Identification of Galaxy Protoclusters Based on the Spherical Top-hat Collapse Theory

Jaehyun Lee, Changbom Park, Juhan Kim, Christophe Pichon, Brad K. Gibson, Jihye Shin, Yonghwi Kim, Owain N. Snaith, Yohan Dubois, C. Gareth Few

The Astrophysical Journal, 960:132 (20pp), 2024 January 10 https://doi.org/10.3847/1538-4357/ad0555

Identification of Galaxy Protoclusters Based on the Spherical Top-hat Collapse Theory

Jaehyun Lee[1,2], Changbom Park[2], Juhan Kim[3], Christophe Pichon[2,4,5], Brad K. Gibson[6], Jihye Shin[1], Yonghwi Kim[7], Owain N. Snaith[8], Yohan Dubois[9], and C. Gareth Few[6]
[1] Korea Astronomy and Space Science Institute, 776 Daedeokdae-ro, Yuseong-gu, Daejeon 34055, Republic of Korea
[2] Korea Institute for Advanced Study, 85 Hoegi-ro, Dongdaemun-gu, Seoul 02455, Republic of Korea
[3] Center for Advanced Computation, Korea Institute for Advanced Study, 85 Hoegiro, Dongdaemun-gu, Seoul 02455, Republic of Korea; kjhan@kias.re.kr
[4] Institut d'Astrophysique de Paris, CNRS and Sorbonne Université, UMR 7095, 98 bis Boulevard Arago, F-75014 Paris, France
[5] IPhT, DRF-INP, UMR 3680, CEA, L'Orme des Merisiers, Bât 774, F-91191 Gif-sur-Yvette, France
[6] E.A. Milne Centre for Astrophysics, University of Hull, Hull HU6 7RX, UK
[7] Korea Institute of Science and Technology Information, 245 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea
[8] University of Exeter, School of Physics and Astronomy, Stocker Road, Exeter EX4 4QL, UK
[9] CNRS and Sorbonne Université, UMR 7095, Institut d'Astrophysique de Paris, 98 bis, Boulevard Arago, F-75014 Paris, France
Received 2023 August 1; revised 2023 October 3; accepted 2023 October 18; published 2024 January 9

Abstract

We propose a new method for finding galaxy protoclusters that is motivated by structure formation theory and also directly applicable to observations. We adopt the conventional definition that a protocluster is a galaxy group whose virial mass $M_{\rm vir} < M_{\rm cl}$ at its epoch, where $M_{\rm cl} = 10^{14}\,M_{\odot}$, but would exceed that limit when it evolves to $z = 0$. We use the critical overdensity for complete collapse at $z = 0$ predicted by the spherical top-hat collapse model to find the radius and total mass of the regions that would collapse at $z = 0$. If the mass of a region centered at a massive galaxy exceeds $M_{\rm cl}$, the galaxy is at the center of a protocluster. We define the outer boundary of a protocluster as the zero-velocity surface at the turnaround radius so that the member galaxies are those sharing the same protocluster environment and showing some conformity in physical properties. We use the cosmological hydrodynamical simulation Horizon Run 5 (HR5) to calibrate this prescription and demonstrate its performance. We find that the protocluster identification method suggested in this study is quite successful. Its application to the high-redshift HR5 galaxies shows a tight correlation between the mass within the protocluster regions identified according to the spherical collapse model and the final mass to be found within the clusters at $z = 0$, meaning that the regions can be regarded as the bona fide protoclusters with high reliability. We also confirm that the redshift-space distortion does not significantly affect the performance of the protocluster identification scheme.

Unified Astronomy Thesaurus concepts: Galaxy kinematics (602); Galaxies (573); Galaxy evolution (594); High-redshift galaxy clusters (2007); Galaxy dynamics (591); Galaxy formation (595)

1. Introduction

Galaxy clusters are typically defined as the objects that are bound and dynamically relaxed with total mass of $M_{\rm tot} > 10^{14}\,M_{\odot}$ (e.g., Overzier 2016). As the progenitors of present-day galaxy clusters, protoclusters must have formed in the densest environments in the early universe, and the majority of the galaxies in protoclusters probably have formed and evolved earlier than those in other environments (Kaiser 1984).

Many observational efforts have been made to search for protoclusters at high redshifts. Deep-field spectroscopic surveys are a direct approach for finding protoclusters (e.g., Steidel et al. 1998, 2000, 2005; Lee et al. 2014c; Cucciati et al. 2014; Lemaux et al. 2014; Toshikawa et al. 2014; Chiang et al. 2015; Diener et al. 2015; Wang et al. 2016; Calvi et al. 2021; McConachie et al. 2022). However, the survey volume should be very large to include many of such rare objects, and spectroscopic observations are currently too time-consuming to carry out large-volume blind surveys for the deep universe. Therefore, large-area imaging surveys have often been conducted to search for overdense regions at high redshifts by utilizing the narrowband photometry for emission-line galaxies or the photo-z/dropout technique (e.g., Shimasaku et al. 2003; Ouchi et al. 2005; Toshikawa et al. 2012, 2016; Cai et al. 2017; Toshikawa et al. 2018; Shi et al. 2019; Yonekura et al. 2022).

Some energetic events are expected to happen in overdense regions at high redshifts. High-z radio galaxies are believed to be the potential progenitors of brightest cluster galaxies, and thus they are assumed as a proxy for protoclusters (Le Fevre et al. 1996; Pascarelle et al. 1996; Venemans et al. 2002, 2004, 2005, 2007; Hatch et al. 2011b, 2011a; Hayashi et al. 2012; Cooke et al. 2014; Shen et al. 2021). Although it is still debated (see Husband et al. 2013; Hennawi et al. 2015), high-z QSOs are also known to trace overdense regions (Djorgovski et al. 2003; Wold et al. 2003; Stevens et al. 2010; Falder et al. 2011; Adams et al. 2015). Lyα blobs can be lit by a huge amount of ionized photons emitted from active galactic nuclei (AGNs) or starburst galaxies in dense regions that still bear sufficient cold gas as a fuel. High-z submillimeter galaxies are regarded as the progenitors of massive ellipticals (e.g., Lilly et al. 1999; Fu et al. 2013; Toft et al. 2014). Therefore, Lyα blobs or overdensity regions of submillimeter galaxies are also used as the indicators of protocluster regions (Stevens et al. 2003; Greve et al. 2007; Prescott et al. 2008; Daddi et al. 2009; Prescott et al. 2012; Umehata et al. 2014, 2015; Oteo et al. 2018; Cooke et al. 2019; Álvarez Crespo et al. 2021; Rotermund et al. 2021). Gas absorption lines are another probe of protoclusters that does not rely on galaxy distribution: high-z overdense regions that still contain plenty of intergalactic

neutral hydrogen can be detected by examining the Lyα forests in the spectra of background QSOs or star-forming galaxies (e.g., Lee et al. 2014b; Stark et al. 2015; Cai et al. 2016, 2017; Newman et al. 2022).

While the observations targeting protoclusters have used a variety of selection techniques, they commonly focus on the identification of overdense regions. The protoclusters that are expected to eventually form massive clusters with the total mass of $M_{\rm tot} > 10^{15}\,M_{\odot}$ have an overdensity of $\delta \sim 10$–12 for typical galaxies or Lyα emitters within an aperture radius of $R \sim 15$ cMpc at $z \sim 2$–3 (e.g., Cucciati et al. 2014; Lemaux et al. 2014; Cai et al. 2017). Toshikawa et al. (2018) identify protocluster candidates in a wide field of >100 deg^2 by selecting the regions that show a galaxy overdensity significance level higher than 4σ within an aperture radius of $R \sim 16$ cMpc at $z \sim 3.8$. This significance level corresponds to the overdensity of the regions that end up forming halos of $M_{\rm halo} \gtrsim 5 \times 10^{14}\,M_{\odot}$. The overdensity significance level is adopted to achieve $\sim$80% reliability, at the cost of completeness (Toshikawa et al. 2016)

Several theoretical studies have been conducted to examine the properties of protocluster regions. Chiang et al. (2013) and Muldrew et al. (2015) investigate the matter and galaxy overdensity in the areas enclosing protoclusters using the semianalytic model of Guo et al. (2011) based on the Millennium simulation (Springel et al. 2005). In the two studies, protoclusters are traced using halo merger trees. They show that the protocluster galaxies are more widespread in larger clusters and that the distribution of protocluter galaxies largely shrinks during $z = 4$–2. Chiang et al. (2013) also show that, in a top-hat box of (15 cMpc)^3, the galaxy overdensity of protoclusters strongly correlates with final cluster mass. Wang et al. (2021) develop a method to identify protoclusters from halo distribution of an N-body simulation using an extension of the friends-of-friends (FoF) algorithm. They show that the approach reasonably recovers protoclusters with high completeness.

Hydrodynamical simulations are also used to study the formation and evolution of clusters of galaxies. Given that the mean separation of rich clusters is $\sim$70 cMpc (Bahcall & West 1992), it is thus necessary to use a simulation box larger than about 1 cGpc^3 to study the formation and evolution of Coma-like clusters accurately and with high statistical significance. However, due to the limitation of the current computing resources, it has been nearly impossible to conduct hydrodynamical simulations in such a large box while keeping a resolution below $\sim$1 kpc. As a compromise between the need for the extremely large dynamic range and the limited computing resources, the zoom-in technique is widely adopted in the hydrodynamical simulations for galaxy clusters (Bahé et al. 2017; Choi & Yi 2017; Truong et al. 2018; Trebitsch et al. 2021; Yajima et al. 2022). In these simulations, cluster regions are pre-identified and zoomed in on from the initial conditions, and protoclusters are traced by using merger trees.

It should be noted that, in the previous studies, protoclusters have been defined inconsistently between observations, theories, and numerical simulations. If a protocluster is defined as the group of all the objects that will eventually collapse into a cluster, their initial distribution typically spans more than tens of cMpc (Chiang et al. 2013; Muldrew et al. 2015, 2018). In this definition, protoclusters can be neither self-bound nor compact, and thus a protocluster is hardly viewed as a physical object in which galaxies are associated with each other in a common environment. Furthermore, diachronic information is not available in observations. Therefore, observers have focused on the identification of sufficiently overdense regions. This is justified by the fact that larger structures in the current universe are more likely to originate from more massive progenitors at high redshifts (Chiang et al. 2013; Muldrew et al. 2015). The range of the overdense region varies between protoclusters. Since the virial radius only encloses the objects that are already bound to the local density peak, it inevitably misses a number of progenitors that are still in the course of infall, outside the virialized regions. Because the proto-objects of larger clusters are more extended (Muldrew et al. 2015), a systematic approach is required to define the boundary (or spatial extent) of protoclusters, which should be based on the physical conditions of specific environments of interest.

This study aims at proposing a new scheme for the identification of protoclusters that is motivated by structure formation theories and also applicable to observations directly. Our prescription is justified and calibrated on a cosmological hydrodynamical simulation Horizon Run 5 (hereafter HR5; Lee et al. 2021; Park et al. 2022). HR5 covers a volume of (1048.6 cMpc)^3 with a spatial resolution down to about 1 kpc. Thanks to its large volume, HR5 enables us to look into the formation and evolution of galaxies in a wide range of environments. By taking advantage of HR5, we derive a scheme applicable to observations to find the centers of protocluster candidates based on the spherical top-hat collapse (SC) model. The scheme also defines the physical region of a given protocluster as the volume within the turnaround radius from their centers. The turnaround radius is the zero-velocity surface at which gravitational infall counterbalances the local Hubble expansion (Gunn & Gott 1972).

This paper is organized as follows. In Section 2, we briefly introduce the HR5 simulation, a structure-finding and tree-building algorithm, and the scheme to identify clusters using a low-resolution version of HR5. In Section 3, we present the methodology to find the candidate regions for protoclusters from the galaxy distribution. The method for finding the boundary of protoclusters is presented in Section 4. We discuss and summarize this study in Section 5. Additional details of structure identification, merger-tree-building schemes, the SC models, and protocluster identification are given in the appendices.

2. Simulation Data

2.1. Horizon Run 5

HR5 is a cosmological hydrodynamical zoomed-in simulation aiming at covering a wide range of cosmic structures in a 1.15 cGpc^3 volume, with a spatial resolution down to $\sim$1 kpc. We adopt cosmological parameters of $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_b = 0.047$, $\sigma_8 = 0.816$, and $h = 0.684$, which are compatible with the Planck data (Planck Collaboration et al. 2016). We generate the initial conditions using the MUSIC package (Hahn & Abel 2011), with a second-order Lagrangian scheme to launch the particles (2LPT; Scoccimarro 1998; L'Huillier et al. 2014). HR5 is conducted using a version of the adaptive mesh refinement code RAMSES (Teyssier 2002) upgraded for an OpenMP plus MPI two-dimensional parallelism (Lee et al. 2021). We generated a number of random sets and selected the one that reproduced the theoretical baryonic acoustic oscillation

features most closely. While the volume of the zoomed-in region is still somewhat insufficient for accurate statistical analyses of the most massive galaxy clusters and the impact of the very large-scale structures, the whole simulation box does manage to encompass the relevant large-scale perturbation modes and provides us with a representative volume corresponding to the input cosmology.

The volume of HR5 is set to have a high-resolution cuboid zoomed-in region of $1048.6 \times 119.0 \times 127.2\,\mathrm{cMpc}^3$ crossing the center of the volume. The effective volume of the region is $\sim(260\,\mathrm{cMpc})^3$. The cosmological box has 256 root cells (level 8, $\Delta x = 4.10$ cMpc) on a side, and the zoomed-in region has 8192 cells (level 13, $\Delta x = 0.128$ cMpc) along the long side in the initial conditions. The high-resolution region initially contains $8192 \times 930 \times 994$ cells and dark matter particles and is surrounded by the padding grids of levels from 12 to 9. The dark matter particle mass is $6.89 \times 10^7\,M_\odot$ in the zoomed-in region and increases by a factor of 8 with a decreasing grid level. The cells are adaptively refined down to $\Delta x \sim 1$ kpc when their density exceeds eight times the dark matter particle mass at level 13. HR5 proceeded through $z = 0.625$.

Physical processes driving the evolution of baryonic components are implemented in subgrid forms in RAMSES. Gas cooling is computed using the cooling functions of Sutherland & Dopita (1993) in a temperature range of 10^4–$10^{8.5}$ K, and fine-structure line cooling is computed down to ∼750 K using the cooling rates of Dalgarno & McCray (1972). RAMSES approximates cosmic reionization by assuming a uniform UV background (Haardt & Madau 1996). The statistical approach of Rasera & Teyssier (2006) is adopted to compute a star formation rate. Supernova feedback affects the interstellar medium in thermal and kinetic modes (Dubois & Teyssier 2008), and AGN feedback operates in radio-jet and quasar modes, relying on the Eddington ratio (Dubois et al. 2012). Massive black holes (MBHs) are seeded with an initial mass of $10^4\,M_\odot$ in grids when gas density is higher than the threshold of star formation and no other MBHs are found within 50 kpc (Dubois et al. 2014a). MBHs grow via accretion and coalescence, and their angular momentum obtained from the feeding processes is traced (Dubois et al. 2014b). Metal enrichment is computed using the method proposed by Few et al. (2012) based on a Chabrier initial mass function (Chabrier 2003), and in particular the abundances of H, O, and Fe are traced individually. One can find further details of HR5 in Lee et al. (2021).

2.2. Identification of Clusters Using a Low-resolution Simulation

We identify FoF halos and self-bound objects embedded in FoF halos using PGalF (Kim et al. 2023). We also construct the merger trees of self-bound objects using ySAMtm (Lee et al. 2014a; Jung et al. 2014) based on stellar particles for galaxies and dark matter particles for halos that contain no stars. The details of the structure-finding and tree-building algorithms are given in Appendix A.

In this study, we define a galaxy cluster as the virialized object that has acquired the total mass of $M_{\rm tot} > 10^{14}\,M_\odot$ at or before $z = 0$. The mass cut is adopted following the conventional mass range of galaxy clusters (e.g., Overzier 2016) and can be varied if a different mass range is necessary. Protoclusters are the progenitors of galaxy clusters that have not reached the cluster-scale mass range yet. By this definition, both clusters and protoclusters can be found at any epoch. According to this definition of a galaxy cluster, we cannot directly identify all the clusters and protoclusters in HR5, as the simulation stopped at $z = 0.625$. At this redshift, we find 63 clusters with $M_{\rm tot} > 10^{14}\,M_\odot$ in the zoomed-in region. Objects having mass contamination higher than 0.7% by the lower-level particles are excluded. However, there can be many structures that are not massive enough to be identified as clusters at $z = 0.625$ but will evolve to cluster-scale halos by $z = 0$.

