application of constrained simulations in understanding the galaxy-halo connection: the case for halo assembly bias

Yen-Ting Lin
ASIAA

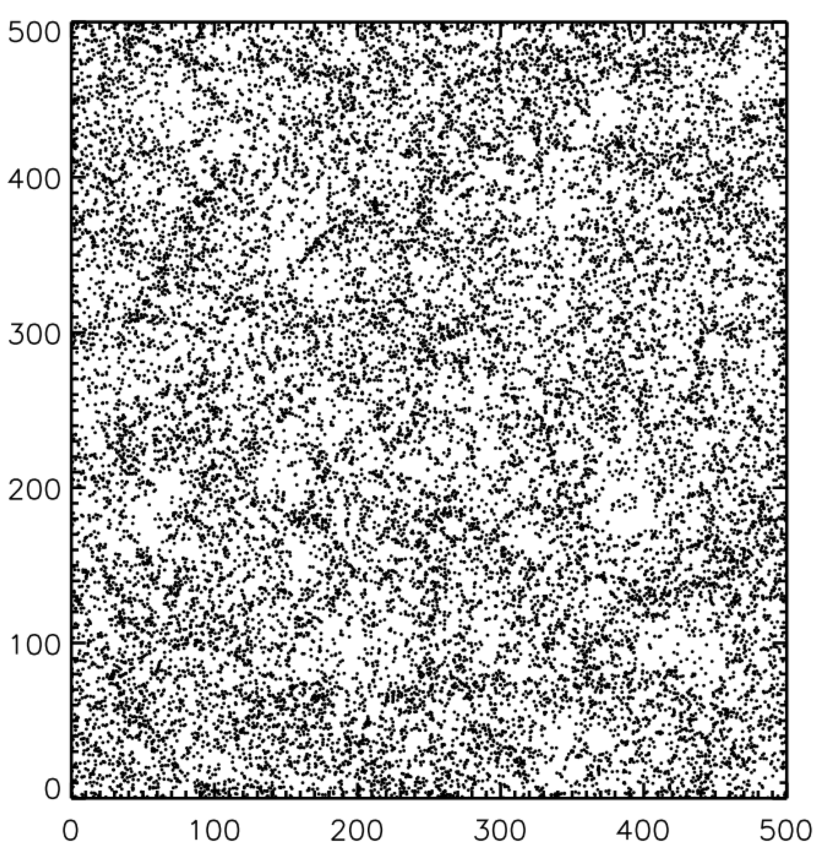
Hironao Miyatake, Hong Guo, Yi-Kuan Chiang, Kai-Feng Chen, Ting-Wen Lan, Yu-Yen Chang

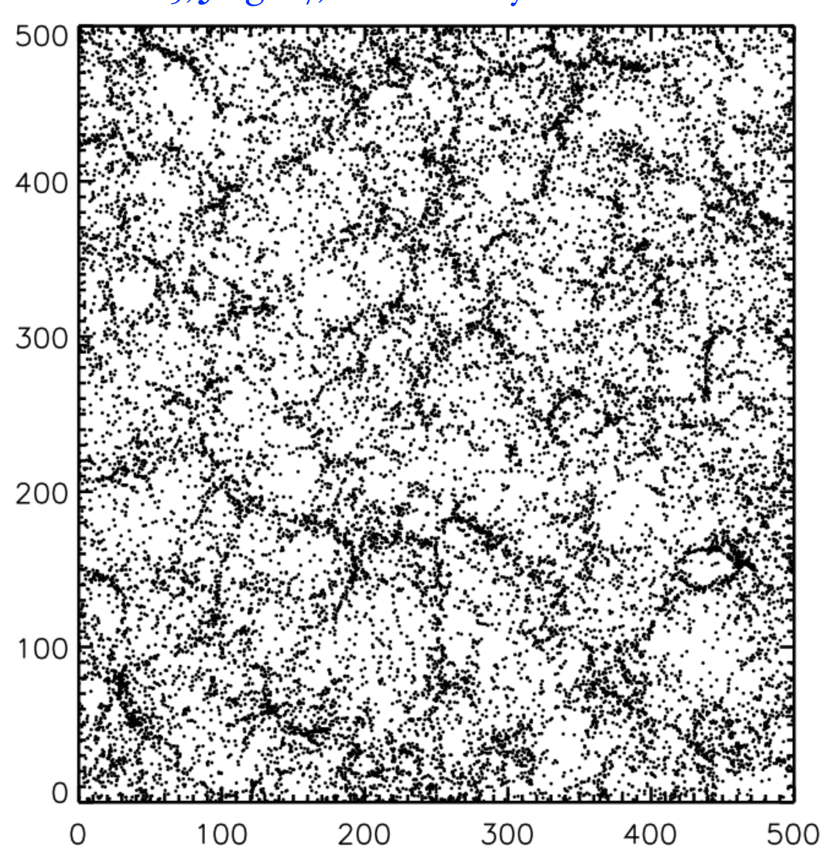
with special thanks to
Huiyuan Wang, Xiaohu Yang, & Houjun Mo

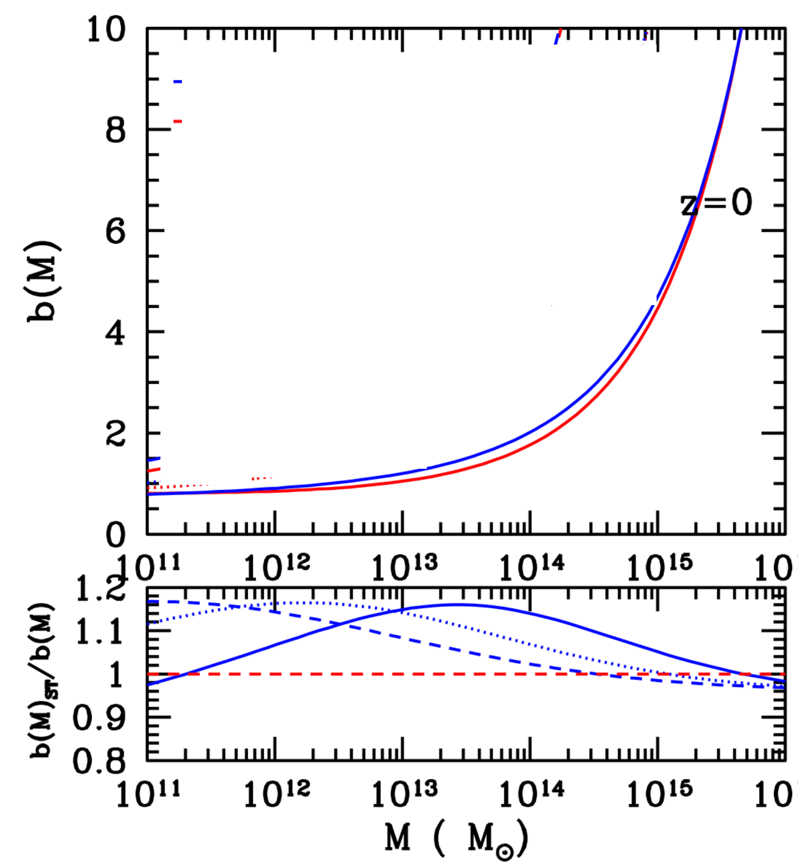
assembly bias?

- 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*: bias also depends on the halo formation time
    - for low mass halos ($\sim10^{12}h^{-1}M_\odot$), 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

Gao+05, Jing+07, Bhattacharya+11



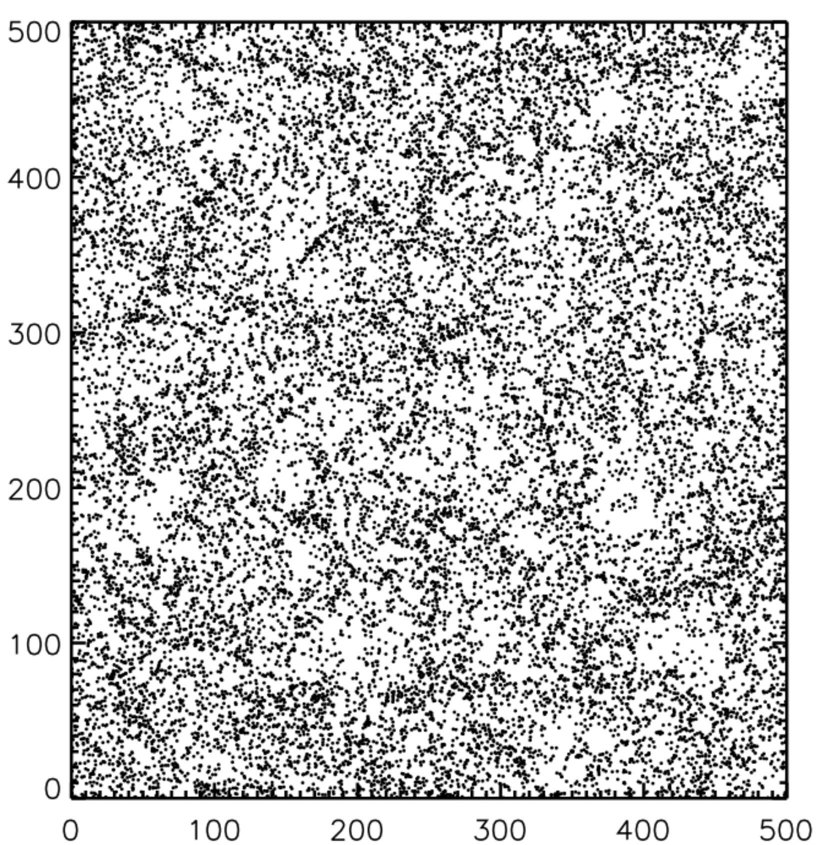


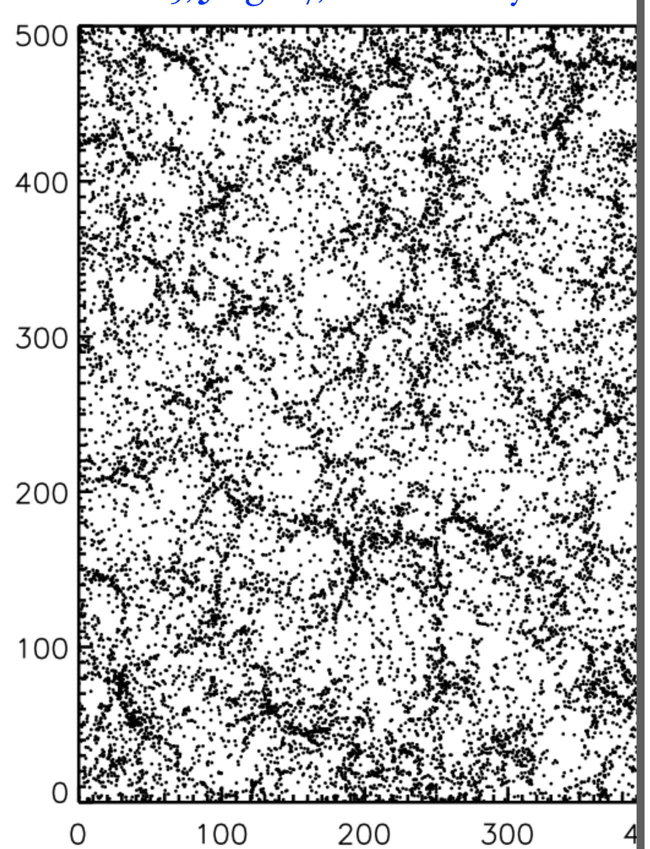


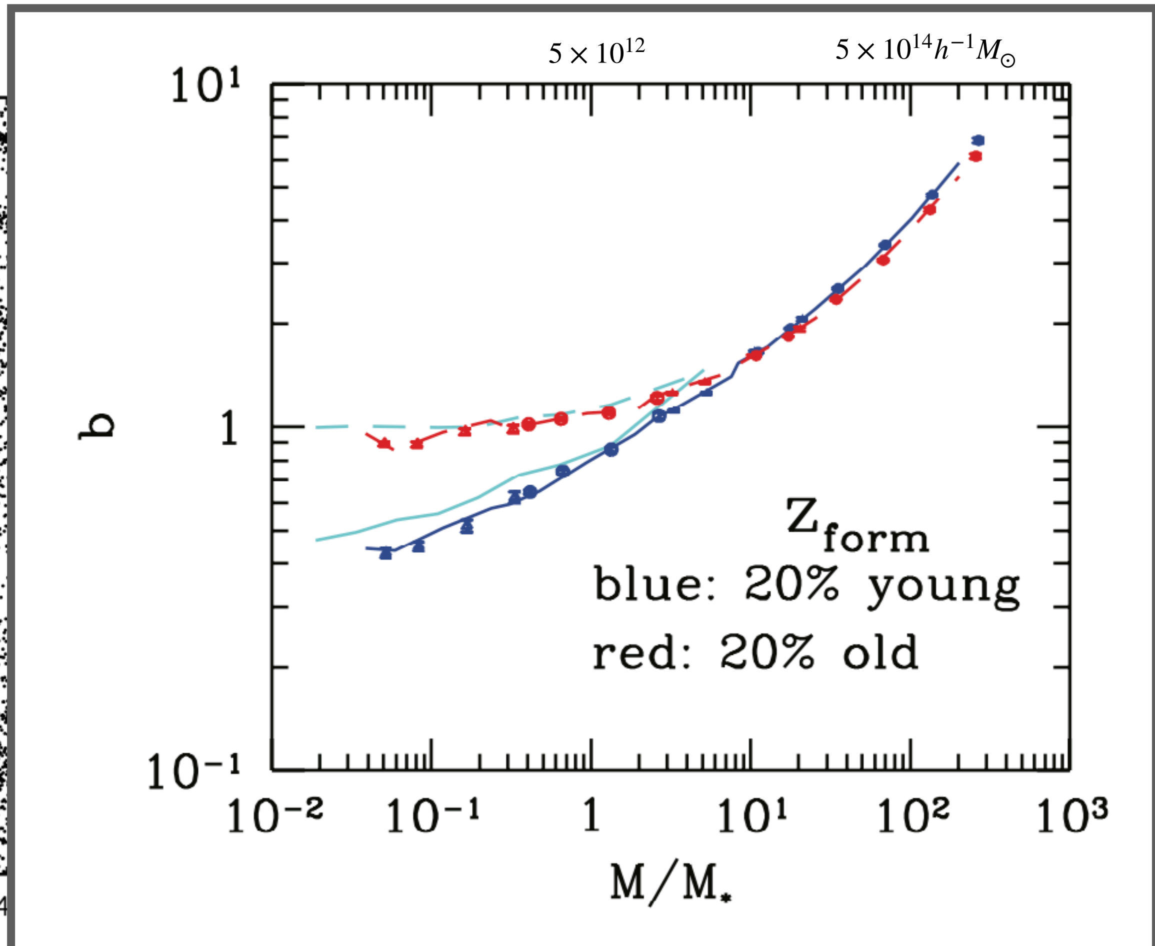
assembly bias?

