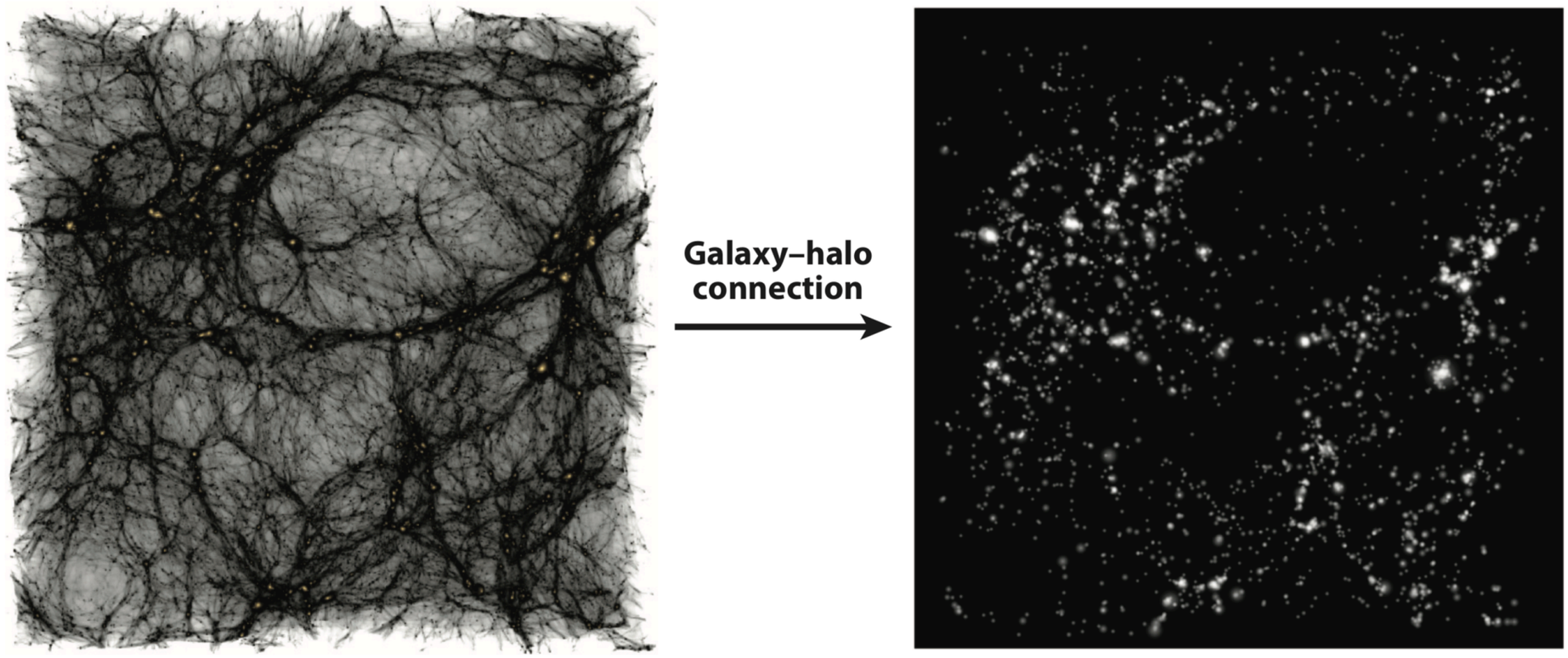
some progress on galaxy–halo connection

Yen-Ting Lin

Institute of Astronomy & Astrophysics, Academia Sinica

Kai-Feng Chen, Chen-Yu Chuang, Tsung-Chi Chen,
Masamune Oguri, Hironao Miyatake, Hong Guo, Yi-Kuan Chiang,
Ting-Wen Lan, Yu-Yen Chang, Christian Jespersen, Shy Genel, Shirley Ho
& HSC collaboration

plan



Wechsler & Tinker 18

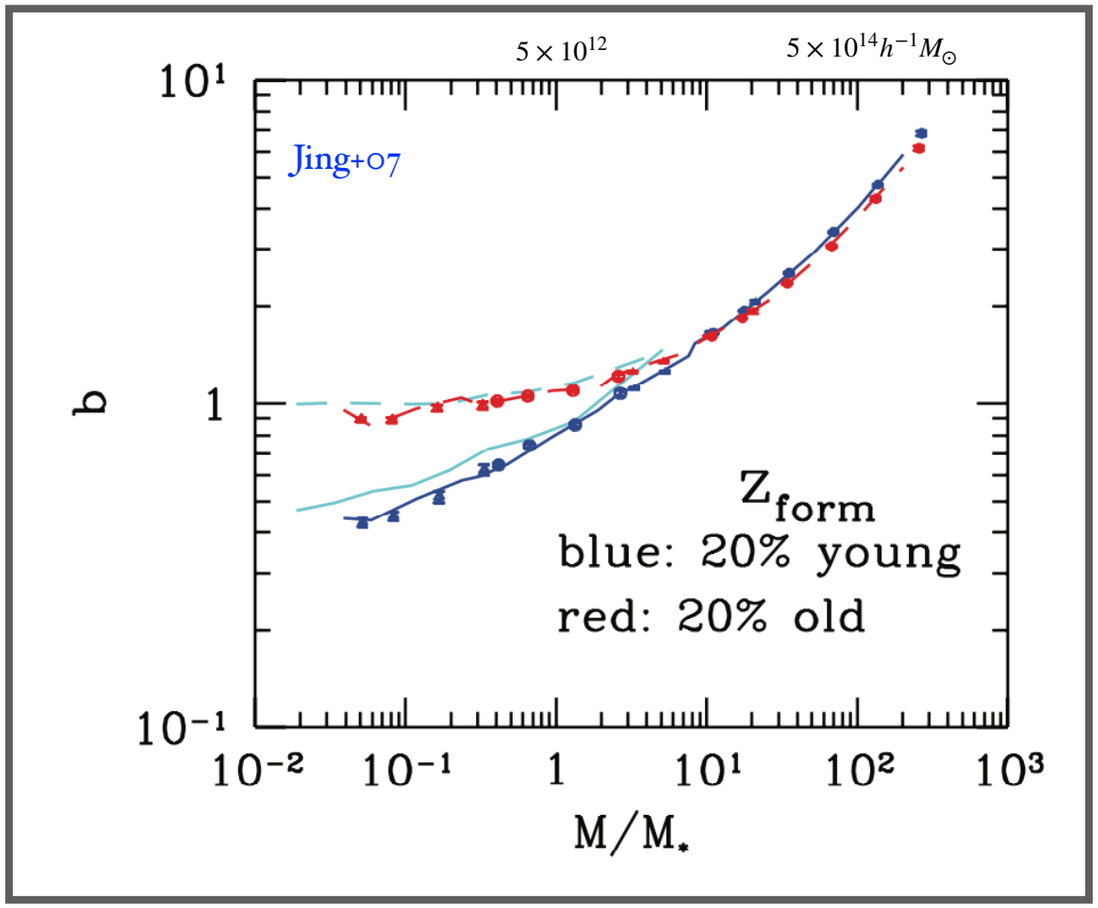
- a new method for detecting/characterizing halo assembly bias at galaxy cluster-scales
- a novel way of quantifying the stellar mass-halo mass relation at the high mass (cluster-scale) end
- a Graph Neural Network (GNN) model for predicting baryonic properties of galaxies purely from full *dark matter* halo merger trees from z=0 out to z=5!

detection of halo assembly bias

halo assembly bias is a generic feature of ΛCDM, yet it has proven difficult to detect the signal observationally

assembly bias?

- dark matter halos biased tracers of matter, with bias primarily as a function of halo mass ⇒ more massive halos are more biased
- a secondary effect is *assembly bias* (AB): bias also depends on the halo formation time
  - for low mass halos ($\sim 10^{12} h^{-1} M_\odot$), those that form earlier would cluster more strongly (having ~40% larger bias)
  - for cluster-scale halos, youngest halos are ~10% more biased than oldest ones



assembly bias?

- dark matter halos biased tracers of matter, with bias primarily as a function of halo mass ⇒ more massive halos are more biased
- a secondary effect is *assembly bias* (AB): bias also depends on the halo formation time
  - for low mass halos ($\sim 10^{12} h^{-1} M_\odot$), those that form earlier would cluster more strongly (having ~40% larger bias)
  - for cluster-scale halos, youngest halos are ~10% more biased than oldest ones

as assembly bias is a robust prediction/feature of ΛCDM
it is important to find observational evidence for it!!