To find clusters and protoclusters in the last snapshot of HR5 (i.e., $z = 0.625$), we additionally conduct a low-resolution simulation HR5-Low($\Delta x \sim 16$ kpc) based on the initial conditions and the model parameters used in HR5. We identify structures from the snapshots of HR5-Low at $z = 0$ and 0.625 using PGalF. At $z = 0$, we find 2794 objects of $M^0_{\rm tot} \geqslant 10^{13}\,M_\odot$ and 189 objects of $M^0_{\rm tot} \geqslant 10^{14}\,M_\odot$ with the contamination tolerance mentioned above. The dark matter particles are traced back to $z = 0.625$ using their IDs, to search for the progenitors of the clusters. We then construct the Lagrangian volume (LV; for details see Oñorbe et al. 2014) of the progenitors using the uniform cubic grids enclosing the dark matter particles finally assembling the clusters. We assume that the LVs constructed from HR5-Low also enclose the clusters or protoclusters in HR5. We present the details of the identification scheme and reliability of this approach in Appendix B. Figure 1 shows the dark matter distribution in three HR5-Low cluster regions at $z = 0$ (left), the same regions of HR5-Low at $z = 0.625$ (middle), and HR5 at $z = 0.625$ (right). The structure colored in yellow is the FoF halo of each cluster (left), its progenitors at $z = 0.625$ (middle), and its counterpart in HR5 (right panels). The grids enclosed by dotted lines mark the LVs of the objects constructed by tracing the dark matter particles. This figure demonstrates that the two simulations are in good agreement despite their different resolutions. The position of a structure may show a slight offset between the two different-resolution simulations (HR5-Low and HR5) at $z = 0.625$, partly due to the adaptive time step in RAMSES.

3. Identification of Protoclusters

We define "protoclusters" as galaxy groups whose total mass within $R_{\rm vir}$ is currently less than $10^{14}\,M_\odot$ at their epochs but would exceed that limit by $z = 0$. The physical extent of a protocluster is defined as the spherical volume within the turnaround radius or the zero-velocity surface. The concept is schematically visualized in Figure 2. A protocluster is located at the center of a sphere that has the mean density $\bar{\delta}$ and encloses the total mass exceeding $10^{14}\,M_\odot$. The critical overdensity $\delta_m^{\rm sc} = \bar{\delta}$ is given by the spherical top-hat theory, and the mass contained is the expected virial mass of the region at $z = 0$. It should be noted that only the galaxies within the turnaround radius are called the protocluster member galaxies and that the cluster progenitor galaxies can be spread out to much larger radii.

We first identify the authentic proto-objects by tracing their merger trees in Section 3.1, and then we present a systematic approach for finding the candidate regions enclosing protoclusters from the galaxy distribution in a snapshot, without diachronic information, in Section 3.2.

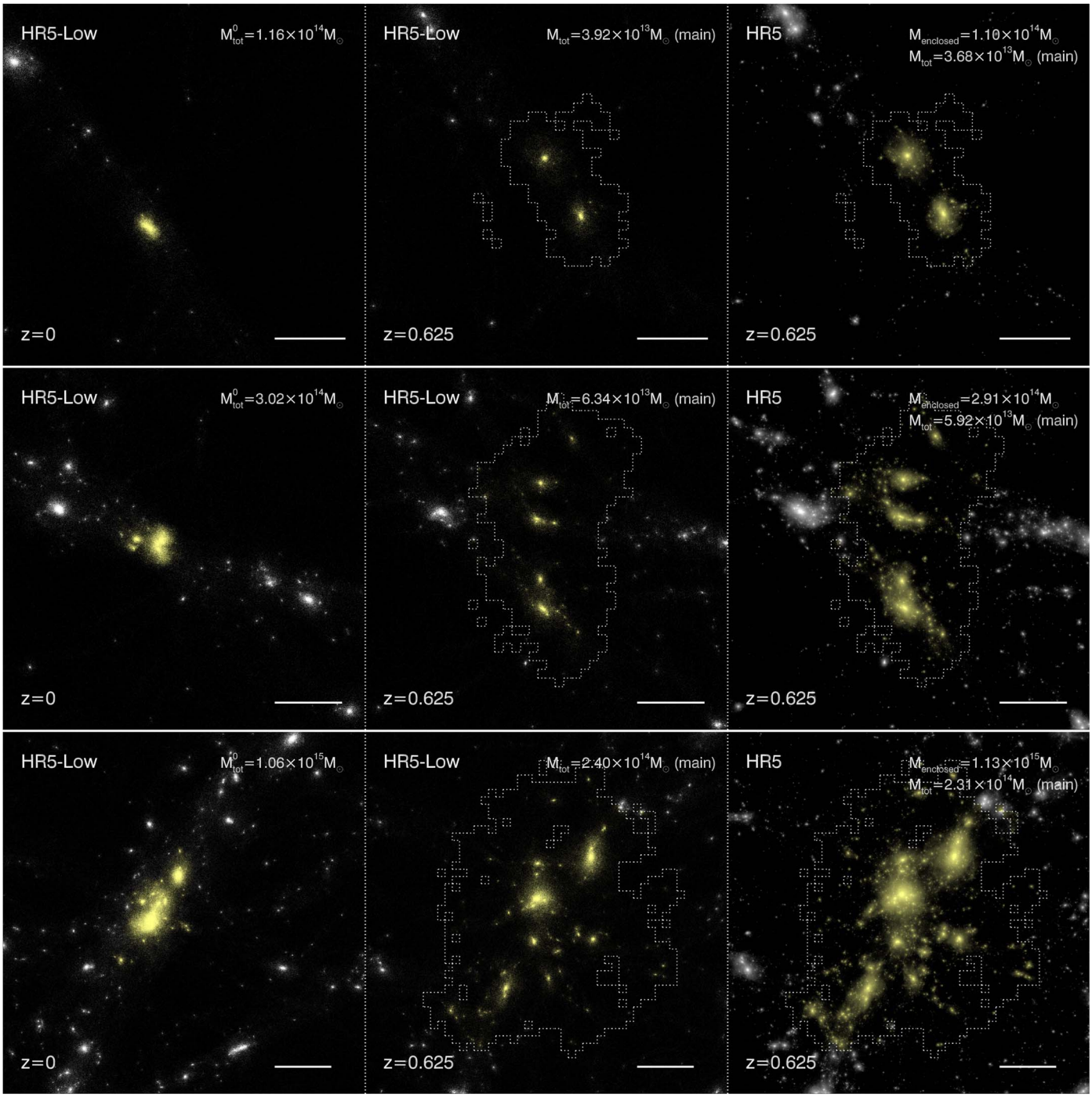

Figure 1. Dark matter particles in three clusters found at $z = 0$ in HR5-Low (left), in their progenitors at $z = 0.625$ (middle), and in the same volumes in HR5 at $z = 0.625$ (right). In the left panels, the halos in yellow are the members of the clusters with $M^0_{\rm tot} \sim 10^{14}\,M_\odot$ (top), $10^{14.5}\,M_\odot$ (middle), and $10^{15}\,M_\odot$ (bottom). White horizontal bars illustrate the scale of 4cMpc. The white dotted lines display the LVs enclosing the dark matter particles that end up forming clusters at $z = 0$ in HR5-Low. We assume that all the objects in HR5 located inside the same LV are the progenitors of the corresponding cluster. The thickness of the projected volume is 8.2 cMpc (top), 13.8 cMpc (middel), and 21.5 cMpc (bottom), fully containing each cluster in the projected direction. In the right panels, $M_{\rm enclosed}$ presents the total mass enclosed by the LV. All the objects inside the LV are traced back to high redshifts using their merger trees in this study.

3.1. Identification of Proto-objects Using Merger Trees

We search for the bona fide progenitors of each cluster or protocluster of HR5 at $z = 0.625$ by tracing backward their merger histories. All the progenitors of each object are identified in all snapshots. Note that we do not call all the progenitors the protocluster galaxies, as protocluster galaxies will be defined as those within the turnaround radius. We define the most massive galaxy among the progenitors in a snapshot as the central galaxy. Thus, the central galaxy of a protocluster may change over time, depending on their mass accretion history.

The bottom panel of Figure 3 shows the distribution of the galaxies belonging to clusters or protoclusters in comoving space at $z = 0.625$. The top four panels show their progenitors that are traced along merger trees. Red, yellow, and blue dots

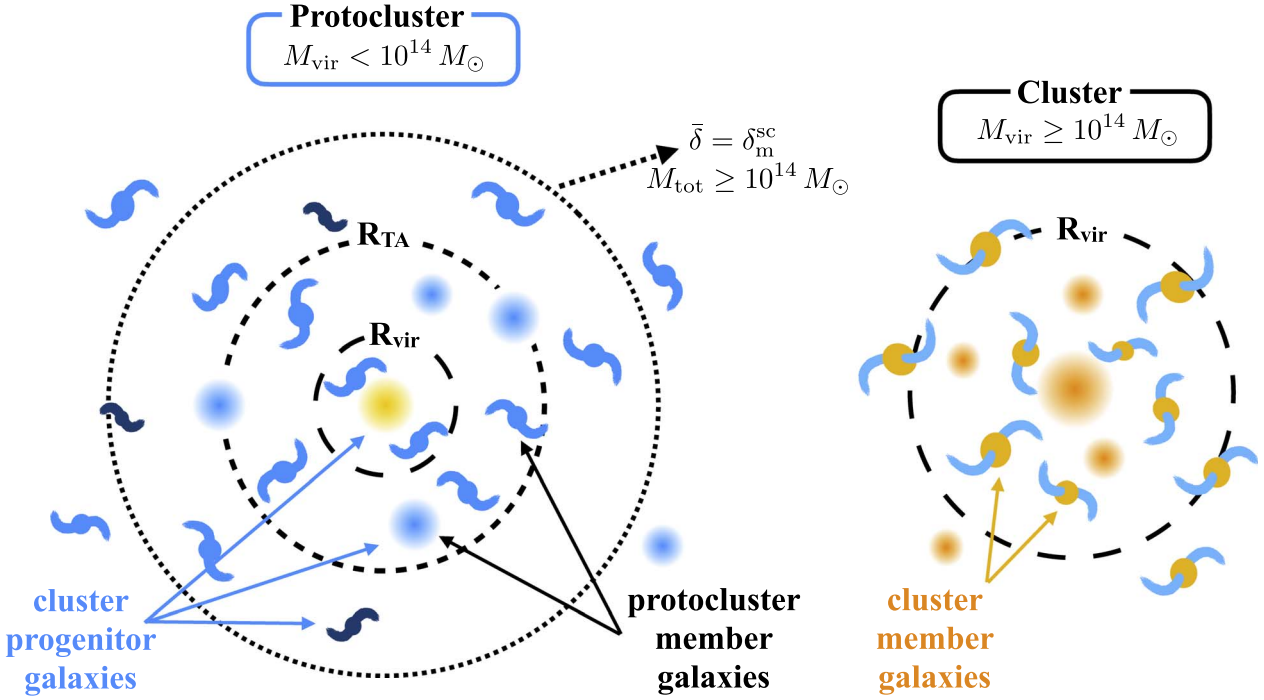

Figure 2. A schematic diagram presenting the definition of galaxy protoclusters and clusters. Clusters are groups of galaxies with $M_{\rm vir}$ currently greater than $10^{14}\,M_\odot$. Protoclusters are those with $M_{\rm vir} < 10^{14}\,M_\odot$ currently, but they will have $M_{\rm vir} \geqslant 10^{14}\,M_\odot$ by $z = 0$. The future virial mass is estimated from the total mass within the region having the mean overdensity $\bar{\delta}$ equal to the critical overdensity $\delta_m^{\rm sc}$ for complete collapse at $z = 0$ predicted by the spherical top-hat theory. The physical volume of protoclusters is defined to be the region within the turnaround radius $R_{\rm TA}$.

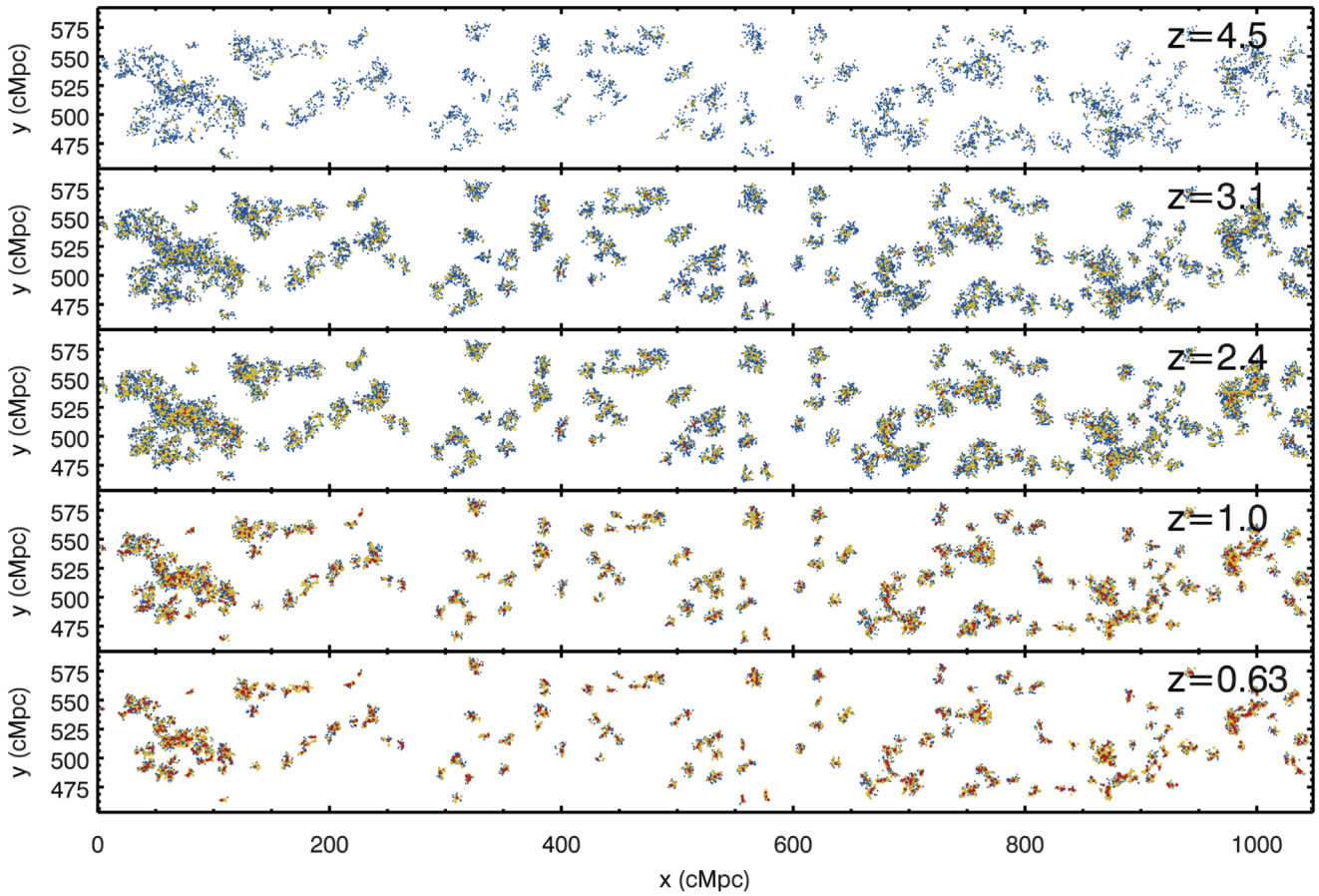

Figure 3. Distribution of cluster progenitor galaxies in HR5 at $z = 0.63$–4.5. Blue, yellow, and red dots mark the locations of the progenitor galaxies with stellar mass of $\log M_\star/M_\odot = 9$–10, 10–11, and >11, respectively.

mark the galaxies with $M_\star > 10^{11}\,M_\odot$, 10^{10}–$10^{11}\,M_\odot$, and 10^9–$10^{10}\,M_\odot$, respectively. It can be seen that the overall locations of protoclusters hardly change over time: the initial conditions are essentially preserved for these massive objects sitting at deep gravitational potential minima. On the other hand, the systems of cluster progenitor galaxies have been monotonically shrinking since $z \sim 2.4$. However, at redshifts higher than $z \sim 2.4$, their extent is roughly static at the value of $R \sim 10$–30 cMpc, and the systems start to fade away. The three redshifts of $z = 2.4$, 3.1, and 4.5 are the target redshifts of the ODIN survey for LAEs at $z = 2.4$, 3.1, and 4.5 (Ramakrishnan et al. 2022). We will discuss the results of this study mainly at these redshifts.

3.2. Identification of Protocluster Candidates Based on the Spherical Top-hat Collapse Model

In this subsection, we propose a systematic method to identify the candidate regions enclosing protoclusters from galaxy distribution based on the SC models.

3.2.1. Overdensity Threshold for Complete Collapse at z = 0

We define protocluster candidate regions as the spherical volumes that enclose total mass greater than $10^{14}\,M_\odot$ and will collapse completely at $z = 0$ according to the overdensity threshold given by the spherical top-hat collapse model. We will search for the centers of protoclusters inside the spherical regions.

In the spherical top-hat collapse model, an overdense region at an epoch will contract into a point at some stage if its overdensity is equal to the critical threshold density. We find this threshold density as a function of redshift for two types of cosmology. In the Einstein−de Sitter (EdS) universe with $\Omega_m = 1$, a homogeneous density sphere that collapses at $z = 0$ reaches its maximum radius at $z = 0.59$ with $\delta_m^{\rm sc} = 9\pi^2/16 - 1 \simeq 4.55$, where $\delta_m^{\rm sc}$ is the spherical top-hat matter overdensity. See Appendix C for more details. For comparison, the linear theory predicts overdensity $\delta_m^{\rm lin} \simeq 1.062$ at $t_{\rm max}$ in the EdS universe.