- 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*: bias also depends on the halo formation time
  - for low mass halos (~$10^{12}h^{-1}M_\odot$), 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

Gao+05, Jing+07, Bhattacharya+11



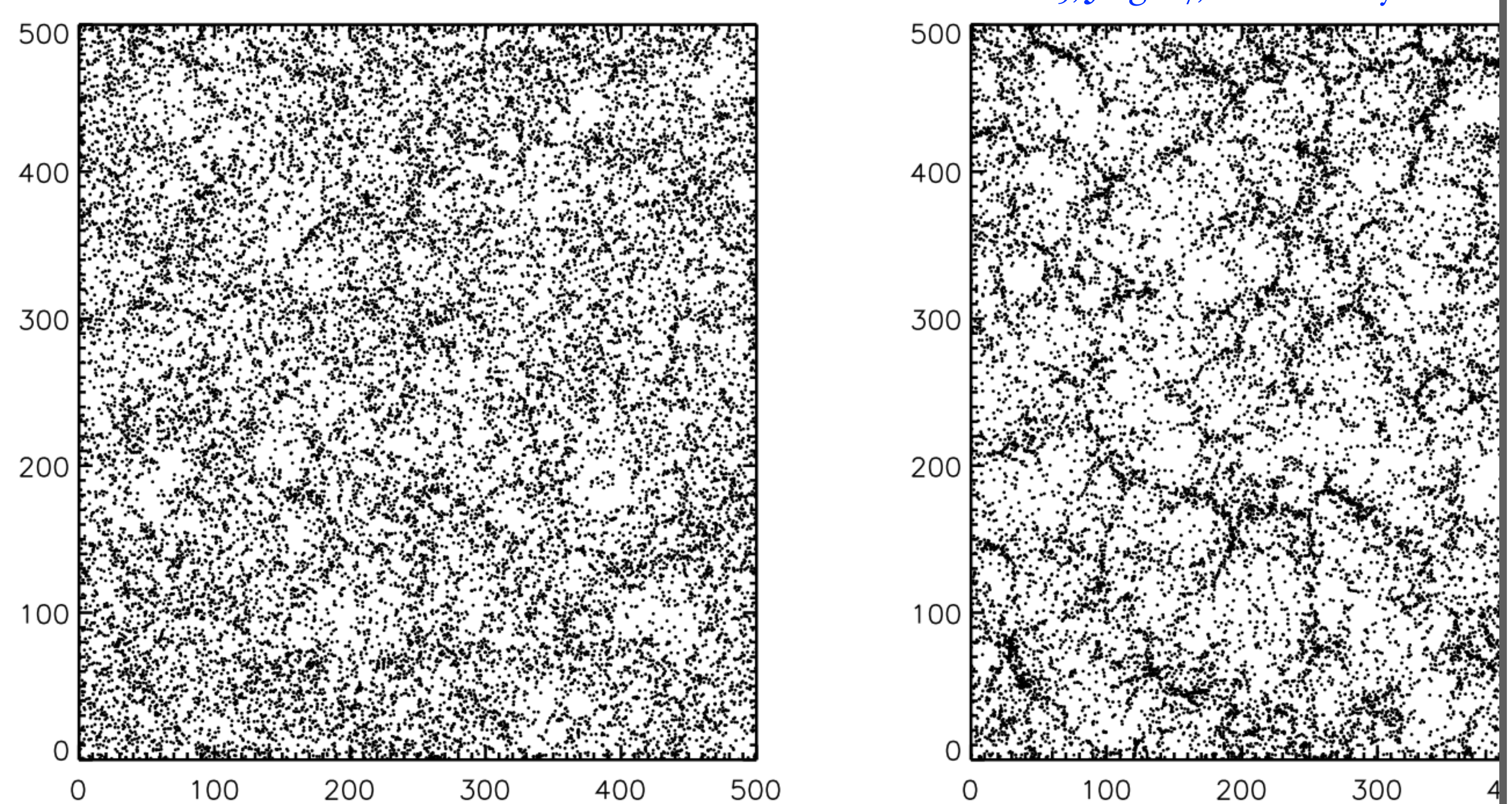


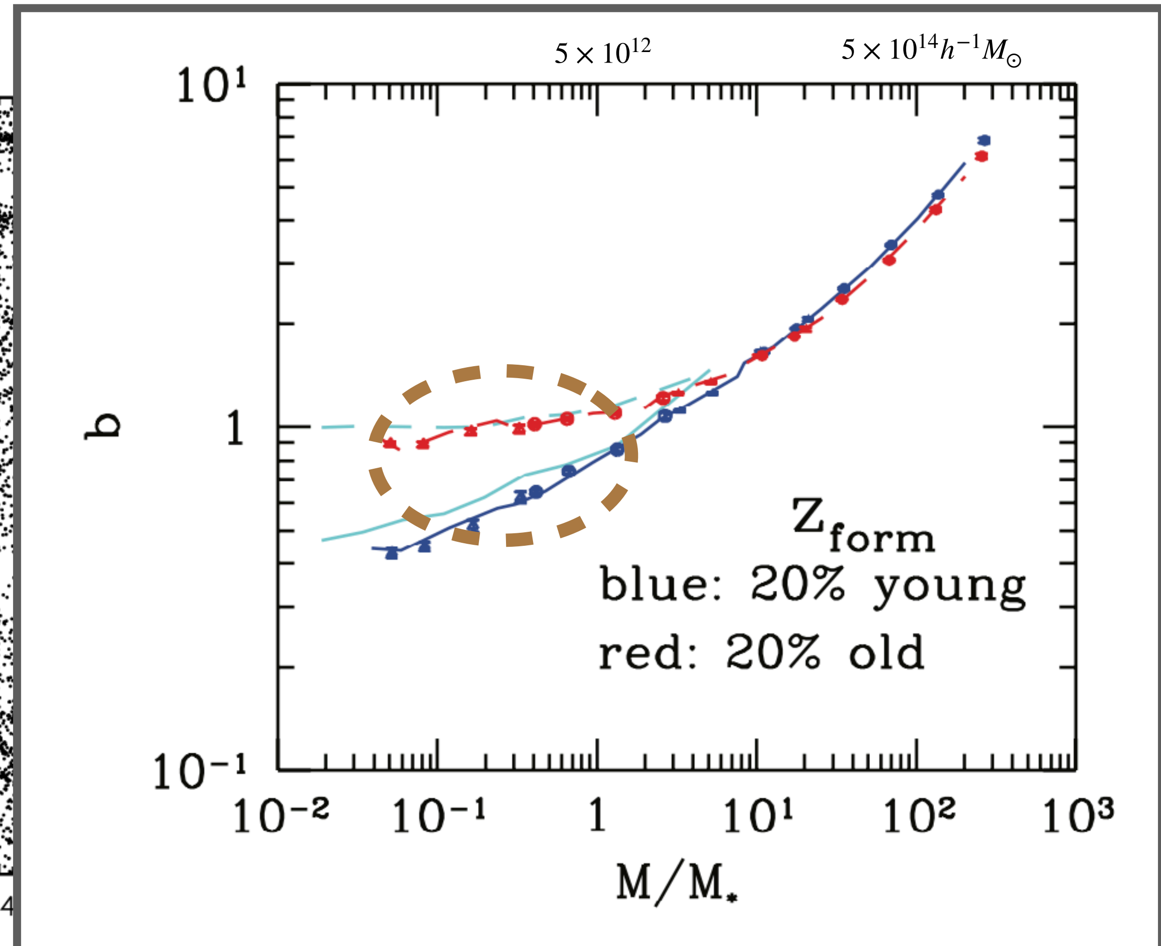


assembly bias?

- 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*: bias also depends on the halo formation time
    - for low mass halos (~$10^{12}h^{-1}M_\odot$), 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

Gao+05, Jing+07, Bhattacharya+11



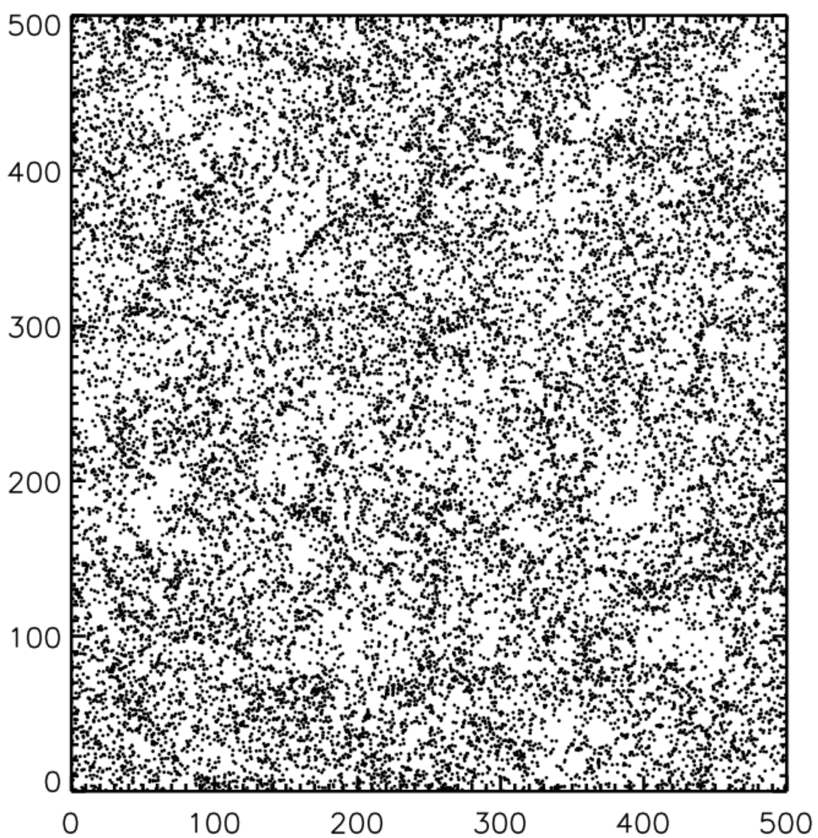


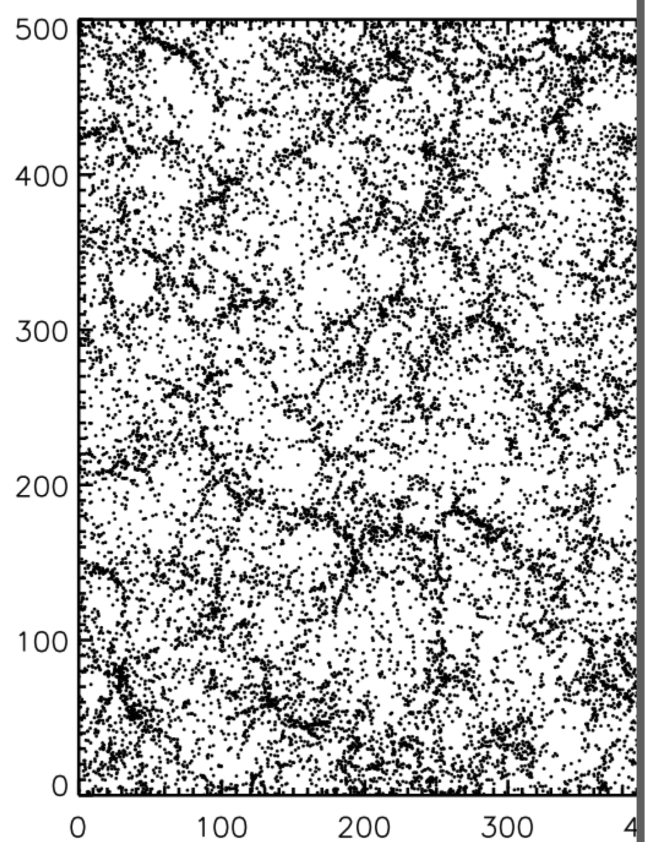
assembly bias?

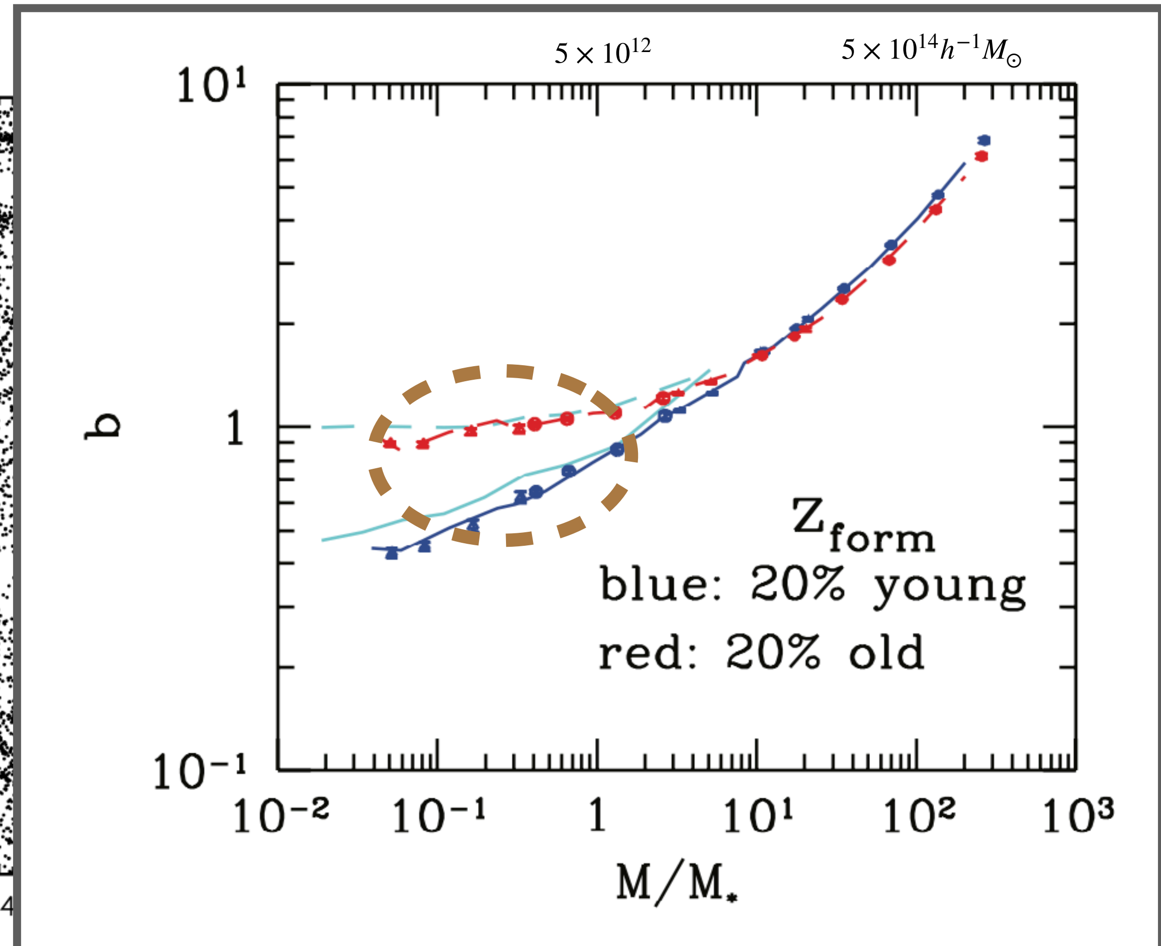
as assembly bias is a robust prediction/feature of ΛCDM
it is important to find observational evidence for it!!

