The Astrophysical Journal, 833:135 (25pp), 2016 December 20 doi:10.3847/1538-4357/833/2/135

MAPPING THE MOST MASSIVE OVERDENSITY THROUGH HYDROGEN (MAMMOTH). I. METHODOLOGY

Zheng Cai[1,2], Xiaohui Fan[2], Sebastien Peirani[3], Fuyan Bian[4,8], Brenda Frye[2], Ian McGreer[2], J. Xavier Prochaska[1], Marie Wingyee Lau[1], Nicolas Tejos[1], Shirley Ho[5], and Donald P. Schneider[6,7]
[1] UCO/Lick Observatory, University of California, 1156 High Street, Santa Cruz, CA 95064, USA; email.arizona.edu
[2] Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA
[3] Institut d'Astrophysique de Paris, 98 bis, Boulevard Arago—F-75014 Paris, France
[4] Research School of Astronomy & Astrophysics, Mount Stromlo Observatory, Cotter Road, Weston ACT 2611, Australia
[5] McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[6] Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA
[7] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
Received 2015 December 21; revised 2016 October 7; accepted 2016 October 12; published 2016 December 13

ABSTRACT

Modern cosmology predicts that a galaxy overdensity (e.g., protocluster) will be associated with a large intergalactic medium gas reservoir, which can be traced by Lyα forest absorption. We have undertaken a systematic study of the relation between Coherently Strong intergalactic Lyα Absorption systems (CoSLAs), which have the highest optical depth (τ) in the τ distribution, and mass overdensities on the scales of $\sim$10–20 h^{-1} comoving Mpc. On such large scales, our cosmological simulations show a strong correlation between the effective optical depth ($\tau_{\rm eff}$) of the CoSLAs and the three-dimensional mass overdensity. In spectra with moderate signal-to-noise ratio, however, the profiles of CoSLAs can be confused with individual high column density absorbers. For $z > 2.6$, where the corresponding Lyβ is redshifted to the optical, we have developed a selection technique to distinguish between these two alternatives. We have applied this technique to $\sim$6000 sight lines provided by Sloan Digital Sky Survey III quasar survey at $z = 2.6$–3.3 with a continuum-to-noise ratio greater than 8, and we present a sample of five CoSLA candidates with $\tau_{\rm eff}$ on 15 h^{-1} Mpc greater than 4.5$\times$ the mean optical depth. At lower redshifts of $z < 2.6$, where the background quasar density is higher, the overdensity can be traced by intergalactic absorption groups using multiple sight lines with small angular separations. Our overdensity searches fully use the current and next generation of Lyα forest surveys, which cover a survey volume of >1 (h^{-1} Gpc)^3. Systems traced by CoSLAs will yield a uniform sample of the most massive overdensities at $z > 2$ to provide stringent constraints to models of structure formation.

Key words: galaxies: high-redshift – intergalactic medium – quasars: absorption lines

1. INTRODUCTION

The most massive large-scale structures at the peak of cosmic star formation, i.e., $z \sim 2$–3, are unique laboratories for understanding cosmic mass assembly. These structures are excellent sites to study the earliest clusters of galaxies, the formation of highly evolved galaxies at high-redshift, and the complex interactions between galaxies and the intergalactic medium (IGM). In addition, the abundance of the extreme tails of the mass overdensity provides a particularly potent constraint on models of structure and galaxy formation. Nevertheless, the task of finding the progenitors of clusters (or so-called protoclusters) at $z > 2$ is challenging. The majority of protoclusters have been found by targeting rare sources, such as quasars (e.g., Hu et al. 1996), radio galaxies (e.g., Venemans et al. 2007), submillimeter galaxies (e.g., Chapman et al. 2004) and Lyα "blobs" (e.g., Yang et al. 2009). The utility of these markers is limited by small duty cycles and strong selection biases, and thus the overdensities traced by these rare sources are highly incomplete. An alternative approach to identifying large-scale structures is by targeting blank fields in galaxy redshift surveys (Steidel et al. 1998; Ouchi et al. 2005; Chiang et al. 2013; Lee et al. 2014). However, the current deep high-z galaxy redshift surveys are limited by small survey area (up to a few deg^2). Owing to these difficulties, the overall number of confirmed protoclusters is still too low to allow for robust comparisons to hierarchical structure formation models or for environmental studies of galaxy properties at different redshifts (Chiang et al. 2013; Lee et al. 2014). A more complete search for galaxy overdensities from a larger volume is highly desirable. We are thus motivated to develop systematic techniques to identify early protoclusters, especially to build a uniform sample of the most biased and evolved examples at $z > 2$. This sample will enable studying how these structures interact with cosmic web filaments, feedback to IGM, and transform into the present-day local clusters, as well as to provide stringent constraints to models of structure formation.

Hydrogen in the IGM maintains a high ionization fraction in the post-reionization epoch under a metagalactic ionizing background originating from star-forming galaxies and active galactic nuclei (AGNs). Lyman-alpha (Lyα) forest absorption marks locations where the quasar line of sight (LOS) intersects intergalactic neutral hydrogen (H I) at the wavelength of the redshifted Lyα resonance lines (e.g., Gunn & Peterson 1965; Lynds 1971; Bi 1993). The optical depth of H I Lyα smoothly traces the dark matter distribution at scales larger than the Jeans scale of the photoionized IGM (Cen et al. 1994; Miralda-Escudé et al. 1996; Rauch 1998; Viel et al. 2012; Lee et al. 2014). Thus, the Lyα forest has been used as a crucial probe of the IGM and underlying mass distribution of $z > 2$ over a sufficiently large scale, of densities ranging from the cosmic mean to highly overdense regions (Lee et al. 2014). Recently, the Lyα forest has been used to measure clustering on large scales of $\sim$100 h^{-1} Mpc, and the Baryonic Acoustic

[8] Stromlo Fellow.

Oscillation (BAO) feature at $z \sim 2.5$ was detected (Slosar et al. 2011, 2013; Busca et al. 2013).

A number of theoretical and observational studies have probed the correlation between Lyα forest absorptions and galaxies on several comoving Mpc (co-Mpc) scales at $z = 2$–3. On the observational side, absorbers with H I column densities $N_{\rm H\,I} \gtrsim 10^{14.5}\,{\rm cm}^{-2}$ are correlated with galaxy positions on ∼Mpc scales, and the association is stronger for the higher column density systems (Rudie et al. 2012). Meanwhile, Adelberger et al. (2003) pointed out that on still larger scales of $10\,h^{-1}$ comoving Mpc, the mean transmission of the Lyα forest tends to be low in volumes that contain an overdensity of Lyman break galaxies (LBGs). On the theoretical side, McDonald et al. (2002) used several hydro-particle-mesh simulations with a box size of $40\,h^{-1}$ comoving Mpc to demonstrate a strong correlation between the mass and the Lyα forest transmitted flux at $10\,h^{-1}$ comoving Mpc scale. At higher redshift, Frye et al. (2008) discovered a strong intergalactic Lyα absorption with the optical depth close to Gunn–Peterson absorption on $30\,h^{-1}$ Mpc in the spectrum of a galaxy at $z = 4.9$. The Lyα optical depth is twice higher than the mean optical depth at $z = 4.9$ on a large scale of $80\,h^{-1}$ comoving Mpc. Matsuda et al. (2010) conducted deep narrow-band imaging centered on this unusual absorption, which indeed revealed a high concentration of Lyα galaxies at $z = 4.87$ on a large scale. The structure expands over a region of ∼20 comoving Mpc × 60 comoving Mpc on the sky with a galaxy overdensity of $\delta \sim 4$. All of the above is evidence that supports the idea that on sufficiently large scales of $\gtrsim 10\,h^{-1}$ comoving Mpc, the IGM H I gas is a good tracer of the underlying mass.

The correlation between Lyα optical depth and galaxies on large scales can be naturally interpreted. At galaxy and cluster scales ($\lesssim$1–2 h^{-1} comoving Mpc), hydrodynamical processes such as AGN feedback, supernova (SN) winds, and shock-heating around the cosmic web should significantly enhance the strength of the local ionizing radiation. Each of these mechanisms can complicate the relation between Lyα forest absorption and three-dimensional mass fluctuation. At the same time, on larger scales of $\gtrsim 5\,h^{-1}$ comoving Mpc, AGN feedback could only have small effects on the mean ionizing background (Kollmeier et al. 2003), and other mechanisms, such as SN winds, shock heating, and metal-line cooling generally have similar or smaller impacts on the Lyα optical depth than AGN feedback (Viel et al. 2013). Moreover, even at small scales, a few simulations suggest that galactic winds tend to escape into the voids, leaving the filaments responsible for the Lyα absorption largely intact (Theuns et al. 2002; Tepper-García et al. 2012). Therefore, it is not a surprise that at scales larger than 5 h^{-1} comoving Mpc, the strong Lyα absorption is highly correlated with the underlying mass overdensity, which can also be traced by a galaxy overdensity.

Guided by the theoretical and observational arguments above, we have developed a new approach for identifying the extreme tail of the matter density distribution at the typical protocluster scale of 10–30 h^{-1} co-Mpc. This approach uses the largest library of quasar spectra currently available from the Baryon Oscillations Spectroscopic Survey (BOSS) (e.g., Dawson et al. 2013). The BOSS project is part of the Sloan Digital Sky Survey III (SDSS-III) (Eisenstein et al. 2011). This largest spectroscopic quasar data set enables one to locate extremely rare, high optical depth H I (Lyα) absorption that is due to IGM overdensities at scales of ∼10–30 h^{-1} co-Mpc. These IGM H I overdensities are in turn expected to trace the most massive early overdense regions. Most of these regions are progenitors of matured clusters at $z = 0$ with $M > 10^{14–15}\,M_{\odot}$, because Chiang et al. (2013) pointed out that the progenitors of such matured clusters at $z = 0$ are characterized by galaxy overdensities ($\delta_g >$ 2–4) on scale of ≈20 Mpc at $z = 2$–3. Compared to overdensities that are traced by biased halos (e.g., QSOs or radio galaxies), this technique allows the coverage of a significantly larger survey volume by using current and next-generation spectroscopic campaigns, such as SDSS-III/BOSS (e.g., Eisenstein et al. 2011; Dawson et al. 2013), SDSS-IV/eBOSS, Dark Energy Spectroscopic Instrument (DESI) (e.g., Flaugher et al. 2014), and Prime Focus Spectrograph (PFS) surveys (e.g., Sugai et al. 2012). This technique could have a higher completeness of tracing overdensities because the H I density is correlated with the underlying dark matter density field over large scales.

This is the first of a series of papers presenting the selection technique of the high effective optical-depth, large-scale intergalactic H I (Lyα) absorption from the SDSS quasar spectral survey. These systems have coherently strong Lyα absorption on scales of ∼10–20 h^{-1} Mpc. Over these scales, we present the strong correlation between the transmitted flux of one-dimensional intergalactic Lyα absorption and three-dimensional mass overdensities. We give detailed procedures for selecting these absorption systems. They trace regions that are excellent candidates of the most massive large-scale structures at $z = 2$–3. In the following, we refer to these systems as the CoSLAs, which stands for "Coherently Strong Lyα Absorption systems." We refer to the whole project as MAMMOTH, which stands for MApping the Most Massive Overdensity Through Hydrogen.

This paper is structured as follows. In Section 2 we introduce the SDSS-III/BOSS Lyα forest sample and the cosmological simulations that are used for theoretical guidance of the SDSS-III/BOSS Lyα forest sample. From simulations, we measure the cross-correlation between the Lyα transmitted flux and mass fluctuation. In Section 3 we make realistic mock spectra by including the high column discrete absorbers (Lyman limit system and damped Lyα absorption systems) in our mock spectra and adding noise as well as considering the continuum fitting errors. In Section 4 we introduce the techniques for selecting CoSLAs, emphasizing the need to eliminate high column density discrete absorbers (HCDs), such as damped Lyα systems (DLAs). In Section 5 we provide a discussion on Lyα absorption systems that are associated with previously confirmed overdensities. In Section 6 we present the sample of CoSLA candidates selected from high signal-to-noise (S/N) spectroscopy obtained with SDSS-III/BOSS and the Multiple Mirror Telescope (MMT). In Section 7 we discuss the implications of our observations and use our observational results to compare with cosmological simulations. When measuring distances in this paper, we refer throughout to comoving separations or distances, measured in Mpc, with $H_0 = 100\ h\,{\rm km\,s^{-1}\,Mpc^{-1}}$. We convert redshifts into distances assuming a ΛCDM cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $h = 0.7$.

2. CORRELATION BETWEEN MASS FLUCTUATION AND LYα TRANSMITTED FLUX AT LARGE SCALES

We explore the general statistical correlation between the transmitted flux ($F = \exp(-\tau_{\rm eff})$) of Ly$\alpha$ forest absorption and the three-dimensional mass fluctuation, and in particular, the scatter between the two. To investigate this cross-correlation, we use two sets of large-scale cosmological simulations to model the Lyα forest. The large-scale Lyα forest simulations are designed to match the properties and guide the observations of large-scale Lyα forest surveys, e.g., the SDSS-III/BOSS Lyα forest sample.

2.1. SDSS-III/BOSS Lyα Forest Sample

We use the Lyα absorption spectra observed in SDSS-III BOSS (Dawson et al. 2013; Ahn et al. 2014). BOSS is one of the four spectroscopic surveys in SDSS-III taken with the 2.5 m Sloan telescope (Gunn et al. 2006). BOSS measures redshifts of 1.5 million luminous red galaxies and Lyα absorption toward 160,000 high-redshift quasars (Bolton et al. 2012; Ross et al. 2012; Dawson et al. 2013). The BOSS spectra have a moderate resolution of $R \sim 2000$ covering 3600–10,400 Å. With a total exposure time of 1 hr for each plate, the BOSS Lyα quasar sample (Lee et al. 2012) has a median S/N of ∼2 per pixel (1 pixel corresponds to ≈1 Å) at rest-frame wavelength $\lambda = 1041$–1185 Å.

The SDSS-III/BOSS DR12 quasar catalog (Pâris et al. 2014) includes 160,000 quasars over 10,000 deg^2, which yields a quasar average density of 1 per (15 arcmin)^2, where (15 arcmin)^2 = $(17\,h^{-1}\,{\rm Mpc})^2$ at $z = 2.5$. When we assume that the typical mass overdensity extends a $15\,h^{-1}$ Mpc and that each quasar probes an average of $\Delta z \sim 0.3$ on the sight line, our survey volume for searching overdensities is approximately $(1.8\,h^{-1}\,{\rm Gpc})^3$. To measure Ly$\alpha$ optical depth, the mean-flux-regulated principal component analysis (MF-PCA) technique is applied for the continuum fitting (Lee et al. 2013). In this technique, PCA fitting is performed redward of the quasar Lyα line in order to provide a prediction for the shape of the Lyα forest continuum. The slope and amplitude of this continuum prediction are then corrected using external constraints for the Lyα forest mean flux (Lee et al. 2012; Becker et al. 2013).

2.2. Large-scale Cosmological Simulations on the Lyα Forest

We used two cosmological simulations to examine the correlation between the Lyα transmitted flux and the mass overdensity. The first cosmological simulation has a box size of $1.5\,h^{-1}$ Gpc and a Plummer equivalent smoothing of $36\,h^{-1}$ kpc. It contains 1500^3 particles (White et al. 2011). The initial particle spacing is $1.0\,h^{-1}$ Mpc. This simulation was originally used for predicting the BAO feature at $\approx 100\,h^{-1}$ Mpc. The deterministic fluctuating Gunn–Peterson approximation (FGPA) was used to generate skewers of optical depth with 4096 pixels for each sight line (e.g., Slosar et al. 2011). A temperature at the mean density of 2×10^4 K and an equation of state $T(\Delta) = T_0\Delta^{\gamma-1}$ with a $\gamma = 1.5$ is assumed, where Δ is the overdensity (e.g., Lee et al. 2015). The optical depth was scaled so that the median transmitted flux is $\bar{F} = \langle \exp(-\tau) \rangle = 0.78$ with $\langle \tau = 0.25 \rangle$, consistent with observations at $z \approx 2.5$ (e.g., Kirkman et al. 2005; Bolton et al. 2009). In the following, we refer to this simulation as the deterministic simulation.

The second simulation was performed with the N-body code GADGET-2 (Springel 2005) and used a box length of $1\,h^{-1}$ Gpc and 1024^3 dark matter particles. The Plummer-equivalent force softening adopted is 5% of the mean interparticle distance $48.8\,h^{-1}$ kpc. From this simulation, we derived a mock Lyα forest at $z \sim 2.5$ using LyMAS (Lyα Mass Association Scheme). The detailed description of this code is given in Peirani et al. (2014). Contrary to the FGPA mapping, which assumes a deterministic relation between dark matter overdensity and continuum-normalized Lyα flux, LyMAS considers a stochastic relation that is described by a conditional probability distribution $P(F_s|\delta_s)$ of the transmitted flux F_s smoothed (one-dimensionally) over the spectral resolution scale (of SDSS-III/BOSS) on the dark matter density contrast δ_s smoothed (three-dimensionally) over a similar scale. The conditional probability distribution has previously been derived from high-resolution hydrodynamic simulations of smaller volumes, including a full treatment of physical processes such as metal-dependent cooling, star formation, photoionization, and heating from a UV background, supernova feedback, and metal enrichment (Peirani et al. 2014). Lochhaas et al. (2015) tested the cross-correlation of the Lyα forest predicted by a LyMAS simulation and found that LyMAS perfectly reproduces the correlation between dark matter and transmitted flux computed from the full hydrodynamic gas distribution. In this paper, we use the extended form of LyMAS, which produces coherent spectra reproducing the one-dimensional power spectrum and one-point flux distribution of the hydro-simulation spectra (in redshift space). LyMAS is expected to be more accurate than a deterministic density-flux mapping. For instance, the deterministic mapping overpredicts the flux correlation function relative to the fully hydrodynamical simulation and LyMAS scheme (Peirani et al. 2014). In the following, we refer to this simulation as LyMAS. We use both simulations to examine the correlation between Lyα absorption and mass overdensities at scales of ten to a few tens of Mpc.

2.3. Strong Correlation Between Mass and Lyα Forest Transmitted Flux over 10–40 h^{-1} Mpc

Using cosmological simulations, we study the large-scale correlation between mass and Lyα transmitted flux as a function of different smoothing lengths (scales).

We define the fluctuation of the transmitted flux $\delta_F = 1 - F/\bar{F}$, where F is the transmitted flux in the Lyα forest after the spectrum has been smoothed with a top-hat filter, and $\bar{F}$ is the mean transmitted flux. We also define the mass perturbation $\delta_m = m/\bar{m}$. The mass m is defined as the total mass of the particles inside the cubes (redshift space) centered on the pixel of the δ_F measurement. Each cube has a linear size equal to the smoothing length.

Figure 1 presents the cross-correlation coefficients of transmitted flux δ_F and mass perturbation δ_m as a function of different smoothing scales. The profile of correlation coefficients predicted by the deterministic simulation is generally consistent with those calculated by the LyMAS simulation. Both simulations suggest a strong correlation between the mass overdensity and Lyα absorption on large scales. The correlation peaks at scales of 10–40 h^{-1} Mpc. From Figure 1, the deterministic scheme gives generally higher correlation coefficients between δ_F and δ_m. For example, at $20\,h^{-1}$ Mpc, the deterministic scheme presents a slightly (∼10%) stronger correlation.

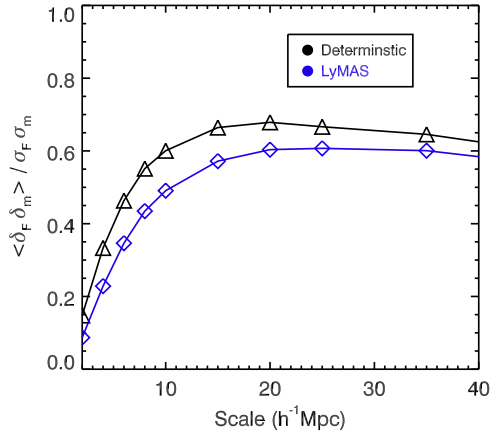

Figure 1. The cross-correlation between three-dimensional mass and one-dimensional transmitted flux as a function of scales. The black curve is calculated from the deterministic simulation (deterministic scheme). Blue represents the LyMAS simulation. The deterministic simulation produces systematically higher correlations than LyMAS. Both cosmological simulations suggest the correlation has a broad peak over the scale of 15–40 h^{-1} Mpc.

The correlation coefficient at $10\,h^{-1}$ Mpc derived by the LyMAS scheme is in excellent agreement with that calculated by Kollmeier et al. (2003), who used a fully hydrodynamical simulation with a box size of $40\,h^{-1}$ Mpc. McDonald et al. (2002) also derived the correlation between the Lyα forest absorption and mass fluctuation on $\sim 5\,h^{-1}$ Mpc. Compared with McDonald et al. (2002) and Kollmeier et al. (2003), this paper extends the cross-correlation to scales from 3 to $40\,h^{-1}$ Mpc. For the following analysis, we mainly use the LyMAS simulation.

2.4. The δm–δ_F Correlation on $15\,h^{-1}$ Mpcs

To better understand the δ_m–δ_F correlation, we choose a specific scale, $15\,h^{-1}$ Mpc, to examine the scatter between m and effective optical depth $\tau_{\rm eff}$ ($\tau_{\rm eff} = {\rm Log}\langle F\rangle$, where $\langle F\rangle$ is the mean transmitted flux). We choose the scale of $15\,h^{-1}$ Mpc because this scale gives a high correlation between the transmitted flux and mass overdensity (Figure 1). Furthermore, $15\,h^{-1}$ Mpc is the typical extent of the large-scale galaxy overdensities at $z > 2$ (e.g., Steidel et al. 1998, 2005; Ouchi et al. 2005; Matsuda et al. 2010; Chiang et al. 2013). We also provide the results on the scales of 10 and $20\,h^{-1}$ Mpc in the appendix. Figure 2 shows a plot of mass versus optical depth on $15\,h^{-1}$ Mpcs. Different colors represent relative number densities of points in the $\tau_{\rm eff}$–m diagram, normalized to systems in regions with the highest density (brown). The black dots indicate systems with the highest effective optical depth. This figure demonstrates that most of the black points, with high $\tau_{\rm eff}$, reside at the most massive end in the mass distribution.