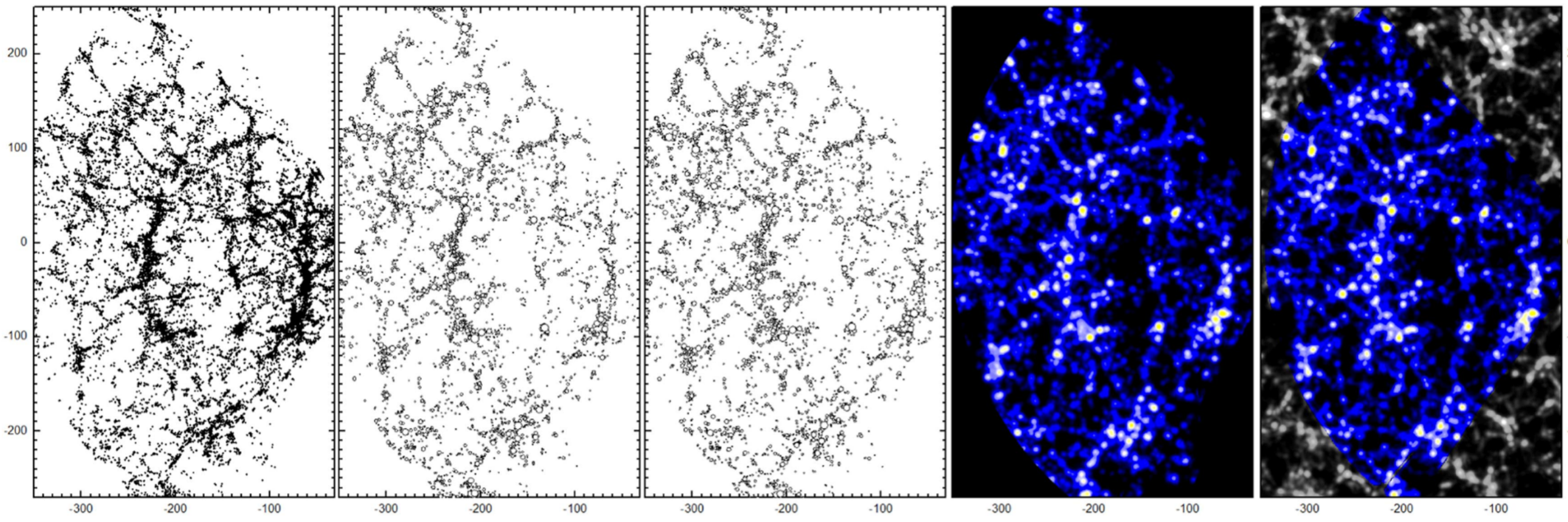
assembly bias?

- dark matter halos biased tracers of matter, with bias primarily as a function of halo mass ⇒ more massive halos are more biased
- a secondary effect is *assembly bias* (AB): bias also depends on the halo formation time
  - for low mass halos ($\sim 10^{12} h^{-1} M_{\odot}$), those that form earlier would cluster more strongly (having ~40% larger bias)
  - for cluster-scale halos, youngest halos are ~10% more biased than oldest ones
- existence of AB posses a "threat" to the halo occupation distribution (HOD) models, for which the central tenet is that halo mass is the *only* factor controlling galaxy formation
- "main stream" way to detect AB is to use *decorated HOD* (Zentner, Hearing & co.) and see how much strength AB should have to explain galaxy clustering and abundance (and weak lensing) measurements
- my approach has been to construct to samples of halos of equal mass but of different ages, then measure the differences in large-scale bias

Elucid for assembly bias

- hard-learned lesson: *reliable* proxy of halo formation time working on *individual* halo basis
- it would be a dream come true if we have the mass accretion history (MAH) of the clusters!
- using the group catalog of Yang et al. (2007), Wang et al. (2016) have run a constrained simulation of the local Universe (SDSS DR7, z<0.12) called *Elucid*

Wang+16



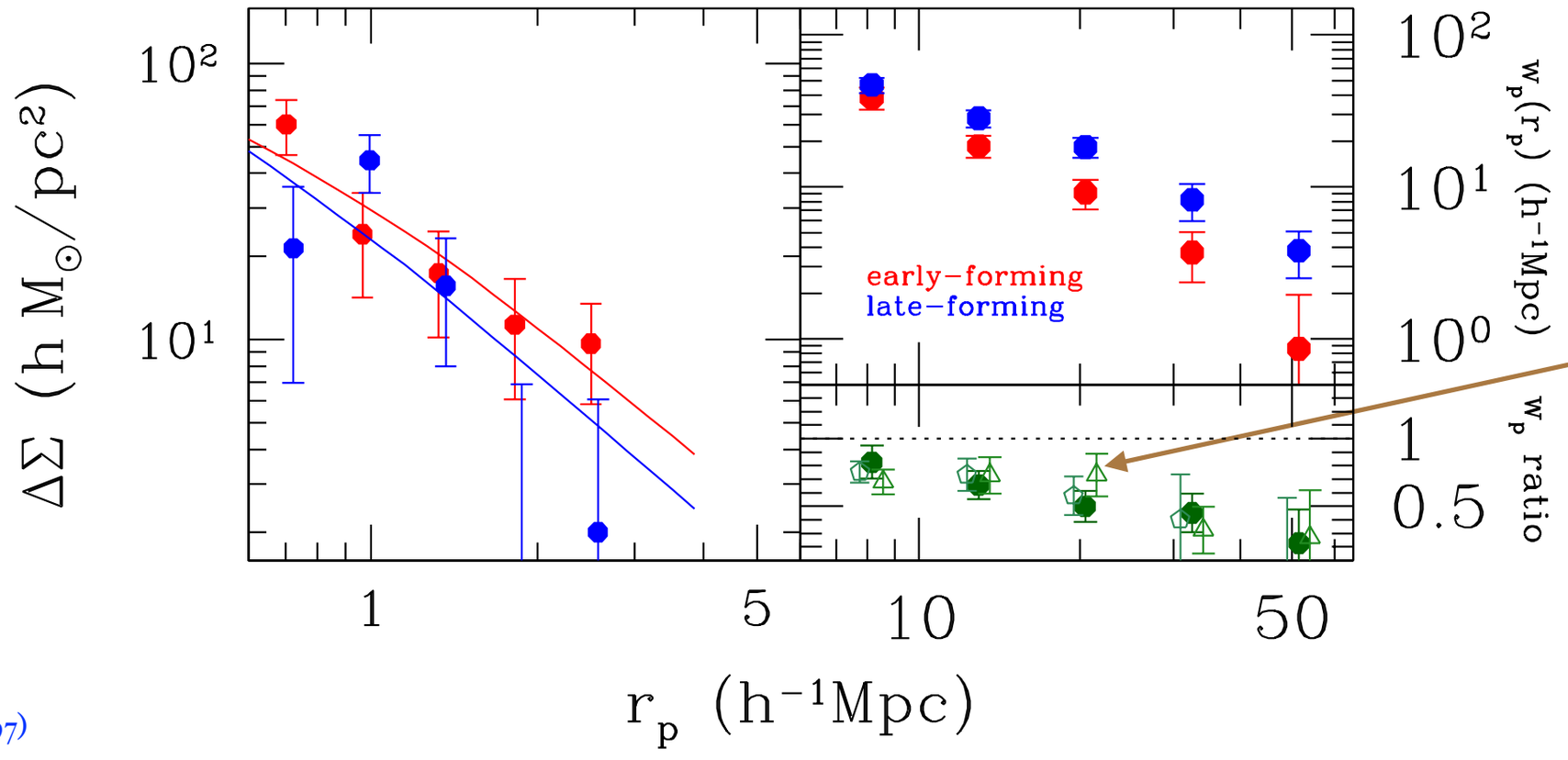
Elucid for assembly bias

- hard-learned lesson: *reliable* proxy of halo formation time working on *individual* halo basis
- it would be a dream come true if we have the mass accretion history (MAH) of the clusters!
- using the group catalog of Yang et al. (2007), Wang et al. (2016) have run a constrained simulation of the local Universe (SDSS DR7, $z<0.12$) called *Elucid*
- for structures larger than $\sim 2h^{-1}$Mpc (typical size of massive clusters), there is very good correspondence between SDSS large scale structures and *Elucid* structures
- we have selected top ~630 most massive clusters at $z<0.12$ from Yang's catalog
- MAH for each cluster is given by the counterpart halo in *Elucid*

result: z_{20}

mass based on Yang+07

- clusters split by extrema in z_{20} *and* limited in mass and redshift: consider oldest and youngest clusters (138 oldest = z_{20}>1.35 ; 121 youngest = z_{20}<0.85) with log M_{200m}=14-14.5 and z=0.06-0.12
- lensing masses *consistent within* 1σ: $M_{200m,e} = (1.3 \pm 0.3) \times 10^{14} h^{-1} M_\odot$; $M_{200m,l} = (1.0 \pm 0.3) \times 10^{14} h^{-1} M_\odot$
- large-scale clustering differs significantly ($p = 1 \times 10^{-10}$) even after accounting for differences in mass: *we have a strong detection!!!*