- a secondary effect is *assembly bias*: bias also depends on the halo formation time
  - for low mass halos (~$10^{12}h^{-1}M_{\odot}$), 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

Gao+05, Jing+07, Bhattacharya+11





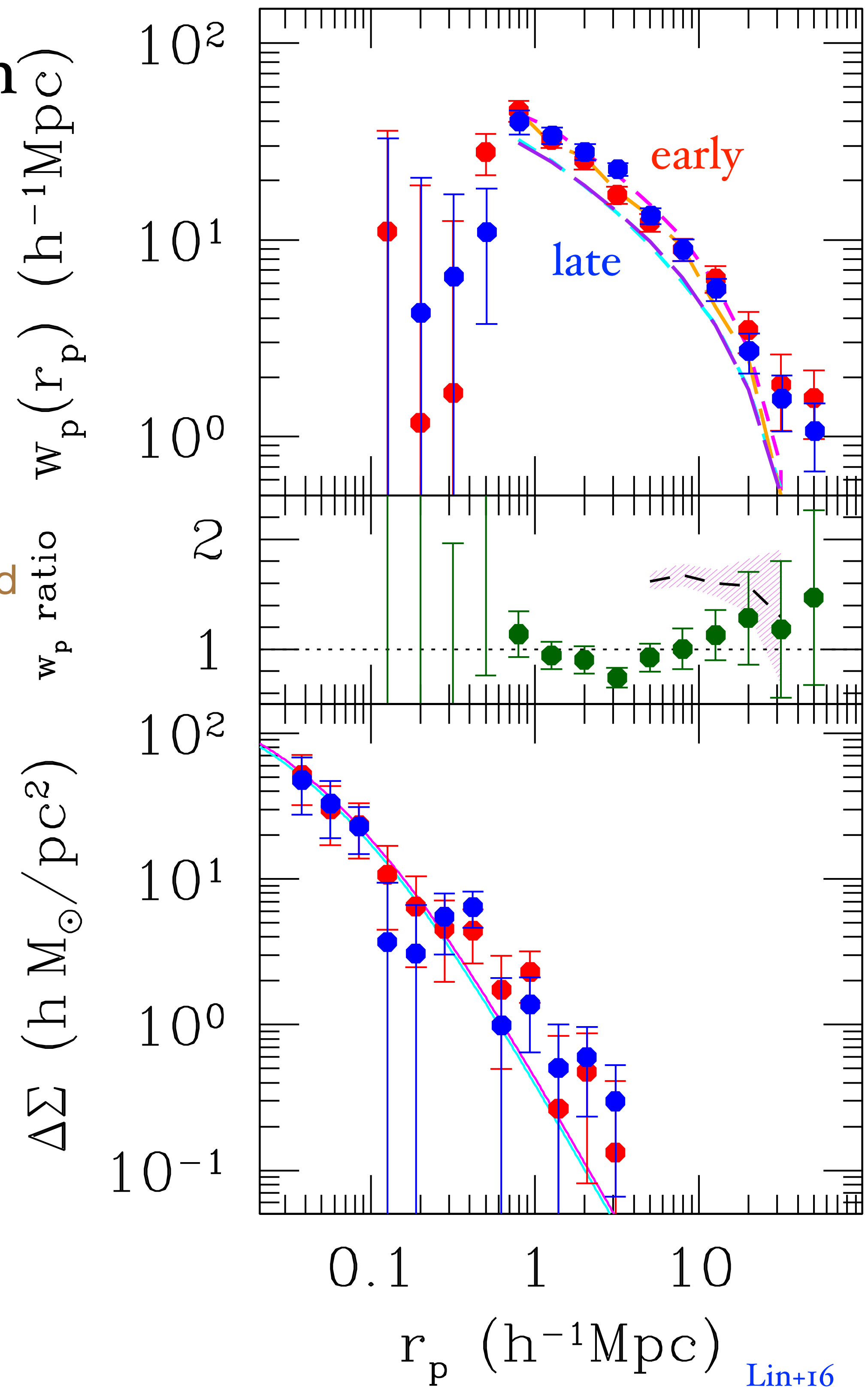


in Milky Way-like halos

non-detection of assembly bias

- we have constructed a pair of early- and late-forming halos, selected by the star formation history (SFH) of the central galaxy
  - assuming SFH of central galaxy correlates well with the formation history of the halo
- masses are $(9\pm2)\times10^{11}h^{-1}M_\odot$ and $(8\pm2)\times10^{11}h^{-1}M_\odot$
- theoretical expectation derived from N-body simulations, taking into account uncertainties in halo mass distribution
  - log-normal form assumed
  - probable values of centroid & width allowed by measured lensing signal
- probability for theory to be consistent with observation is 5×10^{-5}

early-to-late bias ratio squared



Lin+16

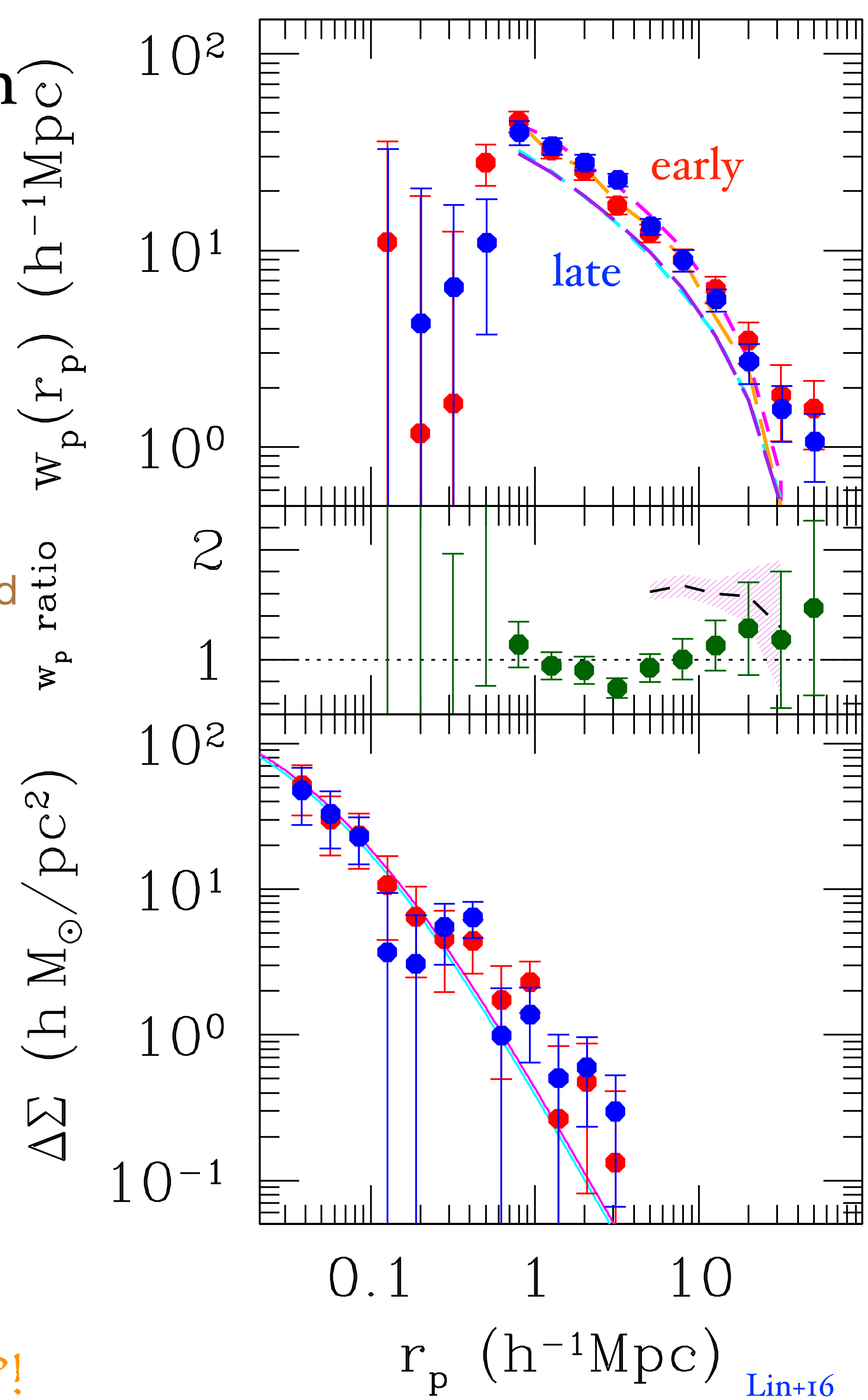
in Milky Way-like halos

non-detection of assembly bias

- we have constructed a pair of early- and late-forming halos, selected by the star formation history (SFH) of the central galaxy
  - assuming SFH of central galaxy correlates well with the formation history of the halo
- masses are $(9\pm2)\times10^{11}h^{-1}M_\odot$ and $(8\pm2)\times10^{11}h^{-1}M_\odot$
- theoretical expectation derived from N-body simulations, taking into account uncertainties in halo mass distribution
  - log-normal form assumed
  - probable values of centroid & width allowed by measured lensing signal
- probability for theory to be consistent with observation is 5×10^{-5}

early-to-late bias ratio squared

insufficient data quality?!