In this figure, the median mass within $15\,h^{-1}$ Mpc boxes in the LyMAS simulation is $2.6 \times 10^{14}\,M_\odot$, denoted by the black horizontal line in Figure 2.

The scatter of the mass–$\tau_{\rm eff}$ relation is primarily due to the three-dimensional geometrical configurations of large-scale structures, e.g., strong one-dimensional Lyα forest absorption is biased toward those lines of sight that are aligned with the major axis of the three-dimensional structures.

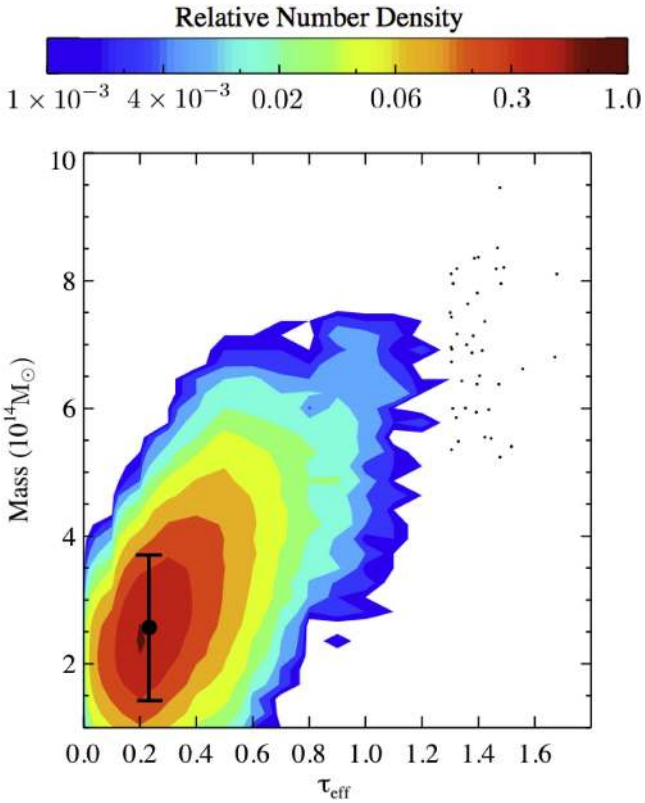

Figure 2. The relation between $\tau_{\rm eff}$ and mass on $15\,h^{-1}$ Mpc derived using the LyMAS simulation. The filled circle with the error bar represents the cosmic median mass within $(15\,h^{-1}\,{\rm Mpc})^3$ and the 1σ error. The colors represent the probability density of systems in this diagram. We normalized to the systems with the highest number density (brown) to unity. Black dots show the mass within $15\,h^{-1}$ Mpc cubes traced by absorptions with the highest effective optical depth. Each black dot represents an indepedent mass overdensity, i.e., there are no two black dots that represent the same system. Most of the black dots reside at the most massive end of the mass distribution.

2.5. Coherently Strong Lyα Absorption Systems

We demonstrated above that there exists a correlation between δ_F ($\tau_{\rm eff}$) and mass on large scales of 10–40 h^{-1} Mpc (Section 2.2). Here, we focus on the mass distribution traced by extreme systems with a largest $\tau_{\rm eff}$ over 10–20 h^{-1} Mpc (15–30 Å in the spectra at $z = 2.5$). We compare the mass overdensities traced by the strongest Lyα forest absorption to those traced by other methods (e.g., quasars) in the simulation.

First, let us define our Lyα forest absorption system sample. We select systems from mock spectra that have an intrinsically high effective optical depth. The red histogram in Figure 3 shows the distribution of $\tau_{\rm eff}$ on the $15\,h^{-1}$ Mpc scale. In the lognormal distribution, the mean optical depth is $\langle\tau_{\rm eff}\rangle = 0.25$, which value is consistent with observations (e.g., Bolton et al. 2009). The standard deviaton of the optical depth is 0.20. We focus on the systems that have $\tau_{\rm eff} > 1.15$, which are beyond the 4σ tail of the lognormal distribution of the effective optical depth. This threshold yields 200 systems in the $(1\,h^{-1}\,{\rm Gpc})^3$ volume, a sample feasible for observational follow-up. In the following, we refer to the Lyα absorption systems having $\tau_{\rm eff} \geqslant 4\sigma$ higher than $\langle\tau_{\rm eff}\rangle$ as CoSLAs.

Figure 4 presents three examples of CoSLAs in the LyMAS simulation. On the $15\,h^{-1}$ Mpc scale, the CoSLAs have $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}} \geqslant 4.5 \times \langle\tau_{\rm eff}\rangle$, 4σ higher than the mean optical depth. The high optical depth is due to the superposition of

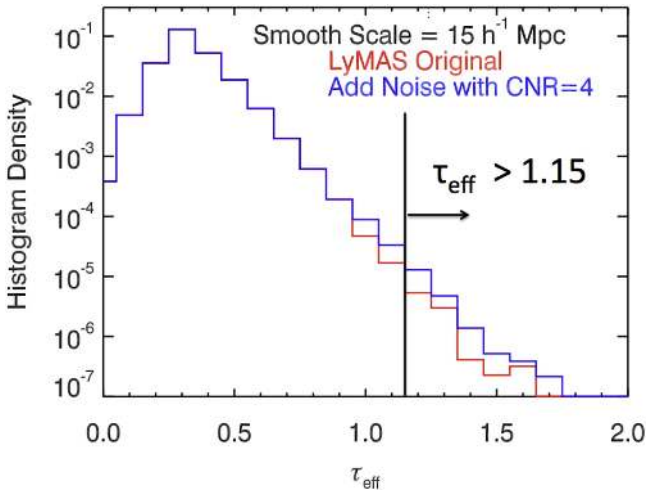

Figure 3. The distribution of the effective optical depth ($\tau_{\rm eff}$) on the scale of 15 h^{-1} Mpc. For the MAMMOTH project, we focus on the strongest absorption systems with the highest effective optical depth ($\tau_{\rm eff} > 1.15$). In the lognormal distribution of optical depth, these systems have $\tau_{\rm eff}$ 4σ higher than the mean optical depth.

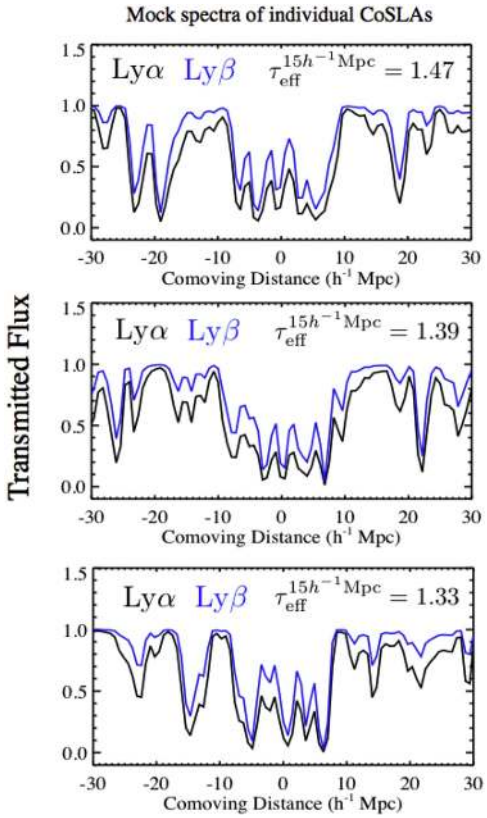

Figure 4. Three examples of the individual mock spectra of CoSLAs selected from the LyMAS simulation. The spectra have been convolved to the SDSS-III/BOSS resolution. Black represents Lyα, and blue shows the corresponding Lyβ. CoSLAs are the superposition of Lyα forest lines. The center is determined on the pixel that gives the lowest transmitted flux on $\pm$7.5 h^{-1} Mpc.

intergalactic Lyα forest. The central pixel of each absorption is determined on the pixel that gives the lowest transmitted flux from $\pm$7.5 h^{-1} Mpc.

Similarly, on the 10 h^{-1} Mpc scale, CoSLAs have $\tau_{\rm eff}^{10\,h^{-1}\,{\rm Mpc}} = 1.40$ (see left panel of the appendix Figure 31),

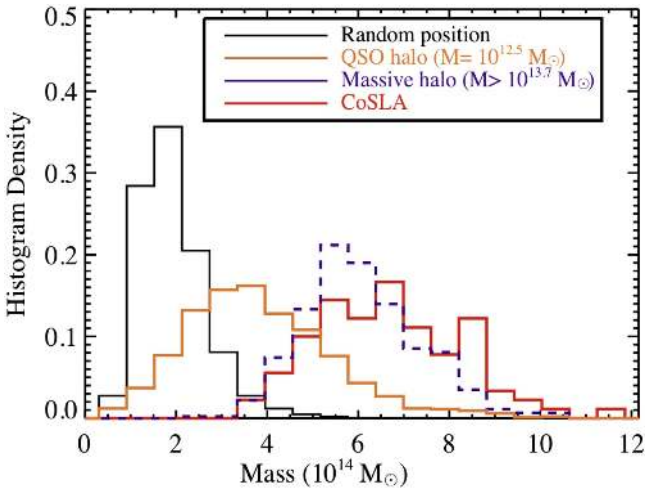

Figure 5. The distribution of mass traced by different objects. The x-axis is the mass within (15 h^{-1} Mpc)^3 cubes. The y-axis is the number of the cubes. Black represents the mass within (15 h^{-1} Mpc)^3 centered on random positions. The yellow histogram shows the mass traced by quasar halos. The purple dash–dotted histogram represents the mass distribution centered on the most massive halos in the LyMAS simulation ($M_{\rm halo} > 10^{13.7}\,M_\odot$). Red is the mass traced by the CoSLAs on 15 h^{-1} Mpc scale, selected from the noise-added mock spectra with a continuum-to-noise ratio of 4.

which is beyond 4σ of $\tau_{\rm eff}$ distribution. At the scale of 20 h^{-1} Mpc, the $\tau_{\rm eff}$ located at the 4σ tail is $\tau_{\rm eff}^{20\,h^{-1}\,{\rm Mpc}} = 1.03$ (see right panel in the appendix Figure 31).

We note that we use the $\tau_{\rm eff}$ rather than $\langle\tau_{\rm eff}\rangle$ to define CoSLAs because the relation between mass and $\tau_{\rm eff}$ is roughly indepedent of the redshift. From Gunn & Peterson (1965), the intergalactic Lyα optical depth τ is proportional to $\rho_{\rm H\,I}/\langle\rho_g\rangle(1+z)\frac{H_0}{H(z)}$, where $\rho_{\rm H\,I}$ is the number density of neutral hydrogen and ρ_g is the gas density. Then, $\tau_{\rm eff}$ is proportional to $\rho_{\rm H\,I}/\langle\rho_{\rm H\,I}\rangle$. For a uniform UV background at a given redshift, $\rho_{\rm H\,I}/\langle\rho_{\rm H\,I}\rangle$ is proportional to a mass overdensity. The relation between $\tau_{\rm eff}$ and the mass overdensities is only dependent on the factor of $(1+z)^3/H(z)$, which factor is roughly a constant in the CDM dominated universe at $z = 2$–3.

2.6. Mass Overdensities Traced by CoSLAs

How effectively does our technique trace large-scale overdensities compared with other tracers? In the following, let us examine these questions using simulations.

Figure 5 shows the distribution of mass traced by the CoSLAs on a 15 h^{-1} Mpc scale in the LyMAS simulation. The x-axis is the mass within (15 h^{-1} Mpc)^3 cubes. The y-axis is the number of such cubes. The different histograms in Figure 5 are explained as follows:

(1) We consider the mass within 15 h^{-1} Mpc in a random distribution. Black represents the cubes centered on random positions in the simulation box. The mass distribution follows a lognormal distribution, with a median mass of 2.6 $\times$ 10^{14} $M_\odot$ and a standard deviation of 1.2 $\times$ 10^{14} $M_\odot$.

(2) We examine the mass distribution traced by quasar halos.[9] The yellow histogram represents the mass within

[9] According to White et al. (2012), the duty cycle of QSOs is about 1%, and the QSO halo is roughly $10^{12.5}\,M_\odot$. In the simulation, we randomly selected 1% of $10^{12.3}$–$10^{12.7}\,M_\odot$ as QSO halos. The median mass of the quasar halo we selected is $10^{12.5}\,M_\odot$ (White et al. 2012).

$15\,h^{-1}$ Mpc, which is centered on quasar halos with $M_{\rm halo} = 2\text{–}3 \times 10^{12}\,M_\odot$ (e.g., White et al. 2012): the median mass in $15\,h^{-1}$ Mpc cubes is $3.8 \times 10^{14}\,M_\odot$, with a standard deviation of $1.6 \times 10^{14}\,M_\odot$ on the logarithmic scale.

(3) The third case uses the most massive halos in the LyMAS simulation as tracers of the mass overdensity at $15\,h^{-1}$ Mpc. The purple histogram (dashed) represents the mass within $15\,h^{-1}$ Mpc cubes centered on halos with $M_{\rm halo} > 10^{13.7}\,M_\odot$, which are more than a factor of ten times more massive than typical quasar halos. They are the top 0.01% most massive halos in the LyMAS simulation, and the $1\,h^{-1}$ Gpc box only contains 256 such halos. We note that $10^{13.7}\,M_\odot$ at $z = 2.5$ is about the mass of the progenitors of Coma-like clusters ($10^{15}\,M_\odot$ at $z = 0$) (Chiang et al. 2013). The median mass within the $(15\,h^{-1}\,{\rm Mpc})^3$ cubes centered on these rarest halos is $6.1 \times 10^{14}\,M_\odot$, with a standard deviation of $1.0 \times 10^{14}\,M_\odot$.

(4) We consider CoSLAs on $15\,h^{-1}$ Mpc. The mass distribution traced by CoSLAs selected from the original mock spectra, i.e., no noise being added to the spectra, has a median mass of $7.0 \times 10^{14}\,M_\odot$, with a standard deviation of $1.6 \times 10^{14}\,M_\odot$. When we add noise according to CNR = 4, the mass distribution traced by CoSLAs has a median mass of $6.7 \times 10^{14}\,M_\odot$, with a standard deviation of $1.6 \times 10^{14}\,M_\odot$. The red solid histogram in Figure 5 shows the mass distribution traced by CoSLAs selected from the noise-added spectra with a CNR = 4.

Figure 5 shows that the coherently strong one-dimensional Lyα absorption (CoSLAs) effectively trace the most massive overdensities. The regions traced by CoSLAs have a mass distribution comparable to those centered on the most massive single halos with $M_{\rm halo} > 10^{13.7}\,M_\odot$ (purple dashed histogram). More than half of the CoSLAs trace systems with halo overdensities of >1.6 on the $15\,h^{-1}$ Mpc scale, representing the top 0.2% most massive structures (>3.3-σ). These systems traced by CoSLAs will collapse to local clusters with a halo mass of $>10^{14}\,M_\odot$ (Chiang et al. 2013). Figure 6 further illustrates that the systems traced by CoSLAs represent the most massive tail of the mass distribution. Table 1 summarizes the mass distribution within $15\,h^{-1}$ that is traced by different objects in the LyMAS simulation.

The noise has a small effect on the mass distribution (Table 1). This result is mainly due to the large smoothing length of $15\,h^{-1}$ Mpc. Each $15\,h^{-1}$ Mpc contains 20 pixels, with the continuum-to-noise ratio (CNR) of 4 on each pixel; the error of the mean transmitted flux over the large scale is <0.05, making the uncertainty of $\tau_{\rm eff}$ small. Therefore, the strongest absorption systems selected from noise-added mock spectra have an intrinsically high optical depth, and they robustly trace the underlying mass overdensities, even though they are selected from data with modest S/N. Similar results have been found at other scales that range from 10 to $20\,h^{-1}$ Mpc (see Appendix).

Figure 7 further demonstrates the stacked mock Lyα absorption spectra that are associated with different mass overdensities ($\delta_m^{15\ {\rm Mpc}}$) at $z = 2.5$. Each spectrum is the median stack of 50 individual simulated spectra associated with similar $\delta_m^{15\ {\rm Mpc}}$. The effective optical depth on $15\,h^{-1}$ Mpc ($\tau_{\rm eff}^{15\,h^{-1}\ {\rm Mpc}} = -\log\langle{\rm Flux}\rangle_{15\,h^{-1}\ {\rm Mpc}}$) increases monotonically

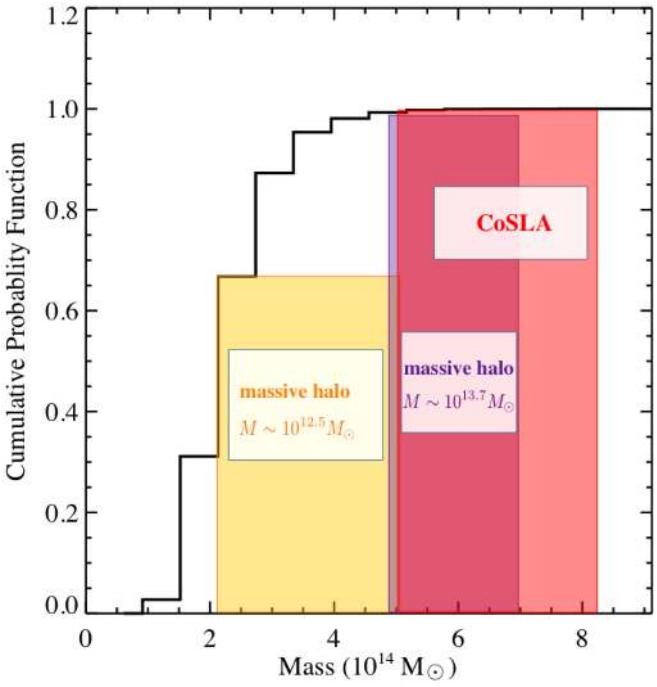

Figure 6. The cumulative probability of mass within the $15\,h^{-1}$ Mpc boxes. The x-axis is the mass within $15\,h^{-1}$ Mpc cubes. The y-axis is the cumulative probability. The yellow shaded area shows the mass distribution ($\pm 1\sigma$) that is traced by quasar halos. The purple shaded area represents the mass distribution centered on the most massive halos in the LyMAS simulation ($M_{\rm halo} > 10^{13.7}\,M_\odot$). The red shaded area represents the mass traced by the CoSLAs on the $15\,h^{-1}$ Mpc scale, selected from the noise-added spectra.

Table 1
Mass in $15\,h^{-1}$ Mpc Cubes Centered on Different Objects

Center	Median Mass ($10^{14}\,M_\odot$)	$\sigma_{15\,h^{-1}\,{\rm Mpc}}$ ($10^{14}\,M_\odot$)
Random	2.6	1.2
Quasars ($M_{\rm halo} = (2\text{–}3) \times 10^{12}\,M_\odot$)	3.7	1.6
Halos with $M_{\rm halo} \geqslant 5 \times 10^{13}\,M_\odot$	6.1	1.0
CoSLA (no noise)	7.0	1.6
CoSLA CNR = 4	6.7	1.6

Note. Summary of mass within $15\,h^{-1}$ using different tracers in the LyMAS simulation. H I (no noise) represents mass traced by CoSLAs selected from the original mock spectra, no noise being added; CoSLA CNR = 4 shows the mass traced by the strongest absorption selected from noise-added spectra.

with mass overdensities. In Figure 8 we present an example of an extreme mass overdensity traced by CoSLAs in the LyMAS simulation. These sight lines of the strongest absorption (bottom panel) all have a separation of $20\,h^{-1}$ Mpc away from each other. They are tracing different overdensities.

3. MOCK SPECTRA—THE INCLUSION OF HIGH COLUMN DENSITY ABSORBERS

In Section 2 we demonstrated that mass overdensities can be effectively traced using the coherently strong Lyα absorption systems (CoSLAs) with $\tau_{\rm eff} \geqslant 4.5 \times \langle\tau_{\rm eff}\rangle$ on $15\,h^{-1}$ Mpc. In real spectra, however, there will be contaminants such as DLAs, LLSs, and BALs. In this section, we explore how we can effectively recover genuine CoSLAs from realistic spectra. We

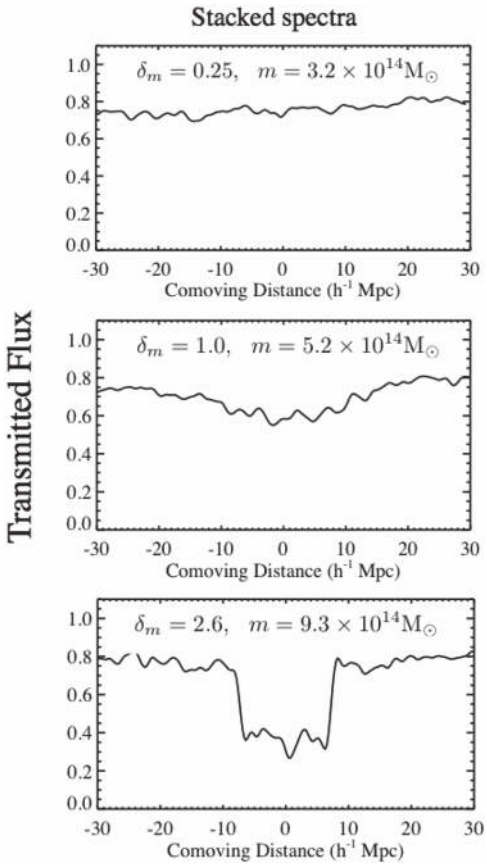

Figure 7. The stacked mock Lyα absorption that are associated with different mass overdensities ($\delta_m^{15\ \mathrm{Mpc}}$) at $z = 2.5$. Each spectrum shown is the result of a median stacking of 50 individual simulated spectra associated with similar $\delta_m^{15\ h^{-1}\ \mathrm{Mpc}}$. The representative spectra show an increasing δ_m from the top to bottom panels.

use the mock spectra to test the selection techniques, and apply these techniques to SDSS-III/BOSS data.