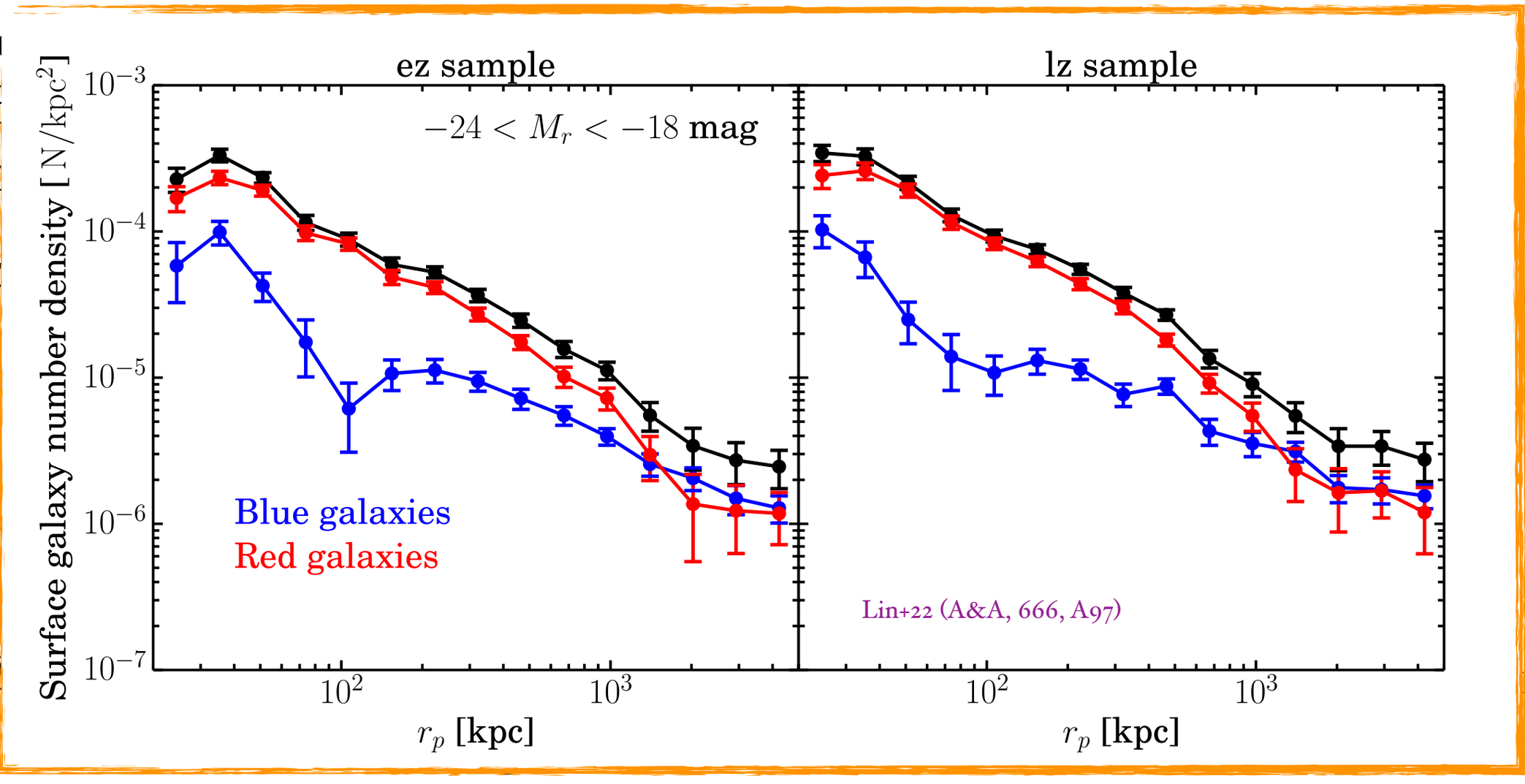


expectations from *Elucid* (p=0.00018): 98 early- and 82 late-forming halos with mean mass of $1.3 \times 10^{14} h^{-1} M_\odot$

Lin+22 (A&A, 666, A97)

properties of galaxy populations

- using a cluster-galaxy cross-correlation technique, we derive surface density profiles of member galaxies of the two cluster samples
  - concentrations of *red* galaxy distribution for the early- and late-forming clusters are $c_e = 7.1 \pm 1.7$ and $c_l = 5.6 \pm 0.6$
- n
- m
- m
- *a*