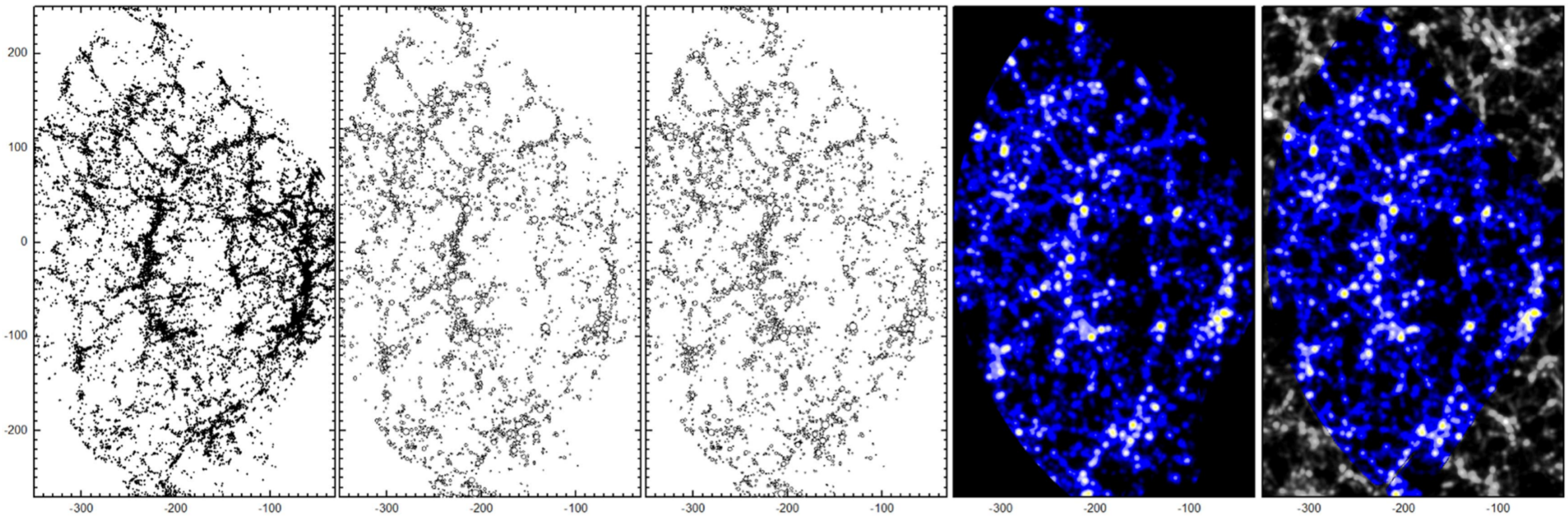


Lin+16

Elucid to the rescue

- 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



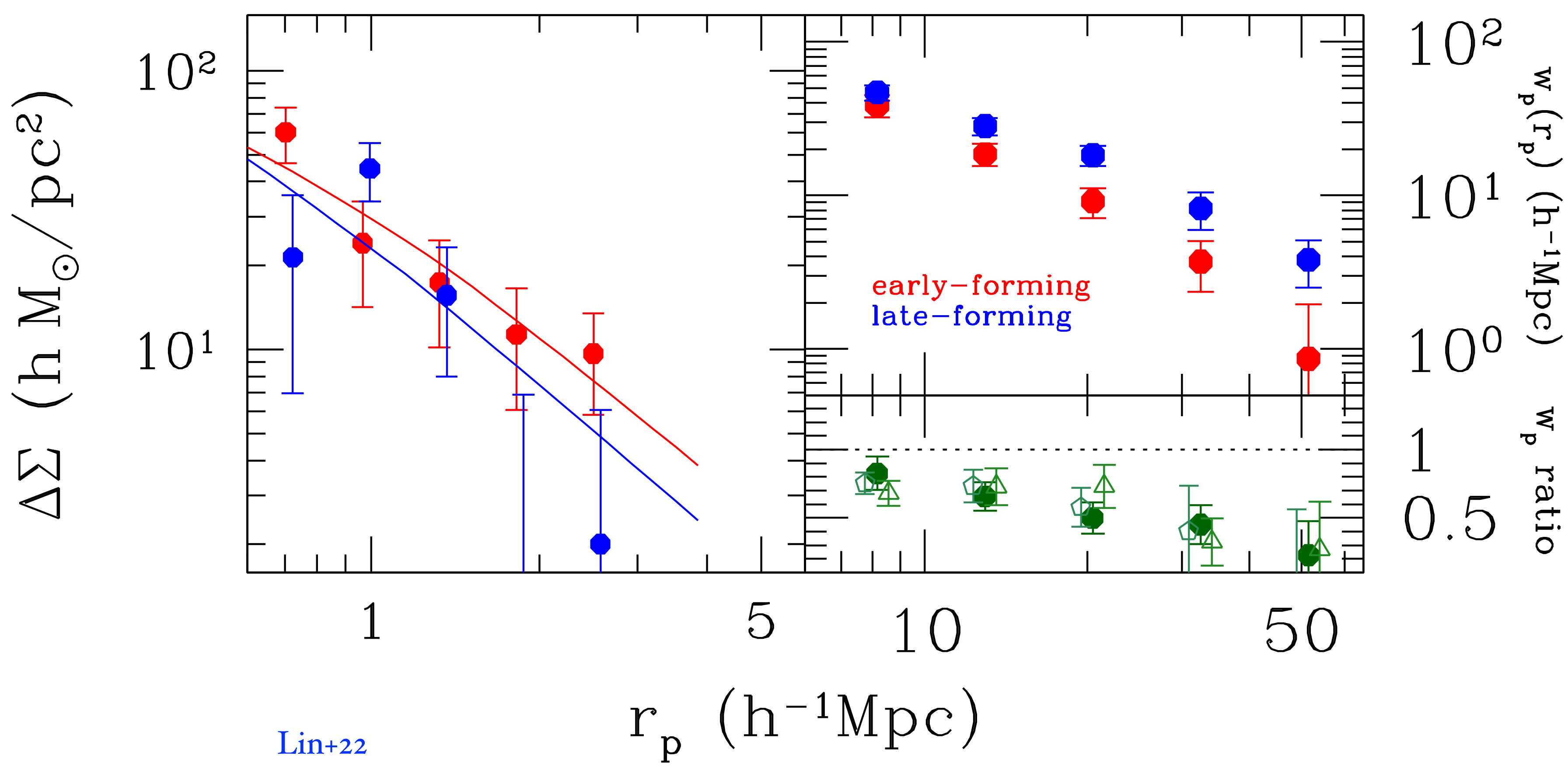
Elucid to the rescue

- 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 ~2 h^{-1}Mpc, 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*

Wang+16

result: z_{20}

- 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*: $M_{200m,e}=(1.3\pm0.3)\times10^{14}\ h^{-1}M_{\odot}$; $M_{200m,l}=(1.0\pm0.3)\times10^{14}\ h^{-1}M_{\odot}$
- large-scale clustering differs significantly ($p=1\times10^{-10}$) even after accounting for differences in mass: *we have a strong detection!!!*

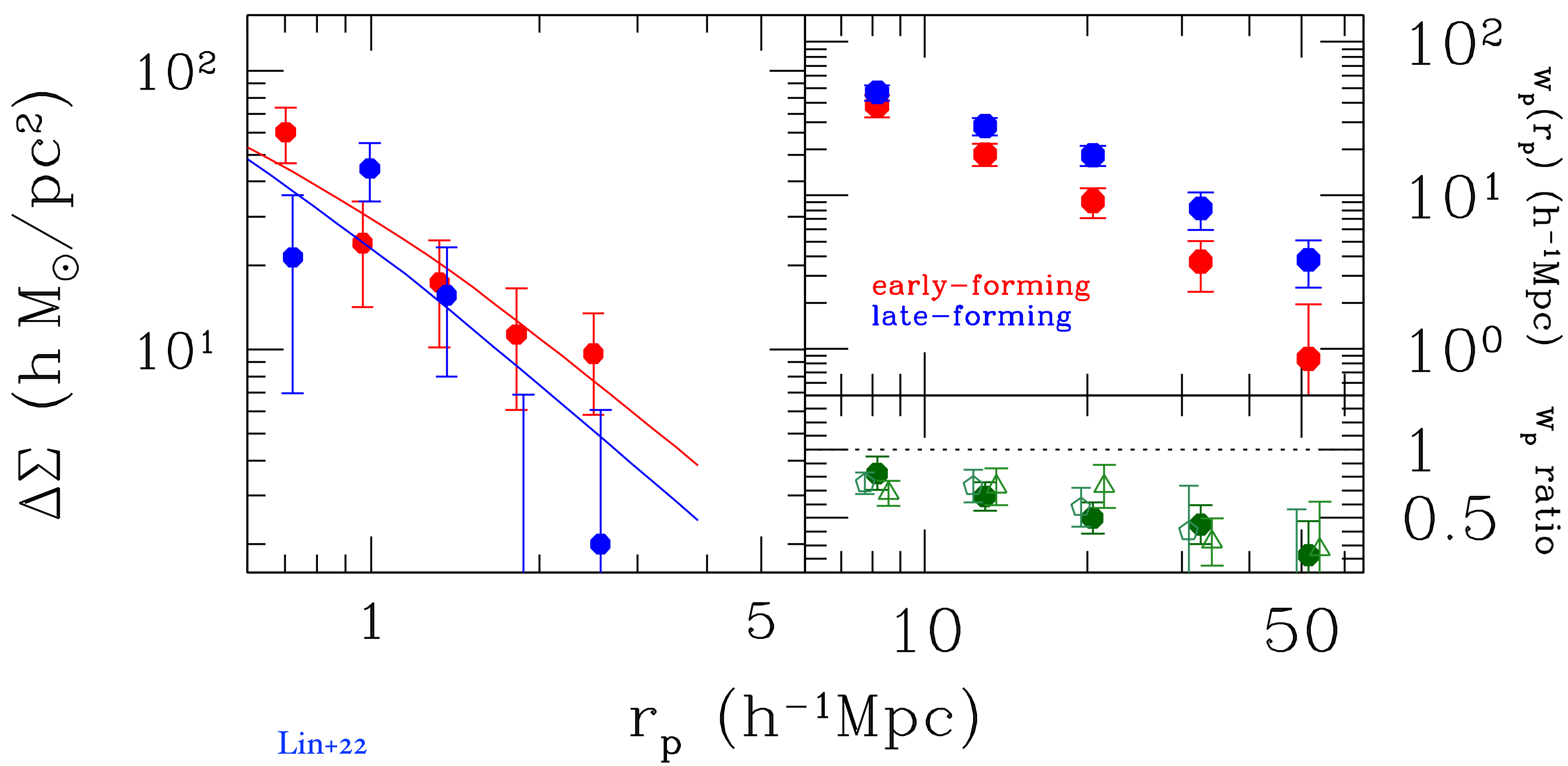


Lin+22

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*: $M_{200m,e}=(1.3\pm0.3)\times10^{14}\ h^{-1}M_\odot$; $M_{200m,l}=(1.0\pm0.3)\times10^{14}\ h^{-1}M_\odot$
- large-scale clustering differs significantly ($p=1\times10^{-10}$) even after accounting for differences in mass: *we have a strong detection!!!*

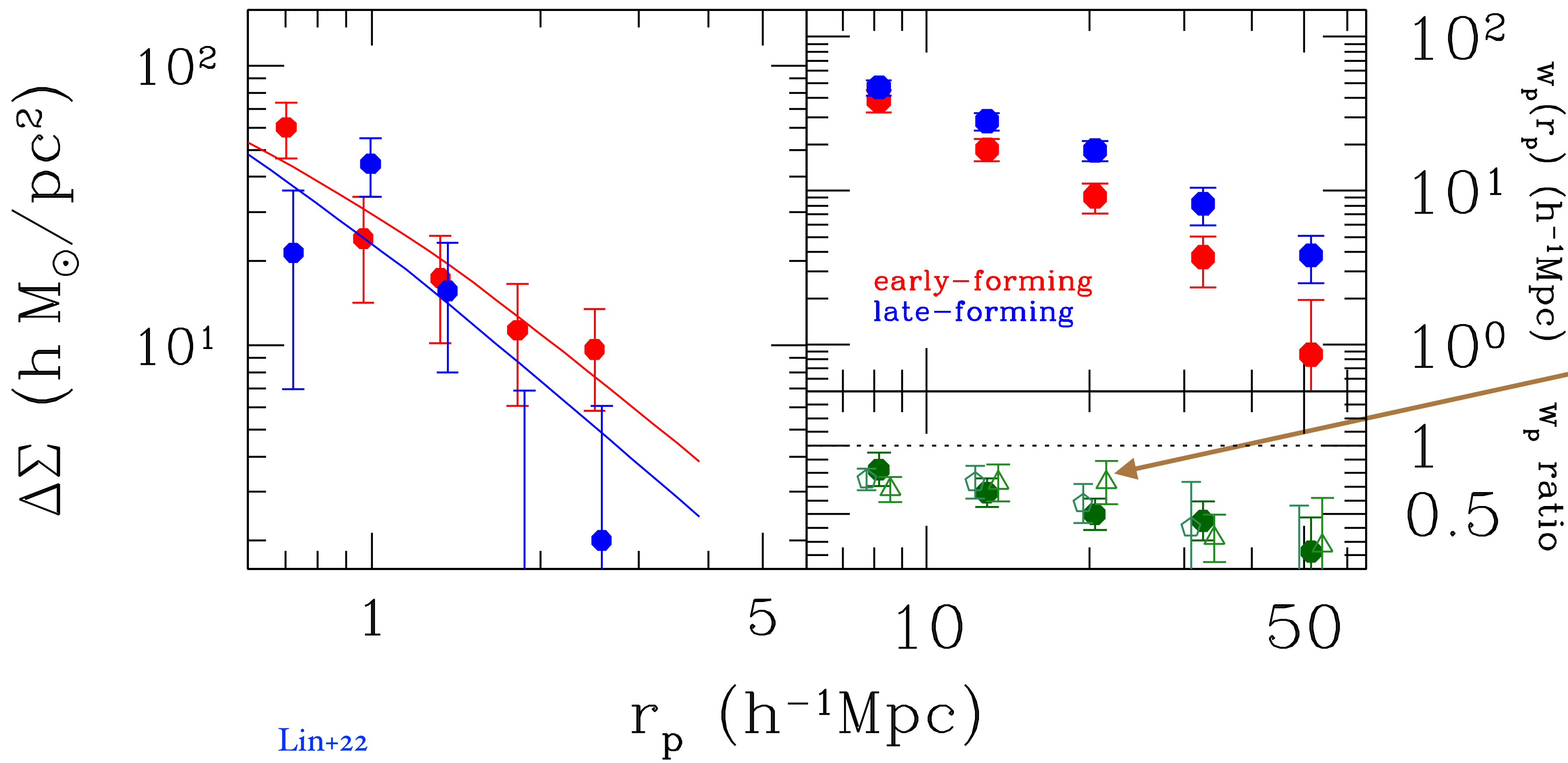


Lin+22

result: z_{20}

mass based on Yang+07

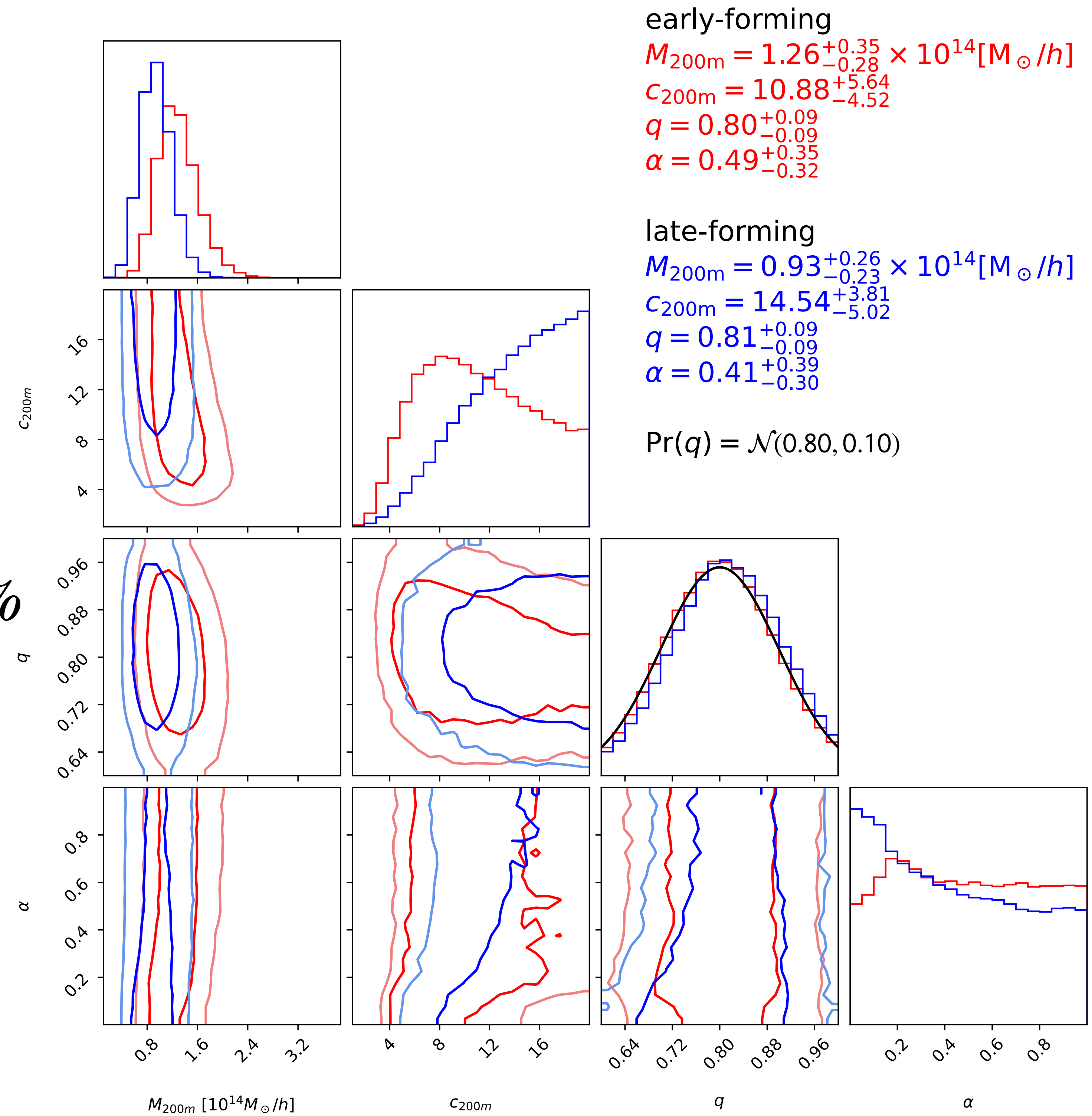
- clusters split by extrema in z_{20} *and* limited in mass and redshift: consider oldest and youngest clusters (138 oldest = $z_{20}>1.35$; 121 youngest = $z_{20}<0.85$) with log M_{200m}=14-14.5 and z=0.06-0.12
- lensing masses *consistent within 1σ*: $M_{200m,e}=(1.3\pm0.3)\times10^{14}\ h^{-1}M_\odot$; $M_{200m,l}=(1.0\pm0.3)\times10^{14}\ h^{-1}M_\odot$
- large-scale clustering differs significantly ($p=1\times10^{-10}$) 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 $1.3\times10^{14}h^{-1}M_\odot$

robustness of WL measurements?