The original mock spectra generated by LyMAS simulations well predicts the intergalactic low-density gas, which in turn produces the optically thin Lyα forest (Peirani et al. 2014). However, these large-scale cosmological simulations lack sufficient resolution and gas self-shielding schemes to reproduce the number of HCDs with $N_{\mathrm{H\,I}} > 10^{17}\,\mathrm{cm}^{-2}$ in fidelity (Miralda-Escudé et al. 1996; Cen et al. 2003). These HCDs are self-shielded: the exterior absorbs the ionizing radiation, and the interior remains mostly neutral (e.g., McDonald et al. 2005). They are clumps of dense gas in galactic or circumgalactic environments (e.g., Rubin et al. 2014) and are observed as damped Lyα systems (DLAs, $N_{\mathrm{H\,I}} > 10^{20.3}\,\mathrm{cm}^{-2}$) and clustered Lyman-limit systems (LLSs, $N_{\mathrm{H\,I}} > 10^{17.2}\,\mathrm{cm}^{-2}$). It is important to take into account the realistic impact of HCDs in the mock spectra for our target selection (McDonald et al. 2005; Font-Ribera et al. 2012).

The CoSLAs, which trace massive overdensities, have profiles that resemble those of DLAs, especially at moderate S/N. This effect is illustrated in Figure 9, which compares a CoSLA in the LyMAS simulation to a DLA with $N_{\mathrm{H\,I}} \sim 10^{20.3}\,\mathrm{cm}^{-2}$. A significant fraction of DLAs resides in low-mass or modest-mass halos with $M < 10^{12}\,M_\odot$ (e.g., Møller et al. 2013), and most of these halos do not trace extremely massive large-scale overdensities. DLAs and strong sub-DLAs (e.g., $N_{\mathrm{H\,I}} \gtrsim 10^{19.5}\,\mathrm{cm}^{-2}$) could be the contaminants. We also need to take into account the

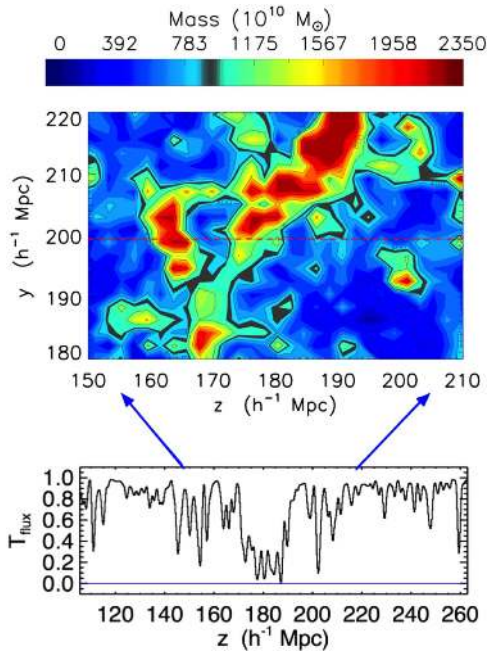

Figure 8. An example of CoSLA in the LyMAS simulation. Upper panel: a large-scale filamentary structure that extends on $\gtrsim 30\,h^{-1}$ Mpc in the simulation. The projected mass distributed within each slice of 0.6 $h^{-1} \times 0.6\ h^{-1} \times 15\ h^{-1}$ Mpc^3 (each cell has a zero-degree orientation along the y-axis). Different colors represent different masses in each cell. Lower panel: the simulated IGM absorption from the LyMAS simulation that traces the large-scale structure (red dashed line) in the upper panel.

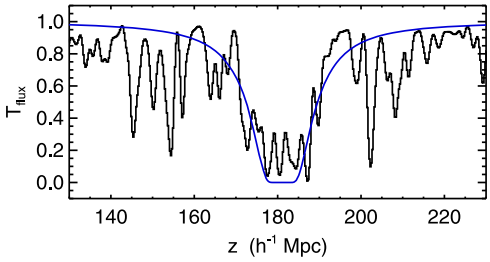

Figure 9. A comparison between the absorption due to a CoSLA and a DLA in the simulated spectra. No noise is added. The black spectrum indicates an extreme IGM absorption system in the LyMAS simulation. Blue shows a DLA system with $N_{\mathrm{H\,I}} = 10^{20.35}\,\mathrm{cm}^{-2}$. At moderate S/N, DLAs or overlapping sub-DLAs can mimic the large-scale IGM Lyα absorption.

clustered LLSs that have the similar equivalent widths (EWs) to those of CoSLAs; this is another population of contaminants.

In the following, we introduce our detailed procedures for generating realistic mock spectra, including inserting HCDs, adding continuum uncertainties, convolving spectra to the resolution of SDSS-III/BOSS, and adding noise. Similar procedures were adapted in previous works (Font-Ribera et al. 2012; Slosar et al. 2013; Delubac et al. 2015).

3.1. Procedures of Inserting HCDs into The Mock Spectra

Below, we give detailed procedures to insert the HCDs into the mock spectra generated using the LyMAS simulation. The LyMAS simulation has 65,536 skewers, with a box size of $1\,h^{-1}$ Gpc on a side. The simulation has a Lyα survey volume that is about the same as that of BOSS DR9, which contains about 30% of the total volume of BOSS DR12.

We assume a distribution of DLAs ($N_{\rm H\,I} > 10^{20.3}\,{\rm cm}^{-2}$) that follows a gamma function (e.g., Prochaska et al. 2005)

$$f_{\rm H\,I}^{\rm DLA}(N, X) = k_g \left(\frac{N}{N_g}\right)^{\alpha_g} \exp(-N/N_g). \quad (1)$$

Following Noterdaeme et al. (2009), at $z \sim 2.5$, we assume $\log k_g = -22.75$, $\log N_g = 21.26$, $\alpha_g = -1.27$, where $dX = \frac{H_0}{H(z)}(1+z)^2 dz$.

The total distance covered by our simulation is 9.4×10^7 Mpc. At $z = 2.5$, this distance corresponds to $X = 9.4 \times 10^7/(23.2) \times 0.066 = 267{,}413$. The total number of DLAs is

$$\int\int f_{\rm H\,I}^{\rm DLA}(N, X) dN dX = l_{{\rm DLA},z=2.5}\, X = 0.055 \times 216,362 = 11{,}899. \quad (2)$$

This is the number of DLAs we placed into the simulation, distributed randomly in the spatial direction along the sight lines.

In reality, DLAs are clustered, but the figuration for DLA host halos is still debated. A number of cross-correlation studies measured the average mass of DLA host halos to be widely distributed around 10^9–$10^{12}\,M_\odot$ (e.g., Cooke et al. 2006; Font-Ribera et al. 2012; Møller et al. 2013; Bird et al. 2014). We argue that neglecting DLA clustering has negligible effects on the selection of CoSLAs:

(1) If the DLA clustering yields an overlapping DLA, the absorption from overlapping DLAs has a greater width than the CoSLAs, and we can easily identify it as such (see Figure 9).

(2) The probability of having a DLA in the same region as a CoSLA is low. From the simulation, we have checked that the 96% of CoSLAs are not associated with particles ($M > 7 \times 10^{10}\,M_\odot$) within the projected distance of 100 kpc. Therefore, the majority of CoSLAs are not associated with DLAs with a host halo with $M > 7 \times 10^{10}\,M_\odot$. However, DLAs with a halo mass of $M < 7 \times 10^{10}\,M_\odot$ cannot be resolved from our simulation. Nevertheless, these low-mass DLAs are expected to have small impact parameters $\lesssim$10 kpc (Font-Ribera et al. 2012). Even if one assumes that the covering fraction of DLA clouds is 100% within 10 kpc around the halo center, we expect that ≈0.2% CoSLAs overlap with DLA contaminants.[10] Therefore, the DLA clustering does not affect our selection of CoSLAs.

Similarly, following the same procedures, we added LLSs to the simulation. Following Prochaska et al. (2005), we assume that $f_{\rm H\,I}(N, X)$ is a single power law over the $10^{17.2}\,{\rm cm}^{-2} < N_{\rm H\,I} < 10^{20.3}\,{\rm cm}^{-2}$ interval:

$$f_{\rm H\,I}^{\rm LLS}(N, X) = k_{\rm LLS}\left(\frac{N}{10^{20.3}\,{\rm cm}^{-2}}\right)^{\alpha_{\rm LLS}}. \quad (3)$$

[10] The DLA halo mass is estimated to be 10^9–$10^{12}\,M_\odot$ (e.g., Cooke et al. 2006). We conservatively assume that all the DLAs have halo masses of between 10^9 and $10^{12}\,M_\odot$. Furthermore, we assume that the covering fraction of DLA clouds is 100% within the impact parameter of 10 kpc. Under these two assumptions, according to Tinker et al. (2010), the number density of DLA halos is $0.8\,h^3$ co-Mpc^{-3}. Thus, the probability of a CoSLA that has a DLA in the cylinder with 10 kpc × 10 kpc × $15\,h^{-1}$ Mpc is 0.002 (0.2%).

At $z \sim 2.5$, $\alpha_{\rm LLS} = -0.9$, $k_{\rm LLS} = 10^{-21.43}$ from SDSS DR3 and DR4 samples (Prochaska et al. 2005), and

$$\int_{10^{17.2}}^{10^{20.3}} f_{\rm H\,I}^{\rm LLS}(N, X) dN dX = 0.5 \int f_{\rm H\,I}^{\rm LLS}(X) dX = 112{,}007. \quad (4)$$

This number yields ≈1.7 LLSs per $1\,h^{-1}$ Gpc path length, which is consistent with the measurements of the mean free path (Prochaska et al. 2010; Worseck et al. 2014). We have inserted 112,007 LLSs and 11,899 DLAs into the mock spectra.

In this paper, we randomly inserted LLSs in the mock spectra. However, LLSs are not randomly distributed. They prefer to reside in the overdense region. Thus, a fraction of CoSLAs should contain LLSs, which superposition affects the selection of CoSLAs (see target selection in Section 4). If we assume that a CoSLA traces an LLS overdenisty of 10 on a $15\,h^{-1}$ Mpc, then this structure is traced by one of the strongest Lyα absorptions (CoSLA). If the LLS overdensity follows the galaxy overdensity, we expect that the LLS overdensity could reach 10 on a $(15\,{\rm Mpc})^3$ volume in the most massive overdensities. If we assume a mean free path of LLSs of $\approx 600\,h^{-1}$ Mpc at $z \approx 2.5$ (Prochaska et al. 2010), we expect 25% of CoSLAs to contain at least 1 LLS within $\pm 7.5\,h^{-1}$ Mpc from the CoSLA center. About 75% of CoSLAs are not associated with LLSs, and the absorption is only due to the superposition of the intergalactic Lyα forest. We note that we assume an overdensity of 10 because it yields a conservative estimate of the genuine CoSLAs that are not associated with LLSs. The most massive structure found to date has an overdensity of ≈10 on a $10\,h^{-1}$ Mpc (Cai et al. 2016), and ≈8 on a $15\,h^{-1}$ Mpc.

3.1.1. Effects of LLS Clustering

The details of LLS clustering along the quasar sight lines are poorly constrained. We treat LLSs to be randomly distributed in our mock spectra. Nevertheless, we should note that the LLS clustering could yield overlapping LLSs that have a similar EW and profiles as those of CoSLAs. The strongly clustered LLSs with $N_{\rm H\,I} = 10^{17\text{–}18.5}\,{\rm cm}^{-2}$ are hard to eliminate using our current selection technique. Using 41 LLSs from the HD-LLS survey (Fumagalli et al. 2015; Prochaska et al. 2015), we roughly estimated how clustered LLSs affect our target selection efficiency. In Figure 10 we present the distribution of the optical depth on $15\,h^{-1}$ Mpc scale ($\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}}$), centered on the LLSs in the HD-LLS sample (black histogram). These observed LLSs have naturally included clustering. For comparison, we plot the optical depth distribution of the LLSs in our mock spectra (blue histogram, LLS clustering not included). From the optical depth distribution of LLSs, we calculate that the $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}} = 1.9 \times \langle\tau_{\rm eff}\rangle$, with a standard deviation of $0.7 \times \langle\tau_{\rm eff}\rangle$. The optical depth distribution follows a lognormal distribution. We estimated that the clustered LLSs with $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}}$ higher than CoSLA threashold occupies 0.05% of the entire LLS population. In a $1\,h^{-1}\,{\rm Gpc}^3$ volume, the number of clustered LLSs that have $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}} > 4.5 \times \langle\tau_{\rm eff}\rangle$ is 66.

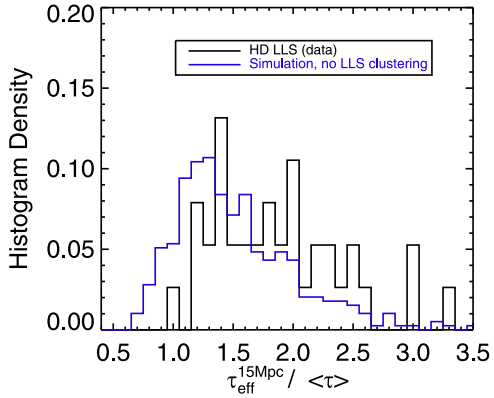

Figure 10. The distribution of $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}}$, centered on the LLSs in mock spectra and HD-LLS sample. The black histogram represents the distribution of the optical depth centered on LLS in HD-LLS sample, with naturally included clustering. The blue histogram represents the optical depth distribution centered on the LLSs in our mock spectra, without LLS clustering.

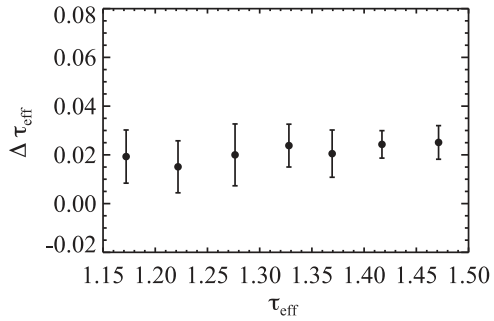

Figure 11. The $\tau_{\rm eff}$ differences of CoSLAs calculated using two sets of continua fitting. In the first set of continua, the optical depth of CoSLAs is calculated with the CoSLAs being masked in the continuum-fitting process. In the second set, $\tau_{\rm eff}$ is calculated without the CoSLAs being masked in the continuum fitting. The presence of CoSLAs makes the MF-PCA fitted continua biased toward a lower level by ≈2%–2.5%. The results are based on the calculation of 2000 mock spectra.

3.2. Convolve the Mock Spectra and Add Noise

After inserting DLAs and LLSs, we convolved the spectra to BOSS resolution using the Gaussian kernel, with the FWHM equal to the actual dispersion of the BOSS spectrum at $\lambda \sim 4250$ Å ($z = 2.5$). We resampled the mock spectra to BOSS pixel scale, using two pixels to populate a resolution element. We then added the noise to the spectra. We produced two sets of mock spectra, one with a CNR of 4 per pixel, the other higher S/N data sets with a CNR = 8 per pixel, a CNR = 10 per pixel, and a CNR = 12 per pixel for comparison and following discussions.

3.3. Uncertainties of Continuum Fitting

The uncertainties of continuum fitting need to be included tin the mock spectra because the observed optical depth of the Lyα forest is calculated based on the continuum. In practice, we use the mean-flux-regulated principal component analysis (MF-PCA) to fit the BOSS quasar continua (Lee et al. 2013). Following the discussion in Lee et al. (2012), the continuum residual $\delta C(\lambda)$ is defined as $\frac{C_{\rm fit}(\lambda_{\rm rest})}{C_{\rm true}(\lambda_{\rm rest})}$. The median r.m.s. continuum fitting error is 4.5% for spectra with 6 <S/N < 10 at $z = 2.5$ and 5.5% for spectra with 4 < S/N < 6. Therefore, we set a 4.5% uncertainty around unity for mock spectra with a CNR = 8, and a 5.5% uncertainty for mock spectra with a CNR = 4 to simulate the fitting errors of continua.

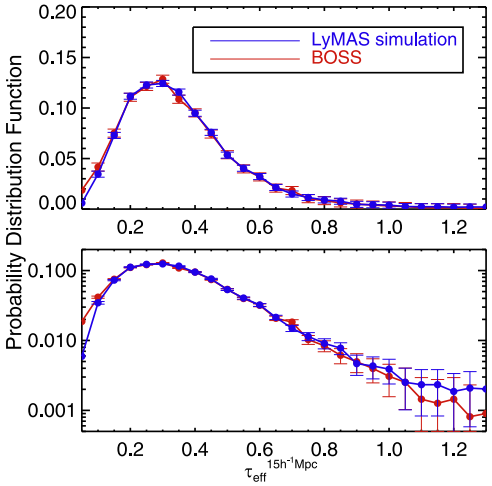

Figure 12. The probability distribution function (PDF) of $\tau_{\rm eff}$ on $15\,h^{-1}$ in BOSS data (red) and LyMAS simulation (blue), after adding noise, inserting HCDs, and including the continnum fitting uncertainties in the LyMAS simulation. We choose 20,000 systems in the LyMAS and the BOSS simulation. The noise is included with bootstrap resamping on the data. The upper panel shows the PDF in a linear scale on the y-axis, and the lower panel shows the PDF in a logarithmic scale.

CoSLAs also bias the quasar continuum to lower amplitude because the MF-PCA technique fits the amplitude based on the mean optical depth of the Lyα forest (Lee et al. 2012). Figure 11 presents the ($\Delta\tau_{\rm eff}$) of 2000 simulated systems, where the ($\Delta\tau_{\rm eff}$) is defined as the optical depth difference using the continua with and without the CoSLAs being masked. Statistically, the CoSLAs reduce the amplitude of the continua by an average level of ≈2.5%. In our target selections from a large data set, we do not know the exact positions of the CoSLAs in advance, and therefore we need to take the continuum bias introduced by CoSLAs into account.

After inserting HCD, adding noise, and including the uncertainty of continua fitting, we compare the probability distribution function (PDF) of $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}}$ of the LyMAS simulation and the BOSS data. We bootstrap 20,000 systems in the LyMAS simulation and the BOSS data. The systems in both LyMAS simulation and BOSS data have a median CNR of 8, and the systems in the BOSS data have redshifts of $z = 2.5 \pm 0.2$ to match the redshift of LyMAS simulation. The two PDFs are generally consistent with each other (Figure 12).

4. SELECTION OF COSLAS

Using these realistic mock spectra, we now present the detailed technique to select CoSLAs on $15\,h^{-1}$ Mpcs. The CoSLAs have a high effective optical depth ($\tau_{\rm eff}$) that is due to the intergalactic H I overdensities. We need to find algorithms to distinguish them from the contaminants that are due to HCDs. The algorithms developed from these realistic mock spectra can be directly applied to SDSS-III/BOSS data to select CoSLAs (Section 6).

4.1. Number of Targets and Contaminants

First, we define the cutoff for the effective optical depth ($\tau_{\rm eff}$) of the CoSLAs. After this initial restriction, the number of

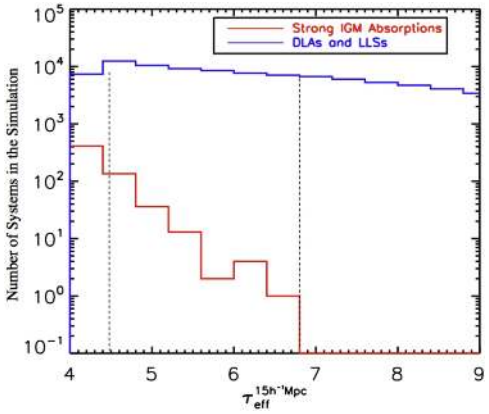

Figure 13. The comparison between the distributions of $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}}$ for strong IGM absorption (including CoSLAs) and DLAs and LLSs in the simulation. We propose a cutoff of optical depth of $4.5 \leqslant \tau_{\rm eff} \leqslant 6.6 \times \langle \tau_{\rm eff} \rangle$ (regions between two black vertical dashed lines). From the simulation, this cut can rule out the DLAs with $N_{\rm H\,I} > 10^{20.5}$ cm^{-2} without eliminating CoSLAs.

contaminant HCDs is significantly higher than the number of CoSLAs. The procedures are as follows:

(1) Cut on $\tau_{\rm eff}$: the $\tau_{\rm eff}$ cutoff is defined for the original mock spectra (no noise being added and no LLSs or DLAs being inserted), where all the highest $\tau_{\rm eff}$ systems are due to intrinsic IGM overdensities (CoSLAs). On a $15\,h^{-1}$ Mpc scale (≈21 Å in observed wavelength at $z \approx 2.5$), we use the top-hat filter to smooth the spectra. We then select the absorption systems with $1.20 < \tau_{\rm eff} < 1.56$ over $15\,h^{-1}$ Mpc ($z = 2.5$). This corresponds to $4.5 \times \langle \tau_{\rm eff} \rangle \leqslant \tau_{\rm eff} \leqslant 6.8 \times \langle \tau_{\rm eff} \rangle$. According to the distribution of the optical depth, these CoSLAs have a $\tau_{\rm eff}$ 4σ beyond the mean optical depth (see Section 2). In Figure 13 we show that this simple cutoff effectively rules out a significant amount of DLAs without eliminating CoSLAs.

(2) Count the number of targets: we added the noise to the mock spectra. These noise-added spectra do not have LLSs or DLAs inserted. From this set of noise-added spectra, we select absorption systems with transmitted flux ($F_{\rm t} = \exp^{-\tau_{\rm eff}}$) within the threshold determined in step (1). In the $1\,h^{-1}$ Gpc LyMAS simulation, a sample of 303 systems is selected using mock spectra with CNR = 4. For mock spectra with a CNR = 8, 289 systems are selected. For a CNR = 10, 278 systems are selected. For a CNR = 12, 275 systems are selected. Figure 5 shows that most of the systems selected from noise-added spectra (CNR = 4) trace large-scale structures with mass >2.6× that in random fields represent $>3.3\sigma$ mass overdensities. This distribution is similar to the CoSLAs selected from the original mock spectra, without noise being added (Table 1). We therefore define our targets as CoSLAs selected from the noise-added mock spectra, but without HCD inserted.