On the other hand, the SC model does not have an exact analytic solution in a flat universe with a nonzero cosmological constant, i.e., $\Omega_m + \Omega_\Lambda = 1$. We thus numerically solve the second-order nonlinear differential equation of the spherical top-hat overdensity $\delta_m^{\rm sc}$ given in Pace et al. (2010):

$$\ddot{\delta}_m^{\rm sc} + \left(\frac{3}{a} + \frac{\dot{E}(a)}{E(a)}\right)\dot{\delta}_m^{\rm sc} - \frac{4}{3}\frac{(\dot{\delta}_m^{\rm sc})^2}{1+\delta_m^{\rm sc}} - \frac{3}{2}\frac{\Omega_m}{a^5 E^2(a)}\delta_m^{\rm sc}(1+\delta_m^{\rm sc}) = 0, \quad (1)$$

where the derivatives are with respect to the expansion factor a, and $E(a) = H(a)/H_0 = \sqrt{\Omega_m/a^3 + \Omega_\Lambda}$, where $H(a)$ and H_0 are the Hubble parameter at the epoch of an expansion factor a and $z = 0 (a = 1)$, respectively. The density parameters of $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ are adopted in this calculation. We numerically search for the initial conditions $\delta_m^{{\rm sc},i}$ and $\dot{\delta}_m^{{\rm sc},i} = \delta_m^{{\rm sc},i}/a_i$ at $a_i = 10^{-3}$ that lead to $\delta_m^{\rm sc} \to \infty$ at $z = 0$ and find a solution $\delta_m^{{\rm sc},i} = 2.16 \times 10^{-3}$. The evolution of $\delta_m^{\rm sc}$ is shown in Figure 4 (dashed line). For the general flat universe with nonzero Ω_Λ, a fitting formula for the numerical solution of the SC model for the objects collapsing at $z = 0$ is given in Appendix C.

3.2.2. Overdensity of the HR5 Regions to be Collapsed

The SC model gives insight into the evolution of overdensities based on a simple assumption of homogeneous density distribution in a spherical region. However, in the real universe, structures are generally not spherical or homogeneous. To examine whether the simple assumption is applicable to practical cases, we compare the critical overdensity predicted by the SC model with the actual overdensity of the spherical region at a high redshift that encloses $M_{\rm tot}^0$, the total mass of each cluster at $z = 0$ measured in the HR5-Low simulation. The sphere is centered at the most massive galaxy among all the cluster progenitors at the redshift.

The open circles in the top panel of Figure 4 show the mean matter overdensity within the radius $R(M_{\rm tot}^0)$ from the most massive progenitor of each of 189 HR5-Low clusters. It should be noted that $\delta_m[R(M_{\rm tot}^0)]$ for HR5 clusters agrees quite well with the prediction of the SC model (dashed line) at all redshifts in the flat ΛCDM universe. This result demonstrates that the SC model is remarkably accurate in the ΛCDM universe at the mass scale of galaxy clusters, and thus the critical density threshold is applicable to identify protocluster regions.

3.2.3. Identification of the Regions Enclosing Protoclusters

We have shown a good agreement between the spherical top-hat overdensity predicted by the SC model and that actually measured for the HR5 clusters. However, to propose a protocluster identification scheme applicable to observations, it is necessary to find the relation between the total mass and stellar mass at the cluster mass scale. For the clusters with $\log M_{\rm tot}^0/M_\odot > 14$ the bottom panel of Figure 4 shows the ratio of stellar mass to total mass $M_\star/M_{\rm tot}$ within the spherical region having the critical overdensity of $\delta_m^{\rm sc}$ at redshift z. Open diamonds are the ratio when only the stars of the galaxies with $M_{{\rm gal},\star} > 2 \times 10^9\,M_\odot$ are used, and open circles are those when all stars are taken into consideration. We provide a fitting formula for the stellar-to-total mass relation in the following form:

$$\log M_\star/M_{\rm tot} = \alpha(1+z)^\beta + \gamma. \quad (2)$$

This formula can fit the ratio well as a function of redshift with $(\alpha,\ \beta,\ \gamma) = (-0.055,\ 1.903,\ -1.915)$ when the galaxies of $M_{{\rm gal},\star} > 2 \times 10^9\,M_\odot$ are used (shown as the dotted curve fitting the diamonds in the bottom panel of Figure 4). When all stellar components are used (open circles), the best fit is made with the parameter set $(-0.057,\ 1.755,\ -1.855)$. We note that the stellar-to-total mass relation is insensitive to mass in the case of the proto-objects of $M_{\rm tot}^0 > 10^{13}\,M_\odot$. This is because the region having the mean overdensity $\delta_m^{\rm sc}$ is typically so large that the ratio converges to a value at a given redshift. The stellar-to-total mass ratio relation can be changed if the parameters of subgrid physics regulating star formation activities are changed. Therefore, the relation needs to be calibrated based on observations.

The protocluster identification starts with finding the candidate regions that enclose protoclusters. At a given epoch, we visit galaxies, starting from the most massive ones, and inspect the spherical volume centered at the galaxy. The radius of the sphere is increased until the overdensity drops to the critical value $\delta_{\rm m}^{\rm sc}$ at that epoch. If the total mass contained

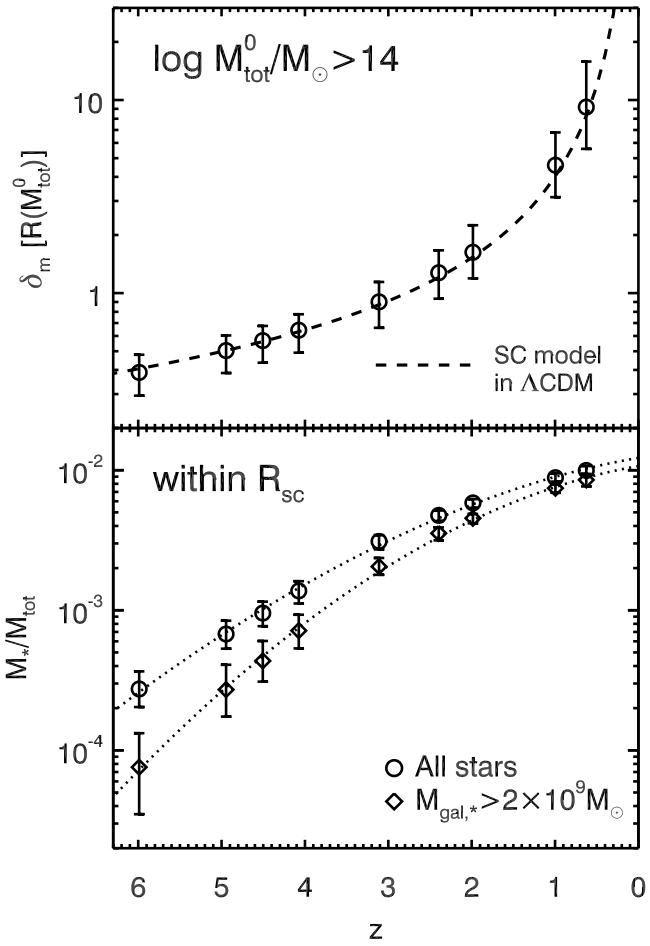

Figure 4. Top: matter overdensity inside the radius enclosing the final mass of protoclusters ($M^0_{\rm tot} > 10^{14}\,M_\odot$) as a function of redshift. Circles are the medians, and scatter bars show 16th–84th percentile distributions. The dashed line is the critical matter overdensity $\delta^{\rm sc}_m$ for collapse at $z = 0$ predicted by the spherical top-hat collapse model in the ΛCDM universe. Bottom: ratio of stellar mass to total mass within the protocluster regions whose mean overdensity is equal to the critical value $\delta^{\rm sc}_m$ of the ΛCDM cosmology. Open circles show the ratios computed from the entire stellar mass, and open diamonds are calculated from the galaxies with $M_{\rm gal,\star} > 2 \times 10^9\,M_\odot$. The dotted curves are the fitting functions given in Equation (2).

within the sphere exceeds $10^{14}\,M_\odot$, the galaxy can be assumed as a candidate for the center of a protocluster. The fitting formula in Equation (2) is used to convert the observed stellar mass to the total mass.

The central candidate galaxies do not always locate at the density peak of each sphere. Thus, we compute the center of mass (CM) from all the galaxies with $M_{\rm gal,\star} > 2 \times 10^9\,M_\odot$ located inside the spherical regions. To find the most representative center of galaxy distribution, we iterate the identification process until the CM converges to $|\boldsymbol{x}_{i-1} - \boldsymbol{x}_i| < \epsilon$, where $\boldsymbol{x}_i$ is the CM at the ith iteration. In this study, we adopt $\epsilon = 0.25$ cMpc for efficient searching since a smaller ϵ does not notably affect the results. A sphere is selected as a region enclosing protoclusters when it finally has $M_{\rm tot} \geqslant 10^{14}\,M_\odot$ after the iteration process.

In a dense environment, the separations between the centers of the protocluster candidates can be very small. We combine a protocluster candidate region i with another one j if $D_{ij}/R_i < 1.0$ or $D_{ij}/R_j < 1.0$, where D_{ij} is the distance between the centers and R_i and R_j are the radii of the spheres within which the mean overdensity meets $\delta^{\rm sc}_m$. In this case, we define the most massive sphere as the central one, and accordingly, $M_{\rm tot}$ of the central one is set as the estimated total mass of a spherical region group (SRG).

3.2.4. Reliability of the Protocluster Identification Scheme

We assume that the objects in the spherical regions of protoclusters identified based on the SC model eventually form cluster-scale objects by $z = 0$. We evaluate the reliability of this approach by comparing the total mass of an SRG ($M^{\rm SRG}_{\rm tot}$) at a redshift z with the mass $M^{0,{\rm SRG}}_{\rm tot}$ that ends up being inside clusters at $z = 0$. The latter is estimated using the final total mass weighted by the stellar mass of the cluster progenitor galaxies found within the SRG as follows:

$$M^{0,{\rm SRG}}_{\rm tot} = \sum_{i=0}^{n} M(G \bigcap P_i)/M(P_i) \times M^0_{{\rm tot},i}, \tag{3}$$

where G is the set of the galaxies enclosed by an SRG, P_i is the set of progenitor galaxies of a cluster i, $M(P_i)$ is the mass sum of P_i, and $M^0_{{\rm tot},i}$ is the final total mass of cluster i. The relation between $M^{\rm SRG}_{\rm tot}$ and $M^{0,{\rm SRG}}_{\rm tot}$ tells us how reliably the spherical top-hat model predicts the final mass of enclosed objects.

It is reasonable to expect that the growth history of an SRG can be affected by its environment and the above relation may depend on the history. So we inspect whether the final mass depends on both $M^{\rm SRG}_{\rm tot}$ and mass growth environment. As a proxy of the environment, we choose $D_1/R_{\rm SRG}$, where D_1 is the distance to the nearest neighbor SRG and $R_{\rm SRG}$ is the radius of the target SRG. An SRG should have the total mass larger than half the total mass of the target SRG of interest to be qualified as a neighbor.

Figure 5 shows the final mass $M^{0,{\rm SRG}}_{\rm tot}$ (encoded by color of large circles) in the $D_1/R_{\rm SRG}$ versus $M^{\rm SRG}_{\rm tot}$ space. Redder color indicates larger final mass. Small circles are the SRGs with $M^{\rm SRG}_{\rm tot} \geqslant 10^{14}\,M_\odot$ at redshift z but with $M^{0,{\rm SRG}}_{\rm tot} \leqslant 10^{14}\,M_\odot$ at $z = 0$, namely failed protocluster candidates. The figure demonstrates a tight correlation of $M^{\rm SRG}_{\rm tot}$ with $M^{0,{\rm SRG}}_{\rm tot}$, which justifies our use of the spherical overdensity criterion for identifying the protocluster centers. In particular, 90% of the SRGs whose $M^{\rm SRG}_{\rm tot}$ is larger than $10^{14.2}\,M_\odot$ end up having $M^{0,{\rm SRG}}_{\rm tot} > 10^{14}\,M_\odot$, indicating that they probably contain the authentic protoclusters. This illustrates the high reliability of our identification scheme. This figure also demonstrates that the final mass to be included in clusters is rather independent of the environment represented by the nearest neighbor SRG distance. We find, however, that the purity slightly improves if we discard the isolated low-mass SRGs with $M^{\rm SRG}_{\rm tot} < 10^{14.15} M_\odot$ and $D_1/R_{\rm SRG} > 2.5$ (the region enclosed by double dotted–dashed lines). Based on these criteria, we examine the purity and completeness of our approach in identifying the bona fide protoclusters in Appendix D. We find that the identification scheme recovers the authentic protoclusters with high reliability. We also show in Appendix E that the redshift-space distortion (RSD) does not significantly affect the performance of the protocluster identification scheme.

4. Protocluster Member Galaxies within Turnaround Radius

In numerical simulations and theories, it is relatively easy to define a protocluster as a group of objects that eventually contracts and forms a cluster. As described in Section 3.1, the

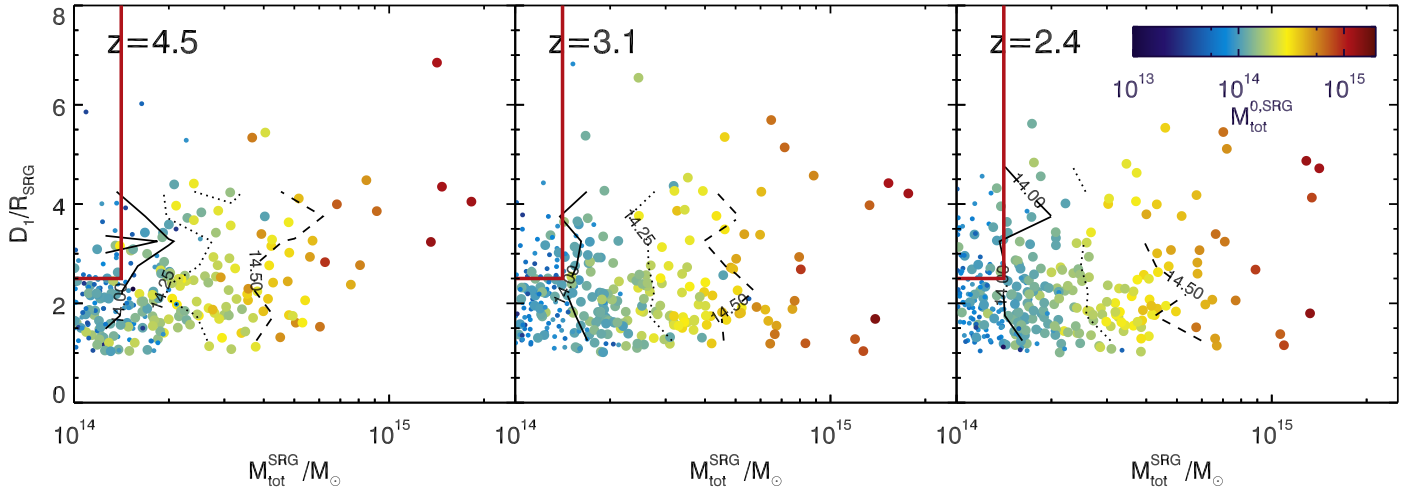

Figure 5. Final total mass $M_{\rm tot}^{0,\rm SRG}$ (encoded by color) as functions of distance to the nearest SRG D_1 (normalized by $R_{\rm SRG}$) and its total mass $M_{\rm tot}^{\rm SRG}$ at each redshift. Each circle indicates an SRG, and color represents the final mass $M_{\rm tot}^{0,\rm SRG}$ measured as in Equation (3). Larger circles are the SRGs with cluster-scale mass ($M_{\rm tot}^{0,\rm SRG} \geqslant 10^{14}\,M_\odot$). The black solid, dotted, and dashed curves delineate the region of average final mass of $M_{\rm tot}^{0,\rm SRG} = 10^{14}$, $10^{14.25}$, and $10^{14.5}\,M_\odot$, respectively. The SRGs in the upper left corner demarcated by red lines are discarded in this work as protocluster candidates.

progenitors of cluster galaxies can be traced using their merger trees in numerical simulations, and the corresponding protoclusters can be identified.

However, as shown in Figure B3, the progenitor galaxies of clusters are widespread up to ∼30 cMpc at high redshifts, and it is not reasonable to adopt all the progenitor galaxies as the physically associated members of protoclusters. Most observations identify protoclusters by finding sufficiently overdense regions of galaxies (see Overzier 2016, and references therein). However, there has been no consensus on the value of the overdensity defining the membership of protocluster galaxies.

Applying the virial radius in identifying protocluster galaxies is not so desirable, as protoclusters are supposed to be the objects still under the process of formation and virialized regions of protoclusters tend to vanish quickly as redshift increases.

We thus propose to define the protocluster member galaxies as those within the zero proper velocity surface from protocluster center. The distance from a density peak to the zero-velocity surface is dubbed the turnaround radius $R_{\rm TA}$. The turnaround radius is the distance to the spherical surface on which the gravitational infall counterbalances the Hubble expansion (Gunn & Gott 1972). The turnaround radius provides a theoretically motivated overdensity for defining the protocluster region and also makes protoclusters physical objects where their member galaxies can have some degree of conformity. In this section we present a scheme for finding $R_{\rm TA}$ from observed galaxy distribution.