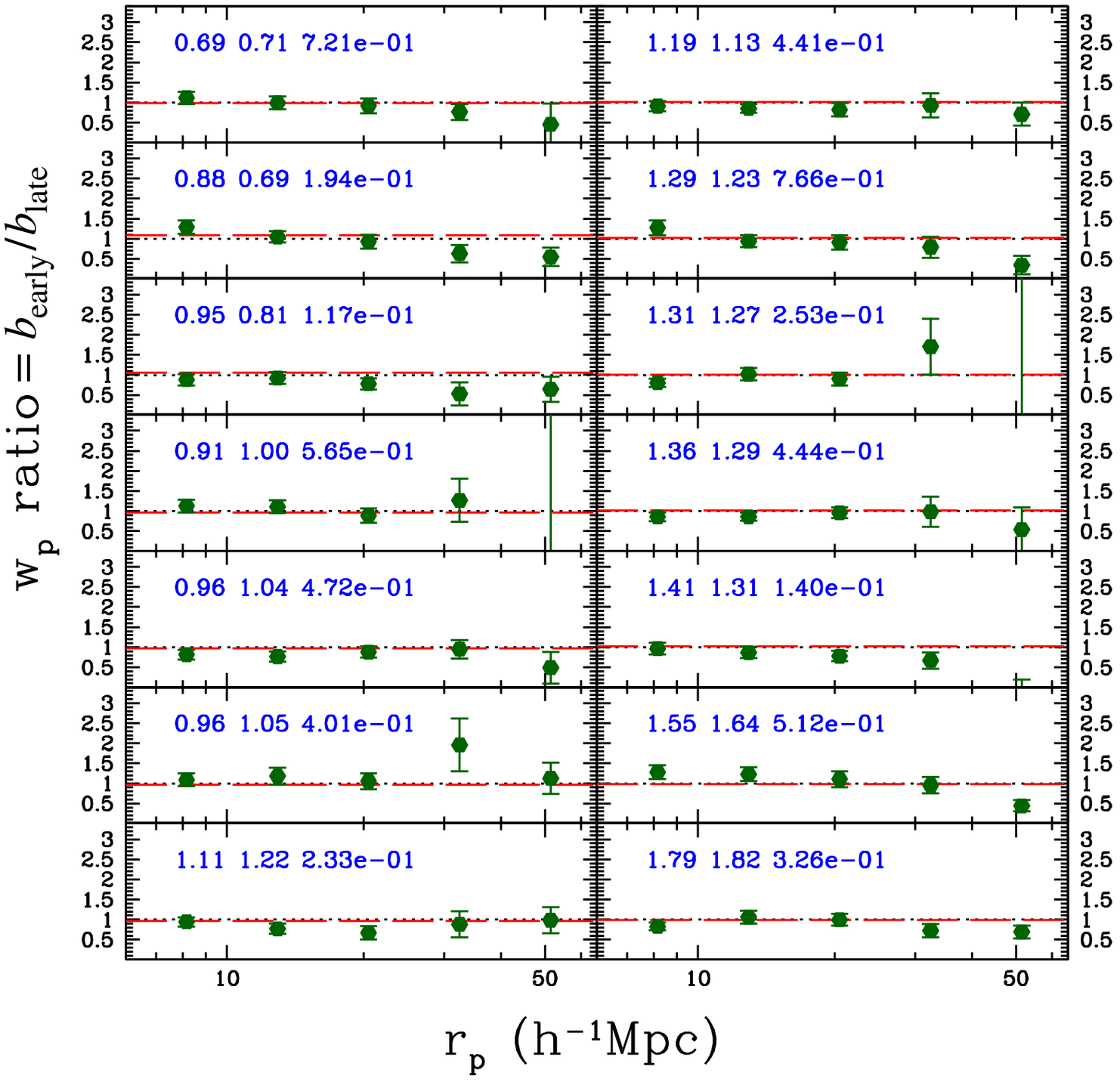
is more masses via galactic cannibalism

properties of galaxy populations

- using a cluster-galaxy cross-correlation technique, we derive surface density profiles of member galaxies of the two cluster samples
  - concentrations of *red* galaxy distribution for the early- and late-forming clusters are $c_e = 7.1 \pm 1.7$ and $c_l = 5.6 \pm 0.6$
- no appreciable differences in mean age of brightest cluster galaxy (BCG) and other member galaxies are detected using full spectral or spectral energy distribution fitting methods
- median magnitude gap between BCG and G2 (2nd most luminous galaxy): $\Delta_e = 0.44 \pm 0.01$, $\Delta_l = 0.38 \pm 0.01$
  - between BCG and G4: $\Delta_e = 0.99 \pm 0.01$, $\Delta_l = 0.87 \pm 0.01$
- median offset of BCG from cluster center: $d_e = (0.11 \pm 0.01) r_{200m}$, $d_l = (0.14 \pm 0.01) r_{200m}$
- *all* of these are consistent with the notion that the early-forming sample is indeed older, allowing BCGs to settle to the center and accrete more masses via galactic cannibalism

null tests

- we have constructed 14 pairs of random cluster samples that have similar distribution in mass & z
  - 3 numbers in blue: masses of early- & late-analog clusters, p for bias ratio to be consistent with theory (red line)
  - *none* shows signals as strong as our samples
- mean BCG offset:
  $d_e = (0.20 \pm 0.01) r_{200m}$,
  $d_l = (0.22 \pm 0.01) r_{200m}$
- median magnitude gap:
  $\Delta_e = 0.42 \pm 0.01$,
  $\Delta_l = 0.42 \pm 0.01$



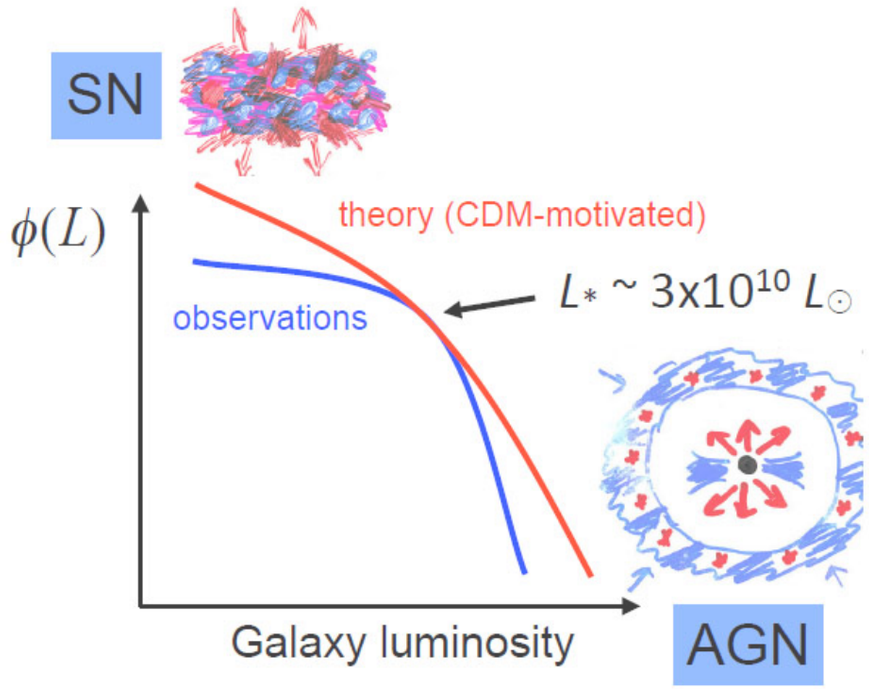
Lin+22 (A&A, 666, A97)

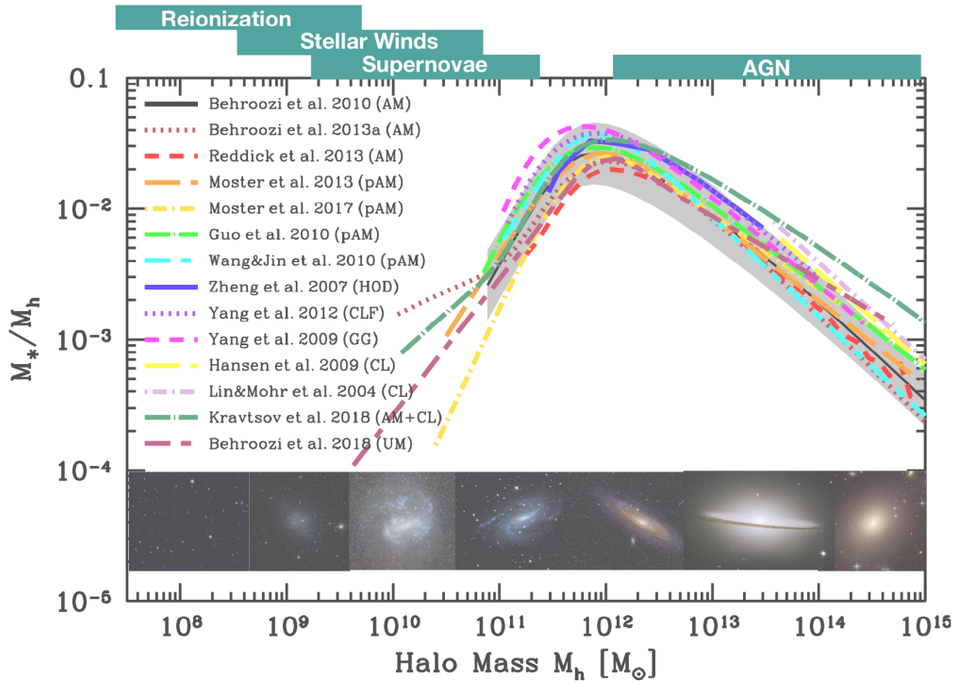
see talks by Fabian Schmidt & Xiaohu Yang

prospects

- among the first group to show a firm detection of assembly bias signal at cluster-scale halos: an important validation of ΛCDM
  - can study other aspects of assembly bias: spin or concentration
  - can further examine differences in intracluster medium
  - hard to detect splashback radius due to small sample size
- *it is still imperative to find ways that are more directly linked to observations to label clusters as early- or late-forming*
  - construct early- and late-forming samples using observable trends found in our study (member galaxy spatial concentration, galaxy number, BCG offset, magnitude gap…)
- forward-modeling techniques like that employed by *Elucid* are becoming popular (e.g., *BORG, TARDIS, COSMIC BIRTH*)
  - rich spectroscopic datasets from DESI and PFS will allow us to do reconstruction at high-z: studying assembly bias/galactic conformity!

please see Lin et al. (2022, A&A, 666, A97) for more details!