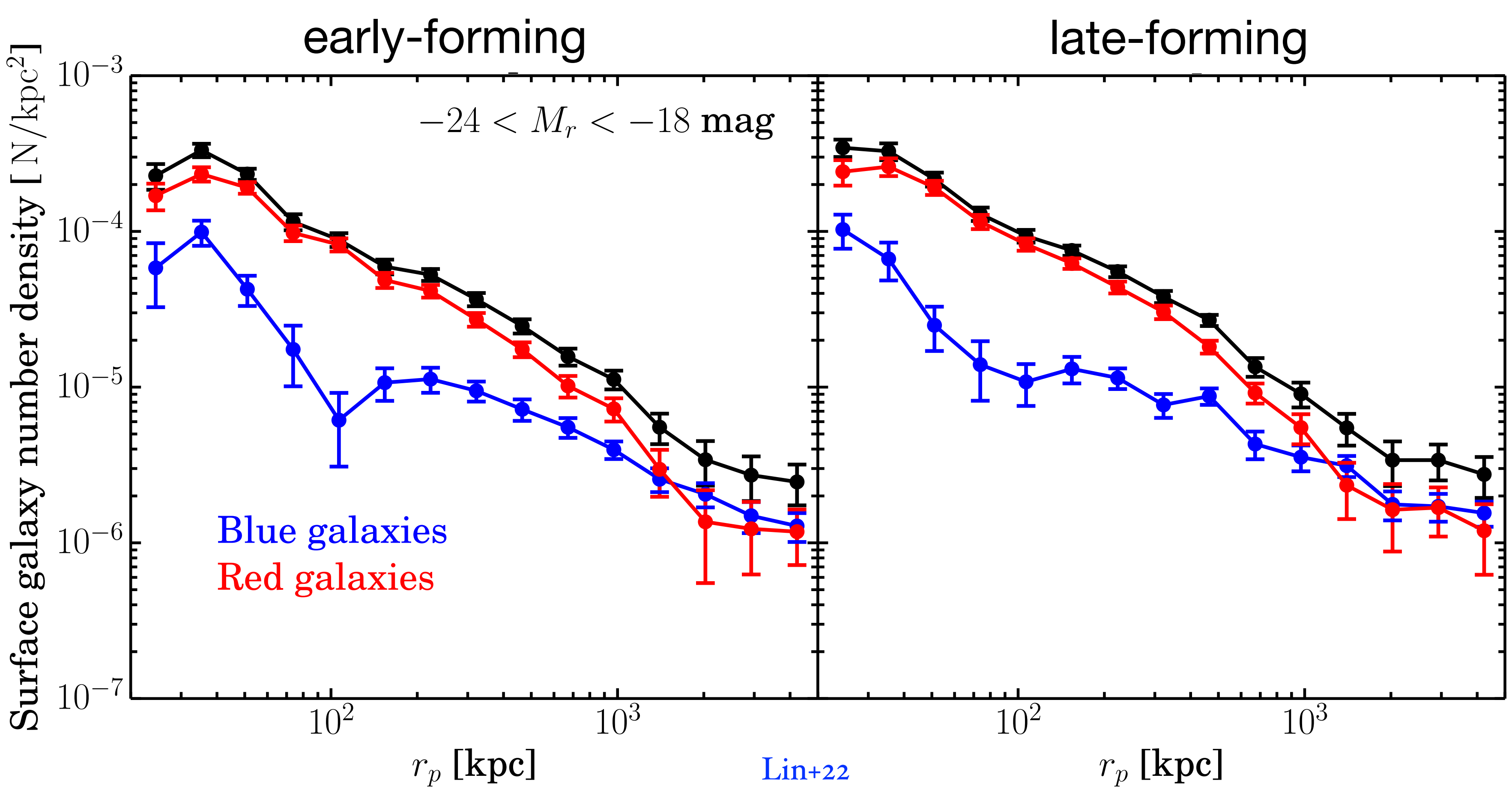
- what if our WL mass measurements were off?
    - maybe the early-forming sample mass is biased high by 1σ, while that of late-forming sample is biased low by 1σ (2.6% chance) ⇒ p=7.5x10^{-5}
    - *or* the late-forming sample mass is biased low by ≥2σ (2.3% chance) ⇒ p=2.4x10^{-5}
    - if we assume 10% uncertainty in the Tinker+08 bias-halo mass relation and artificially decrease the expected biases, the probabilities become 0.0053 and 0.0025 (about 3σ events)



Lin+22

properties of galaxy populations

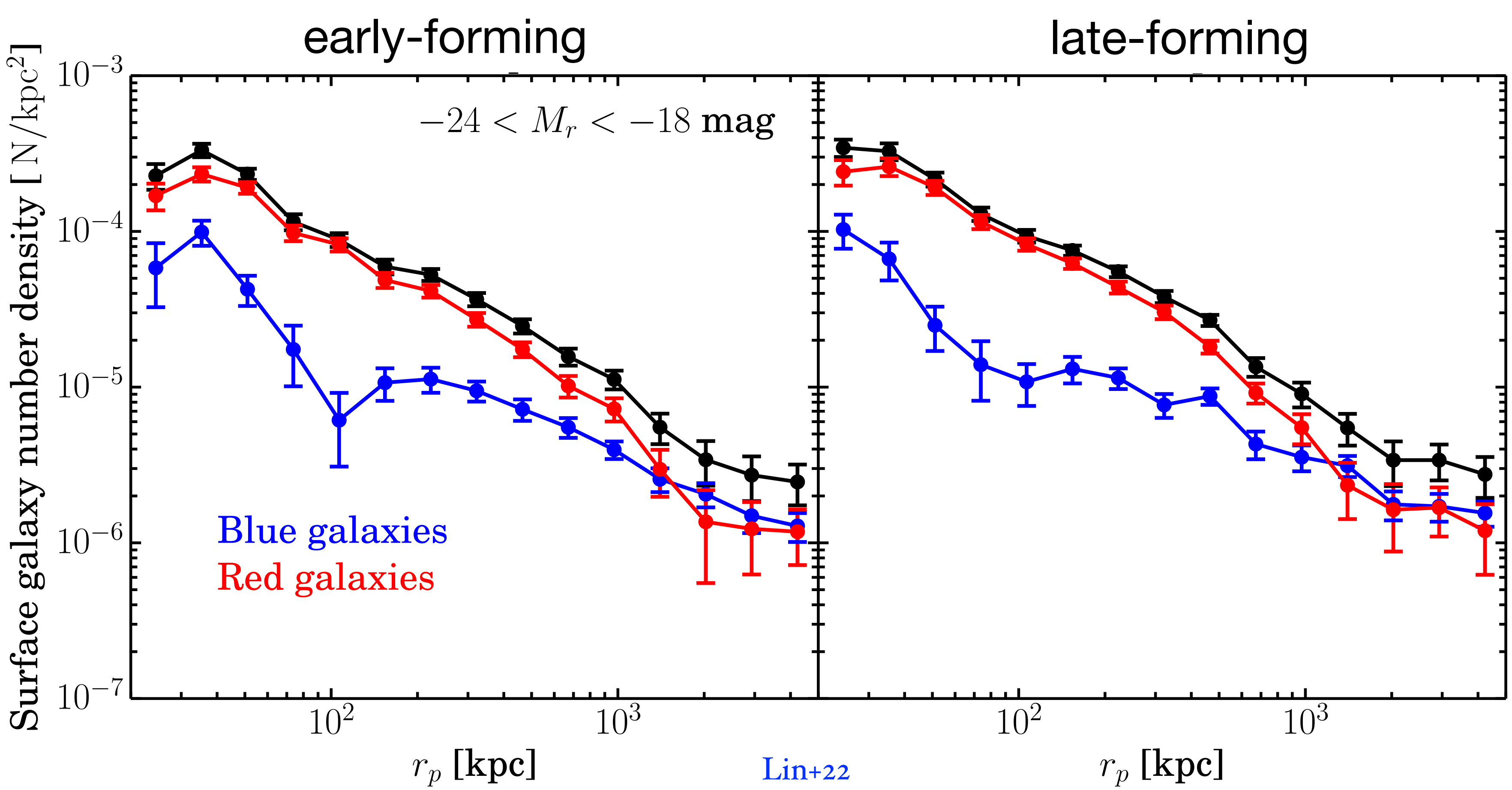
- using a cluster-galaxy cross-correlation technique, we derive surface density profiles of member galaxies of the two samples
  - concentrations of *red* galaxy distribution for the early- and late-forming clusters are c_e=7.1±1.7 and c_l=5.6±0.6