4.1. Turnaround Radius

To measure $R_{\rm TA}$ from the protocluster centers in HR5, we construct the matter (dark matter, gas, and stars) density and peculiar velocity fields on a uniform grid with pixel size of $\Delta x = 0.128$ cMpc. The proper radial velocity v_r at $\boldsymbol{r}_1$ relative to a local density peak at $\boldsymbol{r}_0$ is given as follows:

$$v_r = H(z)|\boldsymbol{r}| + \boldsymbol{e_r}\cdot\boldsymbol{v}, \tag{4}$$

where $\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_0$, $H(z)$ is the Hubble parameter at redshift z, $\boldsymbol{e_r}$ is the unit vector of $\boldsymbol{r}$, and $\boldsymbol{v}$ is the peculiar velocity at $\boldsymbol{r}_1$ relative to the mean velocity of matter within $|\boldsymbol{r}|$. The turnaround radius is measured by finding the radius of a shell on which the average v_r becomes zero.

As an illustration, Figure 6 shows the matter density and velocity fields of an HR5 protocluster region at four redshifts. The blue and yellow circles indicate $R_{\rm vir}$ and $R_{\rm TA}$, respectively, centered at the most massive galaxy in the field at each epoch. Arrows are the proper velocity vectors projected onto a 4 cMpc thick slice centered at the galaxy. The overdensity of the protocluster increases with time, and consequently, both $R_{\rm TA}$ and $R_{\rm vir}$ increase with time too. It can be seen that $R_{\rm vir}$ contains only the very center of the protocluster and becomes uninterestingly too small at high redshifts. On the other hand, $R_{\rm TA}$ is much larger than $R_{\rm vir}$, does separate the inner collapsing region from the outer expanding space, and embraces the high-density region of intersecting filaments of galaxies. In this sense $R_{\rm TA}$ defines the outer boundary of the protocluster, and the galaxies within $R_{\rm TA}$ can be called its "members." Even though protocluster members are identified only within a spherical region, their distribution is quite anisotropic, as the region encloses connecting filaments.

Figure 7 shows $R_{\rm TA}$ of the HR5 protoclusters and protogroups at four redshifts as a function of their final total mass at $z = 0$. $R_{\rm TA}$ has a good correlation with the final mass. The tightness of the correlation increases toward low redshifts. The linear Pearson correlation coefficient is 0.634 at $z = 4.5$, and this increases to 0.81 at $z = 1.0$ in the $\log R_{\rm TA} - \log M_{\rm tot}^0/M_\odot$ plane. We have also checked whether the turnaround radii measured from the most massive galaxies in SRGs are accurate compared to those of bona fide protoclusters, and we find that more than 80% of the SRGs have $R_{\rm TA}$ identical to that of the bona fide protoclusters (Appendix F).

4.2. Correlations between Turnaround Radius, Virial Mass, and Viral Radius

In this section we study the general nature of the turnaround radius by inspecting its relation with the virial mass and radius. The turnaround radius is known to be 3–4 times the virial radius of massive objects in the local universe (Mamon et al. 2004; Wojtak et al. 2005; Rines & Diaferio 2006; Cuesta et al. 2008; Falco et al. 2013). The virial mass of an object is defined

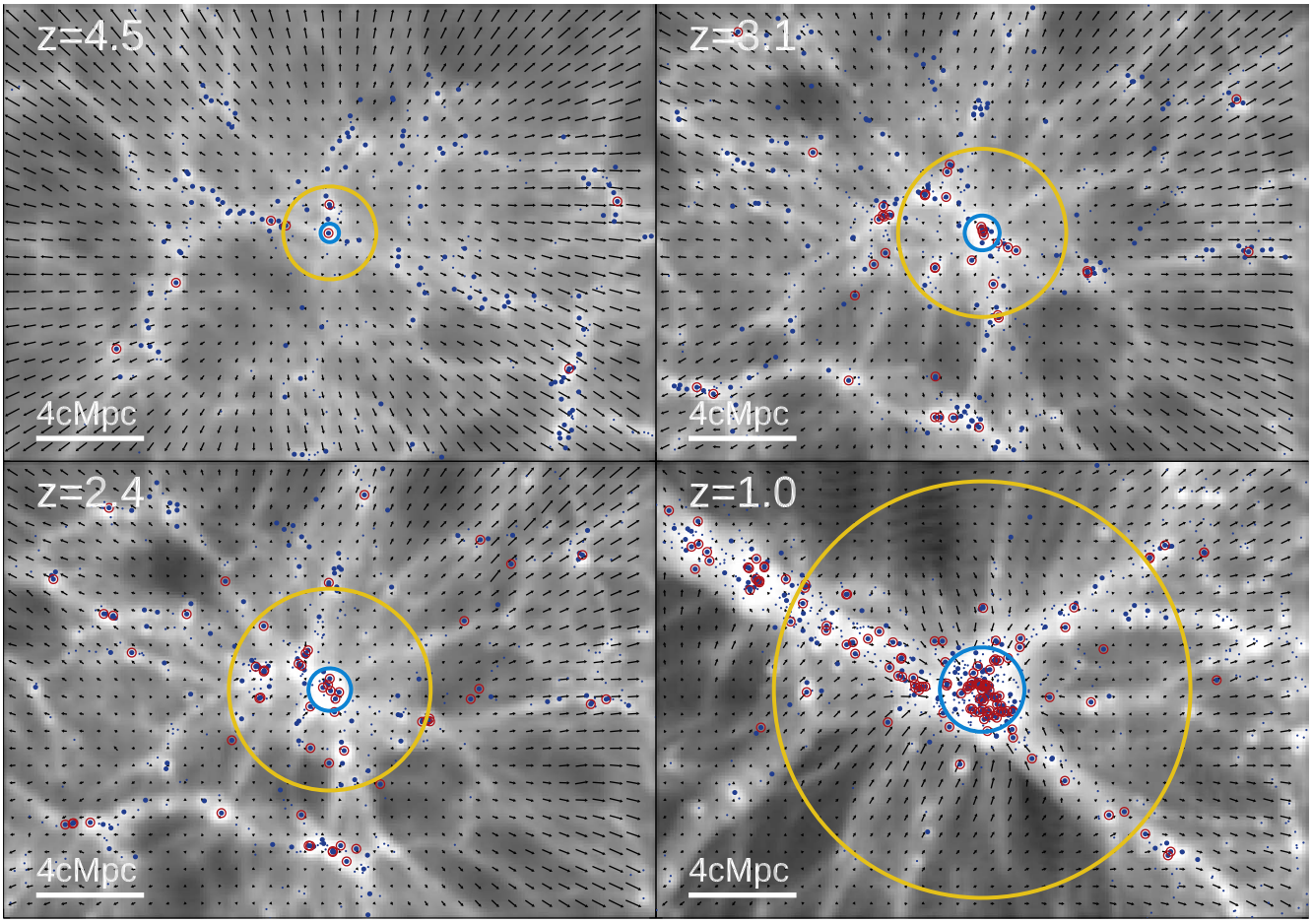

Figure 6. Matter density and velocity fields within and in the vicinity of a protocluster at four epochs. Denser regions are brighter. The panels show matter distribution within ±2 cMpc from the most massive galaxy along the projected direction. Blue and yellow circles indicate $R_{\rm vir}$ and $R_{\rm TA}$ measured from the density peak, respectively. All the blue circles are the gravitationally self-bound objects with the total mass greater than $10^{10}\,M_{\odot}$. Larger blue circles are the cluster progenitor objects, and among them, those with $M_{\star} > 2\times10^{9}\,M_{\odot}$ are marked by red open circles.

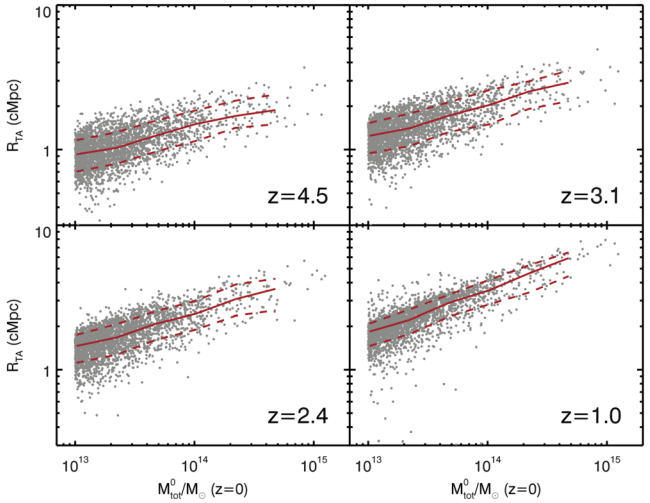

Figure 7. Turnaround radius $R_{\rm TA}$ of the proto-objects as a function of the final total mass $M^{0}_{\rm tot}$ at $z=0$ that is measured in HR5-Low. Contrary to R_{95}, $R_{\rm TA}$ gradually increases as dense regions grow in mass and the Hubble parameter decreases with decreasing redshift.

as $M_{\rm vir} = 4\pi r_{\rm vir}^{3}\Delta_c\rho_c/3$, where $r_{\rm vir}$ is the virial radius within which the mean matter density is Δ_c times the critical density of the universe $\rho_c = 3H^2/8\pi G$, where H is the Hubble parameter at z and Δ_c is computed using the fitting formula derived by Bryan & Norman (1998) for the cosmology with $\Omega_\Lambda > 0$:

$$\Delta_c = 18\pi^2 + 82x - 39x^2, \tag{5}$$

where $x = \Omega_m(z) - 1$.

Meanwhile, the total mean radial velocity at r from the center of a bound object is the sum of the Hubble expansion velocity and mean infall peculiar velocity: $\langle v_r\rangle = H(z)r + \langle v_{\rm infall}(r)\rangle$, where $\langle v_{\rm infall}(r)\rangle$ is the averaged radial velocity of matter in a spherical shell at radius r.

In the region where the Hubble flow starts to dominate and the total mean radial velocity becomes positive, Falco et al. (2014) found a good approximation for the infall velocity profile as follows:

$$\langle v_{\rm infall}\rangle \approx a v_{\rm vir}\left(\frac{r}{r_{\rm vir}}\right)^{-b}, \tag{6}$$

where $v_{\rm vir} = \sqrt{GM_{\rm vir}/r_{\rm vir}}$ is the circular velocity at $r_{\rm vir}$ and a and b are free fitting parameters. The best-fit values found are $a = 0.8\pm0.2$ and $b = 0.42\pm0.16$ at $z=0$ in the N-body simulations of a ΛCDM universe with $\Omega_m = 0.24$ and $h = 0.73$ (Falco et al. 2014). Since $\langle v_r\rangle = 0$ at the turnaround radius, the ratio of $R_{\rm TA}$ to $r_{\rm vir}$ can be reduced to

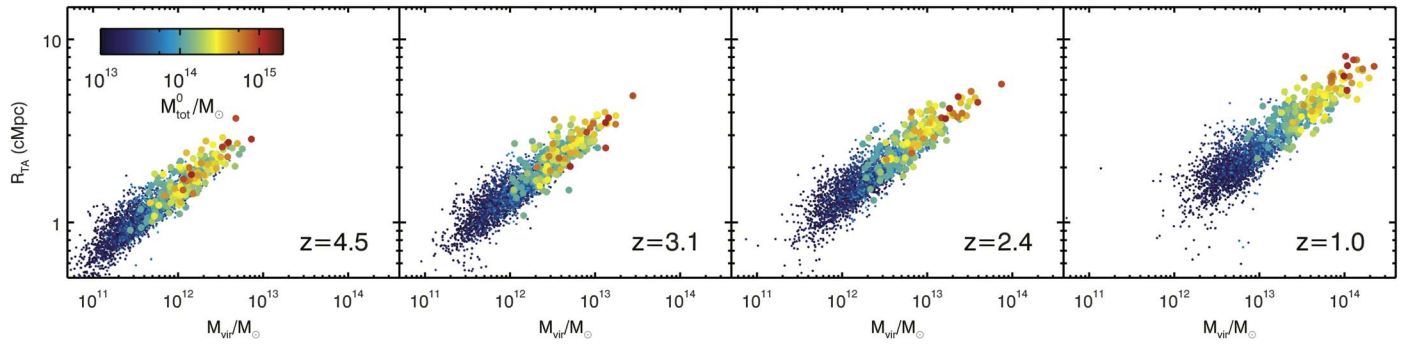

Figure 8. Relations between the turnaround radius $R_{\rm TA}$ and virial mass $M_{\rm vir}$ at four epochs for the proto-objects that will have the final total mass of $M^0_{\rm tot}$ that is measured in HR5-Low. The final total mass is color-coded. Protoclusters are marked by large filled circles, and non-protocluster objects are marked by small circles.

$R_{\rm TA}/r_{\rm vir} = (a\sqrt{\Delta_c/2})^{1/(b+1)}$ by combining the equations above with $r = R_{\rm TA}$ and $\langle v_r \rangle = 0$. Thus, the ratio $R_{\rm TA}/r_{\rm vir}$ is expected to be ∼4.3 and in the range of 3.1–6.2 at $z = 0$.

We now inspect the relation of $R_{\rm TA}$ with $M_{\rm vir}$ or $R_{\rm vir}$ directly for the HR5 protocluster/group regions. Measurements are made relative to the most massive galaxy in each region. Figure 8 demonstrates the tight correlation between $R_{\rm TA}$ and the virial mass at each epoch. Objects are distinguished in color according to their total mass at $z = 0$. Notice that the relation moves slowly downward with time and $R_{\rm TA}$ decreases at the same virial mass at lower redshifts.

A weak evolution of the turnaround-to-virial radius ratio can be seen in Figure 9 for protoclusters (red; $M^0_{\rm tot} \geqslant 10^{14}\,M_\odot$) and the protogroups (blue; $M^0_{\rm tot} = 10^{13}$–$10^{14}\,M_\odot$). The median of the ratio slowly decreases from 4.8 at $z = 6$ to 3.9 at $z = 0.625$ for protoclusters or clusters (red). The decreasing rate of the ratio is higher at $z < 2$ than before as Δ_c significantly lowers. The ratio also decreases a little faster for protogroups. This seems to be caused by the disturbance of velocity field that becomes more severe for lower-mass objects at lower redshifts. The major origin of this weak redshift dependence will be discussed in the next section. Our measurement of $R_{\rm TA}/R_{\rm vir}$ at $z = 0.625$ is consistent with the ratio range of 3.1–6.2 derived based on the semianalytic approach of Falco et al. (2014).

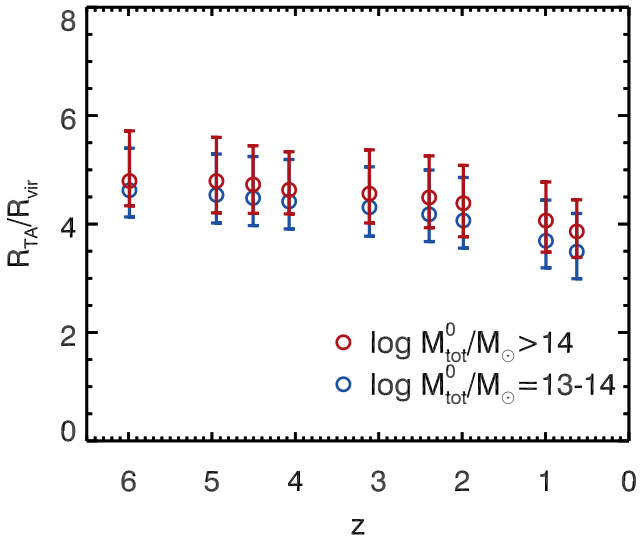

Figure 9. Ratio of the turnaround radius $R_{\rm TA}$ to the virial radius $R_{\rm vir}$ as a function of redshift. The blue and red circles correspond to the structures with $\log M^0_{\rm tot}/M_\odot = 13$–$14$ and $\log M^0_{\rm tot}/M_\odot > 14$ at $z = 0$,measured from HR5-Low, respectively. The scatter bars show 16th–84th percentile distributions. This figure indicates that $R_{\rm TA}/R_{\rm vir}$ evolves very weakly before $z = 2$.

4.3. Matter Overdensity within Turnaround Radius

The tight correlation between $R_{\rm TA}$ and $R_{\rm vir}$ implies nearly constant overdensity within $R_{\rm TA}$ at $z > 2$. We measure the average matter overdensity of the HR5 proto-objects inside the sphere of radius $R_{\rm TA}$. Figure 10 presents the matter overdensity $\delta_m^{\rm TA}$ as a function of $R_{\rm TA}$ for all proto-objects. The large circles mark the protoclusters, and small circles are protogroups with the final mass of $M^0_{\rm tot} = 10^{13}$–$10^{14}\,M_\odot$. The turnaround radius $R_{\rm TA}$ of protoclusters can temporarily decrease and $\delta_m^{\rm TA}$ can jump up when they undergo close encounters with neighbors. In order to mitigate the impact of such temporal events, we choose to use the lower boundary (bottom 5%) of the distribution of $\delta_m^{\rm TA}$ shown in Figure 10 for the threshold overdensity corresponding to $R_{\rm TA}$. When protoclusters have close neighbors, the radius found with the lower boundary will be somewhat larger than the actual turnaround radius directly measured, and the protocluster regions are allowed to overlap. The bottom 5% of the distribution of $\delta_m^{\rm TA}$ are 4.96, 5.04, 5.30, and 6.55 at $z = 4.5$, 3.1, 2.4, and 1.0, respectively. The median and 1σ dispersion are 5.63 ($\sigma = 0.58$), 5.98 (1.01), 6.17 (1.07), and 7.71 (1.91), respectively.