a new way to study the stellar mass-halo mass relation

Silk & Mamon 12, Wechsler & Tinker 18

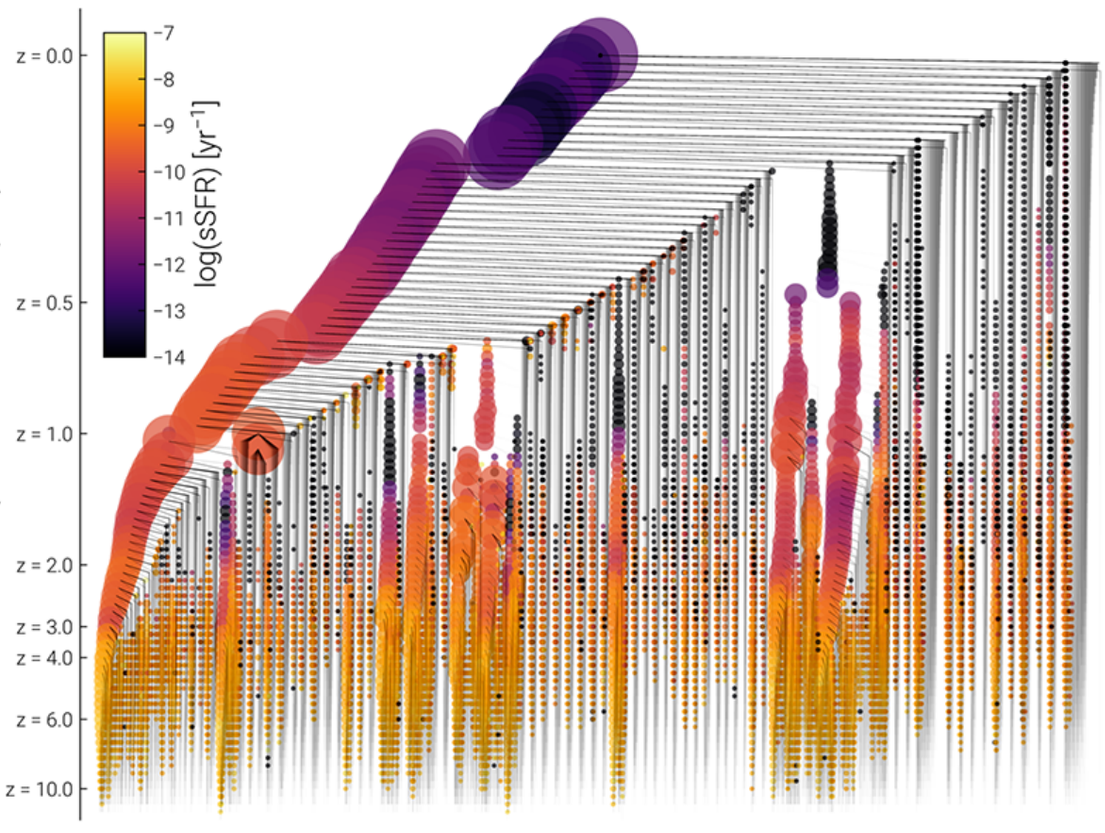
a GNN-based emulator of hydrodynamical simulations

if we succeed, it is easy to "clone" state-of-the-art hydrodynamical simulations and produce realistic mock galaxy catalogs for upcoming sky surveys

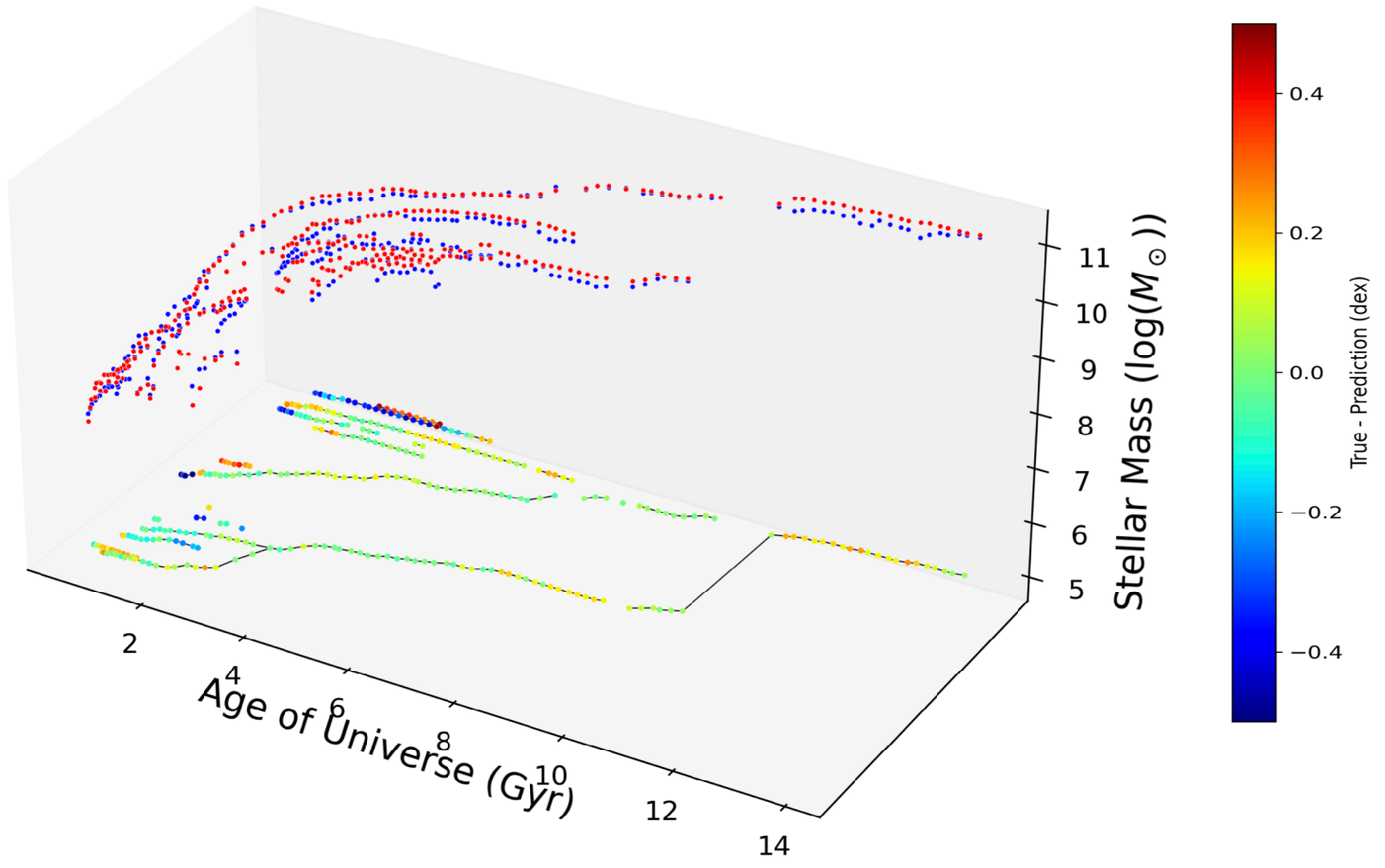
extracting max. info from cosmological simulations?

- we trained a graph neural network (GNN) model that uses *only* dark matter halo properties along the *full* merger trees of galaxies (based on IllustrisTNG300), to *accurately* predict important galaxy properties such as stellar & gas mass, star formation rate, color, stellar & gas metallicity at *z=0-5*, *anywhere* on the tree
- a major improvement over `Mangrove` (Jespersen et al. 2022), which can only do that at z=0