(3) Count the number of contaminants: we smooth over the spectrum with noise being added and LLSs-, DLAs being inserted. We select the absorption systems by the transmitted flux (F_t) that we determined in the step (1). A large number of absorption systems are selected; the vast majority of these systems are due to high column density absorbers (DLAs, sub-DLAs). For spectra with a CNR = 4, 13,635 HCD contaminants and 303 CoSLAs satisfy the cutoff of $\tau_{\rm eff}$ proposed in step (1). For spectra with a CNR = 8, 12,210 HCD contaminants and 289 CoSLAs are selected within this cut (Table 3). For spectra with a CNR = 10, 11,810 HCD contaminants and 278 CoSLAs are selected. For spectra with a CNR = 12, 10,934 HCD contaminants and 275 CoSLAs are selected. HCDs are most likely produced by individual galaxies (Møller et al. 2013) rather than large-scale IGM overdensities, therefore they are treated as contaminants.

After the cuts in optical depth, the number of contaminants is two orders of magnitude higher than the number of targets. We summarize the above results in Table 3. The following section describes our attempts to remove the HCD contaminants.

4.2. Selection Criteria of the Strongest Lyα Absorption Due to IGM Overdensities

We now describe the algorithm to remove the HCD contaminants and select CoSLAs. Figures 9 and 16 show that at a very high S/N, one can identify DLAs either by using the presence of damping wings (Voigt profile) or low-ionization metal lines. However, one should be cautious because this is hard using data with modest S/Ns. In our appendix, we present a few absorption systems that do not have damping wings or low-ionization metal lines in the SDSS data. Our MMT follow-up spectra demonstrate that a larger fraction of them are DLAs, which are shown in the Table 7. In the following, we present the selection algorithm. Our goal is to find selection criteria to eliminate DLAs and sub-DLAs without merely relying on fitting the Voigt profile.

We explored the following criteria that can effectively exclude a significant number of HCD contaminants and efficiently select CoSLAs on $15\,h^{-1}$ Mpc listed above.

(a) $w_{0.8} < 70$ Å, where $w_{0.8}$ is defined as the width of the CoSLA absorption trough in angstroms where the flux is lower than 0.8× the continuum. The $w_{0.8}$ is calculated in the observed frame.

(b) The mean flux of the absorption trough ($F_{\rm trough}$) > 0.15.

(c) Non-detection of low-ionization metal lines associated with the Lyα absorption systems.

(d1) For absorbers with $z > 2.65$: the Lyβ transition contains a series of absorbers, and the EW ratio of Lyβ to Lyα on $15\,h^{-1}$ Mpc is greater than 0.6. (at this redshift range, Lyβ is covered in the optical spectroscopy).

(d2) For absorbers with $z < 2.65$: using the presence of groups of absorption systems in the multiple background QSO sight lines with small angular separations (at $z < 2.4$, the average BOSS quasar density reaches ⩾15 per deg^2). We define an absorption group as ⩾4 absorption systems in a volume of $(20\,h^{-1}$ co-Mpc$)^3$, where each absorption has a $\tau_{\rm eff}$ on $15\,h^{-1}$ Mpc ⩾3× mean optical depth $\langle \tau(z) \rangle$.

In this paper, we mainly consider absorption systems at $z > 2.65$. In the next paper of this series, we focus on the absorption group and the spectroscopic confirmation of an extremely overdense field. In the following, we introduce details for each criterion:

Criterion (a): We define the width of the absorption systems. We eliminate DLAs with a column density $N_{\rm H\,I} > 10^{20.6}$ cm^{-2}. As shown in the simulation (Figure 4), even the most extreme IGM Lyα absorbers selected from a 1 $(h^{-1}$ Gpc$)^3$ volume have $w_{0.8} < 70$ Å. Wider absorbers are due to DLAs.

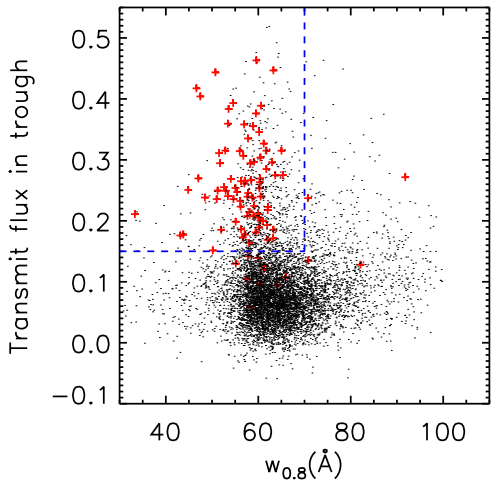

Figure 14. $F_{\rm trough}$–width diagram for absorption systems with a CNR = 8 per pixel. The red points represent CoSLAs. Black points are HCDs that have the same optical depth as CoSLAs. The blue dashed box represents the width selection criterion. This selection criterion selects 70.7% CoSLAs, and 79.4% of contaminants associated with HCDs are excluded by this selection criterion.

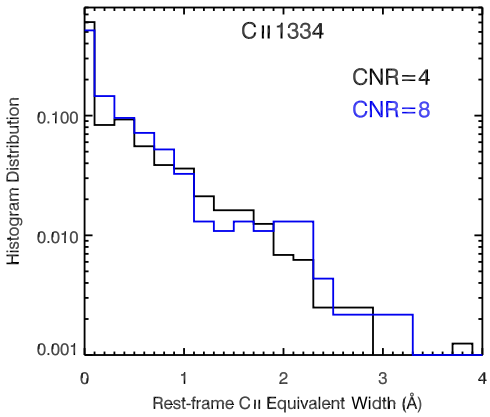

Figure 15. The distribution of C II λ1334 rest-frame equivalent width associated with HCDs with $N_{\rm H\,I} \geqslant 10^{20.0}\,{\rm cm}^{-2}$ in SDSS-III/BOSS data (Noterdaeme et al. 2012). The blue histogram represents HCDs with a CNR = 4, and ≈64% HCDs have detectable C II λ1334 line. The black histogram indicates HCDs with a CNR = 8, and ≈75% HCDs have detectable C II λ1334 absorption line.

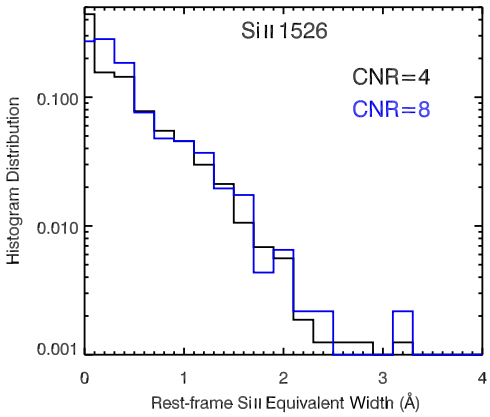

Figure 16. The distribution of Si II λ1526 rest-frame equivalent width associated with HCDs with $N_{\rm H\,I} \geqslant 10^{20.0}\,{\rm cm}^{-2}$ in SDSS-III/BOSS data (Noterdaeme et al. 2012). The blue histogram represents the Si II absorbers associated with HCDs with a CNR > 4, ∼74% HCDs have detectable Si II λ1526 absorption lines. Black indicates DLAs with a CNR > 8, and ∼81% DLAs have detectable Si II λ1526 absorption line.

In practice, we define the absorption center as the wavelength giving the lowest transmitted flux within ±10 Å (±7.5 h^{-1} Mpc at $z = 2.5$). For a given absorption system, the width of an absorption trough within *flux/continuum* = 0.8 ($w_{0.8}$) is calculated as follows: in the left wing, we calculate the median wavelength at a $\frac{\rm flux}{\rm continuum} = 0.80 \pm 0.01$, and denote this wavelength as $w_{l,0.8}$. In the right wing, we also calculate the median wavelength that has a flux to continuum = 0.80 ± 0.01, and denote this wavelength as $w_{r,0.8}$. The width of absorption at a *flux/continuum* = 0.8 ($w_{0.8}$) is then defined as $w_{r,0.8} - w_{l,0.8}$, and we require that $w_{0.8} < 70$ Å.

Criterion (b): This criterion gives the constraint for the mean flux in the dark trough. For DLAs, the expected level in this trough is zero; however, for CoSLAs, the absorption trough has a transmitted flux higher than 0.1.

In practice, we define the flux of the absorption trough ($F_{\rm trough}$) as the mean transmitted flux within the region of ±5 Å from the absorption center. We only select absorption systems with $F_{\rm trough} > 0.15$.

The criteria (a) and (b) are referred to as the width-trough (*w–t*) criteria. Figure 14 shows the separation between HCDs and CoSLAs in a $F_{\rm trough}$–$w_{0.8}$ diagram. With a CNR of 4 per pixel and on applying these two criteria, we can exclude 64.5% of HCD contaminants. Only 26.7% of the CoSLAs are eliminated by applying the same *w–t* criteria. For spectra with $R = 2000$ and a CNR = 8 per pixel, 80.0% of the CoSLAs pass the criteria, and 68.2% of contaminants are ruled out. For a CNR = 10 per pixel, 84.0% of the CoSLAs pass the criteria, and 69.3% of the contaminant can be ruled out. For a CNR = 12, 85.5% of the CoSLAs pass the criteria, and 72.9% of the HCD contaminants can be ruled out.

Therefore, criteria (a) and (b) effectively eliminate contaminants and also preserve high completeness for selecting targets. We summarize this result in Table 4.

Criterion (c): Low-ionization (low-ion) metal lines trace high column density neutral hydrogen (H I) systems in the interstellar and circumgalactic medium (e.g., Ford et al. 2013). However, if the quasar sight line passes through the IGM overdensity, this H I overdensity should not be associated with detectable low-ion metal lines in typical BOSS spectra (e.g., Oppenheimer et al. 2012).

A significant fraction of DLAs can be detected using metal lines (Noterdaeme et al. 2012). In the following, we calculate the fraction of DLAs that can be eliminated using metal absorption lines. We use the results measured from the SDSS data (Noterdaeme et al. 2012). Noterdaeme et al. (2012) presented the EW measurements for metal lines associated with H I absorbers with a column density $N_{\rm H\,I} > 10^{20}\,{\rm cm}^{-2}$ in SDSS-III/BOSS spectra. Using this catalog, we can calculate the fraction of the DLAs and sub-DLAs having detectable low-ion metal lines.

Figure 15 demonstrates that 64% of DLAs with $N_{\rm H\,I} = 10^{20.0}$–$10^{20.4}\,{\rm cm}^{-2}$ have detectable C II absorption lines redward of the quasar Lyα emission. Similarly, Figure 16 shows that 74% of DLAs with $N_{\rm H\,I} = 10^{20.0}$–$10^{20.4}\,{\rm cm}^{-2}$ have Si II absorption lines. For quasar spectra with a CNR = 8, 81% of DLAs with $N_{\rm H\,I} = 10^{20.0}$–$10^{20.4}\,{\rm cm}^{-2}$ can be ruled out using the C II or Si II absorption line. For spectra with a CNR $\gtrsim$ 4, we can remove 75% of DLAs. For spectra with a CNR $\gtrsim$ 8, we can remove 81% of DLAs. For a CNR = 10, we can remove

Table 2
DLA and sub-DLA Metal Line

CNR	C II or Si II Detected
4	75%
8	81%
10	82%
12	84%

Note. Summary of DLAs and sub-DLAs with $N_{\rm H\,I} = 10^{20.0-20.4}$ cm^{-2} that can be ruled out using low-ionization metal lines. Four sets of spectra are presented, and they have a CNR = 4, a CNR = 8, a CNR = 10, and a CNR = 12, respectively.

82% of DLAs by using the presence of the corresponding C II λ1334 and Si II λ1526 low-ionization metal lines. For a CNR = 12, we can remove 84% of DLAs simply by using the presence of the corresponding C II λ1334 and Si II λ1526 low-ionization metal lines.

The metallicity of sub-DLAs at $z \sim 2.5$ with $N_{\rm H\,I} = 10^{19}$–10^{20} cm^{-2}increases as the column density decreases (York et al. 2006; Khare et al. 2007; Péroux et al. 2008). Khare et al. (2007) reported that the average metallicity increases by 0.6 dex for $N_{\rm H\,I} = 10^{19}$–10^{20} cm^{-2} compared to the metallicity of DLAs with $N_{\rm H\,I} = 10^{20}$–$10^{20.5}$ cm^{-2} (e.g., Khare et al. 2007). It has also been reported that ≈50% of the sub-DLAs with 10^{19}–10^{20} cm^{-2} have an Si II rest-frame EW > 0.2 Å. Such Si II absorbers are within the detection limit of the SDSS with a CNR $\geqslant$ 4. Thus, we assume that $\gtrsim$50% sub-DLAs from 10^{19} to 10^{20} cm^{-2} have low-ionization metal lines that can be detected in the SDSS-III/BOSS. For LLSs with column density $N_{\rm H\,I} < 10^{19}$ cm^{-2}, $\gtrsim$20% of the LLSs have strong C II or Si II absorption with rest-frame EW > 0.2 Å and can be detected by BOSS data with a CNR $\geqslant$ 8 (Fumagalli et al. 2015; Prochaska et al. 2015). In our simulations, we conservatively assume that we cannot detect metal lines associated with LLSs in BOSS data.

In Tables 2 and 3 we summarize the results after applying the metal-line selection criteria to our data. Table 3 shows that the number of unidentified DLAs, sub-DLAs, and overlapping LLSs is significantly reduced, but remains more than one order of magnitude higher than the humber of CoSLAs. More criteria are needed to improve the selection efficiency and further eliminate these high column density absorbers.

Criterion (d1): The corresponding Lyβ absorption can be used to further determine the nature of the systems at $z > 2.65$. At $z > 2.65$, Lyβ is covered by BOSS spectra. The comparison between Lyα and Lyβ can help to distinguish whether the absorption system consists of DLAs or of the superposition of Lyα forest lines.

Based on our simulation, the CoSLAs with highest $\tau_{\rm eff}$ on a $15\,h^{-1}$ Mpc contain the superposition of absorbers with $N_{\rm H\,I} \sim 10^{15}$–$10^{17}$ cm^{-2}. The individual Lyα absorbers are in the moderately saturated part of the curve of growth. Using the mock spectra, we have calculated that CoSLAs have an EW ratio of Lyβ to Lyα in the range of 0.70–0.85 (also see Figure 4).

Conversely, without considering for LLS clustering, the major contaminant are high column density absorbers with $N_{\rm H\,I} > 10^{19.5}$ cm^{-2}. The corresponding Lyβ is a single absorber with a much smaller EW (see Figure 17), since it lies on the damped part of the curve of growth. Therefore, even at the BOSS S/N and resolution, the measured EW ratio of

Table 3
Selection of Strongest One-dimensional Lyα Absorption Systems on $15\,h^{-1}$ Mpc with $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}} = 1.15$–$1.56$ at $z = 2.5$ ($4.5 \times \langle\tau_{\rm eff}\rangle \leqslant \tau_{\rm eff} \leqslant 6.8 \times \langle\tau_{\rm eff}\rangle$)

CNR	Procedures	Noise-added Targets	HCD-added Contaminants	Efficiency
4	before w–t selection	303	13,635	···
	after w–t selection	222	4981	···
	after low-ion metal	222	1245	···
	after checking Lyβ	197	336	37%
	considering LLS clustering	197	402	33%
8	before w–t selection	289	12,210	···
	after w–t selection	231	3882	···
	after low-ion metal	231	738	···
	after checking Lyβ	213	140	60%
	considering LLS clustering	213	206	51%
10	before w–t selection	278	11,810	···
	after w–t selection	233	3626	···
	after low-ion metal	233	581	···
	after checking Lyβ	216	93	70%
	considering LLS clustering	216	159	57%
12	before w–t selection	275	10,934	···
	after w–t selection	235	2963	···
	after low-ion metal	235	456	···
	after checking Lyβ	219	70	75%
	considering LLS clustering	219	136	61%

Note. The CoSLA selection efficiency using spectra of $R = 2000$ and different CNRs. The last two columns present the number of absorption systems in noise-added mock spectra and HCD-inserted mock spectra, assuming LLSs are randomly distributed. For data with a CNR = 4, a CNR = 8, a CNR = 10, and a CNR = 12, we summarize our results after applying the w–t and metal-line selection criteria to our data. A significant number of DLA contaminants is excluded after applying these criteria. We note that our mock spectra do not have LLS clustering. We also estimate the number of the strongly clustered LLSs that have $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}} \geqslant 4.5 \times \langle\tau_{\rm eff}\rangle$ (see Section 3.1.1) and provide a rough estimation of the selection efficiency after considering LLS clustering.

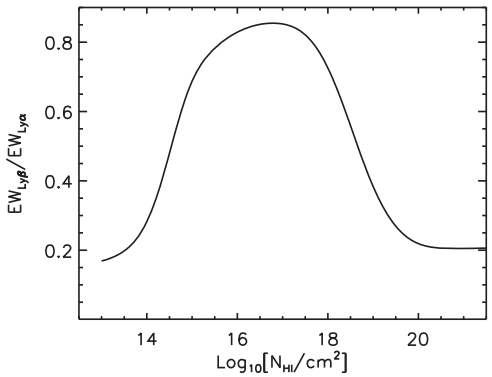

Figure 17. The EW ratio between Lyβ and Lyα for discrete absorbers as a function of H I column densities. When the column density is $N_{\rm H\,I} = 10^{15.5}$–$10^{18.0}$ cm^{-2}, the EW ratio is greater than 0.75. When $N_{\rm H\,I} > 10^{19}$ cm^{-2} or $N_{\rm H\,I} < 14.5$ cm^{-2}, the EW ratio drops below 0.25, regardless of the spectral resolution. This EW ratio can be used to rule out absorbers with $N_{\rm H\,I} > 10^{19.0}$ cm^{-2}.

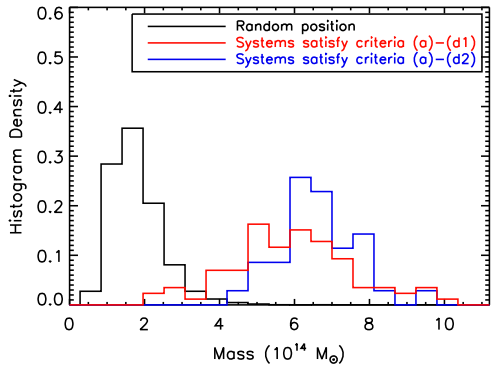

Figure 18. The red histogram indicates the mass distribution traced by absorption systems that pass criteria (a), (b), (c), and (d1) without considering the LLS clustering in our mock spectra. The absorption systems are selected from the HCD-inserted, noise-added mock spectra in the LyMAS simulation. Of the systems, 56% are CoSLAs and trace large-scale structures with mass a factor of 2.5× greater than that in a random field. They represent the $\gtrsim 3.5\sigma$ mass overdensities on 15 h^{-1} Mpc. The blue histogram represents the mass distribution traced by 17 groups of absorption systems (systems pass criteria (a), (b), (c), and (d2). These 17 groups contain strong IGM Lyα absorption, and effectively trace the mass overdensities.

Lyβ/Lyα provides a sensitive test on which part of the curve of growth the absorber lies, and whether it is a CoSLA or a DLA. Figure 17 shows that when the column density $N_{\rm H\,I} = 10^{15.0}$–$10^{18.0}$ cm^{-2}, the EW ratio of Lyβ to Lyα is greater than 0.6. When $N_{\rm H\,I} > 10^{19.5}$ cm^{-2} or $N_{\rm H\,I} < 10^{14.5}$ cm^{-2}, the EW ratio of Lyβ to Lyα significantly drops below 0.25.

Based on the above discussion, we select systems with a ratio of $\rm EW_{Ly\beta}$ to $\rm EW_{Ly\alpha}$ greater than 0.6. For a CNR = 4 and R = 2000, 74% of the contaminants that pass the criteria (a)–(c) can be further ruled out using criterion (d1). However, there are still more contaminants than targets after using criteria (a)–(d1). For R = 2000 and a CNR = 8, over 81% of the systems with $N_{\rm H\,I} > 10^{20}$ cm^{-2} can be identified. For a CNR = 10 and a CNR = 12%, 84%, and 85% of the systems with $N_{\rm H\,I} > 10^{20}$ cm^{-2} can be identified, respectively. We summarize these results in Table 3. In Figure 18 we plot the mass distribution traced by systems with a CNR = 8 that meet criteria (a)–(d1). Of these systems, 56% trace large-scale structures with a median mass of $\geqslant 6.2 \times 10^{14} M_\odot$, a factor of $\gtrsim 2.4\times$ that in random fields represents $\gtrsim 3\sigma$ mass overdensities.

These are the mass overdensity traced by CoSLAs using a single sight line, i.e., without additional 2D information. We note that these results do not include the LLS clustering. From Section 3.1.1, using high-resolution spectra of 41 LLSs with $N_{\rm H\,I} < 10^{19}$ cm^{-2} (Prochaska et al. 2015), we roughly estimate that the LLS clustering yields 0.05% systems with $\tau_{\rm eff}$ higher than our CoSLA selection threshold. Although this is a small fraction, it includes ≈66 clustered LLSs with $N_{\rm H\,I} \leqslant 10^{19}$ cm^{-2}, which are hard to identify from moderate and even high-resolution spectra (see Prochaska et al. 2015). In Table 3, we show that clustered LLSs decrease the selection efficiency.

We apply criteria (a)–(d1) to BOSS data to select the strongest absorption systems at $z > 2.65$. We summarize the sample selected from SDSS-III/BOSS in Section 6.

Criterion(d2): For absorbers at $z < 2.65$, Lyβ is not covered by BOSS spectra. We apply criterion (d2): using groups of Lyα absorption systems to pinpoint mass overdensities.

A considerable fraction of CoSLAs is associated with other nearby IGM Lyα forest systems with $\tau_{\rm eff} > 0.6$ ($3.0 \times \langle\tau\rangle_{z=2.5}$). This effect arises because true IGM overdensities trace filamentary structures that could extend a few tens of Mpc (see Figure 8). However, DLAs or sub-DLAs, which are more likely to trace field galaxies, normally have small H I cross-sections of $\ll$100 kpc × 100 kpc (e.g., Cai et al. 2014, Kashikawa et al. 2014), and there is a small chance of finding a group of HCDs on Mpc scales. For example, with sight-line separations of $\gtrsim 15\,h^{-1}$ Mpc, we estimate that the probability of having a DLA pair is 0.05%.[11] For groups of multiple DLAs within a 15 h^{-1} Mpc, the probability is even lower.

At $z < 2.35$, the SDSS-III/BOSS has reached an average quasar density of $\gtrsim 10$ deg^{-2}. In some subregions of the BOSS area (e.g., Stripe 82), the background quasar density exceeds 20 deg^{-2}. In the SDSS coverage at $z \leqslant 2.35$, ≈32% of the area has a background quasar density high enough to use criterion (d2) to select the massive H I overdensities.