properties of galaxy populations

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

$$N_{\rm early} \approx 0.75 N_{\rm late}$$



Lin+22

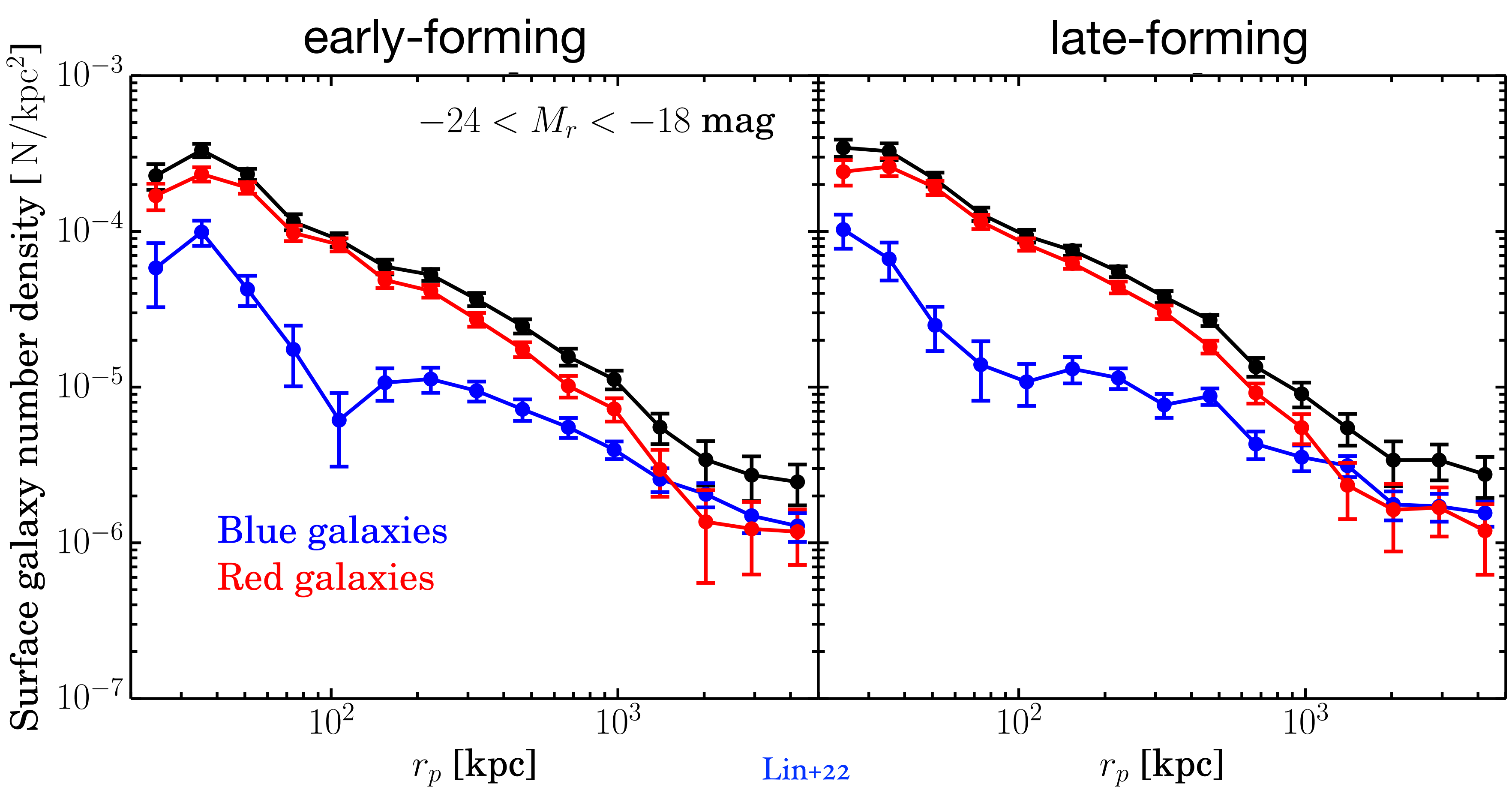
properties of galaxy populations

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

$N_{\mathrm{early}} \approx 0.75 N_{\mathrm{late}}$

$M_{200m,e}=(1.26\pm0.35)\times10^{14}h^{-1}M_\odot$

$M_{200m,l}=(0.95\pm0.26)\times10^{14}h^{-1}M_\odot$



Lin+22

properties of galaxy populations

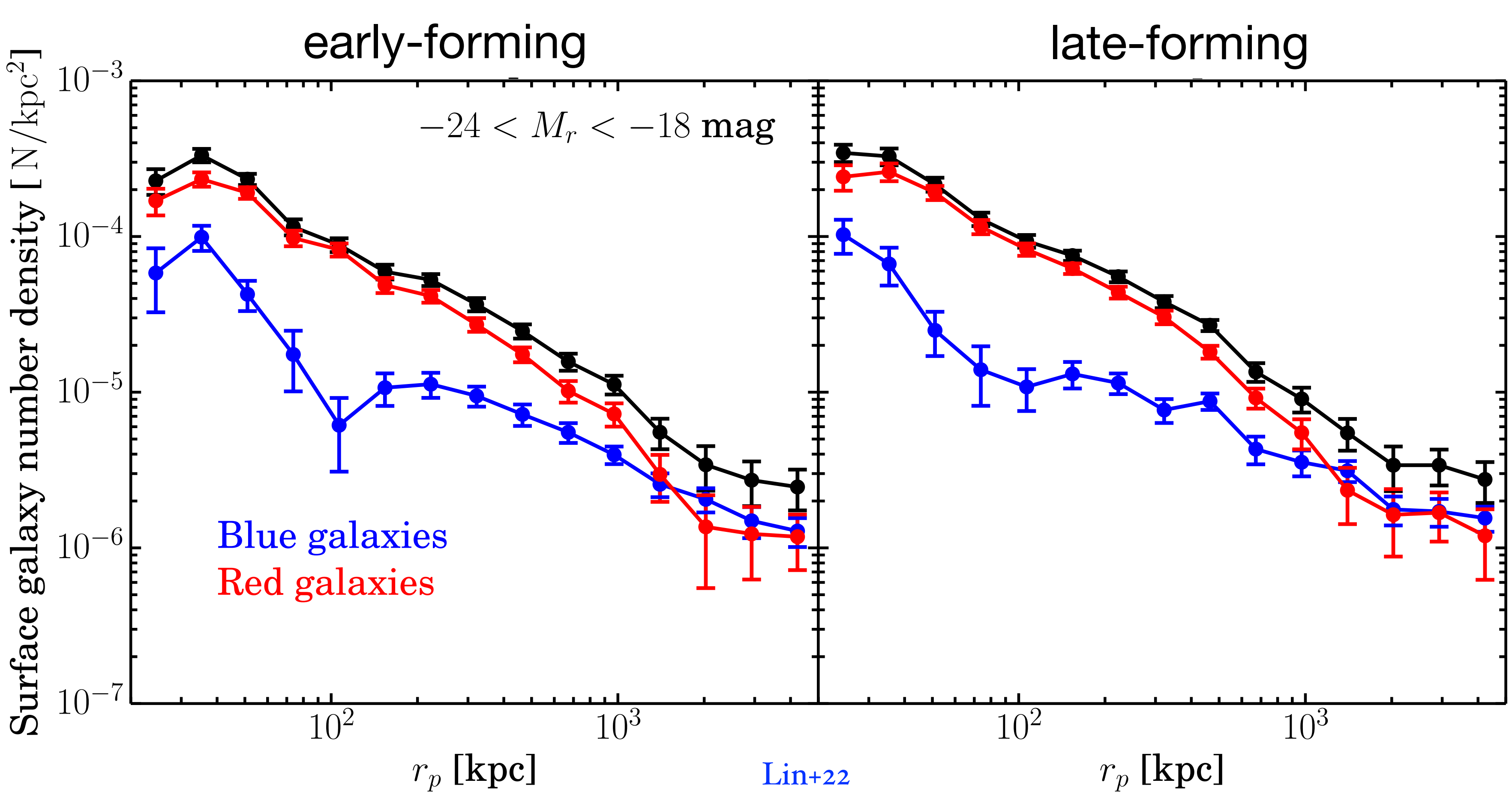
- using a cluster-galaxy cross-correlation technique, we derive surface density profiles of member galaxies of the two samples
  - concentrations of *red* galaxy distribution for the early- and late-forming clusters are c_e=7.1±1.7 and c_l=5.6±0.6

$$N_{\rm early} \approx 0.75 N_{\rm late}$$

$M_{200m,e}$=(1.26±0.35)x$10^{14}h^{-1}M_\odot$

$M_{200m,l}$=(0.95±0.26)x$10^{14}h^{-1}M_\odot$

$$N \propto M^{0.8}$$



Lin+22

properties of galaxy populations

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

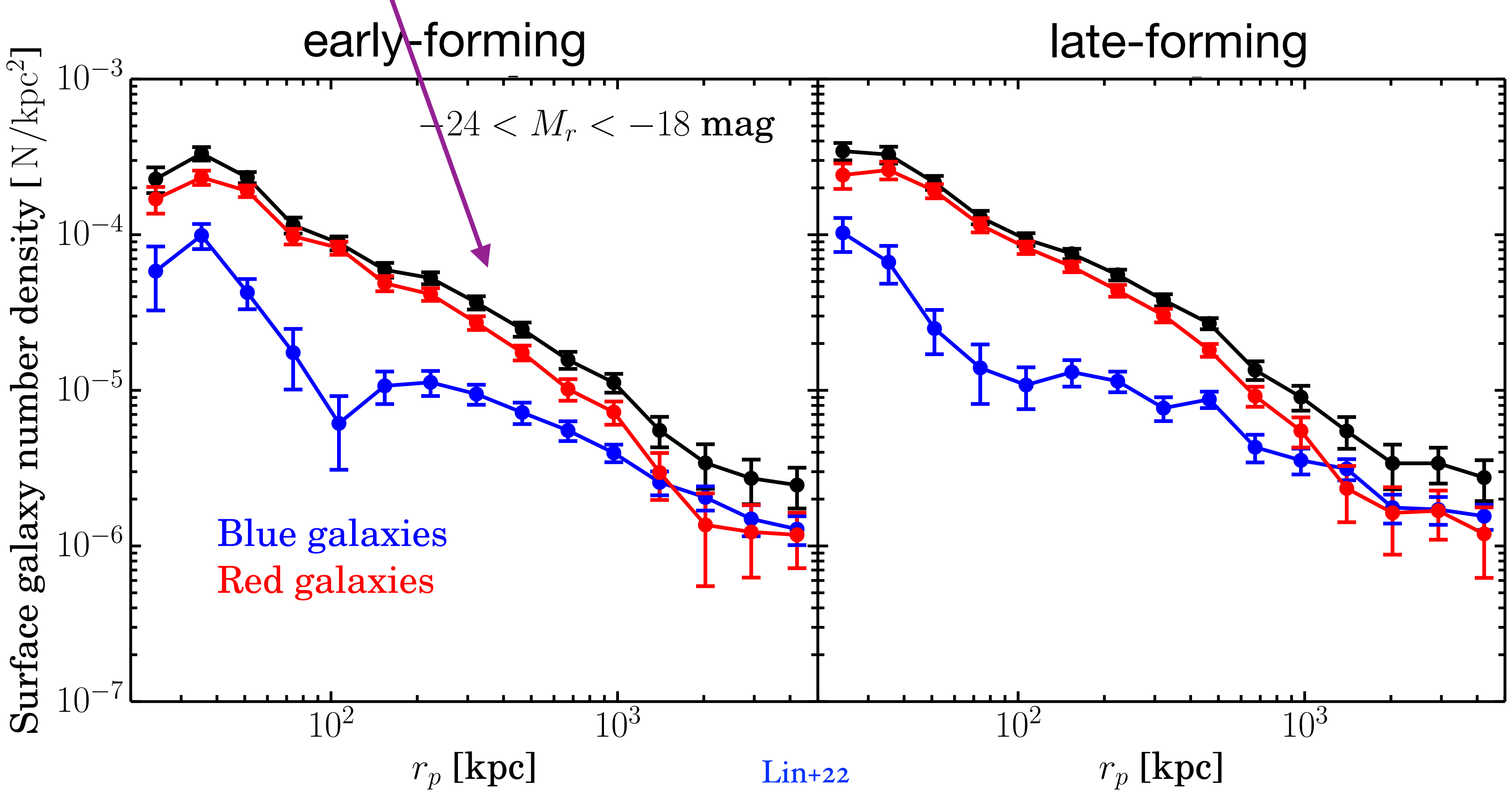
60% **fewer** than late-forming sample (considering cluster mass difference) if there is no AB

$N_{\rm early} \approx 0.75 N_{\rm late}$

$M_{200m,e}$=(1.26±0.35)x10[14]$h^{-1}M_\odot$

$M_{200m,l}$=(0.95±0.26)x10[14]$h^{-1}M_\odot$

$N \propto M^{0.8}$



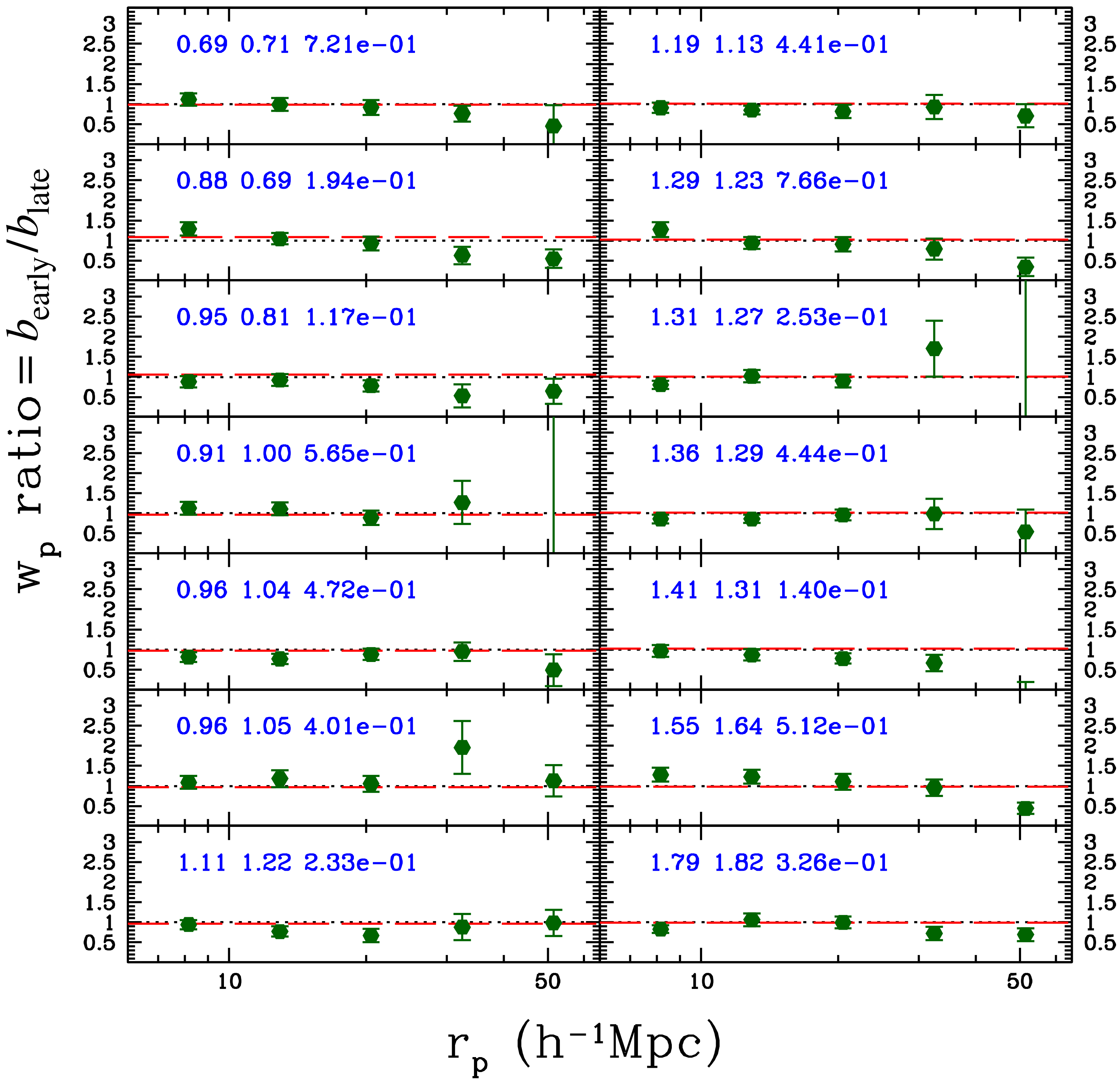
Lin+22

properties of galaxy populations

- using a cluster-galaxy cross-correlation technique, we derive surface density profiles of member galaxies of the two samples
    - concentrations of *red* galaxy distribution for the early- and late-forming clusters are $c_e=7.1\pm1.7$ and $c_l=5.6\pm0.6$
- 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): $\Delta_e=0.44\pm0.01$, $\Delta_l=0.38\pm0.01$
    - between BCG and G4: $\Delta_e=0.99\pm0.01$, $\Delta_l=0.87\pm0.01$
- median offset of BCG from cluster center: $d_e=(0.11\pm0.01)r_{200m}$, $d_l=(0.14\pm0.01)r_{200m}$
- *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 w_p to be consistent with theory (red line)
    - none shows signals as strong as our samples
- mean BCG offset: $d_e=(0.20\pm0.01)r_{200m}$, $d_l=(0.22\pm0.01)r_{200m}$
- median magnitude gap: $\Delta_e=0.42\pm0.01$, $\Delta_l=0.42\pm0.01$