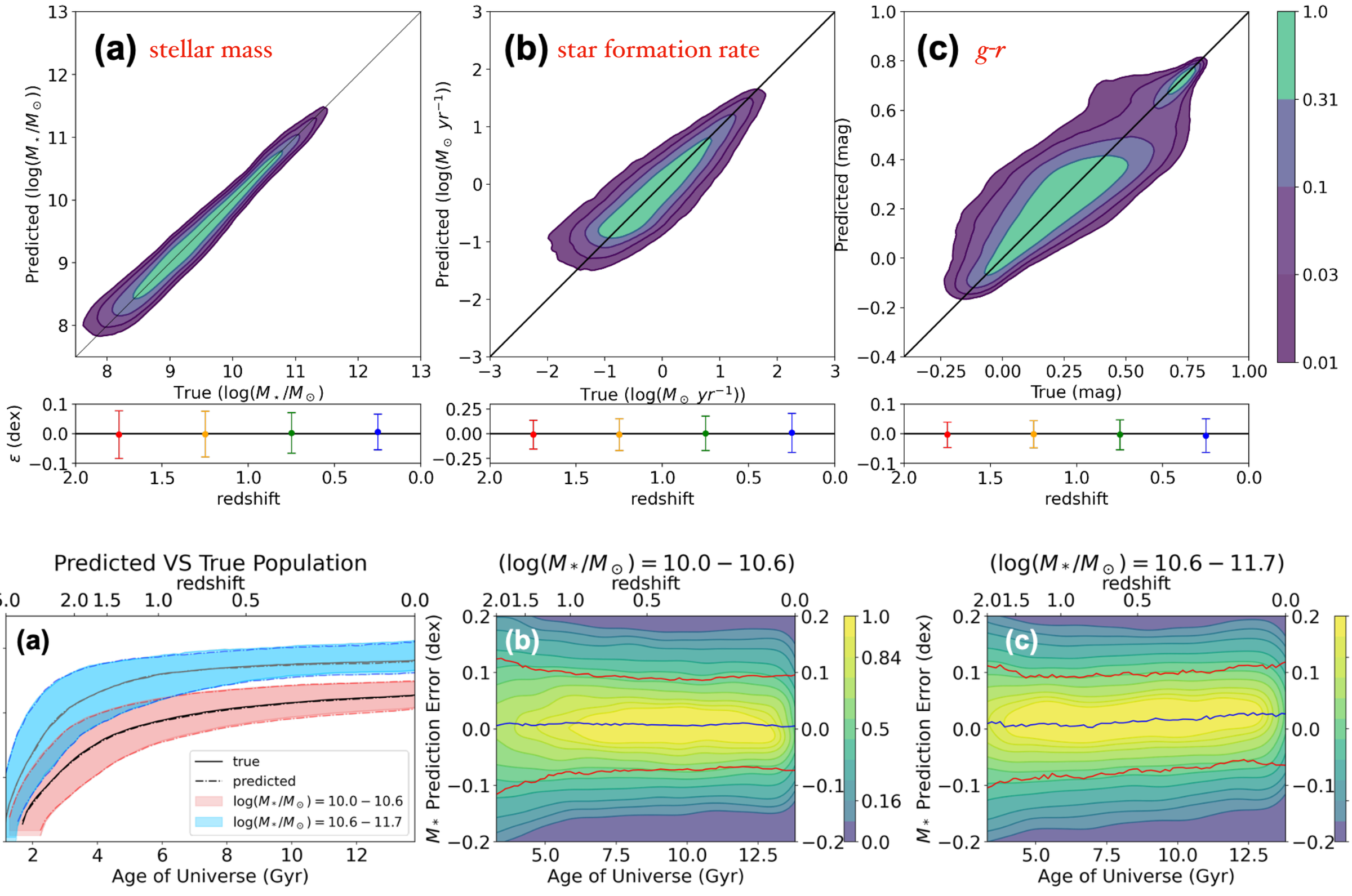
Chuang, Jespersen, Lin, Ho, & Genel (2024, ApJ, 965, 101)