Like $R_{\rm TA}/R_{\rm vir}$, $\delta_m^{\rm TA}$ also weakly evolves over time, with small scatter for protoclusters. It should be noted that $\delta_m^{\rm TA}$ hardly depends on $R_{\rm TA}$ or the final cluster mass of the protoclusters (large circles). On the other hand, $\delta_m^{\rm TA}$ of the low-mass structures with relatively small $R_{\rm TA}$ shows stronger evolution. The scatter of $\delta_m^{\rm TA}$ at small $R_{\rm TA}$ emerges when the field of interest is disturbed by neighboring structures.

Figure 11 illustrates the evolution of four HR5 protoclusters representing different total mass scales at $z = 0$. Dotted circles mark the turnaround radii, and properties shown are stellar mass density and age, gas density, and metallicity. Similar to Figure 6, Figure 11 again shows that the volume within the turnaround radius does encompass the interesting large-scale structures connected to the protocluster cores. Notice in Figure 11 that $R_{\rm TA}$ is not always larger for the protoclusters with larger mass. It is also possible for $R_{\rm TA}$ to decrease temporarily when mergers happen. This is a desirable nature of $R_{\rm TA}$, as it is supposed to define the member galaxies of protoclusters and separate them from approaching nearby objects. However, during close interactions with neighbors, $R_{\rm TA}$ becomes smaller and $\delta_m^{\rm TA}$ tends to increase. The upward scatter of the proto-objects in Figure 10 can be attributed to such events.

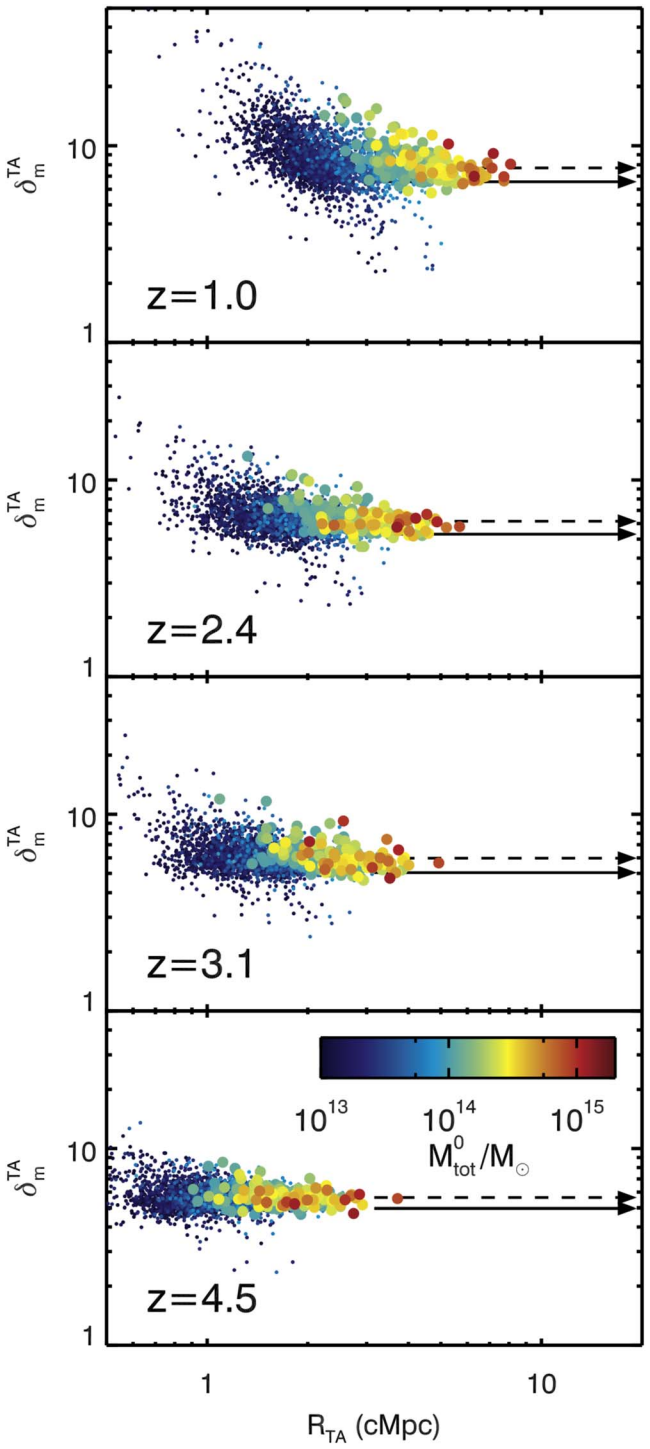

Figure 10. Matter overdensity within the turnaround radius of proto-objects at the four redshifts. Protoclusters ($M^0_{\rm tot} \geqslant 10^{14}\,M_\odot$ at $z=0$, measured in HR5-Low) are marked by large filled circles, and non-protocluster objects are marked by small circles. The color code presents the final total mass of proto-objects. The dashed and solid arrows indicate the medians and bottom 5% of $\delta_m^{\rm TA}$ of protoclusters, respectively. The matter overdensity of protoclusters only weakly increases from $\delta_m^{\rm TA} \approx 5.0$ (bottom 5%) at $z=4.5$ to $\delta_m^{\rm TA} \approx 5.3$ (bottom 5%) at $z=2.4$ (see the text and Figure 12).

4.4. Conversion of Stellar Mass to Total Mass within Turnaround Radius

We define the outer boundary of protoclusters as $R_{\rm TA}$, which is the turnaround radius enclosing the threshold overdensity given by Equation (7) below. We will use an empirical relation between the total mass and stellar mass within $R_{\rm TA}$ so that the definition can be applied to observations. Figure 12 shows the redshift evolution of $\delta_m^{\rm TA}$ of the HR5 protoclusters (top) and the stellar-to-total mass ratio within the turnaround radius, $M_\star^{\rm TA}/M_{\rm tot}^{\rm TA}$, averaged over the HR5 protoclusters. The stellar mass is obtained from all stars (red open circles) or only for the galaxies with $M_{\rm gal,\star} > 2\times10^9\,M_\odot$ (blue open circles).

The overdensity $\delta_m^{\rm TA}$ delineating the bottom 5% of the distribution at z can be fit well by the following formula:

$$\delta_m^{\rm TA}(z) = \frac{a\exp(b(1+z)^c)}{(1+z)^d}, \tag{7}$$

where $(a, b, c, d) = (0.168, 4.068, -0.381, -0.734)$, which is shown as the solid line in the top panel of Figure 12. The error of the fit is smaller than 0.9%. As shown in Section 4.3, $\delta_m^{\rm TA}$ monotonically increases with time on average and reaches a finite maximum at $z=0$. The evolution of $\delta_m^{\rm TA}$ is weak at $z>2$ but becomes rapid at $z<1.5$ owing to a decrease of the Hubble parameter and disturbance by neighboring structures.

The stellar-to-total mass ratio within the turnaround radius can be also fit well by Equation (2) with $(\alpha, \beta, \gamma) = (-0.0092, 2.027, -1.962)$ when all stellar mass is counted, or with $(-0.0128, 1.882, -2.017)$ when only the stellar mass in the galaxies with $M_{\rm gal,\star} > 2\times10^9\,M_\odot$ is used. We use this fitting formula to derive the total mass from the stellar mass within a radius from each protocluster center, and we find the radius within which the mean total mass density reaches the predicted $\delta_m^{\rm TA}$ at the given redshift (i.e., Equation (7)). This gives the estimated turnaround radius.

Figure 13 compares the directly measured $R_{\rm TA}$ with $R_{\rm TA}^{\rm est}$ estimated from stellar mass. They correlate quite well for both cases when all stellar mass is counted in $R_{\rm TA}^{\rm est}$ or only the stellar mass in the galaxies with $M_{\rm gal,\star} > 2\times10^9\,M_\odot$ is used. $R_{\rm TA}^{\rm est}$ tends to be larger than $R_{\rm TA}$ as expected, particularly for relatively lower mass protoclusters, because we use the bottom 5% $\delta_m^{\rm TA}$. At $z=4.5$, when protoclusters have only a few galaxies above our stellar mass threshold, the correlation breaks. This necessitates including the low-mass galaxies with $M_{\rm gal,\star} < 2\times10^9\,M_\odot$ at $z \gtrsim 4$ for accurate estimation of $R_{\rm TA}$ and reliable identification of protocluster environment.

5. Summary and Discussion

In this paper we have proposed a practical method to find galactic protoclusters in observational data and demonstrated its validity to the protoclusters in the cosmological hydrodynamical simulation HR5. We first define "protoclusters" as galaxy groups whose total mass within $R_{\rm vir}$ is currently less than $10^{14}\,M_\odot$ at their epochs but would exceed that limit by $z=0$. Conversely, "clusters" are the groups of galaxies whose virial mass currently exceeds $10^{14}\,M_\odot$. Therefore, there can be a mixture of clusters and protoclusters at $z>0$. The extent of a protocluster is defined as the spherical volume within the turnaround radius or the zero-velocity surface. The future mass that a protocluster would achieve at $z=0$ is estimated using the spherical top-hat collapse model. The whole concept is schematically visualized in Figure 2.

Our protocluster identification method is summarized as follows:

1. Visit galaxies, starting from the most massive ones, and measure the mean total mass density within radius R. The total

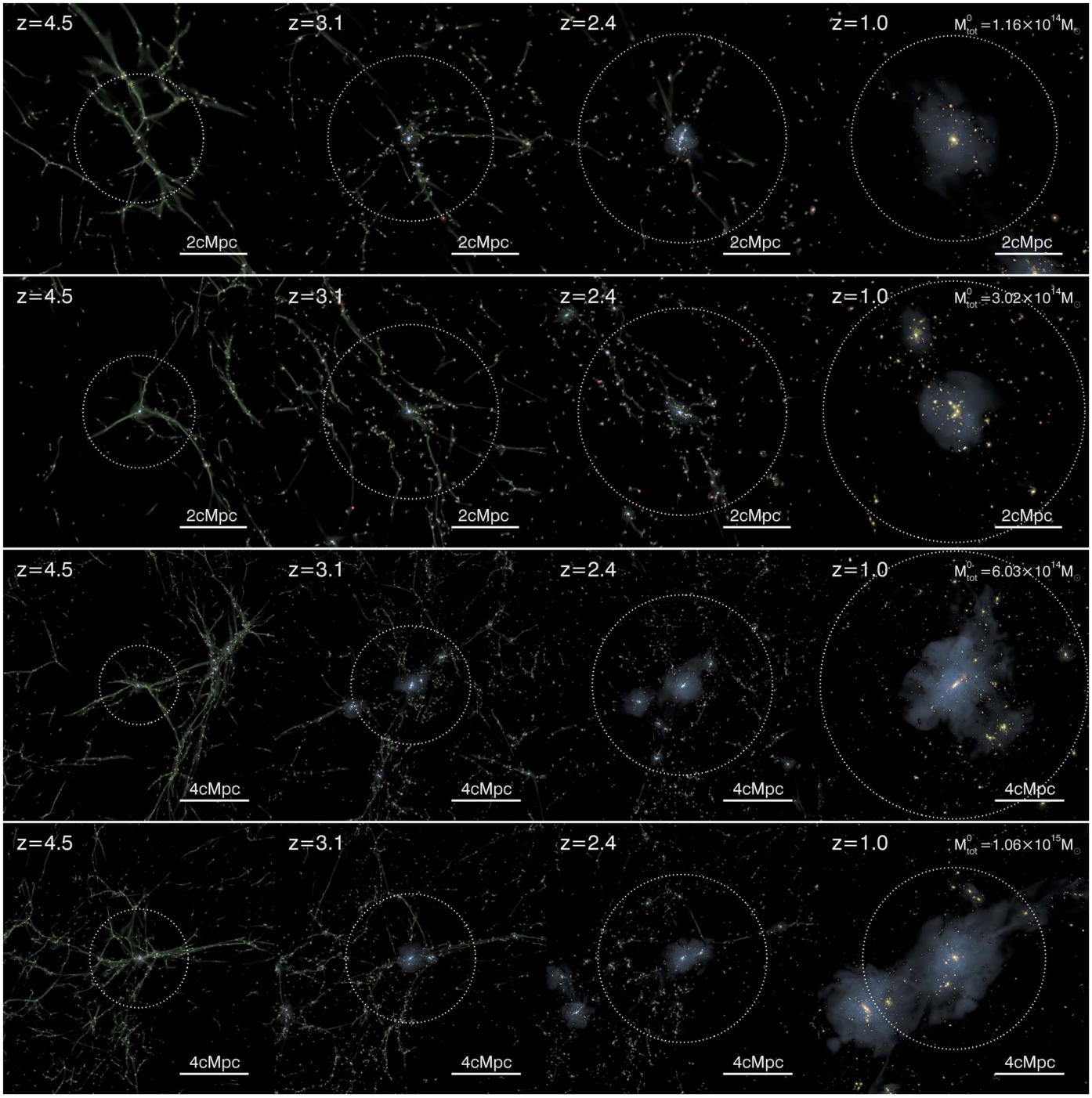

Figure 11. Distribution of gas and stars in the regions of four protoclusters that end up forming clusters with $M^0_{\rm tot} \approx 10^{14}$–$10^{15}\,M_\odot$ at $z = 0$ that is measured in HR5-Low. The dotted circles mark the turnaround radius of the protoclusters. Metal-poor gas is colored in green, and gas color becomes redder with increasing metallicity. Younger stars are shown in blue, and older ones are shown in yellow. Grayish shades display the regions filled with the hot medium with $T > 10^6$ K. The top two panels are relatively zoomed in, as indicated by the scale bars.

mass is obtained from the stellar mass by using the conversion relation in Equation (2).

2. Find the radius where the mean density drops to the threshold density given by the SC model. Equation (C6) is a useful fitting formula for the threshold overdensity $\delta^{\rm sc}_m$.

3. Adopt the galaxy (or nearby density peak) as a protocluster center candidate if the total mass included within the radius is greater than $10^{14}\,M_\odot$. Group the spherical regions if their separation is less than their radii. Protocluster centers are now identified.

4. The protocluster region is defined as the spherical volume from the protocluster center up to the turnaround radius. The turnaround radius is the radius where the mean overdensity drops to the threshold value given by Equation (7). The conversion of stellar mass to total mass within $R_{\rm TA}$ is made using Equation (2), with the parameters given in Section 4.4.

HR5 used in this paper adopts a flat ΛCDM cosmology with $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$. As the threshold density given by the spherical top-hat collapse model is used to find the protocluster centers, it will be useful to check how sensitive the threshold is

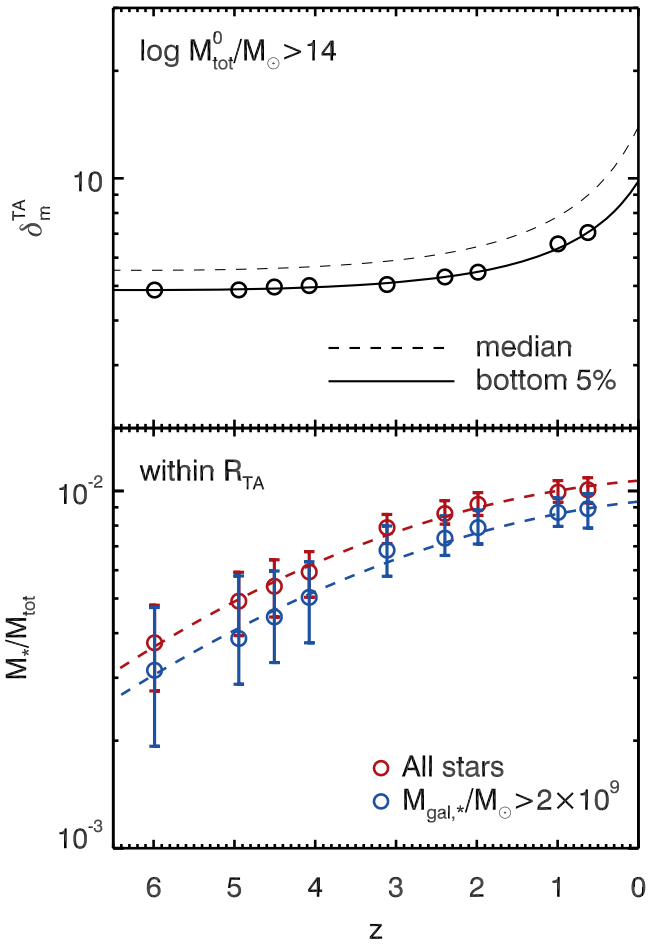

Figure 12. Top: redshift evolution of the overdensity within $R_{\rm TA}$ ($\delta_m^{\rm TA}$) for the protoclusters of $M_{\rm tot}^0 > 10^{14}\,M_\odot$ that is measured in HR5-Low. The open circles indicate the bottom 5% overdensity measured from the HR5 protoclusters, and the dashed and solid lines are the fits to the median and bottom 5%, respectively. Bottom: ratio of stellar to total mass within $R_{\rm TA}$ as a function of redshift. The red and blue open circles denote the ratios measured from all stars and from galaxies with $M_{\rm gal,\star} > 2\times10^9 M_\odot$, respectively. The dashed curves are the fits based on Equation (2) with fitting parameters given in Section 4.4.