Lin+22

a Bayesian way of thinking about this

- are we measuring AB in the real world or only in *Elucid*?
- recast our study as a hypothesis test ⇒ ruling out the null hypothesis that *AB does not exist in the Universe*
- $P(\mathrm{AB}\,|\,\mathrm{data, Elucid}) \propto P(\mathrm{data}\,|\,\mathrm{AB, Elucid})P(\mathrm{AB}\,|\,\mathrm{Elucid})$
  - AB ≡ "AB exists in the Universe"
  - data ≡ properties of our cluster samples (WL, clustering, cluster galaxy properties)
  - prior $P(\mathrm{AB}\,|\,\mathrm{Elucid}) = 1$ or 0 (each 50%)

a Bayesian way of thinking about this

- $P(\mathrm{AB}\,|\,\mathrm{data, Elucid}) \propto P(\mathrm{data}\,|\,\mathrm{AB, Elucid})P(\mathrm{AB}\,|\,\mathrm{Elucid})$
    - AB ≡ "AB exists in the Universe"
    - data ≡ properties of our cluster samples (WL, clustering, cluster galaxy properties)
    - prior $P(\mathrm{AB}\,|\,\mathrm{Elucid}) = 1$ or 0 (each 50%)
    - likelihood $P(\mathrm{data}\,|\,\mathrm{AB, Elucid})$ consists of 4 cases resulting from the combination of (1) whether AB exists in the Universe & (2) whether AB exists in *Elucid* ⇒ comparing our observations with the predicted signals

	AB exists in real Universe	
AB exists in Elucid	yes⊗yes	yes⊗no
	no⊗yes	no⊗no

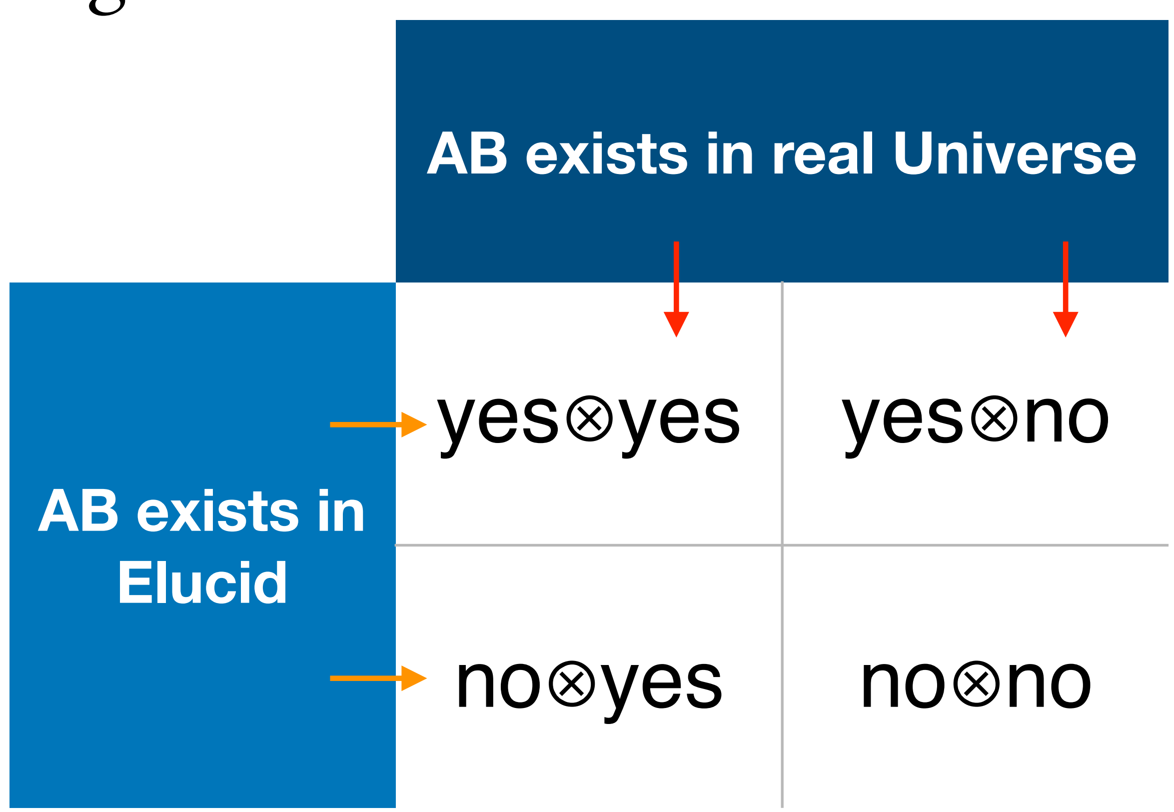
a Bayesian way of thinking about this

- $P(\mathrm{AB}\,|\,\mathrm{data, Elucid}) \propto P(\mathrm{data}\,|\,\mathrm{AB, Elucid})P(\mathrm{AB}\,|\,\mathrm{Elucid})$
  - AB ≡ "AB exists in the Universe"
  - data ≡ properties of our cluster samples (WL, clustering, cluster galaxy properties)
  - prior $P(\mathrm{AB}\,|\,\mathrm{Elucid}) = 1$ or 0 (each 50%)
  - likelihood $P(\mathrm{data}\,|\,\mathrm{AB, Elucid})$ consists of 4 cases resulting from the combination of (1) whether AB exists in the Universe & (2) whether AB exists in *Elucid* ⇒ comparing our observations with the predicted signals

	AB exists in real Universe	
AB exists in Elucid	yes⊗yes	yes⊗no
	no⊗yes	no⊗no

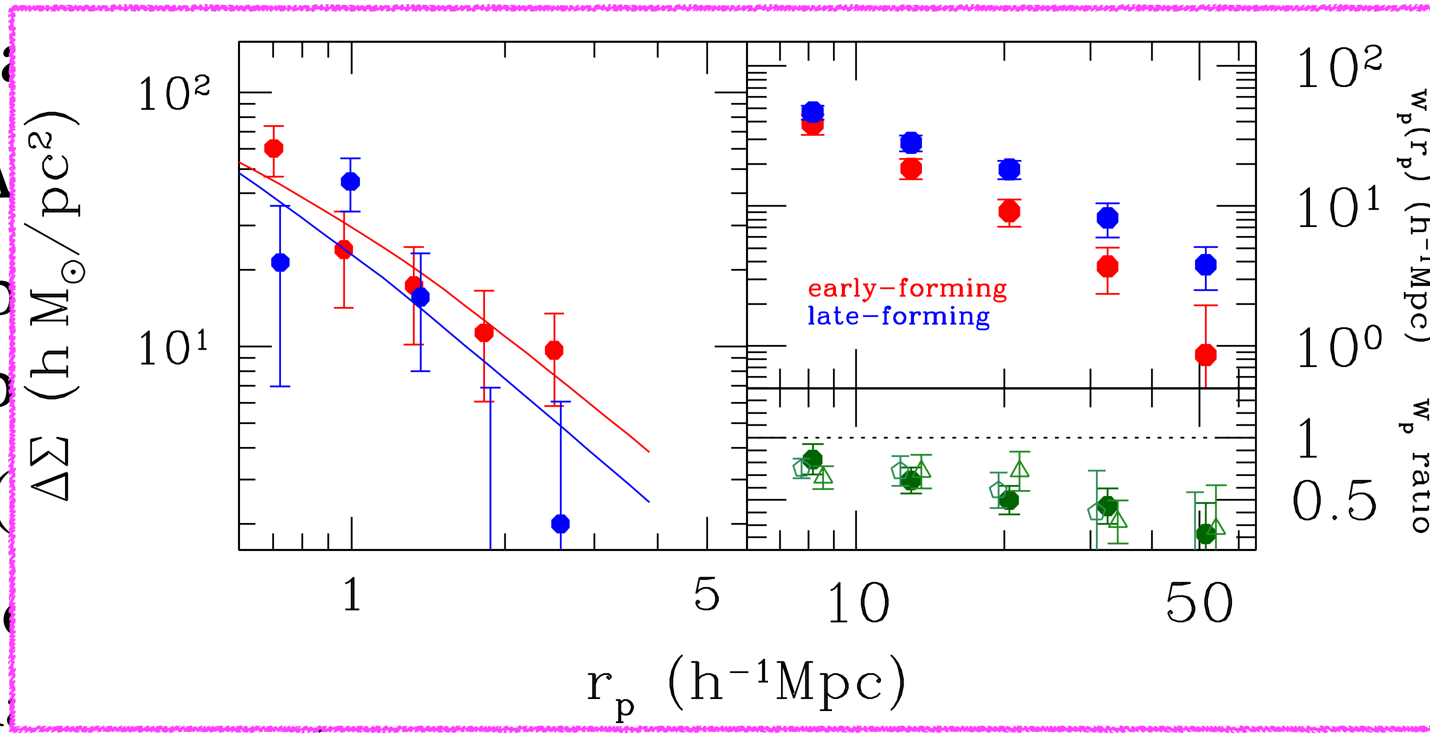
a Bayesian way of thinking about this

- $P(\mathrm{AB}\,|\,\mathrm{data},\mathrm{Elucid}) \propto P(\mathrm{data}\,|\,\mathrm{AB},\mathrm{Elucid})P(\mathrm{AB}\,|\,\mathrm{Elucid})$
  - AB ≡ "AB exists in the Universe"
  - data ≡ properties of our cluster samples (WL, clustering, cluster galaxy properties)
  - prior $P(\mathrm{AB}\,|\,\mathrm{Elucid}) = 1$ or 0 (each 50%)
  - likelihood $P(\mathrm{data}\,|\,\mathrm{AB},\mathrm{Elucid})$ consists of 4 cases resulting from the combination of (1) whether AB exists in the Universe & (2) whether AB exists in *Elucid* ⇒ comparing our observations with the predicted signals



a Bayesian way of thinking about this

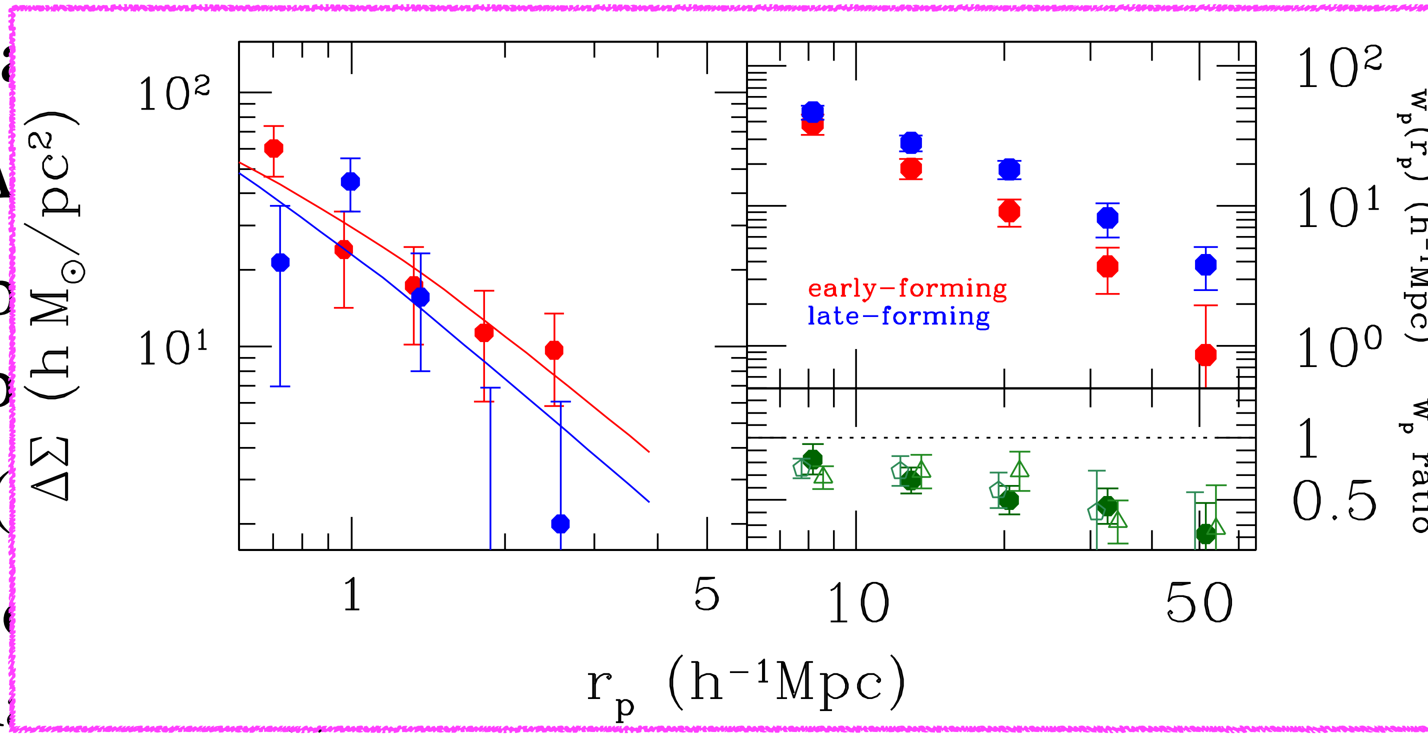
- $P(\mathrm{AB}\,|\,\mathrm{dat}$ … d)
  - AB ≡ "A …
  - data ≡ p … ng, cluster galaxy p …
  - prior $P($ …
  - likeliho … ting from the combina … & (2) whether AB exists in *Elucid* ⇒ comparing our observations with the predicted signals