In the LyMAS simulation, after reducing the density of sight lines to realistic 2D BOSS quasar distribution, 17 groups of absorption systems are selected from a 1 h^{-1} Gpc^3 box, which regions contain $\geqslant$4 absorption systems with $3 \times \langle\tau\rangle < \tau_{\rm eff} < 7 \times \langle\tau\rangle$ in a volume of (20 h^{-1} Mpc)^3. Each of these selected groups contains at least one CoSLA with $\tau_{\rm eff} > 4.5 \times \langle\tau\rangle$. Of these absorption groups, 53% trace >3.3σ overdensities. The three-dimensional mass distribution associated with these 17 groups of absorption systems is presented in the blue histogram of Figure 18. These absorption groups satisfy criteria (a), (b), (c), and (d2).

This method currently works best for BOSS at $z < 2.35$. For $z > 2.35$, the overdensity searches using criterion (d2) could be considerably incomplete because the 2D quasar density is significantly lower.

In our next paper of this series, we will introduce our narrowband imaging results on overdense fields with a group of absorption systems selected from SDSS-III. In this paper, we mainly focus on the spectroscopy results of one-dimensional Lyα absorption systems in single sight lines, and therefore focus on systems at $z > 2.65$.

5. LYα ABSORPTION SYSTEMS AROUND CONFIRMED OVERDENSITIES

In this section, we examine the Lyα absorption signatures of two well-studied galaxy overdensities. There are a few other overdensities at z = 2–3 that can be used to conduct similar tests (e.g., protoclusters in COSMOS, Chiang et al. 2014). The complete results of Lyα absorption around confirmed overdensities are summarized (Mukae et al. 2016).

5.1. SSA22 Protocluster at $z = 3.1$

The SSA22 protocluster is the most intensively studied large-scale galaxy overdensity at high redshift. This overdense field was serendipitously discovered by, e.g., Steidel et al. (1998, 2000) through a deep galaxy redshift survey. Further multiwavelength observations from optical to submillimeter confirmed an overdensity of LBGs, LAEs, LABs, and SMGs in this field at ∼30 h^{-1} Mpc (e.g., Chapman et al. 2004; Matsuda et al. 2005; Tamura et al. 2009). In the SDSS-III/BOSS quasar

[11] We assume that the DLA covering fraction is 100% within the impact parameter of 10 kpc from the halo center. Cooke et al. (2006) reported that the correlation function $\xi_{\rm DLA}(r)$ is ≈30% at r = 15 h^{-1} Mpc. Based on this correlation strength and halo abundance function (Tinker et al. 2010), the probability of finding two DLAs separated by 15 h^{-1} Mpc is approximately 0.05%.

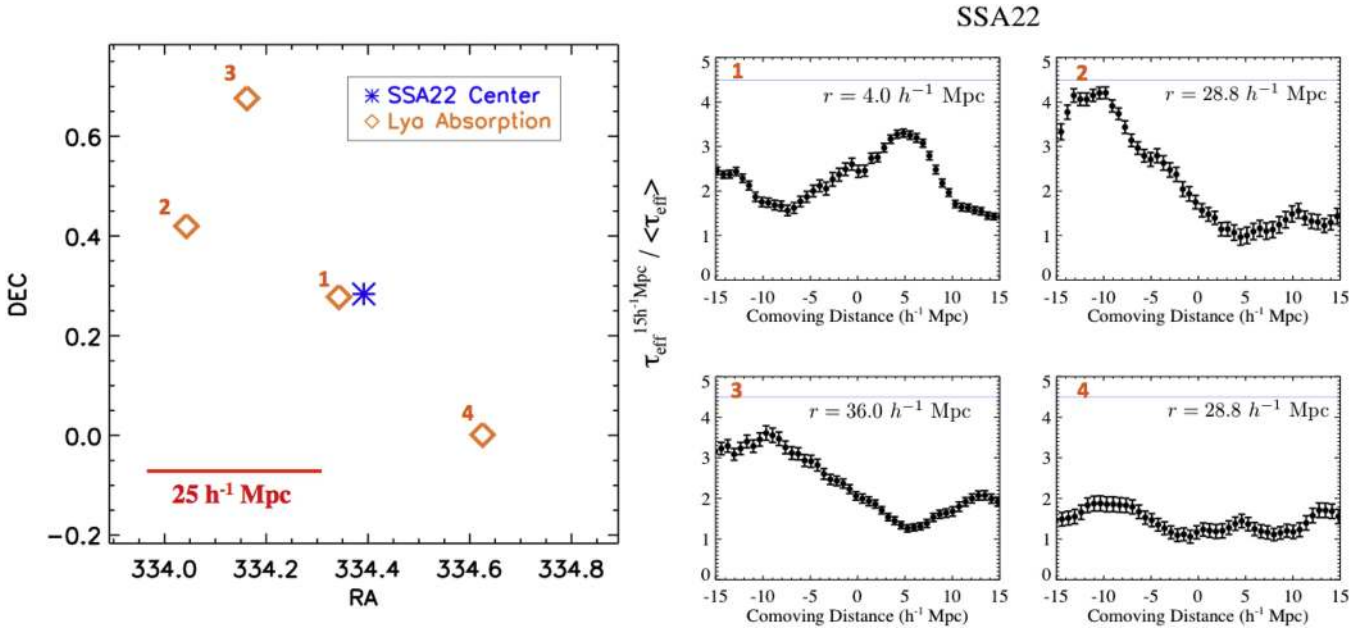

Figure 19. A group of four Lyα absorption systems around the SSA22 overdense field at $z = 3.1$. For all the four panels, we use the same redshift at $z = 3.090$ to define a line-of-sight distance of zero. These four sight lines have impact parameters $\lesssim$35 co-Mpc from the SSA22 field center. Black points with errors represent the optical depth on 15 h^{-1} Mpc ($\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}}$). Blue horizontal lines indicate the CoSLA threshold of 4.5 $\times$ $\langle\tau_{\rm eff}\rangle$. We also give the impact parameter r of each sight line in every panel.

library, we found four background quasars within 30 h^{-1} Mpc from the center of SSA22.

In Figure 19 we present the Lyα absorption around the SSA22 overdense field. The upper left panel presents a background quasar 4.0 h^{-1} co-Mpc away from the field center: α = 22:17:34, δ = +00:17:01 (J2000.0). This strong absorption has also been observed using Keck/HIRES (Adelberger et al. 2005), which confirmed that this absorption is consistent with the superposition of the intergalactic Lyα forest rather than LLSs or DLAs. In the upper right panel, we present an absorption system at similar redshift as the galaxy overdensity ($z = 3.09$), with an effective optical depth on 15 h^{-1} Mpc equal to $\approx$4.0 $\times$ $\langle\tau_{\rm eff}\rangle$. The lower left panel shows a background quasar $\sim$36.0 h^{-1} Mpc away from the center of SSA22 with a high $\tau_{\rm eff} \approx 3.7 \times \langle\tau_{\rm eff}\rangle$ on a 15 h^{-1} Mpc scale. The lower right panel presents a modest Lyα absorption at 28.8 h^{-1} co-Mpc from the field center, with an optical depth $\lesssim$2.0 $\times$ $\langle\tau_{\rm eff}\rangle$. The SSA22 is traced by a group of Lyα absorption systems.

5.2. A Large-scale Lyα Nebula Jackpot Quasar Quartet at $z = 2.05$

Giant Lyα nebulae (also known as Lyα blobs, LABs) are characterized by a high luminosity of the Lyα line emission (L (Lyα) $\gtrsim 10^{43}$ erg s^{-1}) and a spatially large Lyα emitting region from tens of kiloparsec (kpc) up to intergalactic scales of hundreds of kpc (e.g., Matsuda et al. 2005, 2011; Yang et al. 2009, 2010, 2014). Previous studies suggest that LABs are strongly clustered sources. They occupy massive dark matter halos ($\sim 10^{13}\,M_{\odot}$), and represent sites of the most active galaxy formation and large-scale galaxy overdensities (Yang et al. 2009, 2010).

Hennawi et al. (2015) reported a giant and ultraluminous Lyα nebula that is associated with a rare quasar quartet at z = 2.05. We set the position of the brightest quasar (α = 08:41:58.47, δ = 39:21:21.0) as the center of the quasar quartet. This structure is embedded in a substantial overdensity of galaxies (Hennawi et al. 2015). On a large scale of 13 h^{-1} Mpc from this LAB, we found five background quasars in SDSS-III. We excluded one of them because of the CNR < 2. We studied the Lyα absorption on the other four sight lines. Most of these quasar continua have significant excess of Lyα absorptions at $z = 2.05$ (Figure 20).

In Figure 20 we present the Lyα absorption for those sight lines with Lyα absorption $\geqslant 3 \times \langle\tau_{\rm eff}\rangle$ surrounding the quasar quartet. The upper left panel presents a background quasar 110 physical kpc away from the center of the Lyα nebula, and an absorption at the redshift of the quasar quartet ($z = 2.05$, λ = 3710 Å). The lower left panel shows another absorption system 3.9 h^{-1} Mpc away from the LAB. In the upper right panel, we display a background quasar 8.8 h^{-1} Mpc away from the Lyα nebula, and the lower right panel presents another strong absorption 12.5 h^{-1} Mpc away from the LAB. The absorption systems support the anticipation that a large-scale IGM overdensity is associated with this quasar quartet.

The SSA22 overdensity is the most overdense field selected from a $\sim$20,000 Mpc^3 survey volume. The Jackpot nebulae is discovered from a systematic narrowband survey (Flashlight survey, see Arrigoni Battaia et al. 2016) on 15 QSOs. The narrowband survey volume is about 70,000 Mpc^3. Since QSOs are preferred to reside in overdense environments, this Flashlight survey is expected to be biased toward the high-density regions in the universe. We note that finding a CoSLA system needs a survey volume of (100 h^{-1} Mpc)^3. Although these two surveys are small compared to the volume we need to find a CoSLA, the results motivate us to extend the searching to larger survey volumes (see Section 6).

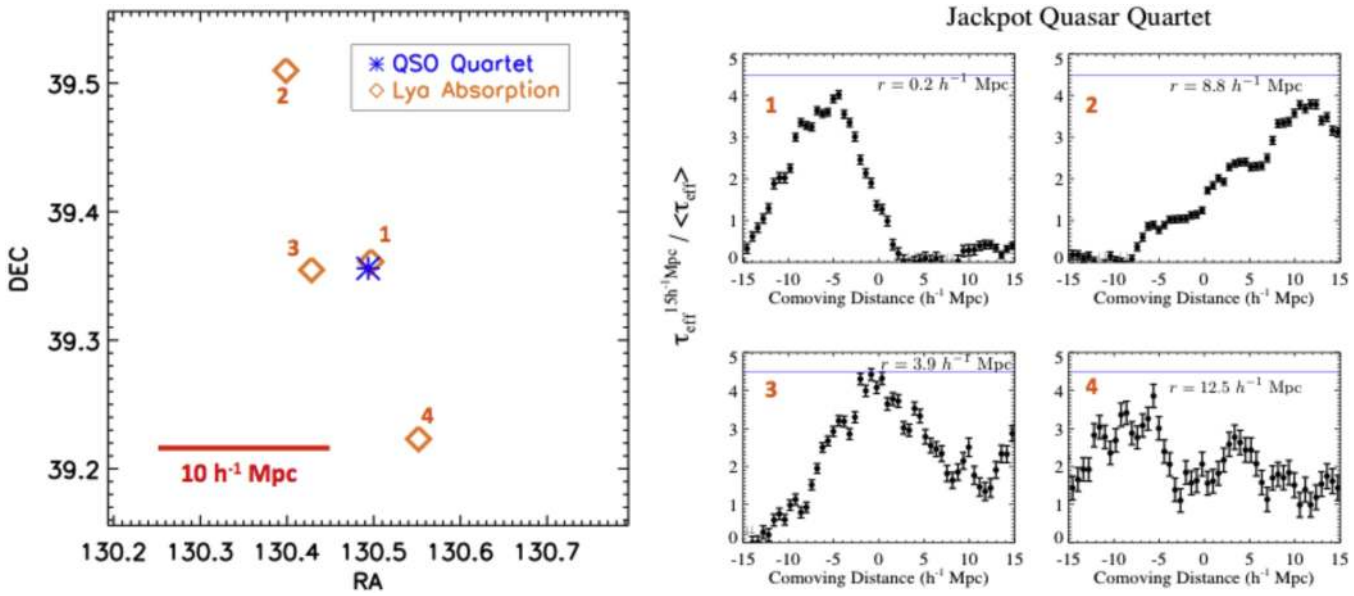

Figure 20. The group of Lyα absorption systems around the brightest QSO in the Jackpot Lyα blob (LAB) at $z = 2.055$. Each sight line has the transverse separations of $\leqslant 15\,h^{-1}$ Mpc from Jackpot quasar quartet. For all the four panels, we use the same redshift at $z = 2.055$ to define a line-of-sight distance of zero. These four sight lines have impact parameters $\lesssim$15 Mpc from the quasar quatet. Black points with errors represent $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}}$. Blue horizontal lines indiciate the CoSLA threshold of $4.5 \times \langle\tau_{\rm eff}\rangle$.

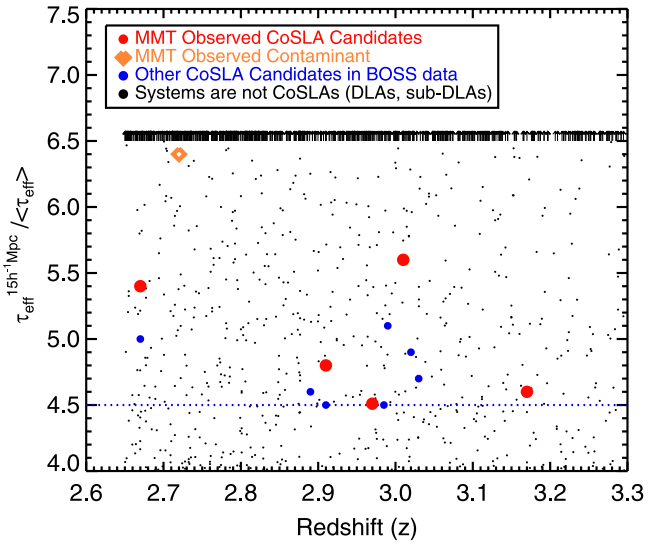

Figure 21. The summary of our data. The five red points represent the CoSLA candidates selected from SDSS-III/BOSS and have been observed by MMT. The seven blue points and one yellow point represent the other CoSLA candidates selected from SDSS-III/BOSS that passed criteria (a)–(d1). The yellow point represents a contaminant that we have identified using our MMT observations (see Appendix). The black dots represent 947 other absorption systems, and the systems with $\tau_{\rm eff}(z_{\rm abs}) > 6.5 \times \langle\tau_{\rm eff}\rangle$ on $15\,h^{-1}$ Mpc are marked by upper arrows, which are most likely to be HCDs. None of the systems represented by the black symbols pass criteria (a)–(d1).

6. COSLA CANDIDATES SELECTED FROM SDSS-III/BOSS DATABASE

6.1. A Sample of Candidate IGM Absorption Systems at $z > 2.65$

In this section, we introduce a sample of CoSLA candidates that are selected from the SDSS-III/BOSS quasar spectral library (see Section 2.1). For each quasar sight line, we search for CoSLAs over a redshift range between $z_{\rm min}$, defined as the redshift where the spectral CNR per pixel reaches 4 for both Lyα and Lyβ, and the $z_{\rm max}$, defined as 3000 km s^{-1} blueward of the quasar redshift. The selected absorption systems have the highest $\tau_{\rm eff}$ over a smoothing distance of $15\,h^{-1}$ Mpc. We apply our selection criteria (a)–(d1) described in Section 5.2 to select absorbers with $z_{\rm abs} > 2.65$, where Lyβ is covered by SDSS-III/BOSS.

Table 4
Mass in $15\,h^{-1}$ Mpc Cubes Centered on Different Objects

Center	Median Mass ($10^{14}\,M_\odot$)	$\sigma_{15\,h^{-1}\,{\rm Mpc}}$ ($10^{14}\,M_\odot$)
random	2.6	1.2
quasars ($M_{\rm halo} = 2$–$3 \times 10^{12}\,M_\odot$)	3.7	1.6
Halos with $M_{\rm halo} > 3 \times 10^{13}\,M_\odot$	6.1	1.0
CoSLAs selected from original mock	7.0	1.6
Systems selected using criteria (a), (b), (c), (d1) CNR = *8*	6.2	1.8
Systems selected using criteria (a), (b), (c), (d2) CNR = *4*	6.4	2.0

Note. Similar to Table 1, the summary of mass within $15\,h^{-1}$ Mpc using different tracers in the LyMAS simulation. "CoSLAs from original mock" represents mass traced by CoSLAs selected from original mock spectra, without noise being added or HCDs being inserted; "Systems selected using criteria (a), (b), (c), and (d1)" refers to the absorption systems selected from realistic mock spectra with HCD-inserted and noise added. We apply criteria (a)−(d1) to select CoSLAs from HCD-inserted realistic mock spectra, without LLSs clustering considered. We also present the mass distribution traced by the absorption groups selected using (a), (b), (c), and (d2).

6.1.1. Survey Volume

Within the R.A. range (R.A. $> 20\,h$ and R.A. $< 12\,h$) that we can reach with our scheduled time in MMT, we probe the CoSLAs from ≈6000 sight lines, a total distance ($d_{\rm sight\_line}$) of $1.53 \times 10^6\,h^{-1}$ Mpc along the lines of sight. The significant

Table 5
Summary of The Targets that Satisfy Criteria (A)–(D1) and have MMT Follow-up Observations

Name	z_{QSO}	z_{abs}	SDSS τ_{eff}	MMT τ_{eff}	MMT R	MMT Seeing	ExptIme (minutes)	SDSS CNR	MMT CNR	$\delta_m^{15\,h^{-1}\,Mpc}$ Predicted[a]
J025252.07+025704.0	3.06	2.91	$1.60^{+0.13}_{-0.12}$	$1.64^{+0.09}_{-0.09}$	3000	1.″0	3 × 20	9	15	1.2
J081103.27+281621.0	3.39	2.97	$1.86^{+0.12}_{-0.10}$	$1.79^{+0.09}_{-0.09}$	3000	1.″2	3 × 20	10	14	1.4
J084259.37+365704.3[b]	3.45	3.17	$2.42^{+0.31}_{-0.26}$	$2.38^{+0.25}_{-0.16}$	3000	1.″0	3 × 30	8	12	1.4
J113647.76+192633.9[c]	3.47	3.01	$1.84^{+0.16}_{-0.13}$	$1.80^{+0.12}_{-0.10}$	2000	1.″5	3 × 20	7	12	1.8
J122615.09+110543.4[d]	3.04	2.67	$1.60^{+0.15}_{-0.14}$	$1.55^{+0.11}_{-0.11}$	2000	1.″5	2 × 30	8	12	1.7

Notes.
[a] The predicted mass enclosed by the overdensity on 15 h^{-1} Mpc, assuming no LLS contamination.
[b] J084259.37+365704.3 is associated with two strong Lyα absorbers within 20 h^{-1} Mpc.
[c] J113647.76+192633.9 is associated with two quasars within 20 h^{-1} Mpc.
[d] J122615.09+110543.4 is associated with group of three quasars within 20 h^{-1} Mpc.

Table 6
A Contaminant Confirmed Using MMT

Name	z_{QSO}	z_{abs}	τ_{eff} ($\langle\tau_{eff}\rangle$)	T_{trough}	$w_{0.8}$ (Å)	$w_{0.5}/w_{0.8}$	SDSS CNR	MMT CNR	Exptime (minutes)	$N_{H\,I}$ (cm^{-2})
J161052.40+362333.1	3.24	2.72	$6.4^{+0.33}_{-0.28}$	0.15 ± 0.04	54	0.8	5	8	3 × 20	$10^{20.0}$

overdensity of H I gas is expected to extend to at least $\pm 5\,h^{-1}$ Mpc (e.g., see Figure 8), and each quasar probes a cylinder with a volume of $\sim 10\,h^{-1} \times 10\,h^{-1} \times d_{sight\_line}$ Mpc^3. Overall, we have probed a volume of $\sim 10 \times 10 \times 1.53\,h^{-3}$ $Mpc^3 = 0.15\ (h^{-1}\ Gpc)^3$.

From this volume and by applying selection criteria (a)–(d1), we select 13 CoSLA candidates from 947 absorption systems with $\tau_{eff}(z_{abs}) > 4.5 \times \langle\tau_{eff}(z_{abs})\rangle$ on $15\,h^{-1}$ Mpc. We summarize our data in Figure 21. We have obtained high S/N observations for 6 of these 13 absorption systems at MMT (red points in Figure 21, see Section 6.2). One of the 6 absorption systems is identified as a sub-DLA contaminant (see Appendix A.3).

We do not find any CoSLA with $\tau_{eff} > 7 \times \langle\tau_{eff}(z_{abs})\rangle$, where $\langle\tau_{eff}\rangle = 0.29$–$0.42$ at $z = 2.65$–3.40 (Bolton et al. 2009; Faucher-Giguère et al. 2008). This result is consistent with the LyMAS simulation. This result indicates that there are probably no CoSLAs with $\tau_{eff} > 7 \times \langle\tau_{eff}\rangle$ in a 0.15 Gpc^3 survey volume.

In the following, we present high S/N spectra of five CoSLA candidates. In Tables 5 and 6 we list the properties of this sample.

6.2. MMT Spectroscopy of a Sample of IGM Absorption Systems at $z = 2.7$–3.4

The MMT follow-up spectroscopy increases the CNRs in both Lyα and Lyβ regime. The MMT follow-up spectroscopy increases the CNRs that can either identify whether the CoSLA candidates are HCD contaminants or can enhance the CoSLA selection rate. Spectroscopic observations of CoSLA candidates were obtained using the blue-channel spectrograph onboard the 6.5 m MMT in 2012–2014 January. This section presents the sample of five CoSLA candidates at $z = 2.7$–3.4 that are strongly suggested by both BOSS and MMT spectra (also see Table 5).