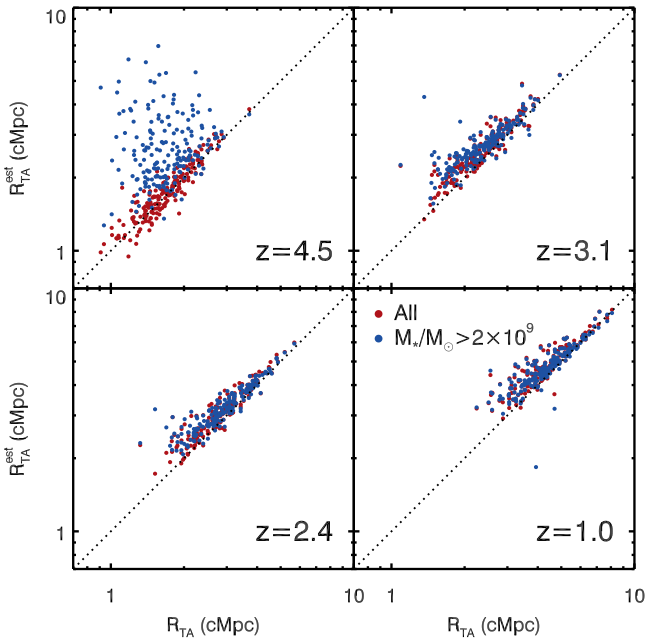

Figure 13. The turnaround radius estimated from the stellar mass distribution using the relations shown in Figure 12 vs. the directly measured turnaround radius of protoclusters. The former is based on the $\delta_m^{\rm TA}$ of the bottom 5%. The turnaround radii estimated by using all stars and by using only the galaxies more massive than $2\times10^9\,M_\odot$ are marked by red and blue circles, respectively.

to the cosmology adopted. We examine how $\delta_m^{\rm sc}$ changes depending on the matter density parameter while keeping the geometry of the universe flat and fixing the dark energy equation of state parameter to -1. Our choice of HR5 is based on the Planck data (Planck Collaboration et al. 2016). This is close to the recent measurement of Dong et al. (2023), who used the extended Alcock–Paczyski test to obtain $\Omega_m = 0.285^{+0.014}_{-0.009}$. In Figure 14, $\delta_m^{\rm sc}$ for four choices of Ω_m, i.e., 0.25, 0.3, 0.35, and 1, safely bracketing the recent observational values, are plotted. The figure shows that $\delta_m^{\rm sc}$ differs on average only by $\sim$12% at $z=6-2$ among the flat ΛCDM models with Ω_m from 0.25 to 0.35. Therefore, the threshold density used for finding the protocluster centers is not very sensitive to the choice of the matter density parameter, when the current tight constraint on the parameter is taken into account.

To estimate the reliability of this prescription, we use the clusters at $z=0$ with $M_{\rm tot}^0 \geqslant 10^{14}\,M_\odot$ and groups with $10^{13}\,M_\odot < M_{\rm tot}^0 < 10^{14}\,M_\odot$ identified in HR5-Low, a low-resolution version of HR5. There are 2794 objects with $M_{\rm tot}^0 > 10^{13}\,M_\odot$ in the zoomed-in region of HR5, and among them, 189 are clusters. Merger trees are constructed for these objects, and all progenitor galaxies are identified. We apply our protocluster identification scheme to the galaxy distributions at four simulation snapshots of $z=4.5$, 3.1, 2.4, and 1, being motivated by the ODIN survey of Lyα emitters. We find a tight correlation between the mass within the protocluster regions identified in accordance with the SC model and the final mass to be situated within clusters at $z=0$. In particular, it is highly likely (probability $\gtrsim$90%) for a protocluster region to evolve to a cluster if the region contains a total mass greater than about $2\times10^{14}\ M_\odot$, meaning that the region is likely to be the authentic protocluster.

We have defined the outer boundary of protoclusters as the zero-velocity surface at the turnaround radius. Even though protocluster members are identified within a spherical region, their distribution is quite anisotropic, as the region encloses numerous filaments beaded with galaxies. The definition would make sense if the galaxies within the turnaround radius do share some physical properties, which is not found for those outside. In the next study, we will examine the physical properties and evolution of the protocluster galaxies based on the definition proposed in this study.

Acknowledgments

We thank the anonymous referee for constructive comments and careful reading of the manuscript. J.L. is supported by the National Research Foundation of Korea (NRF-2021R1C1C2011626). C.P. and J.K. are supported by KIAS Individual Grants (PG016903, KG039603) at Korea Institute for Advanced Study. B.K.G. acknowledges the support of STFC through the University of Hull Consolidated Grant ST/R000840/1, access to VIPER, the University of Hull High Performance Computing Facility, and the European Union's

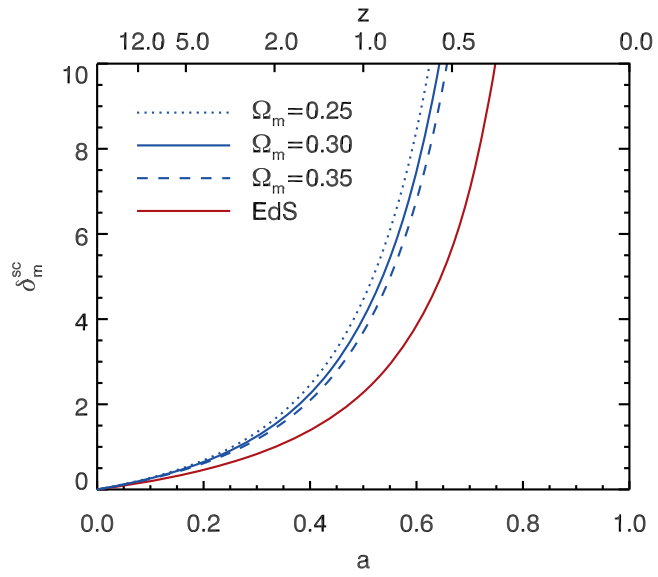

Figure 14. The critical overdensity for complete collapse at $z = 0$ given by the spherical top-hat collapse model in the EdS universe (red) and the flat ΛCDM universes with three different matter density parameters.

Horizon 2020 research and innovation program (ChETEC-INFRA—project No. 101008324). Y.K. is supported by Korea Institute of Science and Technology Information (KISTI) under the institutional R&D project (K-23-L02-C04). O.S. acknowledges support from an ERC Consolidator Grant (Grant Agreement ID 101003096) and STFC Consolidated Grant (ST/V000721/1). This work benefited from the outstanding support provided by the KISTI National Supercomputing Center and its Nurion Supercomputer through the Grand Challenge Program (KSC-2018-CHA-0003, KSC-2019-CHA-0002). This research was also partially supported by the ANR-19-CE31-0017 (http://www.secular-evolution.org). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT, 2022M3K3A1093827). This work is partially supported by the National Science Foundation under grant No. NSF PHY-1748958. Large data transfer was supported by KREONET, which is managed and operated by KISTI. This work is also supported by the Center for Advanced Computation at Korea Institute for Advanced Study.

Appendix A
Structure Finding and Merger Trees

We use a galaxy finder `PGalF` introduced by Kim et al. (2023) to extract self-bound and stable galaxies from the snapshots of `HR5`. `PGalF` is devised to identify the FoF group of particles from the distribution of heterogeneous particles, i.e., star, MBH, gas, and dark matter in `HR5`. For the mixture of various types of particles, `PGalF` uses an adaptive linking length to connect a pair of particles of different species or masses. `PGalF` identifies self-bound substructures in the FoF halos. We classify a substructure as a galaxy when it contains stellar particles. To find galaxies from an FoF halo, `PGalF` first constructs an adaptive stellar density field and hierarchically determines the membership of the particles bound to the galaxies centered at stellar density peaks. A bound particle is eventually assigned to a galaxy when it is located inside the tidal boundary of the galaxy. We note that a galaxy identified in this process is generally composed of heterogeneous particles. For the substructures with no stellar particles, a similar process is conducted for the rest of the matter species (dark matter and gas). For a full description of the method, refer to Kim et al.(2023).

Since stellar or dark matter particles carry their own unique identification numbers (IDs) throughout the simulation runs, we are able to trace the progenitors/descendants of substructures between two time steps. A branch of a merger tree is described using the binary relation between the two sets of all stellar particles in two snapshots, motivated by the Set theory. First, we define $\mathcal{S}_i$ as a set of all stellar particles at time step, t_i. Then,

$$\mathcal{S}_i = \mathcal{S}_{i-1} \bigcup \{s|\text{new stars born in}(t_{i-1}, t_i]\}, \tag{A1}$$

where "new stars" are those created between time steps t_{i-1} and t_i. We define $\mathcal{G}_i^j$ as the group of star particles of the j'th galaxy at time step i. Because a stellar particle is never destroyed in `HR5`, $\mathcal{S}_{i-1} \subseteq \mathcal{S}_i$. Our galaxy finder dictates that $\mathcal{G}_i^j \cap \mathcal{G}_i^k = \varnothing$ for $j \neq k$. The relation below is also satisfied; $\bigcup_{j=1}^n \mathcal{G}_i^j \subseteq \mathcal{S}_i$, where n is the total number of galaxies identified in time step i. The left-hand and right-hand sides of the equation are not always equal owing to stray stellar particles that are not bound to any galaxies.

We associate galaxies between two snapshots by mapping a set of stellar particles (a galaxy) at a time step into sets of stellar particles (galaxies) at the next time step using `ySAMtm` (Lee et al. 2014a; Jung et al. 2014). In `ySAMtm`, we define the j'th galaxy as the main descendant of the k'th galaxy when satisfying the mapping:

$$f\colon k \underset{\text{desc}}{\mapsto} \underset{j}{\text{argmax}}[P(\mathcal{G}_{i+1}^j|\mathcal{G}_i^k)], \tag{A2}$$

where $P(\mathcal{G}_{i+1}^j|\mathcal{G}_i^k)$ is the fractional number of stellar particles of the k'th galaxy to be found in the j'th galaxy. Multiple galaxies in time step i are allowed to have a common main descendant in time step $i+1$ once the mapping is satisfied, or in short $f(j) = f(k)$ for $j \neq k$.

Now we consider the reverse mapping as

$$g\colon k \underset{\text{prog}}{\mapsto} \underset{j}{\text{argmax}}[P(\mathcal{G}_{i-1}^j|\mathcal{G}_i^k)], \tag{A3}$$

which denotes that the j'th galaxy in time step $i-1$ is the main progenitor to the k'th galaxy in time step i. Unlike the mapping f for the main descendant, in principle, multiple galaxies in time step i cannot have a common main progenitor in time step $i-1$. Hence, in this case $g(j) \neq g(k)$ for all $j \neq k$. This is because we assume that a galaxy cannot be fragmented into multiple descendants in `ySAMtm`.

The mapping f is the left inverse mapping of g; it can be defined more formally as

$$(f \circ g)(j) \equiv f(g(j)) = j, \tag{A4}$$

$$(g \circ f)(j) \equiv g(f(j)) \neq j. \tag{A5}$$

Here Equation (A4) means that the main descendant of a main progenitor is the galaxy itself. One the other hand, g is not *left* inverse mapping of f (Equation (A5)) because of the case when the j'th galaxy is merged into its descendant.

Our tree-building scheme does not allow two galaxies to have the same main progenitor (or $g(j) \neq g(k)$ for $j \neq k$), but this usually happens when a galaxy flies by a more massive galaxy. To circumvent such cases, we remove the main

progenitor mapping of the less massive galaxy (the flying-by one) and trace back its previous history until its actual main progenitor is found, using the most bound particle (MBP). The MBP is a particle that has the largest negative total energy in the galaxy (Hong et al. 2016), and thus we assume that the MBPs trace density peaks of galaxies. We use dark matter particles as the MBPs because, unlike stellar particles, dark matter particles do not disappear or form anew throughout snapshots. We also use the MBP scheme to trace the substructures with no stellar particles. The merger trees of substructures are constructed by connecting the progenitor–descendant relations across the all snapshots. The progenitor/descendant relation of FoF halos is traced based on the merger trees of their most massive substructures. Further details of the tree-building algorithm are given in Park et al. (2022).

Appendix B
Identification of Cluster Progenitors in HR5

In this appendix, we describe the details of the identification process of the clusters in HR5 using its low-resolution simulation HR5-Low. While HR5 achieves a spatial resolution down to $\Delta x \sim 1$ kpc and minimum dark matter particle mass of $m_p \simeq 6.89 \times 10^7\,M_\odot$, HR5-Low is set to have a spatial resolution down to $\Delta x \sim 16$ kpc with a minimum dark matter particle mass of $m_p \simeq 3.02 \times 10^9\,M_\odot$. Because the main purpose of HR5-Low is to identify structures at $z = 0$, we use the parameters and initial conditions of HR5 without any modification or calibration. We identify structures from the snapshot at $z = 0$ and 0.625 of HR5-Low using PGalF. At $z = 0$, we find 2794 halos in $M^0_{\rm tot} \geqslant 10^{13}\,M_\odot$ and 189 halos in $M^0_{\rm tot} \geqslant 10^{14}\,M_\odot$, with the number fraction of lower level particles less than 0.1%, which ensures a mass contamination lower than 0.7%. The dark matter particles of the clusters are traced back to $z = 0.625$ using their IDs, to search for the progenitors of halos of $M^0_{\rm tot} \geqslant 10^{13}\,M_\odot$.

We measure the LV of a cluster in terms of the Cartesian grids. In HR5-Low, we place a mesh of uniform cubic grids with $\Delta l = 0.512$ cMpc over the entire volume of interest (the simulated zoomed-in region). To build a density field, we use the dark matter particles of cluster halos at $z = 0$. When dark matter particles in a grid do not belong to (or are not members to) a single cluster, the grid is finally associated with the cluster that contributes most to the grid mass. By utilizing the LV method with the HR5-Low data, we are able to define protocluster regions at an arbitrary redshift.

In the subsequent analysis we assume that the LVs of HR5 clusters are identical to the LVs of corresponding HR5-Low clusters. In the last snapshot of HR5, therefore, we are able to find structures inside the LVs directly imported from the HR5-Low clusters. We only use grids having mass larger than $10^{10}\,M_\odot$ because 97.5% of galaxies with $M_\star \geqslant 10^9\,M_\odot$ have $M_{\rm tot} > 10^{10}\,M_\odot$. This mass cut helps us minimize the contamination by noncluster progenitors in the LVs of the cluster progenitors at $z = 0.625$.

Figure B1 presents the relation between the cluster mass in HR5-Low at $z = 0$ and the corresponding LV mass $M_{\rm LV}$ in HR5 at $z = 0.625$. The two masses are nearly the same, with a median scattering of ~6%. The mass difference may be caused by matter that happens to be enclosed in the LVs but would not fall into the cluster at $z = 0$.

To examine the consistency or similarity in particle distributions between HR5-Low and HR5 especially on halo scales at $z = 0.625$, we identify an HR5 FoF halo that is spatially closest to the main progenitor of each HR5-Low cluster. Here the progenitor of a cluster is determined by the scheme described in Section 2.2.

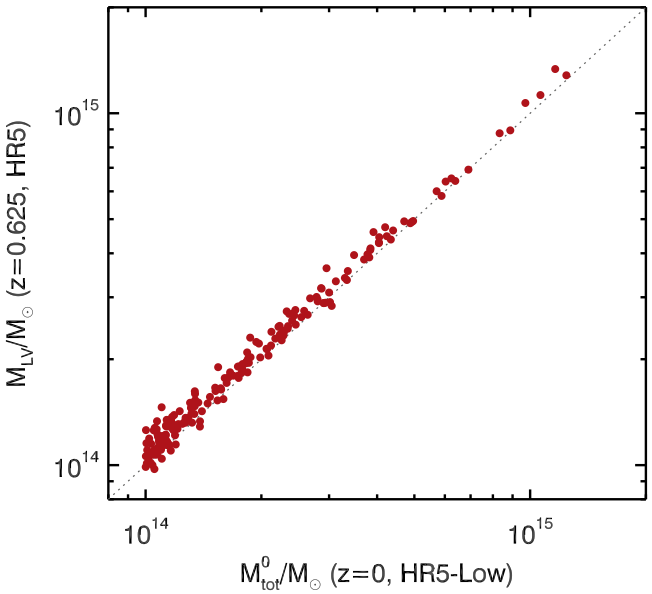

Figure B1. The relation between the total mass of the clusters found at $z = 0$ in HR5-Low and the LV mass at $z = 0.625$ in HR5. The LV mass is on average ~6% higher than the cluster mass owing to the matter that is contained in voxels at the epoch but will not form the clusters.