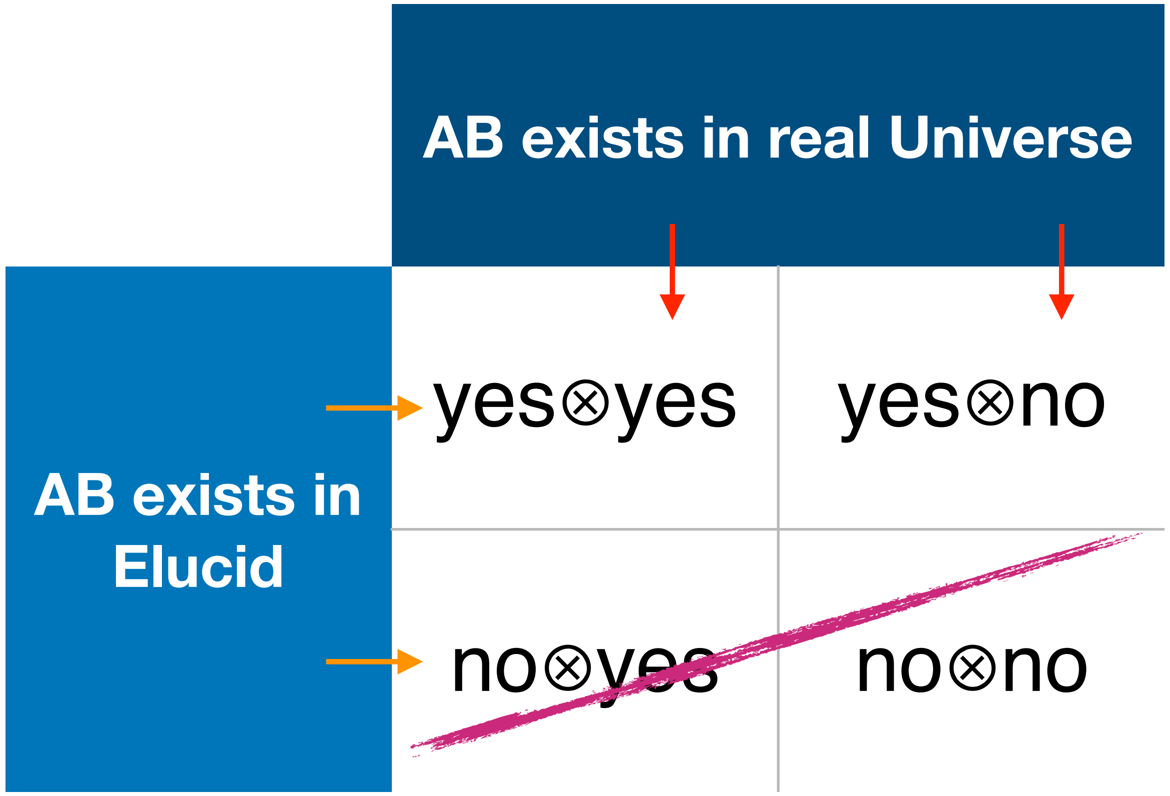


	AB exists in real Universe	
AB exists in Elucid	yes⊗yes	yes⊗no
	no⊗yes	no⊗no

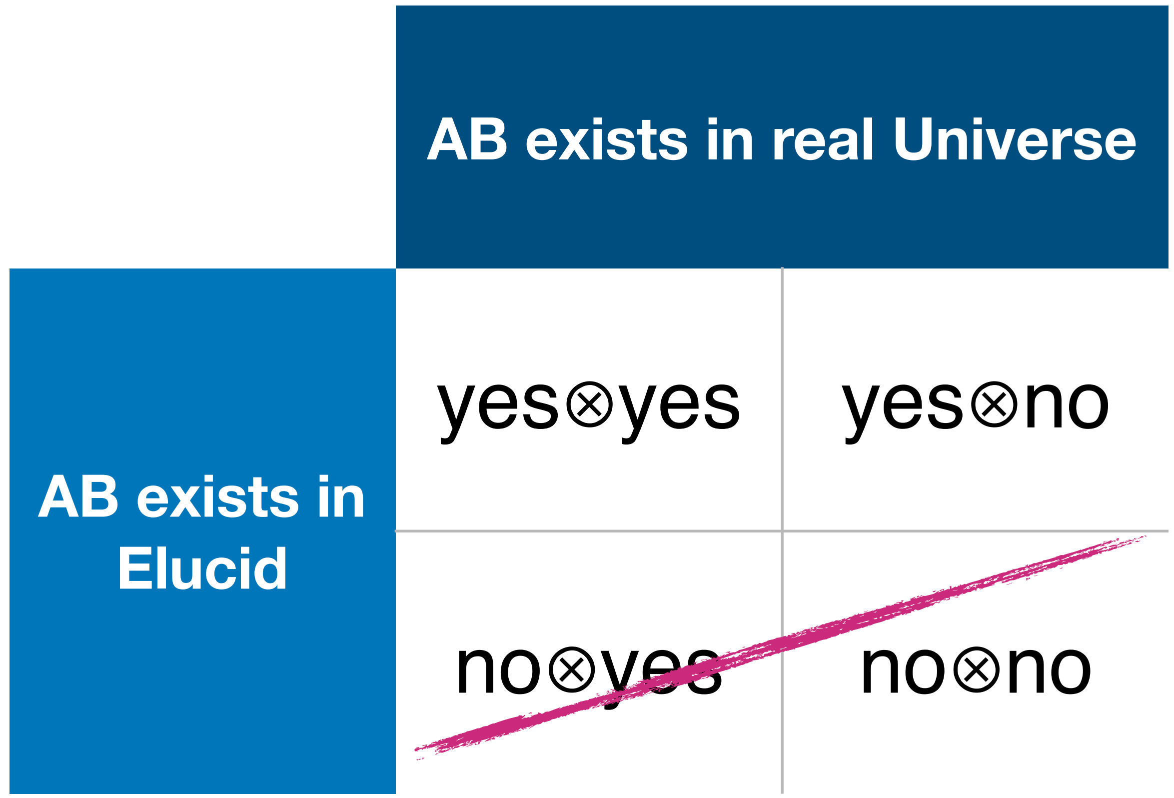
a Bayesian way of thinking about this

- P(AB | dat … d)
    - AB ≡ "A
    - data ≡ p … ng, cluster galaxy p
    - prior P(
    - likeliho … ting from the combin … & (2) whether AB exists in *Elucid* ⇒ comparing our observations with the predicted signals





a Bayesian way of thinking about this



Observable	True AB	Observed trend	Spurious AB due to circularity	Spurious AB due to incorrect cluster mass
Concentration	$c_e > c_l$	$c_e > c_l$	$c_e = c_l$	$c_e \approx c_l$
Galaxy number	$N_e < N_l$	$N_e < N_l$	$N_e = N_l$	$N_e \lesssim N_l$
BCG offset	$d_e < d_l$	$d_e < d_l$	$d_e = d_l$	$d_e > d_l$
Magnitude gap	$\Delta_e > \Delta_l$	$\Delta_e > \Delta_l$	$\Delta_e = \Delta_l$	$\Delta_e > \Delta_l$

- circularity: *Elucid* is built from the density field based on group and cluster catalog of Yang+07, so *Elucid* halos are expected to be in LSS similar to Yang+07 clusters ⇒ z_{20} only meaningful in *Elucid*

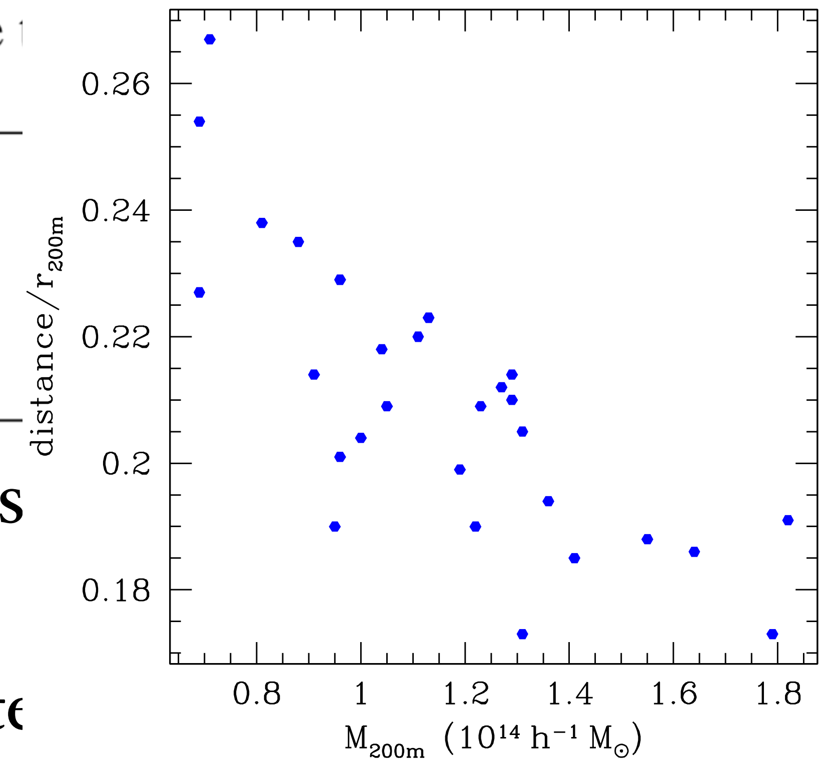
a Bayesian way of thinking about this

Observable	True AB	Observed trend	Spurious AB due to circularity	Spurious AB due to incorrect cluster mass
Concentration	$c_e > c_l$	$c_e > c_l$	$c_e = c_l$	$c_e \approx c_l$
Galaxy number	$N_e < N_l$	$N_e < N_l$	$N_e = N_l$	$N_e \lesssim N_l$
BCG offset	$d_e < d_l$	$d_e < d_l$	$d_e = d_l$	$d_e > d_l$
Magnitude gap	$\Delta_e > \Delta_l$	$\Delta_e > \Delta_l$	$\Delta_e = \Delta_l$	$\Delta_e > \Delta_l$

- if cluster mass of our late-forming sample is severely underestimated
  - concentration varies very weakly with cluster mass $\Rightarrow c_e \approx c_l$
  - $N \propto M^{0.8}$ (but with large scatter) $\Rightarrow N_e \lesssim N_l$
  - BCG offset decreases with increasing cluster mass $\Rightarrow d_e > d_l$
  - mag gap is found to decrease with cluster mass (Lin+10) $\Rightarrow \Delta_e > \Delta_l$

a Bayesian way of thinking about this

Observable	True AB	Observed trend	Spurious AB due circularity
Concentration	$c_e > c_l$	$c_e > c_l$	$c_e = c_l$
Galaxy number	$N_e < N_l$	$N_e < N_l$	$N_e = N_l$
BCG offset	$d_e < d_l$	$d_e < d_l$	$d_e = d_l$
Magnitude gap	$\Delta_e > \Delta_l$	$\Delta_e > \Delta_l$	$\Delta_e = \Delta_l$



- if cluster mass of our late-forming sample is underestimated
  - concentration varies very weakly with cluste
  - $N \propto M^{0.8}$ (but with large scatter) $\Rightarrow N_e \lesssim N_l$
  - BCG offset decreases with increasing cluster mass $\Rightarrow d_e > d_l$
  - mag gap is found to decrease with cluster mass (Lin+10) $\Rightarrow \Delta_e > \Delta_l$

a Bayesian way of thinking about this

Observable	True AB	Observed trend	Spurious AB due to circularity	Spurious AB due to incorrect cluster mass
Concentration	$c_e > c_l$	$c_e > c_l$	$c_e = c_l$	$c_e \approx c_l$
Galaxy number	$N_e < N_l$	$N_e < N_l$	$N_e = N_l$	$N_e \lesssim N_l$
BCG offset	$d_e < d_l$	$d_e < d_l$	$d_e = d_l$	$d_e > d_l$
Magnitude gap	$\Delta_e > \Delta_l$	$\Delta_e > \Delta_l$	$\Delta_e = \Delta_l$	$\Delta_e > \Delta_l$

- if cluster mass of our late-forming sample is severely underestimated
  - concentration varies very weakly with cluster mass $\Rightarrow c_e \approx c_l$
  - $N \propto M^{0.8}$ (but with large scatter) $\Rightarrow N_e \lesssim N_l$
  - BCG offset decreases with increasing cluster mass $\Rightarrow d_e > d_l$
  - mag gap is found to decrease with cluster mass (Lin+10) $\Rightarrow \Delta_e > \Delta_l$

a Bayesian way of thinking about this

Observable	True AB	Observed trend	Spurious AB due to circularity	Spurious AB due to incorrect cluster mass
Concentration	$c_e > c_l$	$c_e > c_l$	$c_e = c_l$	$c_e \approx c_l$
Galaxy number	$N_e < N_l$	$N_e < N_l$	$N_e = N_l$	$N_e \lesssim N_l$
BCG offset	$d_e < d_l$	$d_e < d_l$	$d_e = d_l$	$d_e > d_l$
Magnitude gap	$\Delta_e > \Delta_l$	$\Delta_e > \Delta_l$	$\Delta_e = \Delta_l$	$\Delta_e > \Delta_l$

- if cluster mass of our late-forming sample is severely underestimated
  - concentration varies very weakly with cluster mass $\Rightarrow c_e \approx c_l$
  - $N \propto M^{0.8}$ (but with large scatter) $\Rightarrow N_e \lesssim N_l$
  - BCG offset decreases with increasing cluster mass $\Rightarrow d_e > d_l$
  - mag gap is found to decrease with cluster mass (Lin+10) $\Rightarrow \Delta_e > \Delta_l$
- none of the 14 pairs of control/random samples passes these tests

a Bayesian way of thinking about this

	AB exists in real Universe	
AB exists in Elucid	yes⊗yes	~~yes⊗no~~
	~~no⊗yes~~	~~no⊗no~~

...purious AB due to circularity	Spurious AB due to incorrect cluster mass
$c_e = c_l$	$c_e \approx c_l$
$N_e = N_l$	$N_e \lesssim N_l$
$d_e = d_l$	$d_e > d_l$
$\Delta_e = \Delta_l$	$\Delta_e > \Delta_l$

...ing sample is severely

...kly with cluster mass $\Rightarrow c_e \approx c_l$

- $N \propto M^{0.8}$ (but with large scatter) $\Rightarrow N_e \lesssim N_l$
- BCG offset decreases with increasing cluster mass $\Rightarrow d_e > d_l$
- mag gap is found to decrease with cluster mass (Lin+10) $\Rightarrow \Delta_e > \Delta_l$

- none of the 14 pairs of control/random samples passes these tests
- we live in the "yes" ⊗ "yes" box!

a Bayesian way of thinking about this

	AB exists in real Universe	
AB exists in Elucid	yes⊗yes	~~yes⊗no~~
	no⊗yes	~~no⊗no~~

purious AB due to circularity	Spurious AB due to incorrect cluster mass
$c_e = c_l$	$c_e \approx c_l$
$N_e = N_l$	$N_e \lesssim N_l$
$d_e = d_l$	$d_e > d_l$
$\Delta_e = \Delta_l$	$\Delta_e > \Delta_l$

ing sample is severely

kly with cluster mass $\Rightarrow c_e \approx c_l$

- $N \propto M^{0.8}$ (but with large scatter) $\Rightarrow N_e \lesssim N_l$
- BCG offset decreases with increasing cluster mass $\Rightarrow d_e > d_l$
- mag gap is found to decrease with cluster mass (Lin+10) $\Rightarrow \Delta_e > \Delta_l$

- none of the 14 pairs of control/random samples passes these tests
- we live in the "yes" ⊗ "yes" box!
- *even if the mean mass of our late-forming sample is truly severely underestimated, the difference in masses is still far from sufficient to explain the huge difference in biases ⇒ something like AB at work*

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, Miyatake et al. (2022, A&A, 666, A97) for more details!