Depending on the central wavelength of the absorption, we either use the 800 lines mm^{-1} ($R = 2000$) or the 1200 lines mm^{-1} ($R = 3000$) gratings, with the selection made for the following two reasons: (1) to match the central wavelength of the absorption with the most sensitive part of the grating and spectrograph, and (2) to better resolve the absorption feature. Typically, 90–120 minutes of on-source exposures were taken for each target, which varied according to the weather conditions and the quasar flux density. We divided long exposures into a series of single 20–30 minute individual exposures. Quality checks were conducted to meet the CNR requirement.

Data reduction was completed with the automatic pipeline during the observations in order to decide whether more exposures were needed. The wavelength coverage varies according to the central wavelength of the specific absorption system. The airmass of the observations did not exceed 1.4, and we used position angles close to the parallactic angle. Spectrophotometric standard stars were observed for flux calibration, and a CuAr arc lamp was used for wavelength calibration. In the region of the absorption system, the typical CNR per pixel ($\sim 50\,km\,s^{-1}$) is ≈ 10.

6.2.1. J025252.07+025704.0, $z_{abs} = 2.91$

In the upper panel of Figure 22 we show the SDSS-III/BOSS spectra of the CoSLA candidate J025252.07+025704.0 at $z = 2.91$. The absorption system (yellow shaded area) well satisfies the selection criteria of CoSLAs on $15\,h^{-1}$ Mpc scale centered at 4755 Å. From the BOSS data, the effective optical depth of this system is $\tau_{eff} = 1.60^{+0.13}_{-0.12}$, a factor of 4.8× the mean optical depth at $z = 2.91$. The effective optical depth is given in the middle panel on $15\,h^{-1}$ Mpc $\tau_{eff}^{15\,h^{-1}\,Mpc}$, which is greater than our selection threshold (blue horizontal line) of $4.5 \times \langle\tau_{eff}\rangle$. This CoSLA candidate is associated with a QSO at the same redshift of $z = 2.91$. The projected distance between the QSO ($\alpha = 02{:}52{:}22.37$, $\delta = +02{:}55{:}25.0$) and the CoSLA candidate is 3 physical Mpc at $z = 3.91$.

The lower panel presents the MMT follow-up observations of this target with 3 × 30-minute exposures using a grating of

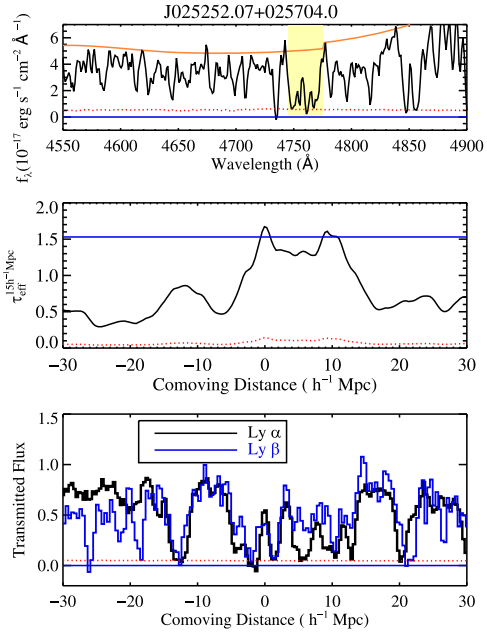

Figure 22. The spectra of CoSLA candidate J025252.07+025704.0. The upper panel shows the BOSS spectrum, with absorption marked in the yellow shaded area. Orange shows the continuum using a mean optical depth regulated PCA fit (Lee et al. 2012). The middle panel presents the $\tau_{\rm eff}$ over 15 h^{-1} Mpc centered on the absorption center. The blue horizontal lines indicate the threshold of CoSLAs, which is a 4.5× the mean optical depth. The lower panel presents MMT spectra (R = 3000), expanding the Lyα absorption. Black presents the Lyα absorption, and blue indicates Lyβ absorption. We define the comoving distance (x-axis) = 0 as the wavelength that gives the highest $\tau_{\rm eff}$ on ±7.5 Å. From the equivalent width comparison between the Lyα (black) and Lyβ absorption (blue), this system contain multiple absorbers, with the column density of each absorber $N_{\rm H\,I} \sim 10^{16}$–$10^{17}$ cm^{-2}.

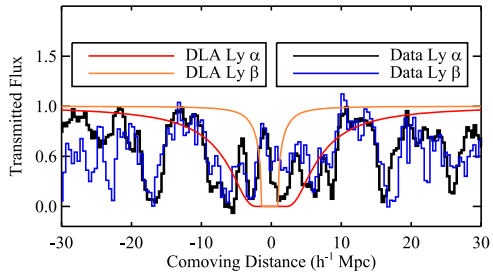

Figure 23. A comparison between the CoSLA candidate J025252.07 +025704.0 and a sub-DLA (super LLS) with $N_{\rm H\,I} = 10^{20.0}$ cm^{-2}. The red spectrum is the DLA Lyα; and orange is DLA Lyβ, overplotted with the Lyα absorption (black) and Lyβ (blue) of J025252.07+025704.0.

1200 lines mm^{-1}. With MMT spectra, we can resolve any Lyα and Lyβ absorbers with a rest-frame Doppler parameter $b > 100/(1+z) = 25\,{\rm km\,s^{-1}}$. From the lower panel, the EW ratio between Lyβ (blue) and Lyα (black) is $\frac{\rm EW_{Ly\beta}}{\rm EW_{Ly\alpha}} = 0.88^{+0.03}_{-0.03}$, which suggests that this absorption consists of the superposition of a series of individual absorbers with $N_{\rm H\,I} \sim 10^{15\text{–}18.5}$ cm^{-2} (Figure 17). This absorption system is similar to the strongest intergalactic Lyα absorption system predicted in the LyMAS simulation. The result in Figure 23 furthermore suggests that the absorption J025252.07 +025704.0 strongly deviates from DLA or sub-DLA absorption.

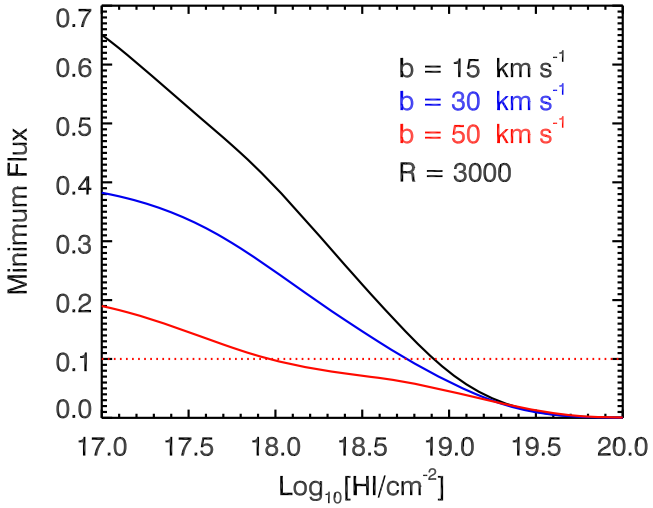

Figure 24. The minimum flux as a function of column densities at a resolution of R = 3000. The black curve represents a single Lyα absorber with a Doppler parameter b = 15 km s^{-1}. Blue indicates a single absorber with b = 30 km s^{-1}, and red shows an absorber with b = 50 km s^{-1}.

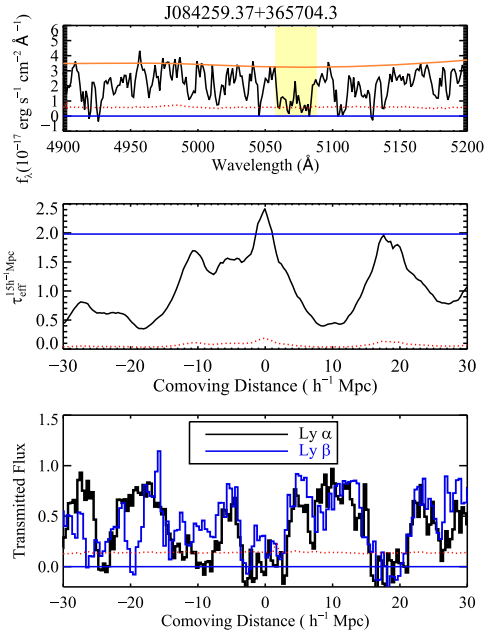

Figure 25. Similar to the format of Figure 22, we show the CoSLA candidate J084259.37+365704.3. The spectra suggest that this strong Lyα absorption system is associated with the superposition of the Lyα forest in the range 10^{15} cm^{-2} < $N_{\rm H\,I}$ < $10^{18.5}$ cm^{-2}.

The distribution of the Doppler parameter b in the Lyα forest is characterized by a Gaussian function with a median $b \approx 30\,{\rm km\,s^{-1}}$ and $\sigma = 10\,{\rm km\,s^{-1}}$, and cropped below $b \approx 20\,{\rm km\,s^{-1}}$ (e.g., Rudie et al. 2012; Pieri et al. 2014). Figure 24 shows that at an MMT resolution of 100 km s^{-1} bins (or wider bins, such as the BOSS resolution), the Lyα forest line of the typical Doppler parameters does not reach the zero level, regardless of its column density. Only single lines with Log $\left[\frac{\rm NHI}{\rm cm^{-2}}\right] \gtrsim 19$ or unusually high Doppler parameters ($\gtrsim$50 km s^{-1}) reach a minimum flux of $F \leqslant 0.15$ (also see Pieri et al. 2014). It is true that the absorption reaches a minimum transmitted flux of $F \leqslant 0.1$. Nevertheless, we argue that such a strong absorption is due to the superposition of the

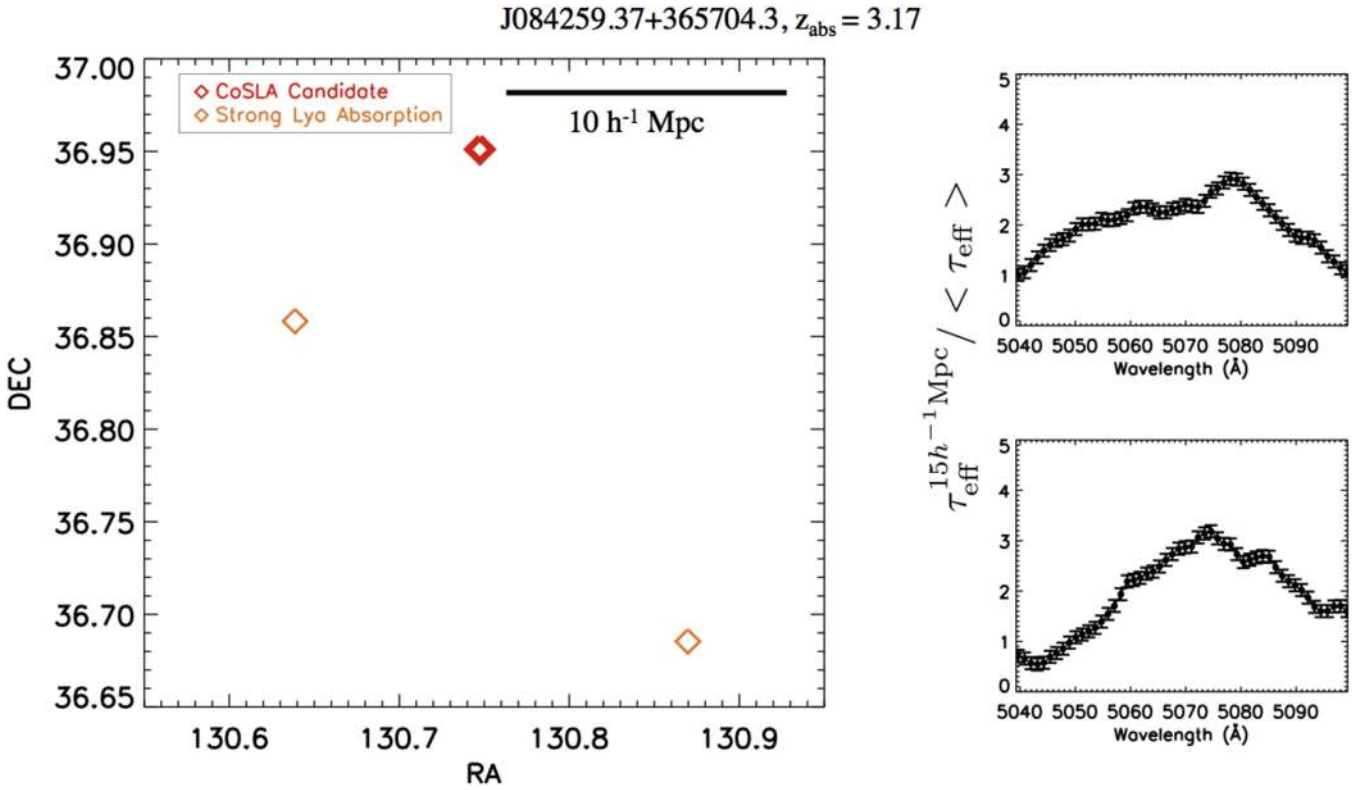

Figure 26. Left: the projected two-dimensional diagram of the CoSLA candidate J084259.37+365704.3 at $z = 3.13$ (red diamonds, also see Figure 25) and the absorptions in the other two sight lines (orange diamonds) within the 20 h^{-1} Mpc scale. Both sight lines (orange diamonds) contain strong Lyα absorption (see right panel). Right: the ratio of $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}}$ to $\langle\tau_{\rm eff}\rangle$ of the Lyα absorption in nearby sight lines. Only two sight lines are associated with this CoSLA candidate, and both of them have strong Lyα absorption. In the upper panel we show the middle sight line (middle orange diamond in the left panel). In the lower panel we present the bottom sight line (bottom orange diamond in the left panel). Both these two sight lines contain strong Lyα absorption with $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}}/\langle\tau_{\rm eff}\rangle > 3$.

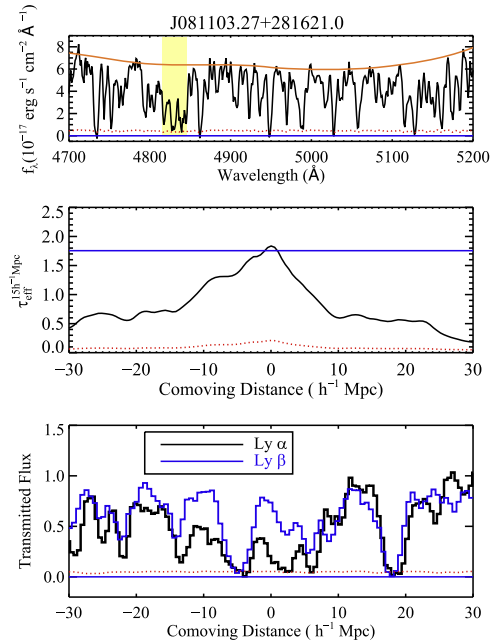

Figure 27. Same format as the previous Figure 22, presenting the CoSLA candidate J081103.27+281621.0.

Lyα forest line with $N_{\rm H\,I} < 10^{19}\,{\rm cm}^{-2}$. If a strong absorber exists with $N_{\rm H\,I} > 10^{19}\,{\rm cm}^{-2}$, then the Ly$\beta$ to Lyα ratio is expected to be lower than 0.2 (Figure 15). However, this absorption has an EW ratio greater than 0.8 in both MMT and BOSS spectra. The solution is that the Lyα absorption consists of the superposition of Lyα forest lines (also see Pieri et al. 2014). Thus, we conclude that the nature of strong Lyα absorption in J025252.07+025704.0 is highly likely due to the superposition of Lyα lines.

A natural quastion to ask is what the underlying mass of such strong absorption is. We define the mass overdensity within 15 h^{-1} comoving Mpc as

$$M_{15\,h^{-1}\,{\rm Mpc}} = (1 + \delta_m) \times \langle M_{15\,h^{-1}\,{\rm Mpc}}\rangle, \qquad (5)$$

where $\langle M_{15\,h^{-1}\,Mpc}\rangle$ is 2.6 $\times$ $10^{14}\,M_\odot$, the average mass within a $(15\,h^{-1}\,{\rm Mpc})^3$ box. The quantity δ_m is the mass overdensity. On the 15 h^{-1} Mpc scale, the effective optical depth of this absorption is about 4.8 $\times$ $\langle\tau_{\rm eff}(z = 2.9)\rangle$, where the mean optical depth at $z = 2.9$ $\langle\tau_{\rm eff}(z = 2.9)\rangle = 0.34$, 35% higher than that at $z = 2.5$ (Faucher-Giguère et al. 2008; Bolton et al. 2009). Assuming the $\tau_{\rm eff}$–mass relation (Figure 2) at $z = 2.9$, it also follows that at $z = 2.5$, J025252.07+025704.0 corresponds to a system with $\tau_{\rm eff} = 1.20$ at $z = 2.5$. From Figure 2, the median mass traced by CoSLAs with $\tau_{\rm eff} = 4.8 \times \langle\tau\rangle$ is around 6.5 $\times$ $10^{14}\,M_\odot$, equivalent to an overdensity about 1.5 on the large scale of 15 h^{-1} Mpc.

The true mass distribution associated with a CoSLA candidate should be qualitatively similar to the simulation results (Figure 8). We know that a single overdensity contains some subregions that have a high mass concentration. Figure 8 presents the projected matter distribution in the x–z plane, where the z-axis is along the LOS direction. The density peaks

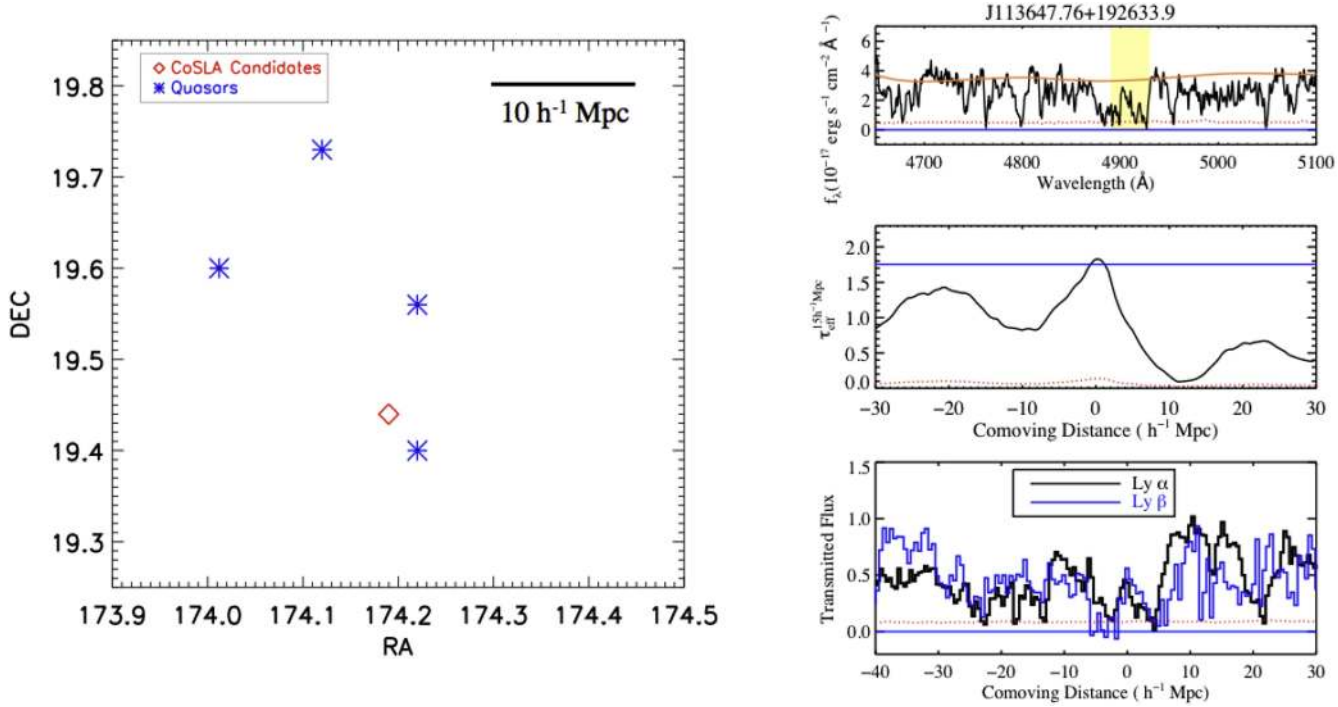

Figure 28. Left: the projected two-dimensional diagram of the CoSLA candidate J113647.76+192633.9 at $z = 3.03$ (red diamonds) and QSOs at $z = 3.00$–3.04 (blue asterisks) within the $20\,h^{-1}$ Mpc scale. Right: same format as the previous Figure 22, presenting the CoSLA candidate J113647.76+192633.9.

on the $(5\,h^{-1}\,\mathrm{Mpc})^3$ volume could have mass overdensities reaching $\delta_m \approx 10$. Such high overdensity peaks mark regions of the most overdense environments at high redshifts (e.g., Steidel et al. 2005; Chiang et al. 2014).

6.2.2. J084259.37+365704.3, $z_{\rm abs} = 3.17$

In the upper panel of Figure 25 we present the SDSS-III/BOSS spectrum of CoSLA candidate J084259.37+365704.3 at $z_{\rm abs} = 3.17$. This CoSLA candidate is centered at 5076 Å and has an effective optical depth of $\tau_{\rm eff} = 2.42^{+0.31}_{-0.26}$, which is a factor of 5.4× the mean optical depth at $z = 3.17$ (Dall'Aglio et al. 2008; Faucher-Giguère et al. 2008; Bolton et al. 2009). The middle panel presents the effective optical depth on $15\,h^{-1}$ Mpc ($\tau_{\rm eff}^{15\,h^{-1}\,\rm Mpc}$). The lower panel displays the MMT observations on this target with 3 × 30-minute exposures using 1200 lines mm^{-1} grating. The black spectrum indicates the Lyα transition, overplotted with the corresponding Lyβ transition in blue. Over a $15\,h^{-1}$ scale, this absorption has a rest-frame EW ratio between Lyβ to Lyα $\left(\frac{\rm EW_{Ly\beta}}{\rm EW_{Ly\alpha}}\right) = 0.83^{+0.03}_{-0.03}$. Based on the EW ratio, this strong Lyα absorption is not due to DLAs, but is likely to arise from the superposition of Lyα forest with EW ratio between Lyβ to Lyα $\gtrsim 0.8$. Based on Figure 17, our fits suggest that this system consists of multiple H I absorbers with column densities $N_{\rm H\,I} = 10^{15}$–$10^{18.5}$ cm^{-2}.