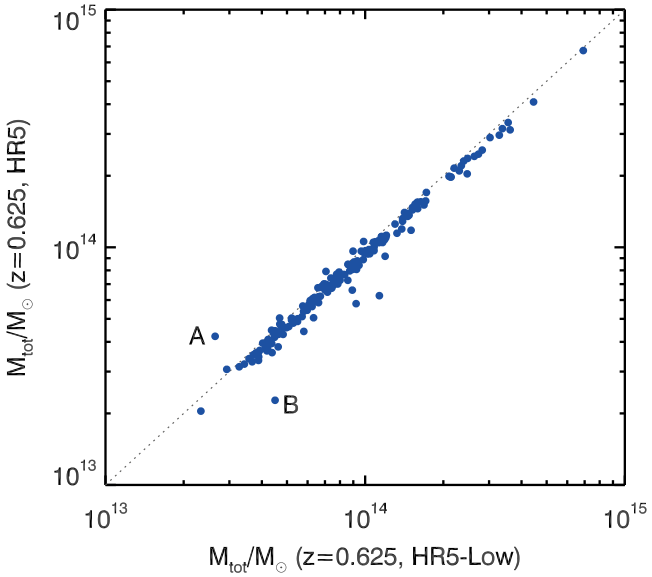

Figure B2. Relation between the total mass of the main progenitors ($z = 0.625$) of the clusters found at $z = 0$ in HR5-Low and the total mass of their counterparts in HR5. Halo A is the one that is identified as two separate structures in HR5-Low, while a smaller one already becomes a substructure of the halo in HR5. Halo B is the opposite case. The halos in HR5 are ~9% less massive than their counterparts in HR5-Low because their small neighboring structures are not well resolved in HR5-Low.

Figure B2 shows the relation of FoF halo masses between the main progenitors of clusters in HR5-Low and their counterparts in HR5 at $z = 0.625$. Except for two cases marked by A and B, all FoF halos in the two simulations have nearly the same mass. We slightly overestimate the mass of FoF halos in HR5-Low compared to HR5 because of the purer mass

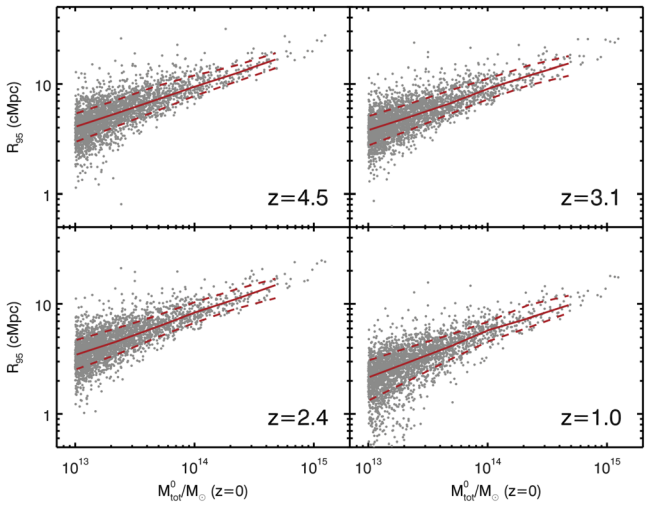

Figure B3. Radius that encloses 95% of the stellar mass of the proto-objects of the FoF halos identified at $z = 0$ as a function of their final mass that is measured from HR5-Low. The radius measurement is centered at the most massive galaxy in each proto-object. Red dashed and solid lines mark 16th and 84th percentiles and the median of R_{95} at a given final mass, respectively.

resolution, which tends to more easily destroy clumpy structures in the outskirts of halos. Here A and B are the cases when substructures are distinguishable only within either HR5 or HR5-Low. We assume that, although rare, the adaptive linking length may cause the different FoF halo identification between two simulations at different resolutions. Alternatively, the different-resolution simulations may, of course, produce different particle distributions more often in the outskirts of halos, especially around a close binary or a multiple system of halos.

Figure B3 shows R_{95}, the radius enclosing 95% of stellar mass in cluster progenitors, as a function of the final total mass. The progenitors of more massive halos tend to have larger R_{95}. The range of R_{95} is consistent with Muldrew et al. (2015), who measure R_{90} of protoclusters using the semianalytic model of Guo et al. (2011). In this study, we suggest the turnaround radius as the physical size of protoclusters instead of R_{95} because R_{95} merely measures the spatial extent of the distribution of progenitor galaxies.

Appendix C
Spherical Top-hat Overdensity in the ΛCDM and Einstein −de Sitter Universe

In the EdS universe with $\Omega_m = 1$, the outermost radius R of a sphere of mass M evolves over time t as follows:

$$\ddot{R} = -\frac{GM}{R^2}, \tag{C1}$$

where G is the gravitational constant. This equation has the cycloidal solution:

$$t = \frac{t_{\rm max}}{\pi}(\theta - \sin\theta),$$
$$R = \frac{R_{\rm max}}{2}(1 - \cos\theta), \tag{C2}$$

where $t_{\rm max}$ is the time when the sphere reaches a maximum radius $R_{\rm max}$. In this solution, the spherical region collapses at the collapse time $t_c = 2t_{\rm max}$ ($\theta = 2\pi$). The overdensity of the

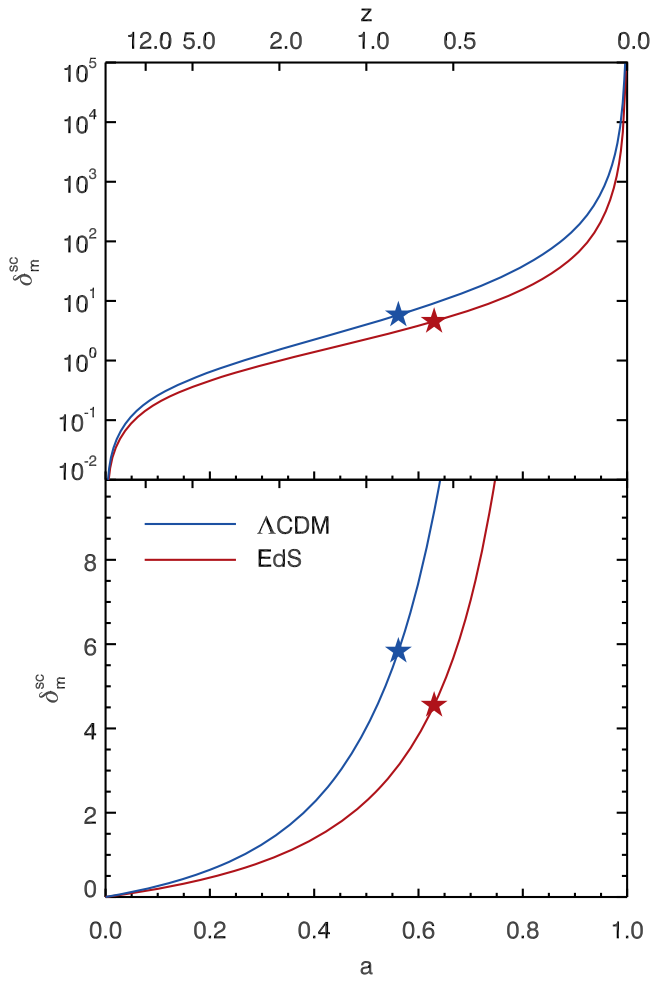

Figure C1. The critical overdensity of a homogeneous top-hat sphere collapsing at $z = 0$ predicted by the spherical top-hat collapse model in the ΛCDM (blue) and EdS (red) universe in logarithmic (top) and linear (bottom) scales. Stars indicate the epoch and overdensity when the sphere reaches its maximum radius in each universe.

sphere at a given epoch derived from the analytic solution is given by (e.g., Peebles 1980; Suto et al. 2016)

$$\delta_m^{\rm sc} \equiv \frac{\rho_m^{\rm sc}}{\bar{\rho}_m} - 1 = \frac{9}{2}\frac{(\theta - \sin\theta)^2}{(1 - \cos\theta)^3} - 1. \tag{C3}$$

A homogeneous density sphere that collapses at $z = 0$ reaches its maximum radius at $z = 0.59$ with $\delta_m^{\rm sc} = 9\pi^2/16 - 1 \simeq 4.55$ in the EdS universe. For comparison, the linear theory predicts overdensity $\delta_m^{\rm lin} \simeq 1.062$ at $t_{\rm max}$ in the EdS universe.

In the flat universe with nonzero Ω_Λ, the expansion factor of maximum radius $a_{\rm max}$ can be derived using the formula (Peebles 1984; Eke et al. 1996)

$$a_{\rm max} = \left[\frac{\exp(3\sqrt{\omega}I(\omega)) - 1}{2\sqrt{\omega}}\right]^{\frac{2}{3}} \exp(-\sqrt{\omega}I(\omega)), \tag{C4}$$

where $\omega = \Omega_\Lambda/\Omega_{\rm m}$ and $I(\omega)$ is given from

$$I(\omega) = \frac{1}{2}\int_0^{a_c} \frac{\sqrt{a}}{\sqrt{\omega a^3 + 1}} da, \tag{C5}$$

where a_c is the expansion factor at the time of collapse. These equations give $a_{\rm max} = 0.56$ in the case of $a_c = 1.0 (z = 0)$, and the overdensity at the epoch is interpolated as $\delta_m^{\rm sc} = 5.85$ for

our choice of the ΛCDM universe. When $\Omega_m = 1.0$ and $\Omega_\Lambda = 0$, Equation (1) has a solution that is equal to the exact solution of the EdS universe case derived above.

Figure C1 shows the overdensity evolution in a homogeneous sphere that collapses at $z = 0$ in the EdS (blue) and ΛCDM (red) universe. The two filled stars indicate the overdensities at the epochs of maximum radius. Because dark energy counteracts gravitational collapse and the growth of overdensity is relatively slower, the sphere should have higher overdensity in the universe with $\Omega_\Lambda > 0$ than in the EdS universe, to be able to collapse by $z = 0$.

Since $\delta_m^{\rm sc}$ does not have an exact analytic solution in the ΛCDM universe adopted, we find a formula that fits the numerical solution of the SC model for the objects collapsing at $z = 0$:

$$\delta_m^{\rm sc}(z) = \frac{0.0224\exp(5.39\,z^{-0.246})}{(1+z)^{0.294}}. \quad \text{(C6)}$$

This formula has an error <0.3% in the redshift range of $z = [0.5, 6.0]$.

Appendix D
Performance of the Protocluster Identification Scheme Based on the SC Model

Figure D1 demonstrates the completeness and purity of our protocluster identification scheme (top) and the relation between $M_{\rm tot}^{0,\rm SRG}$ and $M_{\rm tot}^{\rm SRG}$ (bottom). We define the purity as the number fraction of SRGs enclosing bona fide protoclusters (those identified based on merger trees) to all SRGs more massive than a given mass. The completeness is the number fraction of the authentic protoclusters enclosed by SRGs above a given mass. In these statistics, we assume that an SRG recovers a protocluster when the most massive galaxy of the SRG is the member of the protocluster and half the galaxy mass of the protocluster is enclosed by the SRG. In this scheme, an SRG can be associated with only one protocluster. The color code in the bottom panels indicates the $D_1/R_{\rm SRG}$ parameter. Colored circles show the distribution of the entire SRG sample, and black concentric circles mark the SRGs with $M_{\rm tot}^{\rm SRG} > 10^{14.15}$ or $D_1/R_{\rm SRG} < 2.5$. These two different mass definitions are overall in good agreement, particularly at $z < 4$. Their correlation becomes tighter with decreasing redshift as structures form and develop further. The completeness and purity show that more than 80% of protoclusters can be recovered by our scheme with ∼60% purity at $z \sim 2$–3. The purity increases to 80% in $M_{\rm tot}^{\rm SRG} \geqslant 2 \times 10^{14}\,M_\odot$. At $z = 4.5$, however, these statistics are inevitably poorer than at lower z because galaxies have not had time to develop yet. We note that the purity and completeness are enhanced by ∼10% if an SRG is allowed to associate with all the protoclusters in which half their galaxy mass is enclosed by the SRG.

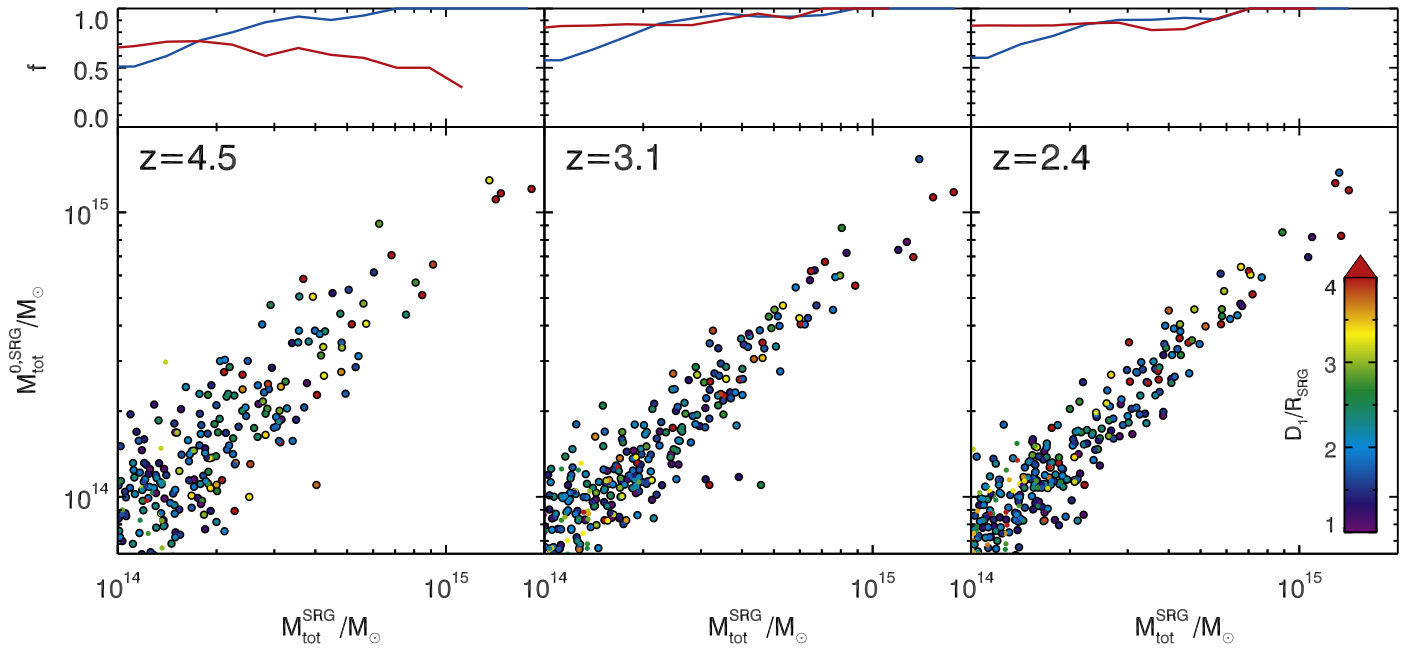

Figure D1. Bottom: final mass of SRGs estimated from the final mass of bona fide protoclusters (those identified based on merger trees) embedded in the SRGs ($M_{\rm tot}^{0,\rm SRG}$) as a function of the total mass of SRGs ($M_{\rm tot}^{\rm SRG}$). The color code denotes $D_1/R_{\rm SRG}$. Colored circles mark the entire SRG sample, and black concentric circles indicate the SRGs with $M_{\rm tot}^{\rm SRG} > 10^{14.15}$ or $D_1/R_{\rm SRG} < 2.5$. As also seen in Figure 5, most protoclusters have $D_1/R_{\rm SRG} \lesssim 4$. We note that $M_{\rm tot}^{0,\rm SRG}$ is an estimated mass to examine the prediction accuracy of $M_{\rm tot}^{\rm SRG}$. Top: purity (blue) and completeness (red) of the bona fide protoclusters in the spherical regions found by the SC model as a function of $M_{\rm tot}^{\rm SRG}$. The purity is the number fraction of the SRGs enclosing bona fide protoclusters to the entire sample of SRGs above a given mass. The completeness is the number fraction of the authentic protoclusters that are recovered by SRGs and more massive than a given mass.

Appendix E
Redshift-space Distortion Effect on the Protocluster Identification

The peculiar velocities of galaxies distort the distribution of galaxies in redshift space (see, e.g., Guzzo et al. 1997; Hamilton 1998). We examine the impact of RSD on the protocluster identification scheme. In this test, we assume that a virtual observer has the line of sight aligned with the major axis of the HR5 zoom-in region. The redshift of a snapshot is assigned to the center of the zoomed-in region, and the cosmological redshifts of the galaxies in the snapshot are computed from the distance relative to a virtual observer at $z = 0$. The Doppler redshifts induced by the peculiar velocities of galaxies are added to the cosmological redshifts, and the distances to the galaxies are reestimated from the combined redshifts. The standard deviations of the differences between the intrinsic and redshift-distorted distances are 2.3, 3.0, and 3.6 cMpc at $z = 2.4$, 3.1, and 4.5, respectively. Figure E1 presents the impact of the RSDs on the protocluster identification scheme. Since the large-scale peculiar velocity vector tends to point toward overdense regions, the galaxy distribution near a protocluster is statistically flattened along the line of sight in redshift space (Kaiser 1987). This results in slight overestimation of the overdensity and size of the top-hat spheres of dense regions. The final impact is that the completeness increases, at higher redshifts in particular, while the purity slightly decreases. The bottom panels of Figure E1 show that the RSD effect slightly increases the SRG mass, but the overall distribution is similar between the cases with and without the RSD effects. These statistics are computed based on the assumption that an SRG is only associated with a protocluster. The purity and completeness can change if an SRG is allowed to recover multiple protoclusters. This result demonstrates that the RSD effect does not have significant impact on the protocluster identification scheme.