IllustrisTNG



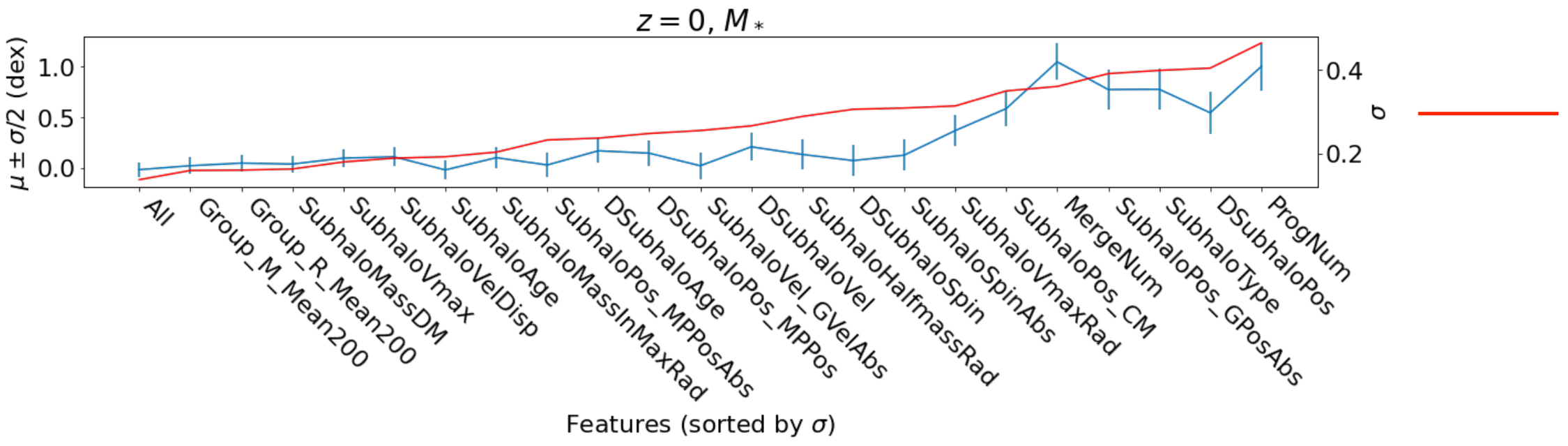
using GNN to predict galaxy properties

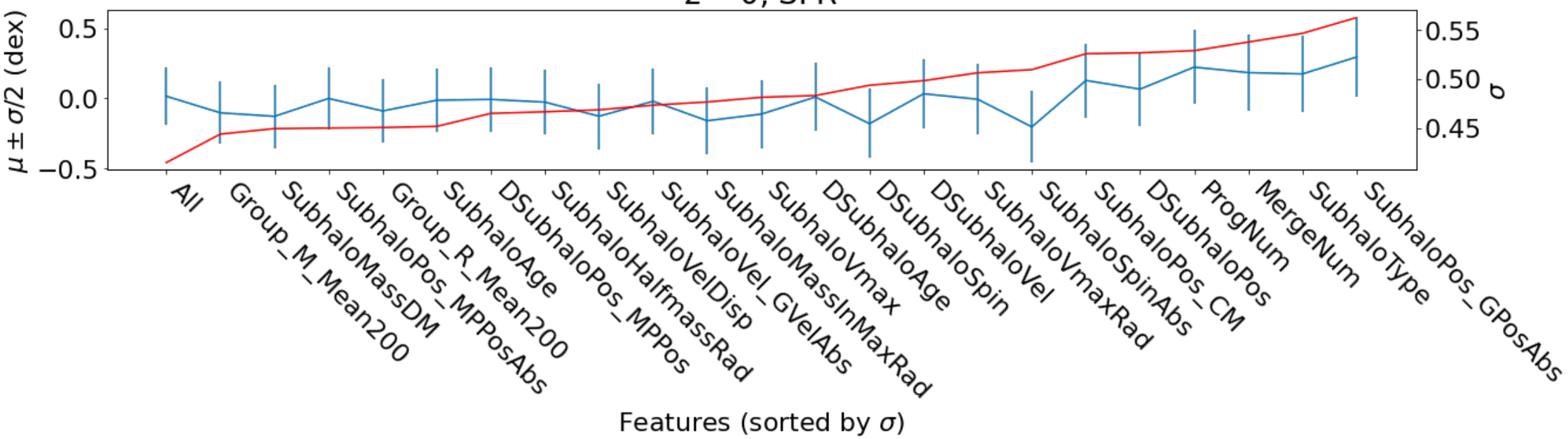


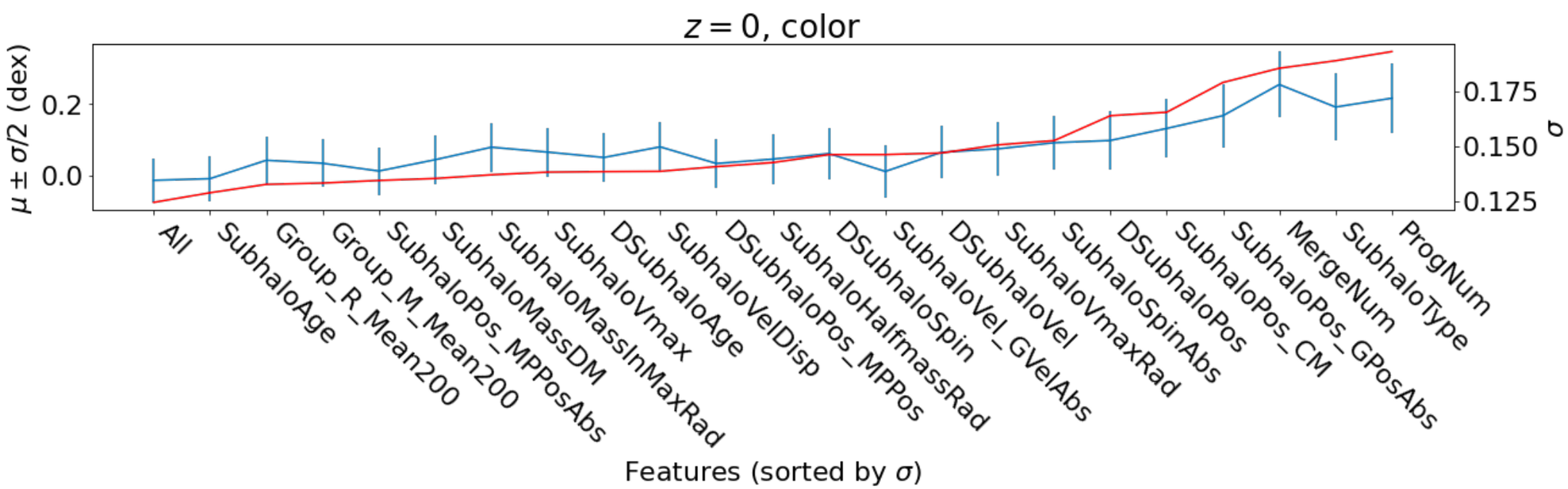
Chuang, Jespersen, Lin, Ho, & Genel (2024, ApJ, 965, 101)

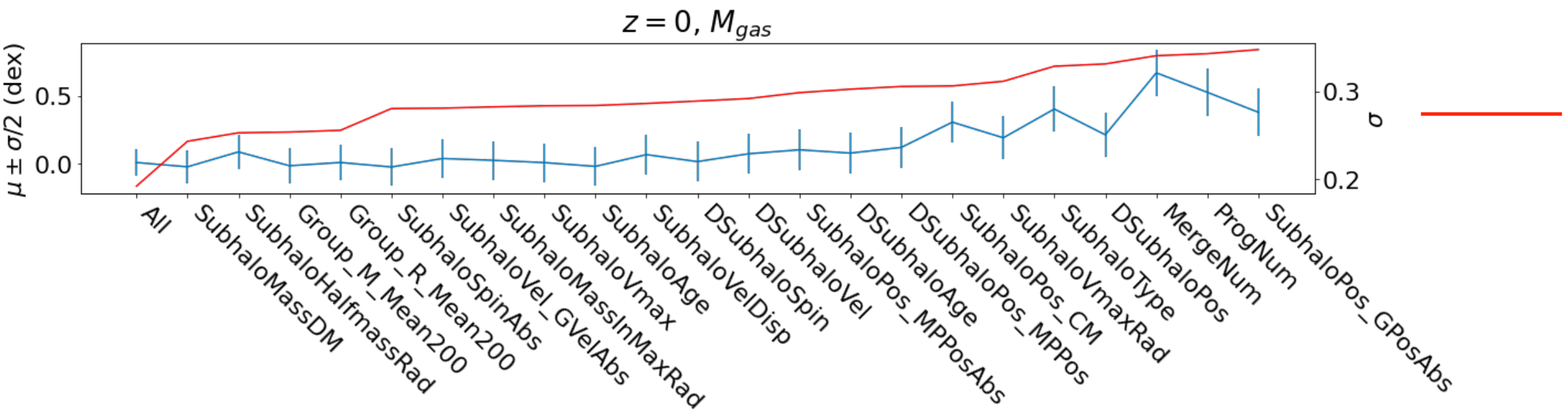
learning physics of galaxy formation

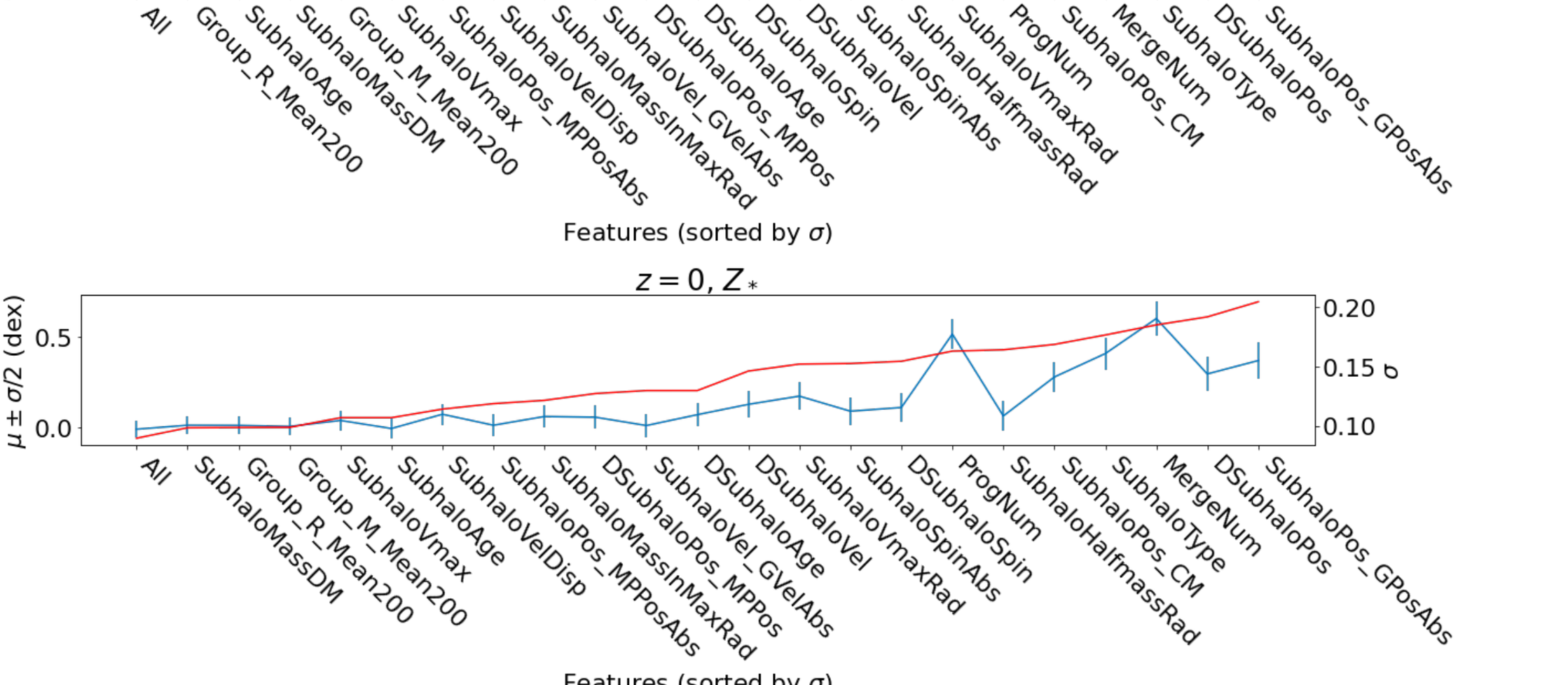
- features used in the model include
  - dark matter halo mass, half-mass radius, V_{max}, radius where V_{max} is, mass within r_{Vmax}, velocity dispersion, spin, relative position of subhalo with respect to halo center, velocity offset between subhalo and halo
  - difference in position of subhalo between different snapshots, difference in velocity of subhalo between different snapshots, difference in spin of subhalo between different snapshots
  - for a given tree: number of progenitors, number of mergers
- by systematically removing the features, the model performance should degrade
  - the ones that affect the performance most should be most relevant for galaxy formation


$z = 0$, M_*
$\mu \pm \sigma/2$ (dex)
1.0
0.5
0.0
0.4
0.2
σ
All
Group_M_Mean200
Group_R_Mean200
SubhaloMassDM
SubhaloVmax
SubhaloVelDisp
SubhaloAge
SubhaloMassInMaxRad
SubhaloPos_MPPosAbs
DSubhaloAge
DSubhaloPos_MPPos
SubhaloVel_GVelAbs
DSubhaloVel
SubhaloHalfmassRad
DSubhaloSpin
SubhaloSpinAbs
SubhaloVmaxRad
SubhaloPos_CM
MergeNum
SubhaloPos_GPosAbs
SubhaloType
DSubhaloPos
ProgNum
Features (sorted by σ)


