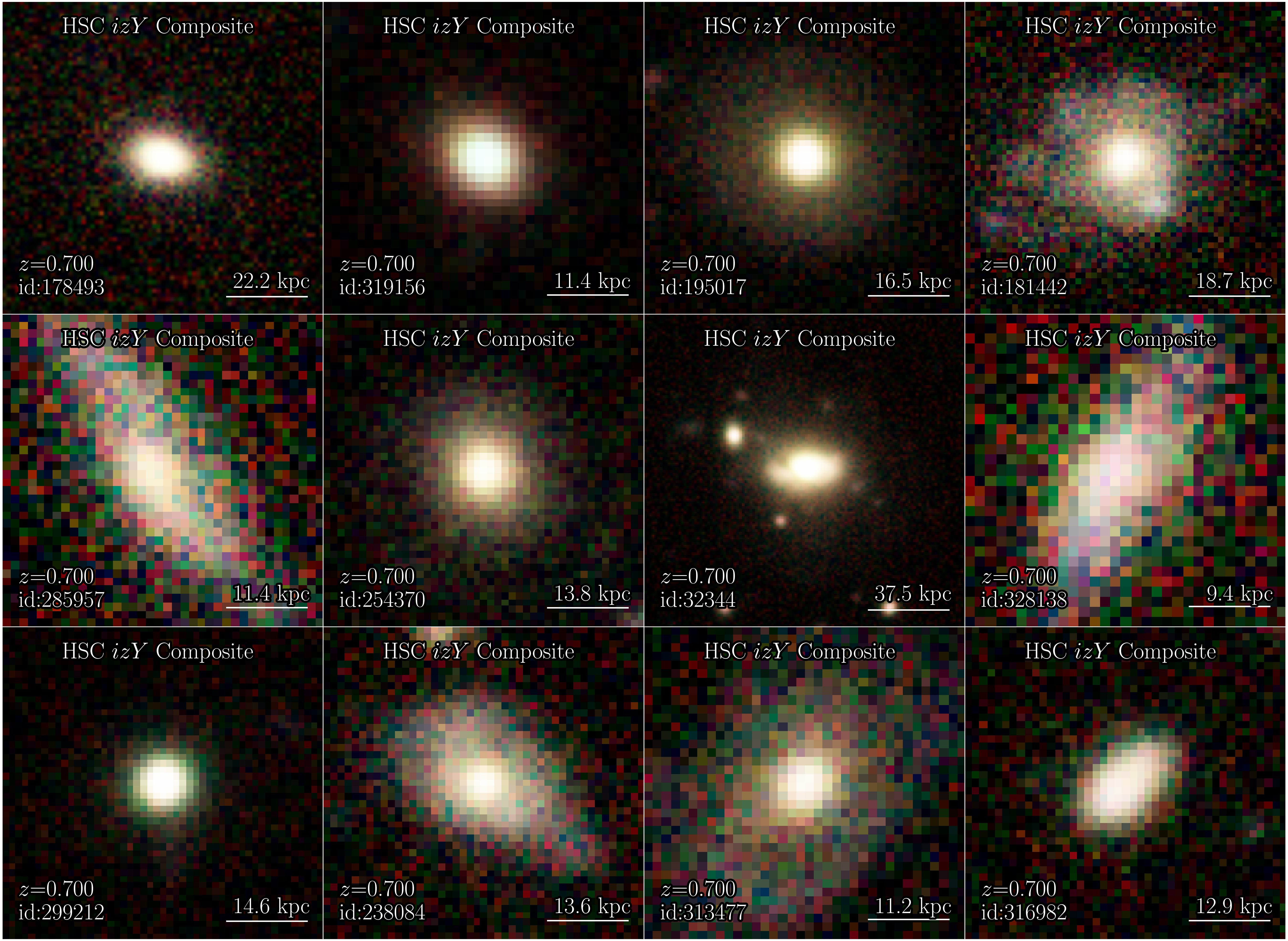

HSC riz Composite
z=0.099
id:491055
16.8 kpc
HSC riz Composite
z=0.099
id:463266
11.4 kpc
HSC riz Composite
z=0.099
id:407765
22.6 kpc
HSC riz Composite
z=0.099
id:0
37.5 kpc
HSC riz Composite
z=0.099
id:305021
12.6 kpc
HSC riz Composite
z=0.099
id:199226
37.5 kpc
HSC riz Composite
z=0.099
id:398581
21.1 kpc
HSC riz Composite
z=0.099
id:272228
34.5 kpc
HSC riz Composite
z=0.099
id:119443
24.9 kpc
HSC riz Composite
z=0.099
id:429137
14.3 kpc
HSC riz Composite
z=0.099
id:391508
10.7 kpc
HSC riz Composite
z=0.099
id:474178
14.4 kpc