In Figure 26 we show that this CoSLA candidate is associated with two sight lines with a projected separation of $20\,h^{-1}$ co-Mpc at $z = 3.17$ on the sky. Strong IGM absorption is present in both sight lines: J084328.73+364107.4 and J084233.26+365129.9, with the transverse separations of 10 and $20\,h^{-1}$ Mpc from this CoSLA candidate, respectively. The two strong Lyα absorptions have $\tau_{\rm eff}^{15\,h^{-1}\rm Mpc} \geqslant 1.2$ at $z = 3.17$, $\gtrsim$3× the mean optical depth at $z = 3.1$.

Assuming that the $\tau_{\rm eff}$–M relation at $z = 3.1$ is similar to that at $z = 2.5$ (Figure 2), this absorption system corresponds to a system having $\tau_{\rm eff} = 1.15$ at $z = 2.5$. The median mass overdensity traced by CoSLAs with $\tau_{\rm eff} = 4.6\times$ mean optical depth is 1.4.

6.2.3. J081103.27+281621.0, $z_{\rm abs} = 2.97$

Figure 27 presents the spectra of CoSLA candidate J081103.27+281621.0 at $z = 2.97$. The yellow shaded area of the upper panel presents the BOSS spectrum of this CoSLA candidate centered at 4830 Å. From the BOSS data, the effective optical depth of this system $\tau_{\rm eff} = 1.86^{+0.12}_{-0.10}$, a factor of 4.6× higher than the mean optical depth at $z = 2.97$. The middle panel presents the $\tau_{\rm eff}^{15\,h^{-1}\,\rm Mpc}$ of the absorption system. The red dotted line represents the noise of the $\tau_{\rm eff}^{15\,h^{-1}\,\rm Mpc}$. The lower panel presents the follow-up MMT observations on this target with a 3 × 30-minute exposure using 1200 lines mm^{-1} grating. The EW ratio between Lyβ and Lyα is $0.61^{+0.02}_{-0.02}$. Based on the EW comparison between the Lyα (black) and Lyβ absorption (blue line), this system consists of the blending of Lyα absorbers with a column density $N_{\rm H\,I} = 10^{15}$–$10^{18.5}$ cm^{-2}.

The minimum transmitted fluxes of Lyα and Lyβ both reach almost zero. Figure 24 suggests that at a resolution of 100 km s^{-1}, an LLS with $N_{\rm H\,I} < 10^{19}$ cm^{-2} and typical Doppler parameter does not reach the zero level (Pieri et al. 2014). Combined with the EW ratio of Lyβ to Lyα, the most likely scenario is that this CoSLA candidate consists of a blending of Lyα forest absorbers with $N_{\rm H\,I} = 10^{15}$–$10^{18.5}$ cm^{-2}.

When we assume that the $\tau_{\rm eff}$–M relation at $z = 3.0$ is similar to that at $z = 2.5$ (Figure 2), this system corresponds to an absorption with $\tau_{\rm eff} = 1.15$ at $z = 2.5$. Figure 2 suggests that

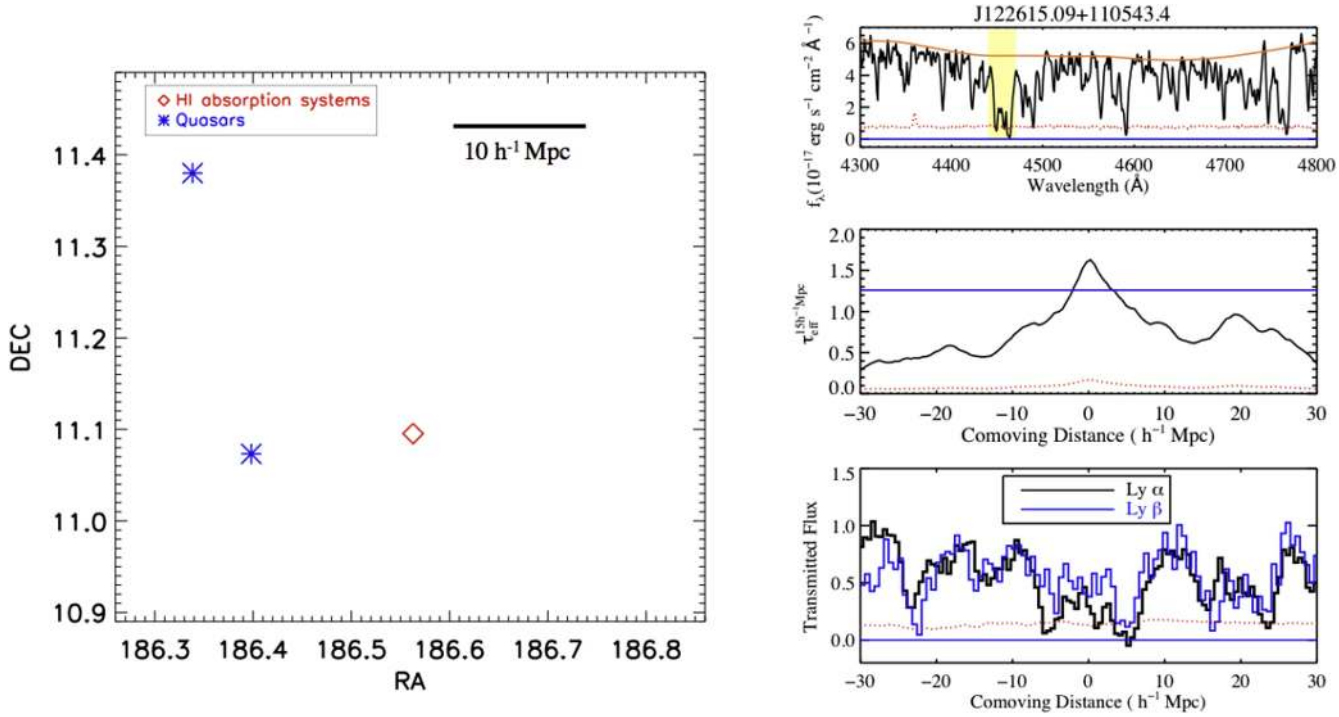

Figure 29. Left: the projected two-dimensional diagram of the CoSLA candidate J1 at $z = 2.67$ (red diamonds) and QSOs at $z = 2.66$–2.68 (blue asterisks) within a $25\,h^{-1}$ Mpc scale. Right: same format as Figure 20, showing strong Lyα absorption of the CoSLA candidate J122615.09+110543.4.

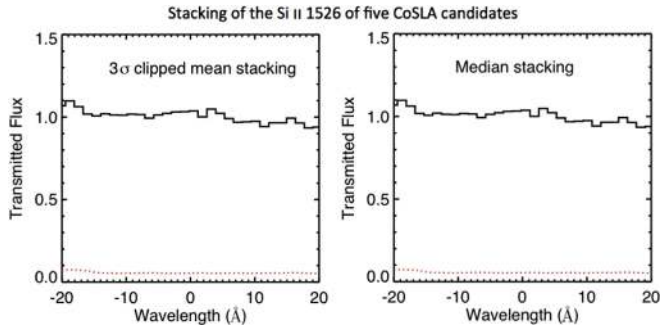

Figure 30. The stacking of the Si IIλ1526 of these five CoSLA candidates, using 3σ clipped mean (left panel) and median stacking (right panel). The observed 2σ $EW_{Si\ II} = 0.24$ Å corresponds to the rest-frame Si II EW of $EW_{Si\ II} = 0.06$ Å.

the overdensities traced by systems with $\tau_{\rm eff} = 1.15$ have a median mass overdensity of 1.4. We note that the overdensity is estimated under the assumption that no LLSs are included in the CoSLA candidate. The presence of LLSs should bias the overdensity to a lower level.

6.2.4. J113647.76+192633.9, $z_{\rm abs} = 3.03$

Figure 28 presents the CoSLA candidate J113647.76 +192633.9 at $z = 3.03$, which is associated with a rare quasar group at similar redshifts of $z = 3.02 \pm 0.02$. The upper panel presents the BOSS spectrum, which contains a CoSLA candidate on $15\,h^{-1}$ Mpc centered at 4912 Å (yellow shaded area). There is also another strong Lyα absorption system centered at 4885 Å. At 4912 Å, judging from the BOSS data, the effective optical depth of this system is $\tau_{\rm eff} = 1.84^{+0.16}_{-0.13}$, a factor of 4.6× higher than the mean optical depth at $z = 3.0$. The lower panel presents the follow-up MMT observations on this target with a 2×30 minutes exposure using a grating of 800 lines mm^{-1} ($R = 2000$).

The observed EW ratio between the Lyβ to Lyα is $0.96^{+0.06}_{-0.06}$. In addition, Lyα and Lyβ both reach transmitted fluxes below 0.1. As with the previous discussions, this system is highly likely to be consisting of the superposition of Lyα forest lines with H I column density $N_{\rm H\,I} = 10^{15}$–$10^{18.5}$ cm^{-2}.

This absorption system is associated with a rare quasar group at the same redshifts: J113630.91+194337.6 at $z = 3.04$ with a transverse separation of $2\,h^{-1}$ Mpc, J113653.23+192346.3 at $z = 2.99$ with a transverse distance $6\,h^{-1}$ Mpc from the absorption system, and J113602.86+193557.8 at $z = 3.01$, with a transverse separation of $16\,h^{-1}$ Mpc. This absorption system, together with this rare quasar group at $z = 3$, traces a massive overdense region.

J113647.76+192633.9 corresponds to a similar system at $z = 2.5$ with $\tau_{\rm eff} = 5.6 \times \langle \tau_{\rm eff}(z = 2.5) \rangle = 1.40$. Figure 2 shows that such absorption systems trace structures over a large scale of $15\,h^{-1}$ Mpc.

6.2.5. J122615.09+110543.4, $z_{\rm abs} = 2.67$

Figure 29 presents the spectra of the strong IGM absorption system J122615.09+110543.4 at $z = 2.67$. This absorption system is associated with two quasars at the same redshift with a transverse separation of $30\,h^{-1}$ Mpc. The upper panel presents the BOSS spectrum. Based on this figure, the system satisfies the criteria of IGM absorption over $15\,h^{-1}$ Mpc centered at 4454 Å.

The middle panel presents $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}}$ and the lower panel shows the MMT observations on this absorption with 2×30-minute exposures using a grating of 800 lines mm^{-1}. This absorption has an effective optical depth $\tau_{\rm eff} = 1.60^{+0.15}_{-0.14}$, a factor of 5.4 ×higher than the mean optical depth at $z = 2.66$. The EW ratio between Lyβ and Lyα is $0.69^{+0.04}_{-0.04}$. Figure 17

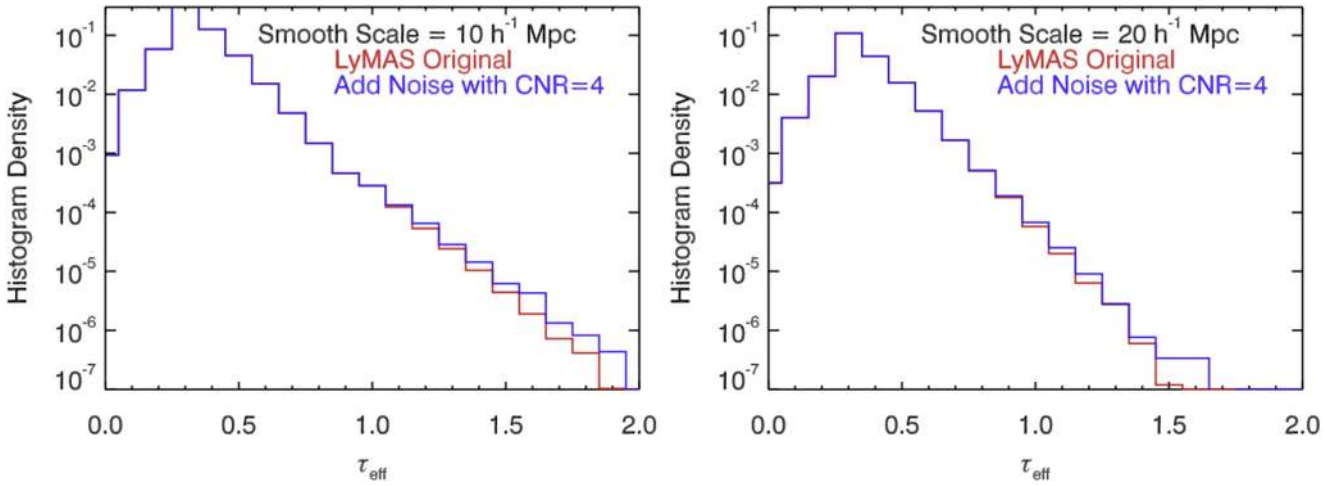

Figure 31. Similar plot to Figure 5. Left: the distribution of $\tau_{\rm eff}$ on the scale of $10\,h^{-1}$ Mpc. Right: the distribution of $\tau_{\rm eff}$ on $20\,h^{-1}$ Mpc.

shows that this system consists of the superposition of the Lyα forest with a column density $N_{\rm H\,I} = 10^{15}$–$10^{18.5}$ cm^{-2}.

This Lyα absorption is associated with two quasars at the same redshifts as the CoSLA candidate. The two quasars are J122535.57+110423.9 at $z = 2.68$ and J122521.12+112248.4 at $z = 2.67$, which have transverse separations of $11\,h^{-1}$ Mpc and $23\,h^{-1}$ Mpc, respectively, from the CoSLA candidate J122615.09+110543.4. The CoSLA candidate J122615.09 +110543.4 together with the two quasars at $z = 2.67 \pm 0.01$ could trace a large-scale structure.

We did not detect a low-ionization metal line in these five CoSLAs from SDSS-III/BOSS data. We conducted a stacking of the Si IIλ1526 of these five CoSLA candidates, using both 3σ clipped mean and median stacking (Figure 30). We did not detect the stacked Si II λ1526 absorption, giving a stringent upper limit of the observed 2-σ $\mathrm{EW}_{\rm Si\,II}$ of 0.24 Å, corresponding to the rest-frame 2σ upper limit of $\mathrm{EW}_{\rm Si\,II} = 0.06$ Å. These stacking results further support that the absorption of the CoSLA candidates is due to the IGM overdensity.

We have carried out narrowband (NB403) and broadband (Bw) imaging on multiple fields traced by CoSLAs using the KPNO-4 m Mayall and LBT/LBC. We have used LBT/MODS to spectroscopically confirm one of these overdense fields. We will present these observational results in the next paper of this series (Cai et al. 2016).

7. SUMMARY

Local galaxy clusters are identified by overdensities of galaxies, dark matter, and hot ICM (e.g., Fabian et al. 2006). Galaxy kinematics and gravitational lensing studies show that clusters of galaxies are embedded in a massive halo of dark matter with mass $\gtrsim 10^{14}\,M_{\odot}$ (e.g., Carlberg et al. 1997). The progenitors of such galaxy clusters can be identified at high redshifts at $z > 2$, given large-scale ($\sim$10–40 h^{-1} Mpc) density contrasts compared to random fields (e.g., Hu et al. 1996; Steidel et al. 1998; Matsuda et al. 2005, 2010; Ouchi et al. 2005; Venemans et al. 2007). At high redshift, galaxies are believed to interact with the surrounding IGM, and galaxies assemble their gas from the intergalactic H I gas (Adelberger et al. 2005; Frye et al. 2008; Matsuda et al. 2010; Rudie et al. 2012; Tejos et al. 2014). A large-scale mass overdensity is associated with an H I reservoir in the IGM, which in turn traces a large galaxy overdensity. Using the Lyα tomography, Lee et al. (2014, 2015) and Stark et al. (2015) further reported that protoclusters can be recovered from the large-scale flux decrements on a $10\,h^{-1}$ Mpc scale. The average separation of the sight lines is about a 3 Mpc scale.

In this paper, we systematically studied the correlation between mass overdensities and Lyα absorption on scales of 10–30 h^{-1} Mpc, which scales correspond to typical extents of the large-scale galaxy overdensities at $z > 2$ (e.g., Steidel et al. 1998, 2005; Matsuda et al. 2005; Ouchi et al. 2008). Using cosmological simulations, we showed that the most massive overdensity can be traced by a single strong absorption system, i.e., CoSLA. This technique allows the coverage of a significantly larger survey volume compared with current galaxy redshift surveys. We provide our summary below.

(a) Our cosmological simulations suggest that a strong correlation exists between the mass and Lyα transmitted flux over large scales, and this correlation peaks at scales of 15–30 h^{-1} Mpc (see Section 3). This strong correlation suggests that Lyα absorptions could be used to probe high-z overdensities. Using the SDSS-III/BOSS quasar data set, we have confirmed that the groups of Lyα absorption systems exist in two well-studied overdense fields: SSA22 and the Jackpot nebula field.

(b) We focused on the study of Coherently Strong Lyα Absorption systems (CoSLAs). These CoSLAs have the highest $\tau_{\rm eff}$ on $\sim$15 h^{-1} Mpc scales. These absorptions are due to intergalactic H I overdensity rather than high column density absorbers (e.g., DLAs and sub-DLAs), which are mainly due to the ISM/CGM of galaxies (Section 3).

(c) Using the cosmological simulation on the $(1\,h^{-1}\,\mathrm{Gpc})^3$ volume, we defined CoSLAs to have effective optical depths $\geqslant 4.5\times$ the mean optical depth on $15\,h^{-1}$ Mpc, corresponding to systems beyond 4σ in the optical depth distribution. Of these absorption systems, 52% trace structures with mass overdensities of $\delta_m > 1.6$, $>3.3\sigma$ beyond the density fluctuation in random fields (Figures 5–32). In particular, suggested by simulations, the CoSLAs have an upper limit of $\tau_{\rm eff} \approx 1.6$ at $z \approx 2.5$, $\approx 7 \times \langle \tau_{\rm eff} \rangle$ on the scale of $15\,h^{-1}$ Mpc. In a $1\,h^{-1}$ Gpc $\times$ $1\,h^{-1}$ Gpc $\times$ $1\,h^{-1}$ Gpc volume, any absorber with a $\tau_{\rm eff}$ higher than this upper limit is highly likely to be associated with HCDs.

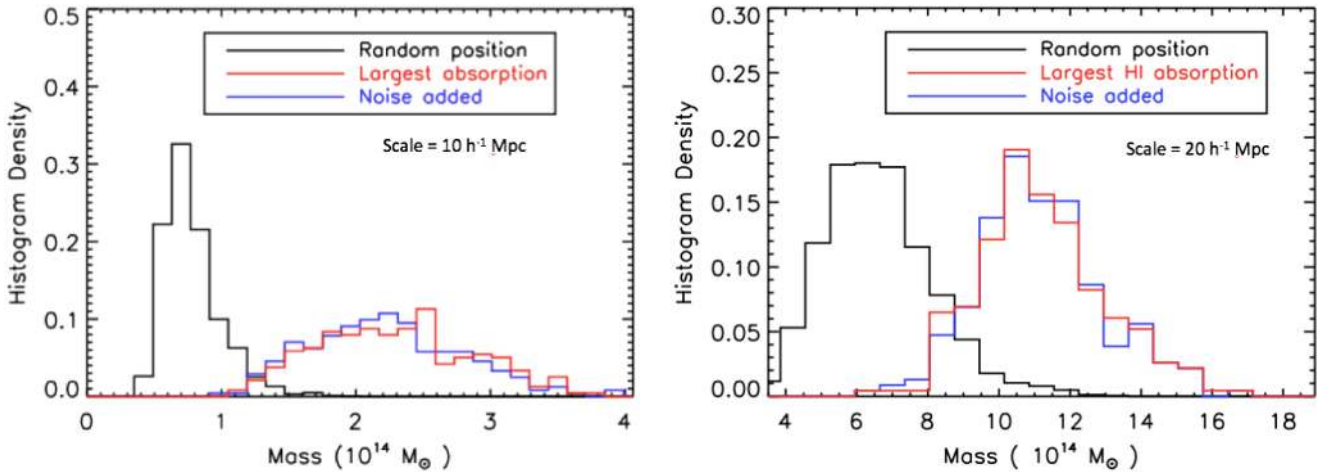

Figure 32. The distribution of mass traced by CoSLAs on 10 h^{-1} Mpc (left) and 20 h^{-1} Mpc (right) scale in LyMAS simulation. In both panels, the black histogram represents mass centered on random positions. Red represents CoSLAs selected from the original mock spectra (no noise added). Blue shows mass traced by CoSLAs selected from noise-added mock spectra, with a CNR of 4 per pixel. The figure shows that the CoSLAs effectively trace a three-dimensional large-scale structure. On 10 h^{-1} Mpc (left), most of the overdensities traced by CoSLAs contain mass a factor of 3.6× cosmic mean, representing 4σ mass overdensities. On 20 h^{-1} Mpc, CoSLAs trace mass a factor of 1.7× cosmic mean, representing 3σ mass overdensities.

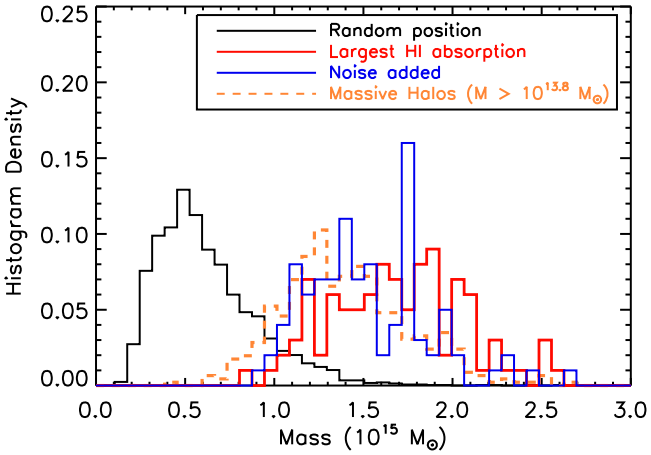

Figure 33. Mass distribution within 20 h^{-1} Mpc scale from the deterministic simulation with (1.5 h^{-1} Gpc)^3. x-axis is the mass within the 20 h^{-1} Mpc cube. The y-axis is the number of the cubes. The black histogram presents random distribution. Yellow shows the mass distribution centered on the most massive halos ($M_{\rm halo} > 10^{13.8}\,M_{\odot}$). Red is the mass traced by the CoSLAs on 20 h^{-1} Mpc scale, selected from the original mock spectra without noise. Blue is largest Lyα absorption selected from the noise-added mock spectra. The noise is added according to a CNR of 4.