$z = 0$, SFR
$\mu \pm \sigma/2$ (dex)
0.5
0.0
−0.5
0.55
0.50
0.45
σ
All
Group_M_Mean200
SubhaloMassDM
SubhaloPos_MPPosAbs
Group_R_Mean200
SubhaloAge
DSubhaloPos_MPPos
SubhaloHalfmassRad
SubhaloVelDisp
SubhaloVel_GVelAbs
SubhaloMassInMaxRad
DSubhaloVmax
DSubhaloAge
DSubhaloSpin
SubhaloVel
SubhaloVmaxRad
SubhaloSpinAbs
DSubhaloPos_CM
ProgNum
DSubhaloPos
MergeNum
SubhaloType
SubhaloPos_GPosAbs
Features (sorted by σ)


$z = 0$, color
$\mu \pm \sigma/2$ (dex)
0.2
0.0
0.175
0.150
0.125
σ
All
SubhaloAge
Group_R_Mean200
Group_M_Mean200
SubhaloPos_MPPosAbs
SubhaloMassDM
SubhaloMassInMaxRad
SubhaloVmax
DSubhaloAge
SubhaloVelDisp
DSubhaloPos_MPPos
SubhaloHalfmassRad
DSubhaloSpin
SubhaloVel_GVelAbs
SubhaloVel
SubhaloVmaxRad
DSubhaloPos
SubhaloSpinAbs
SubhaloPos_CM
MergeNum
SubhaloPos_GPosAbs
SubhaloType
ProgNum
Features (sorted by σ)


$z = 0$, M_{gas}
$\mu \pm \sigma/2$ (dex)
0.5
0.0
0.3
0.2
σ
All
SubhaloMassDM
SubhaloHalfmassRad
Group_M_Mean200
Group_R_Mean200
SubhaloSpinAbs
SubhaloVel_GVelAbs
SubhaloMassInMaxRad
SubhaloVmax
SubhaloAge
DSubhaloVelDisp
DSubhaloSpin
SubhaloVel
DSubhaloPos_MPPosAbs
DSubhaloAge
SubhaloPos_MPPos
SubhaloPos_CM
SubhaloVmaxRad
DSubhaloType
DSubhaloPos
MergeNum
ProgNum
SubhaloPos_GPosAbs
Features (sorted by σ)


$z = 0$, Z_{gas}
$\mu \pm \sigma/2$ (dex)
0.50
0.25
0.00
0.25
0.20
σ
All
Group_R_Mean200
SubhaloAge
SubhaloMassDM
Group_M_Mean200
SubhaloVmax
SubhaloPos_MPPosAbs
SubhaloVelDisp
SubhaloMassInMaxRad
SubhaloVel_GVelAbs
DSubhaloPos_MPPos
DSubhaloAge
DSubhaloSpin
SubhaloVel
SubhaloSpinAbs
SubhaloHalfmassRad
ProgNum
SubhaloVmaxRad
SubhaloPos_CM
MergeNum
SubhaloType
DSubhaloPos
SubhaloPos_GPosAbs
Features (sorted by σ)

$z = 0$, Z_*
$\mu \pm \sigma/2$ (dex)
0.5
0.0
0.20
0.15
0.10
σ
All
SubhaloMassDM
Group_R_Mean200
Group_M_Mean200
SubhaloVmax
SubhaloAge
SubhaloVelDisp
SubhaloPos_MPPosAbs
SubhaloMassInMaxRad
DSubhaloPos_MPPos
SubhaloVel_GVelAbs
DSubhaloAge
SubhaloVel
SubhaloVmaxRad
DSubhaloSpin
SubhaloSpinAbs
ProgNum
SubhaloHalfmassRad
SubhaloPos_CM
SubhaloType
MergeNum
DSubhaloPos
SubhaloPos_GPosAbs
Features (sorted by σ)

learning physics of galaxy formation

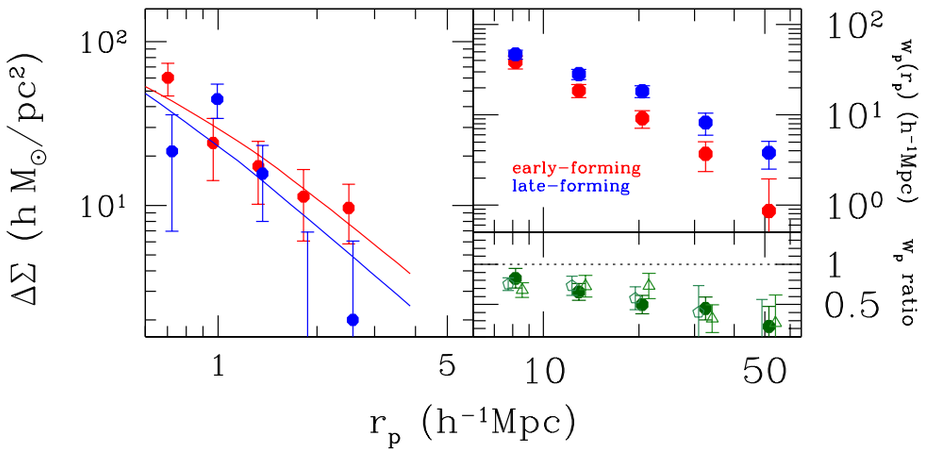
- features used in the model include
  - dark matter halo mass, half-mass radius, V_{max}, radius where V_{max} is, mass within r_{Vmax}, velocity dispersion, spin, relative position of subhalo with respect to halo center, velocity offset between subhalo and halo
  - difference in position of subhalo between different snapshots, difference in velocity of subhalo between different snapshots, difference in spin of subhalo between different snapshots
  - for a given tree: number of progenitors, number of mergers
- by systematically removing the features, the model performance should degrade
  - the ones that affect the performance most should be most relevant for galaxy formation
  - it appears that *halo mass & subhalo mass* play most critical roles!

prospects

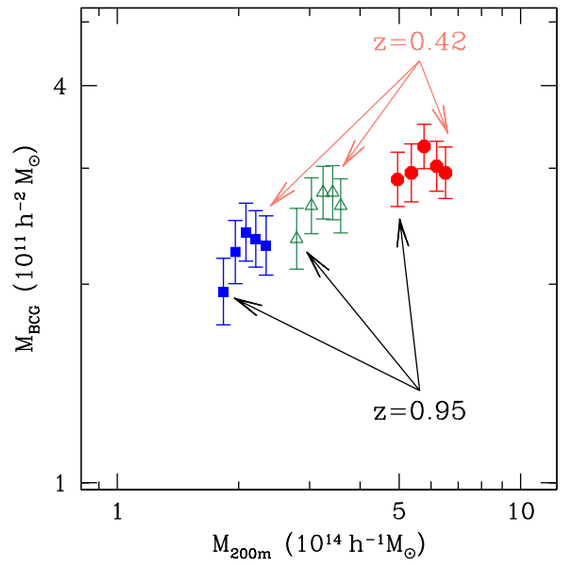
- our GNN model is fairly close to what is theoretically achievable for emulating the TNG results
- we can "port" TNG (or any other state-of-the-art) galaxy formation model to *big* N-body simulations of the same resolution easily ⇒ allowing for fast production of TNG-like mock galaxy catalogs/lightcones
- can apply our model to a high-fidelity, high-resolution constrained N-body simulation of the local Universe, and compare mock galaxies with real ones to examine the physical model of TNG, as well as to infer the star formation history of real galaxies
- may also be applied to simulation based inference framework?
- *caveat*: all of these is based on the assumption that TNG is a good enough model of the Universe
- please see Chuang et al. (2024, ApJ, 965, 101) for more details!

summary

- constrained simulation Elucid has enabled a detection of assembly bias at cluster scales



- a new approach to constrain cluster galaxy evolution, taking into account of stochasticity of hierarchical structure formation and scatter in observable-halo mass relation



- a GNN model that can serve as an emulator of *any* kind of simulations, and can be used to understand galaxy formation, as well as for generating accurate mock galaxy catalogs rapidly