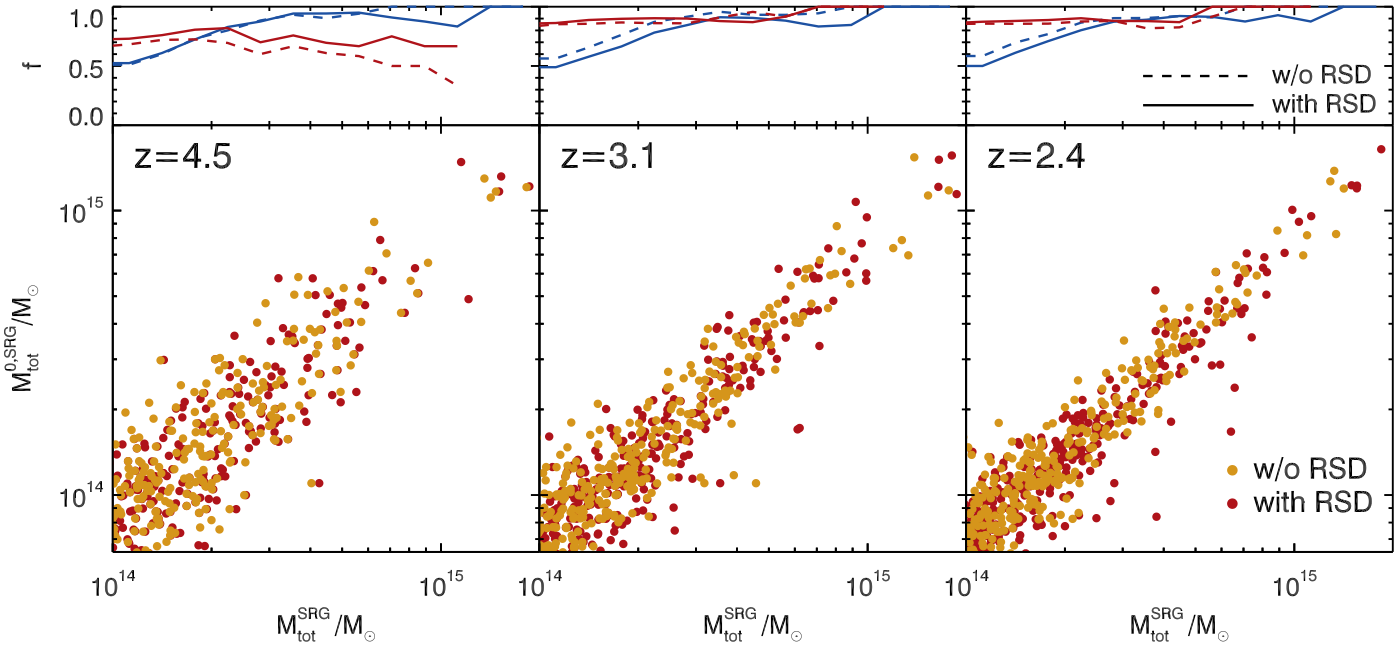

Figure E1. Same as Figure D1, but for the cases with and without the RSD effect. In the bottom panels, the scatter between $M_{tot}^{0,SRG}$ and M_{tot}^{SRG} is similar between the two cases with and without the RSD effect. Top panels show that the RSD effect lowers the purity while slightly enhancing the completeness at given mass. This is caused by the RSD effect, which makes overdense regions look flattened in the redshift space (Kaiser 1987), resulting in the overestimation of the SRG radius.

Appendix F
Turnaround Radius of Spherical Region Groups

We examine whether the turnaround radius is reasonably recovered in the SRGs. For consistency with the $R_{\rm TA}$ measurement for the bona fide protoclusters, we measure the turnaround radius relative to the most massive galaxy in an SRG ($R_{\rm TA}^{\rm SRG}$) and compare it with $R_{\rm TA}$ of the protocluster that hosts the most massive galaxy of the SRG and shares half its total galaxy mass with the SRG. Figure F1 shows the $R_{\rm TA}$ $-R_{\rm TA}^{\rm SRG}$ relation at the four redshifts. In this comparison, more than 80% of the SRGs have $R_{\rm TA}^{\rm SRG}$ identical to $R_{\rm TA}$. The scatter is caused when the most massive galaxy in an SRG is not the most massive one in its host protocluster.

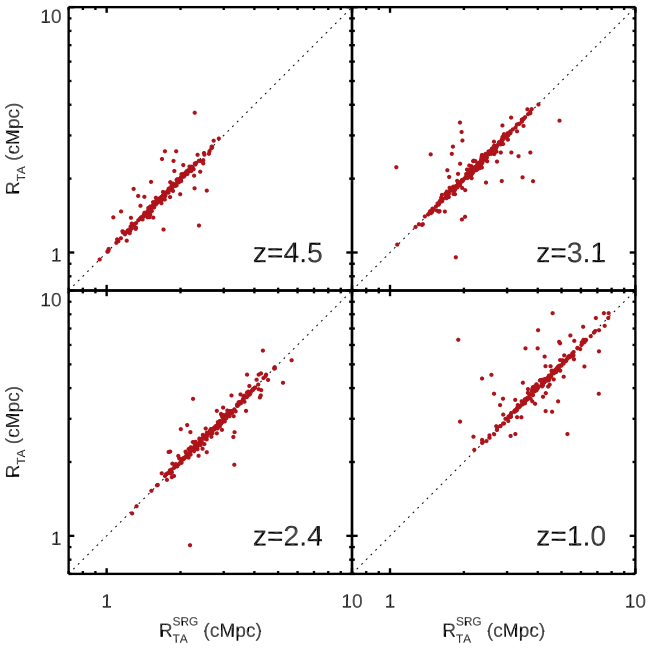

Figure F1. Relation of turnaround radius of bona fide protoclusters ($R_{\rm TA}$) to turnaround radius measured from the most massive galaxies in SRGs ($R_{\rm TA}^{\rm SRG}$). A protocluster is assumed to be associated with an SRG when half its galaxy mass is enclosed by the SRG. Scatter is caused when the most massive galaxy of an SRG is not the most massive one of its host protocluster. We find that ∼80% of SRGs recover the $R_{\rm TA}$ of enclosed protoclusters.

ORCID iDs

Jaehyun Lee https://orcid.org/0000-0002-6810-1778
Changbom Park https://orcid.org/0000-0001-9521-6397
Juhan Kim https://orcid.org/0000-0002-4391-2275
Christophe Pichon https://orcid.org/0000-0003-0695-6735
Brad K. Gibson https://orcid.org/0000-0003-4446-3130
Jihye Shin https://orcid.org/0000-0001-5135-1693
Yonghwi Kim https://orcid.org/0000-0003-4164-5414
Yohan Dubois https://orcid.org/0000-0003-0225-6387
C. Gareth Few https://orcid.org/0000-0003-3940-7687

References

Adams, S. M., Martini, P., Croxall, K. V., Overzier, R. A., & Silverman, J. D. 2015, MNRAS, 448, 1335
Álvarez Crespo, N., Smolić, V., Finoguenov, A., Barrufet, L., & Aravena, M. 2021, A&A, 646, A174
Bahcall, N. A., & West, M. J. 1992, ApJ, 392, 419
Bahé, Y. M., Barnes, D. J., Dalla Vecchia, C., et al. 2017, MNRAS, 470, 4186
Bryan, G. L., & Norman, M. L. 1998, ApJ, 495, 80
Cai, Z., Fan, X., Bian, F., et al. 2017, ApJ, 839, 131
Cai, Z., Fan, X., Peirani, S., et al. 2016, ApJ, 833, 135
Calvi, R., Dannerbauer, H., Arrabal Haro, P., et al. 2021, MNRAS, 502, 4558
Chabrier, G. 2003, ApJL, 586, L133
Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2013, ApJ, 779, 127
Chiang, Y.-K., Overzier, R. A., Gebhardt, K., et al. 2015, ApJ, 808, 37
Choi, H., & Yi, S. K. 2017, ApJ, 837, 68
Cooke, E. A., Hatch, N. A., Muldrew, S. I., Rigby, E. E., & Kurk, J. D. 2014, MNRAS, 440, 3262
Cooke, E. A., Smail, I., Stach, S. M., et al. 2019, MNRAS, 486, 3047
Cucciati, O., Zamorani, G., Lemaux, B. C., et al. 2014, A&A, 570, A16
Cuesta, A. J., Prada, F., Klypin, A., & Moles, M. 2008, MNRAS, 389, 385
Daddi, E., Dannerbauer, H., Stern, D., et al. 2009, ApJ, 694, 1517
Dalgarno, A., & McCray, R. A. 1972, ARA&A, 10, 375
Diener, C., Lilly, S. J., Ledoux, C., et al. 2015, ApJ, 802, 31
Djorgovski, S. G., Stern, D., Mahabal, A. A., & Brunner, R. 2003, ApJ, 596, 67
Dong, F., Park, C., Hong, S. E., et al. 2023, ApJ, 953, 98
Dubois, Y., Devriendt, J., Slyz, A., & Teyssier, R. 2012, MNRAS, 420, 2662
Dubois, Y., Pichon, C., Welker, C., et al. 2014a, MNRAS, 444, 1453
Dubois, Y., & Teyssier, R. 2008, A&A, 477, 79
Dubois, Y., Volonteri, M., & Silk, J. 2014b, MNRAS, 440, 1590
Eke, V. R., Cole, S., & Frenk, C. S. 1996, MNRAS, 282, 263
Falco, M., Hansen, S. H., Wojtak, R., et al. 2014, MNRAS, 442, 1887
Falco, M., Mamon, G. A., Wojtak, R., Hansen, S. H., & Gottlöber, S. 2013, MNRAS, 436, 2639
Falder, J. T., Stevens, J. A., Jarvis, M. J., et al. 2011, ApJ, 735, 123
Few, C. G., Courty, S., Gibson, B. K., et al. 2012, MNRAS, 424, L11
Fu, H., Cooray, A., Feruglio, C., et al. 2013, Natur, 498, 338
Greve, T. R., Stern, D., Ivison, R. J., et al. 2007, MNRAS, 382, 48
Gunn, J. E., & Gott, J. R. I. 1972, ApJ, 176, 1
Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101
Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, ApJ, 489, 37
Haardt, F., & Madau, P. 1996, ApJ, 461, 20
Hahn, O., & Abel, T. 2011, MNRAS, 415, 2101
Hamilton, A. J. S. 1998, in Astrophysics and Space Science Library, Volume 231: The Evolving Universe, ed. D. Hamilton (Dordrecht: Springer), 185
Hatch, N. A., De Breuck, C., Galametz, A., et al. 2011a, MNRAS, 410, 1537
Hatch, N. A., Kurk, J. D., Pentericci, L., et al. 2011b, MNRAS, 415, 2993
Hayashi, M., Kodama, T., Tadaki, K.-I., Koyama, Y., & Tanaka, I. 2012, ApJ, 757, 15
Hennawi, J. F., Prochaska, J. X., Cantalupo, S., & Arrigoni-Battaia, F. 2015, Sci, 348, 779
Hong, S. E., Park, C., & Kim, J. 2016, ApJ, 823, 103
Husband, K., Bremer, M. N., Stanway, E. R., et al. 2013, MNRAS, 432, 2869
Jung, I., Lee, J., & Yi, S. K. 2014, ApJ, 794, 74
Kaiser, N. 1984, ApJL, 284, L9
Kaiser, N. 1987, MNRAS, 227, 1
Kim, J., Lee, J., Laigle, C., et al. 2023, ApJ, 951, 137
Le Fevre, O., Deltorn, J. M., Crampton, D., & Dickinson, M. 1996, ApJL, 471, L11
Lee, J., Shin, J., Snaith, O. N., et al. 2021, ApJ, 908, 11
Lee, J., Yi, S. K., Elahi, P. J., et al. 2014a, MNRAS, 445, 4197
Lee, K.-G., Hennawi, J. F., White, M., Croft, R. A. C., & Ozbek, M. 2014b, ApJ, 788, 49
Lee, K.-S., Dey, A., Hong, S., et al. 2014c, ApJ, 796, 126
Lemaux, B. C., Cucciati, O., Tasca, L. A. M., et al. 2014, A&A, 572, A41
L'Huillier, B., Park, C., & Kim, J. 2014, NewA, 30, 79
Lilly, S. J., Eales, S. A., Gear, W. K. P., et al. 1999, ApJ, 518, 641
Mamon, G. A., Sanchis, T., Salvador-Solé, E., & Solanes, J. M. 2004, A&A, 414, 445
McConachie, I., Wilson, G., Forrest, B., et al. 2022, ApJ, 926, 37
Muldrew, S. I., Hatch, N. A., & Cooke, E. A. 2015, MNRAS, 452, 2528
Muldrew, S. I., Hatch, N. A., & Cooke, E. A. 2018, MNRAS, 473, 2335
Newman, A. B., Rudie, G. C., Blanc, G. A., et al. 2022, Natur, 606, 475
Oñorbe, J., Garrison-Kimmel, S., Maller, A. H., et al. 2014, MNRAS, 437, 1894
Oteo, I., Ivison, R. J., Dunne, L., et al. 2018, ApJ, 856, 72
Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2005, ApJL, 620, L1
Overzier, R. A. 2016, A&ARv, 24, 14
Pace, F., Waizmann, J. C., & Bartelmann, M. 2010, MNRAS, 406, 1865
Park, C., Lee, J., Kim, J., et al. 2022, ApJ, 937, 15
Pascarelle, S. M., Windhorst, R. A., Keel, W. C., & Odewahn, S. C. 1996, Natur, 383, 45

Peebles, P. J. E. 1980, The Large-scale Structure of the Universe (Princeton, NJ: Princeton Univ. Press)
Peebles, P. J. E. 1984, ApJ, 284, 439
Planck Collaboration, Ade, P. A. R., & Aghanim, N. E. A. 2016, A&A, 594, A13
Prescott, M. K. M., Dey, A., Brodwin, M., et al. 2012, ApJ, 752, 86
Prescott, M. K. M., Kashikawa, N., Dey, A., & Matsuda, Y. 2008, ApJL, 678, L77
Ramakrishnan, V., Moon, B., Dey, A., et al. 2022, AAS Meeting, 54, 428.05
Rasera, Y., & Teyssier, R. 2006, A&A, 445, 1
Rines, K., & Diaferio, A. 2006, AJ, 132, 1275
Rotermund, K. M., Chapman, S. C., Phadke, K. A., et al. 2021, MNRAS, 502, 1797
Scoccimarro, R. 1998, MNRAS, 299, 1097
Shen, L., Lemaux, B. C., Lubin, L. M., et al. 2021, ApJ, 912, 60
Shi, K., Huang, Y., Lee, K.-S., et al. 2019, ApJ, 879, 9
Shimasaku, K., Ouchi, M., Okamura, S., et al. 2003, ApJL, 586, L111
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Natur, 435, 629
Stark, C. W., White, M., Lee, K.-G., & Hennawi, J. F. 2015, MNRAS, 453, 311
Steidel, C. C., Adelberger, K. L., Dickinson, M., et al. 1998, ApJ, 492, 428
Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2000, ApJ, 532, 170
Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2005, ApJ, 626, 44
Stevens, J. A., Ivison, R. J., Dunlop, J. S., et al. 2003, Natur, 425, 264
Stevens, J. A., Jarvis, M. J., Coppin, K. E. K., et al. 2010, MNRAS, 405, 2623
Sutherland, R. S., & Dopita, M. A. 1993, ApJS, 88, 253
Suto, D., Kitayama, T., Osato, K., Sasaki, S., & Suto, Y. 2016, PASJ, 68, 14
Teyssier, R. 2002, A&A, 385, 337
Toft, S., Smolčić, V., Magnelli, B., et al. 2014, ApJ, 782, 68
Toshikawa, J., Kashikawa, N., Ota, K., et al. 2012, ApJ, 750, 137
Toshikawa, J., Kashikawa, N., Overzier, R., et al. 2014, ApJ, 792, 15
Toshikawa, J., Kashikawa, N., Overzier, R., et al. 2016, ApJ, 826, 114
Toshikawa, J., Uchiyama, H., Kashikawa, N., et al. 2018, PASJ, 70, S12
Trebitsch, M., Dubois, Y., Volonteri, M., et al. 2021, A&A, 653, A154
Truong, N., Rasia, E., Mazzotta, P., et al. 2018, MNRAS, 474, 4089
Umehata, H., Tamura, Y., Kohno, K., et al. 2014, MNRAS, 440, 3462
Umehata, H., Tamura, Y., Kohno, K., et al. 2015, ApJL, 815, L8
Venemans, B. P., Kurk, J. D., Miley, G. K., et al. 2002, ApJL, 569, L11
Venemans, B. P., Röttgering, H. J. A., Miley, G. K., et al. 2005, A&A, 431, 793
Venemans, B. P., Röttgering, H. J. A., Miley, G. K., et al. 2007, A&A, 461, 823
Venemans, B. P., Röttgering, H. J. A., Overzier, R. A., et al. 2004, A&A, 424, L17
Wang, K., Mo, H. J., Li, C., & Chen, Y. 2021, MNRAS, 505, 3892
Wang, T., Elbaz, D., Daddi, E., et al. 2016, ApJ, 828, 56
Wojtak, R., Łokas, E. L., Gottlöber, S., & Mamon, G. A. 2005, MNRAS, 361, L1
Wold, M., Armus, L., Neugebauer, G., Jarrett, T. H., & Lehnert, M. D. 2003, AJ, 126, 1776
Yajima, H., Abe, M., Khochfar, S., et al. 2022, MNRAS, 509, 4037
Yonekura, N., Kajisawa, M., Hamaguchi, E., Mawatari, K., & Yamada, T. 2022, ApJ, 930, 102