HSC izY Composite
z=0.700
id:178493
22.2 kpc
HSC izY Composite
z=0.700
id:319156
11.4 kpc
HSC izY Composite
z=0.700
id:195017
16.5 kpc
HSC izY Composite
z=0.700
id:181442
18.7 kpc
HSC izY Composite
z=0.700
id:285957
11.4 kpc
HSC izY Composite
z=0.700
id:254370
13.8 kpc
HSC izY Composite
z=0.700
id:32344
37.5 kpc
HSC izY Composite
z=0.700
id:328138
9.4 kpc
HSC izY Composite
z=0.700
id:299212
14.6 kpc
HSC izY Composite
z=0.700
id:238084
13.6 kpc
HSC izY Composite
z=0.700
id:313477
11.2 kpc
HSC izY Composite
z=0.700
id:316982
12.9 kpc



HSC izY Composite
z=0.700
id:178493
70.2 kpc
HSC izY Composite
z=0.700
id:319156
36.1 kpc
HSC izY Composite
z=0.700
id:195017
52.2 kpc
HSC izY Composite
z=0.700
id:181442
59.2 kpc
HSC izY Composite
z=0.700
id:285957
36.0 kpc
HSC izY Composite
z=0.700
id:254370
43.7 kpc
HSC izY Composite
z=0.700
id:32344
118.8 kpc
HSC izY Composite
z=0.700
id:328138
29.8 kpc
HSC izY Composite
z=0.700
id:299212
46.1 kpc
HSC izY Composite
z=0.700
id:238084
43.2 kpc
HSC izY Composite
z=0.700
id:313477
35.5 kpc
HSC izY Composite
z=0.700
id:316982
40.9 kpc


HSC izY Composite
z=0.700
id:178493
70.2 kpc
HSC izY Composite
z=0.700
id:319156
36.1 kpc
HSC izY Composite
z=0.700
id:195017
52.2 kpc
HSC izY Composite
z=0.700
id:181442
59.2 kpc
HSC izY Composite
z=0.700
id:285957
36.0 kpc
HSC izY Composite
z=0.700
id:254370
43.7 kpc
HSC izY Composite
z=0.700
id:32344
118.8 kpc
HSC izY Composite
z=0.700
id:328138
29.8 kpc
HSC izY Composite
z=0.700
id:299212
46.1 kpc
HSC izY Composite
z=0.700
id:238084
43.2 kpc
HSC izY Composite
z=0.700
id:313477
35.5 kpc
HSC izY Composite
z=0.700
id:316982
40.9 kpc


HSC izY Composite
z=0.700
id:178493
22.2 kpc
HSC izY Composite
z=0.700
id:319156
11.4 kpc
HSC izY Composite
z=0.700
id:195017
16.5 kpc
HSC izY Composite
z=0.700
id:181442
18.7 kpc
HSC izY Composite
z=0.700
id:285957
11.4 kpc
HSC izY Composite
z=0.700
id:254370
13.8 kpc
HSC izY Composite
z=0.700
id:32344
37.5 kpc
HSC izY Composite
z=0.700
id:328138
9.4 kpc
HSC izY Composite
z=0.700
id:299212
14.6 kpc
HSC izY Composite
z=0.700
id:238084
13.6 kpc
HSC izY Composite
z=0.700
id:313477
11.2 kpc
HSC izY Composite
z=0.700
id:316982
12.9 kpc

Data Science at the Crossroads of Observational and Computational Galaxy Astronomy

Machine learning can be leveraged to connect the observable features of galaxies to their current merger status and formation histories in simulations.

Recent *merger remnants can be identified* by their visual appearance or stellar kinematics using deep classification models.

The *time-since-merger* for any galaxy can be estimated using deep regression models with reasonable accuracies up to many Gyr later.

Bridging the models to observations requires more realistic synthetic observations that include observational nuisance factors (e.g. atmosphere).

A still-growing dataset of over 50K synthetic HSC-SSP images of TNG50 galaxies has been made for HSC Project 405.

CONNOR BOTTRELL
(KAVLI IPMU | connor.bottrell@ipmu.jp)