(d) Guided by our simulation, we developed techniques to select these CoSLAs from the SDSS-III/BOSS spectra. Using the absorption troughs, absorption wings, corresponding metal lines, and the corresponding Lyβ absorptions, we can effectively select the CoSLAs and rule out contaminant DLAs and sub-DLAs. These selection criteria work best for absorbers with $z \geqslant 2.65$, where corresponding Lyβ absorption is covered with BOSS spectra (Section 4).

We also propose that with the selection of the Lyα absorption groups, one can effectively pinpoint massive galaxy protoclusters (Figure 18) without examining the nature of Lyβ absorption. At $z \lesssim 2.35$, ≈30% of the BOSS area have an average quasar density high enough to use absorption groups (see Section 4).

(e) Based on the selection criteria we proposed in Section 4, we selected a sample of CoSLA candidates from SDSS-III/BOSS by examining the absorption spectra of ≈6000 sight lines provided by the SDSS-III quasar survey at $z =$ 2.6–3.3 with a CNR $\gtrsim 5$ (see Section 6). We found a sample of five CoSLA candidates and ruled out one contaminant sub-DLAs (see Section 6 and appendix). These CoSLA condidates are consistent with the predictions in LyMAS simulation and are expect to pinpoint massive overdensities over ∼15 h^{-1} Mpc. Some of them are associated with strong Lyα absorptions and/or QSOs at the same redshifts in the nearby sight lines (Section 6). We expect that our technique can also be applied to future-generation spectroscopic campaigns, such as the DESI (e.g., Flaugher et al. 2014) and Subaru PFS surveys (e.g., Sugai et al. 2012).

The authors acknowledge the anonymous referee for the insightful comments that have signicantly improved the paper. Z.C., X.F., and I.M. acknowledge the support from the US NSF grant AST 11-07682. Z.C. and J.X.P. acknowledge support from NSF AST-1412981. Z.C. acknowlendges the insightful comments from Nobunari Kashikawa, Martin White, Linhua Jiang, Ann Zabludoff, Richard Green, Ran Wang, Brant Robertson, Masami Ouchi, Daniel Stark, and Yun-Hsin Huang. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, University of Cambridge, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah,

Table 7
Summary of the Confirmed DLAs that Do Not Have Damping Wings or Do Not Have Dark Troughs in their SDSS-III/BOSS Spectra

Name	z_{QSO}	z_{abs}	τ_{eff} ($\langle\tau_{eff}\rangle$)	T_{trough}	$w_{0.8}$ (Å)	$w_{0.5}/w_{0.8}$	SDSS CNR	MMT CNR	Exptime (minutes)	$N_{H\,I}$ (cm^{-2})
J010349.82+032856.1	2.99	2.64	7.4	0.13 ± 0.06	62	0.53	5	8	3 × 20	$10^{20.3}$
J021222.01+042745.3	2.52	2.29	13.3	0.22 ± 0.04	74	0.57	3	6	6 × 20	$10^{20.7}$
J081453.64+392828.6	2.61	2.21	16.7	0.02 ± 0.05	78	0.72	3	6	3 × 20	$10^{20.9}$
J091813.67+205623.7	2.67	2.41	18.0	0.04 ± 0.06	40	0.64	5	5	6 × 20	$10^{21.7}$
J104033.69+355247.9	2.37	2.22	18.0	0.02 ± 0.05	72	0.68	3	5	3 × 20	$10^{21.0}$
J131956.21+363624.1	2.55	2.24	7.3	0.11 ± 0.05	55	0.70	3	8	3 × 30	$10^{20.2}$
J145337.15+000410.0	3.04	2.49	7.5	0.12 ± 0.04	56	0.63	10	10	1 × 30	$10^{19.8}$
J143003.15+065719.1	3.35	2.66	9.3	0.24 ± 0.05	44	0.77	4	7	3 × 20	$10^{20.0}$
J154511.76+165630.5	2.82	2.44	9.5	0.06 ± 0.04	80	0.41	5	10	3 × 20	$10^{20.5}$

a

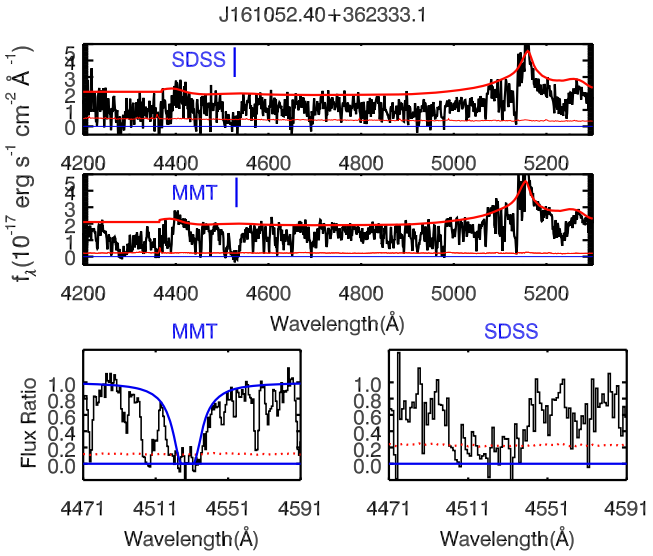

Figure 34. The contaminant in our MMT follow-up sample. The upper panel presents the SDSS-III spectrum. From the SDSS-III, the effective optical depth of this system $\tau_{eff} = 1.8^{+0.2}_{-0.2}$, 6× the mean optical depth at $z = 2.7$. The middle panel presents the spectrum using the MMT blue channel with 800 lines/mm ($R = 2000$). The lower left panel shows the zoom-in of Lyα absorption. The high S/N MMT observation clearly indicate that this absorption system is largely contributed by a sub-DLA with column density $N_{H\,I} = 10^{20.0}\,cm^{-2}$.

Vanderbilt University, University of Virginia, University of Washington, and Yale University.

APPENDIX

A.1. Mass Distribution at 10 and 20 h^{-1} Smoothing Scale

In this appendix, we present the mass traced by coherently strong Lyα absorption at 10 and 20 h^{-1} Mpc scales.

On 10 h^{-1} and 20 h^{-1} Mpc, mass overdensities can be effectively traced by coherently strong Lyα absorption. The left and right panels of Figure 32 (appendix) show mass distributions traced by strong Lyα absorption at scales of 10 and 20 h^{-1} Mpc in the LyMAS simulation, respectively. Again, we choose systems where the lowest transmitted flux is defined in the second paragraph of Section 2.2 (systems beyond 4.5σ in the τ_{eff} distribution). Similar to Figure 5, red represents mass traced by CoSLAs with the highest τ_{eff} selected from the original mock spectra (no noise added). Blue shows CoSLAs selected from noise-added mock spectra, with a CNR of 4 per pixel.

The deterministic simulation has a larger simulation box with 1.5 h^{-1} Gpc. In the deterministic scheme, Figure 33 presents similar results: IGM Lyα absorption systems most effectively trace the overdensities over a large scale. Using deterministic simulation, we study the mass overdensities traced by CoSLAs.

In Figure 33, black represents the cubes centered at the random positions in the simulation box: the cosmic mean mass in a 20 h^{-1} Mpc cube is $5 \times 10^{14}\,M_{\odot}$ with a standard deviation of $2.3 \times 10^{14}\,M_{\odot}$ on the logarithmic scale. Yellow represents the mass within 20 h^{-1} Mpc, which is centered on the most massive halo in a volume of $(1.5\,h^{-1}\,\mathrm{Gpc})^3$, with $M_{halo} > 10^{13.8}\,M_{\odot}$, 0.2 dex larger than the most massive halos in LyMAS simulation because of the higher box volume.

Red and blue lines present cases where masses are traced by the CoSLAs selected from $(1.5\,h^{-1}\,\mathrm{Gpc})^3$ box. Red presents mass overdensities traced by CoSLAs at 20 h^{-1} Mpc selected from the original mock spectra, without adding noise. More than half of the largest Lyα absorption traces the top 0.2% most massive overdensities (>3.0-σ) on 20 h^{-1} Mpc scale. Similar to red, blue presents mass distribution traced by CoSLAs selected from noise-added mock spectra with a CNR of 4 per pixel.

Therefore, both simulations support the hypothesis that extreme mass overdensities over ~10 h^{-1}–20 h^{-1} Mpc can be traced by the largest Lyα absorption systems.

A.2. Strong Absorption Systems

The cosmological simulations suggest that there is a low probability of finding CoSLAs with $\tau_{eff}^{15\,h^{-1}\mathrm{Mpc}} \geqslant 7 \times \langle\tau_{eff}\rangle$ in a $\leqslant$1 Gpc^3 survey volume (criterion b). Our survey suggests that systems with $\tau_{eff} > 7 \times \langle\tau_{eff}\rangle$ are most likely to be associated with HCDs rather than CoSLAs (see Figure 12). In Table 7 we present a few high optical depth absorptions that do not show DLA damping wings, dark trough, or low-ionization metal lines in the SDSS data. We have obtained MMT spectra for all of them. Most of the systems have $\tau_{eff} > 7 \times \langle\tau_{eff}\rangle$, greater than the strongest IGM absorption system in our cosmological simulation. Our follow-up MMT observations have confirmed that all of these absorbers contain DLAs.

A.3. Contaminants

We present a contaminant absorption of J161052.40+362333.1 in Table 5 and Figure 34. This absorption has an

optical depth $\tau_{\rm eff}^{15\,h^{-1}\,{\rm Mpc}} = 6.4 \times \langle \tau_{\rm eff} \rangle$. Its SDSS-III/BOSS spectrum satisfies our proposed selection criteria (a), (b), (c), and (d1). This absorption has a CNR of 5. Our follow-up spectra use the MMT blue channel with 800 lines/mm ($R = 2000$). We achieved a higher CNR of 8. The high S/N MMT observations clearly indicate that this absorption system is largely contributed by a sub-DLA with a column density of $N_{\rm H\,I} = 10^{20.0}\,{\rm cm}^{-2}$. This result is consistent with our simulations. From Table 3, our simulation indicates that 67% of CoSLA candidates with CNR = 4 spectra are DLAs, and 41% of the CoSLA candidates can be identified as the DLAs in the CNR = 8 spectra. In our MMT follow-up observations, this is the only CoSLA candidate selected from CNR = 5 spectra in the SDSS.

A.4. Requirements for Imaging Follow-up CoSLA Candidates

After identifying the CoSLA candidates, the next step is to confirm them with follow-up imaging observations. We use the star-forming galaxies to quantify the overdensities associated with the extreme IGM Lyα absorption systems. Here, we list a few observation requirements for the imaging follow-up. We introduce our imaging follow-up observational results in the next paper of this series.

Through the multicolor broadband imaging with U, G, and R, we can select LBGs at $z = 2$–3.5 (Steidel et al. 2004). Although the LBG selection technique is only sensitive to a wide redshift range ($2.0 < z < 2.7$, $z = 3.1 \pm 0.3$) and the overdensity could be smeared by the foreground and background galaxies, if the galaxy overdensity is 10 over a $15\,h^{-1}$ Mpc distance down to the depth of L^* (${\rm mag_{i\text{-}band}} = 25.0$), we can detect LBGs 1.5× denser than the cosmic mean. Another follow-up strategy is to select the Lyα emitting galaxies (LAEs) that are to be mapped and quantify the overdensity. Compared with the broadband selection, the narrowband-selected LAE candidates normally have a higher selection efficiency on mapping the structure. A narrowband selection usually produces a smaller survey volume, however.

To obtain sufficient galaxies to map the overdensities, the depth of the broadband imaging should at least reach L (UV) $\sim L^*$(UV)) at $z \sim 3$ (Reddy et al. 2008, Bian et al. 2013). In addition, the depth of the narrowband (NB) imaging should reach $NB \sim 25$ (L(Lyα) $\sim L^*$(Lyα)) at $z \sim 3$ (Ciardullo et al. 2012). Ideally, sufficiently wide field cameras with fields of view of $\gtrsim 30'$ are preferred for conducting the imaging observations to quantify the galaxy overdensity on $\gtrsim 20\,h^{-1}$ Mpc. After deep imaging, multislit spectroscopy (MOS) observations are needed to fully map and quantify the massive overdensities. We will introduce our imaging follow-up in the next paper of this series.

REFERENCES

Adelberger, K. L., Shapley, A. E., Steidel, C. C., et al. 2005, ApJ, 629, 636
Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, ApJ, 584, 45
Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2014, ApJS, 211, 17
Arrigoni Battaia, F., Hennawi, J. F., Cantalupo, S., & Prochaska, J. X. 2016, ApJ, 829, 3
Becker, G. D., Hewett, P. C., Worseck, G., & Prochaska, J. X. 2013, MNRAS, 430, 2067
Bi, H. 1993, ApJ, 405, 479
Bian, F., Fan, X., Jiang, L., et al. 2013, ApJ, 774, 28
Bird, S., Vogelsberger, M., Haehnelt, M., et al. 2014, MNRAS, 445, 2313
Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144
Bolton, J. S., Oh, S. P., & Furlanetto, S. R. 2009, MNRAS, 396, 2405
Busca, N. G., Delubac, T., Rich, J., et al. 2013, A&A, 552, A96
Cai, Z., Fan, X., Bian, F., et al. 2016, arXiv:1609.02913
Cai, Z., Fan, X., Noterdaeme, P., et al. 2014, ApJ, 793, 139
Cantalupo, S., Arrigoni-Battaia, F., Prochaska, J. X., Hennawi, J. F., & Madau, P. 2014, Natur, 506, 63
Carlberg, R. G., Yee, H. K. C., Ellingson, E., et al. 1997, ApJL, 485, L13
Cen, R., Miralda-Escudé, J., Ostriker, J. P., & Rauch, M. 1994, ApJL, 437, L9
Cen, R., Ostriker, J. P., Prochaska, J. X., & Wolfe, A. M. 2003, ApJ, 598, 741
Chapman, S. C., Scott, D., Windhorst, R. A., et al. 2004, ApJ, 606, 85
Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2013, ApJ, 779, 127
Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2014, ApJL, 782, L3
Ciardullo, R., Gronwall, C., Wolf, C., et al. 2012, ApJ, 744, 110
Cooke, J., Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2006, ApJ, 652, 994
Dall'Aglio, A., Wisotzki, L., & Worseck, G. 2008, A&A, 491, 465
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
Delubac, T., Bautista, J. E., Busca, N. G., et al. 2015, A&A, 574, A59
Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72
Fabian, A. C., Sanders, J. S., Taylor, G. B., et al. 2006, MNRAS, 366, 417
Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, ApJL, 682, L9
Flaugher, B., & Bebek, C. 2014, Proc. SPIE, 9147, 91470S
Font-Ribera, A., Arnau, E., Miralda-Escudé, J., et al. 2013, JCAP, 5, 018
Font-Ribera, A., Miralda-Escudé, J., Arnau, E., et al. 2012, JCAP, 11, 059
Ford, A. B., Oppenheimer, B. D., Davé, R., et al. 2013, MNRAS, 432, 89
Frye, B. L., Bowen, D. V., Hurley, M., et al. 2008, ApJL, 685, L5
Fumagalli, M., O'Meara, J. M., Prochaska, J. X., Rafelski, M., & Kanekar, N. 2015, MNRAS, 446, 3178
Gunn, J. E., & Peterson, B. A. 1965, ApJ, 142, 1633
Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332
Hennawi, J. F., Prochaska, J. X., Cantalupo, S., & Arrigoni-Battaia, F. 2015, Sci, 348, 779
Hu, E. M., McMahon, R. G., & Egami, E. 1996, ApJL, 459, L53
Kashikawa, N., Misawa, T., Minowa, Y., et al. 2014, ApJ, 780, 116
Khare, P., Kulkarni, V. P., Péroux, C., et al. 2007, A&A, 464, 487
Kirkman, D., Tytler, D., Suzuki, N., et al. 2005, MNRAS, 360, 1373
Kollmeier, J. A., Weinberg, D. H., Davé, R., & Katz, N. 2003, ApJ, 594, 75
Le Fèvre, O., Adami, C., Arnouts, S., et al. 2014, Msngr, 155, 33
Le Fevre, O., Tasca, L. A. M., Cassata, P., et al. 2014, arXiv:1403.3938
Lee, K.-G., Bailey, S., Bartsch, L. E., et al. 2013, AJ, 145, 69
Lee, K.-G., Hennawi, J. F., Spergel, D. N., et al. 2015, ApJ, 799, 196
Lee, K.-G., Hennawi, J. F., White, M., Croft, R. A. C., & Ozbek, M. 2014, ApJ, 788, 49
Lee, K.-G., Suzuki, N., & Spergel, D. N. 2012, AJ, 143, 51
Lee, K.-S., Dey, A., Hong, S., et al. 2014, ApJ, 796, 126
Lehner, N., Howk, J. C., Tripp, T. M., et al. 2013, ApJ, 770, 138
Liske, J., Webb, J. K., Williger, G. M., Fernández-Soto, A., & Carswell, R. F. 2000, MNRAS, 311, 657
Lochhaas, C., Weinberg, D. H., Peirani, S., et al. 2015, arXiv:1511.04454
Lynds, R. 1971, ApJL, 164, L73
Matsuda, Y., Richard, J., Smail, I., et al. 2010, MNRAS, 403, L54
Matsuda, Y., Yamada, T., Hayashino, T., et al. 2004, AJ, 128, 569
Matsuda, Y., Yamada, T., Hayashino, T., et al. 2005, ApJL, 634, L125
Matsuda, Y., Yamada, T., Hayashino, T., et al. 2011, MNRAS, 410, L13
McDonald, P., Miralda-Escudé, J., & Cen, R. 2002, ApJ, 580, 42
McDonald, P., Seljak, U., Cen, R., et al. 2005, ApJ, 635, 761
Miralda-Escudé, J., Cen, R., Ostriker, J. P., & Rauch, M. 1996, ApJ, 471, 582
Møller, P., Fynbo, J. P. U., Ledoux, C., & Nilsson, K. K. 2013, MNRAS, 430, 2680
Mukae, S., Ouchi, M., Kakiichi, K., et al. 2016, arXiv:1605.00379
Noterdaeme, P., Petitjean, P., Carithers, W. C., et al. 2012, A&A, 547, L1
Noterdaeme, P., Petitjean, P., Ledoux, C., & Srianand, R. 2009, A&A, 505, 1087
Noterdaeme, P., Petitjean, P., Pâris, I., et al. 2014, A&A, 566, AA24
Oppenheimer, B. D., Davé, R., Katz, N., Kollmeier, J. A., & Weinberg, D. H. 2012, MNRAS, 420, 829
Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2005, ApJL, 620, L1
Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, ApJS, 176, 301
Pâris, I., Petitjean, P., Aubourg, É., et al. 2014, A&A, 563, A54
Peirani, S., Weinberg, D. H., Colombi, S., et al. 2014, ApJ, 784, 11
Péroux, C., Meiring, J. D., Kulkarni, V. P., et al. 2008, MNRAS, 386, 2209
Pieri, M. M., Mortonson, M. J., Frank, S., et al. 2014, MNRAS, 441, 1718
Prochaska, J. X., Herbert-Fort, S., & Wolfe, A. M. 2005, ApJ, 635, 123
Prochaska, J. X., O'Meara, J. M., Fumagalli, M., Bernstein, R. A., & Burles, S. M. 2015, ApJS, 221, 2
Prochaska, J. X., O'Meara, J. M., & Worseck, G. 2010, ApJ, 718, 392

Rauch, M. 1998, ARA&A, 36, 267
Reddy, N. A., Steidel, C. C., Pettini, M., et al. 2008, ApJS, 175, 48
Ross, N. P., Myers, A. D., Sheldon, E. S., et al. 2012, ApJS, 199, 3
Rubin, K. H. R., Hennawi, J. F., Prochaska, J. X., et al. 2014, arXiv:1411.6016
Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012, ApJ, 750, 67
Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65
Slosar, A., Font-Ribera, A., Pieri, M. M., et al. 2011, JCAP, 9, 1
Slosar, A., Iršič, V., Kirkby, D., et al. 2013, JCAP, 4, 26
Springel, V. 2005, MNRAS, 364, 1105
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. 2003, ApJ, 592, 728
Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2005, ApJ, 626, 44
Steidel, C. C., Shapley, A. E., Pettini, M., et al. 2004, ApJ, 604, 534
Sugai, H., Karoji, H., Takato, N., et al. 2012, Proc. SPIE, 8446, 84460Y
Tamura, Y., Kohno, K., Nakanishi, K., et al. 2009, Natur, 459, 61
Tejos, N., Morris, S. L., Finn, C. W., et al. 2014, MNRAS, 437, 2017
Tepper-García, T., Richter, P., Schaye, J., et al. 2012, MNRAS, 425, 1640
Theuns, T., Viel, M., Kay, S., et al. 2002, ApJL, 578, L5
Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878
Venemans, B. P., Röttgering, H. J. A., Miley, G. K., et al. 2007, A&A, 461, 823
Viel, M., Markovič, K., Baldi, M., & Weller, J. 2012, MNRAS, 421, 50
Viel, M., Schaye, J., & Booth, C. M. 2013, MNRAS, 429, 1734
White, M., Blanton, M., Bolton, A., et al. 2011, ApJ, 728, 126
White, M., Myers, A. D., Ross, N. P., et al. 2012, MNRAS, 424, 933
Worseck, G., Prochaska, J. X., O'Meara, J. M., et al. 2014, MNRAS, 445, 1745
Yang, Y., Zabludoff, A., Eisenstein, D., & Davé, R. 2010, ApJ, 719, 1654
Yang, Y., Zabludoff, A., Jahnke, K., & Davé, R. 2014, ApJ, 793, 114
Yang, Y., Zabludoff, A., Tremonti, C., Eisenstein, D., & Davé, R. 2009, ApJ, 693, 1579
York, D. G., Khare, P., Vanden Berk, D., et al. 2006, MNRAS, 367, 945