A Lyman-α protocluster at redshift 6.9

Weida Hu1,2✉, Junxian Wang1,2✉, Leopoldo Infante3,4,5, James E. Rhoads6, Zhen-Ya Zheng7, Huan Yang3, Sangeeta Malhotra6, L. Felipe Barrientos4, Chunya Jiang7, Jorge González-López3,5, Gonzalo Prieto4, Lucia A. Perez8, Pascale Hibon9, Gaspar Galaz4, Alicia Coughlin8, Santosh Harish8, Xu Kong1,2, Wenyong Kang1,2, Ali Ahmad Khostovan6, John Pharao8, Francisco Valdes10, Isak Wold8, Alistair R. Walker11 and XianZhong Zheng8,9

Protoclusters, the progenitors of the most massive structures in the Universe, have been identified at redshifts of up to 6.6 (refs. 1–4). Besides exploring early structure formation, searching for protoclusters at even higher redshifts is particularly useful to probe the reionization. Here we report the discovery of the protocluster LAGER-z7OD1 at a redshift of 6.93, when the Universe was only 770 million years old and could be experiencing rapid evolution of the neutral hydrogen fraction in the intergalactic medium5. The protocluster is identified by an overdensity of 6 times the average galaxy density, and with 21 narrowband selected Lyman-α galaxies, which among which have been spectroscopically confirmed. At redshifts similar to or above this record, smaller protogroups with fewer members have been reported3,4. LAGER-z7OD1 shows an elongated shape and consists of two subprotoclusters, which would have merged into one massive cluster with a present-day mass of $10^{13.7}$ solar masses. The local volume of the identified bubbles generated by its member galaxies is found to be comparable to the volume of the protocluster itself, indicating that we are witnessing the merging of the individual bubbles and that the intergalactic medium within the protocluster is almost fully ionized. LAGER-z7OD1 thus provides a unique natural laboratory to investigate the reionization process.

High-redshift Lyman-α (Lyα)-emitting galaxies (LAEs) are star-forming galaxies with strong Lyα lines, which can be effectively selected with narrowband imaging surveys11–13. Aiming to build a statistical sample of LAEs at redshift $z \approx 7$, we are carrying out a deep narrowband imaging survey, Lyman Alpha Galaxies in the Epoch of Reionization (LAGER), utilizing the Dark Energy Camera (DECam, with a field of view of ~3 deg²) on the Cerro Tololo Inter-American Observatory (CTIO) Blanco 4 m Telescope and a customized narrowband filter DECam-NB964. The central wavelength and full-width at half-maximum of the filter are ~9,642 Å and 92 Å, respectively (Fig. 1), corresponding to a redshift range of 6.89–6.97 and a line-of-sight (LOS) scale of 26 cMpc. See Methods for more details. In the LAGER Cosmic Evolution Survey (COSMOS) field, we obtained 47.25 h narrowband exposure reaching a 5σ detection limit of 25.2 magnitude and a Lyα sensitivity of $10^{42.65}$ erg s⁻¹. Combining the deep narrowband image with the ultradepth broadband images from the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP), we uniformly selected 49 $z \approx 7$ LAEs11, See Methods and refs. 11–13 (hereafter Z17 and H19, respectively) for more details about the LAE selection.

As narrowband imaging can constrain the redshift of LAEs to a very narrow range $\Delta z < 0.1$, corresponding to a LOS distance of 30–45 cMpc at $z = 6–8$, it is also a promising approach to search for overdense structures, for example, protoclusters, in the early Universe14. Figure 2b shows the spatial distribution (blue circles) and number density (blue contours) of 49 LAGER $z \approx 7$ LAEs in the whole COSMOS field as presented in H19. A high-number-density region (as marked by the black dashed rectangle) is clearly revealed, containing 14 uniformly selected LAEs in H19 (see Supplementary Table 1 for the catalogue). This overdense region (LAGER-z7OD1) has a scale of 26.4×12, and a three-dimensional volume of 6.6×30×26 cMpc³. We calculate the galaxy overdensity of LAGER-z7OD1 following $\delta_g = (n - \bar{n})/\bar{n}$, where $n$ and $\bar{n}$ are the average number densities of LAEs in the LAGER-z7OD1 and the COSMOS field, respectively. We obtain the galaxy overdensity of LAGER-z7OD1 to be $\delta_g = 5.11^{+1.06}_{-1.03}$, which indicates LAGER-z7OD1 is a heavily overdense region, compared with the average galaxy number density. See Methods for more details. Up to now, candidate LAEs have been selected in four LAGER fields, that is, COSMOS, Chandra Deep Field South (CDFS), Wide 12 h field (WIDE12) and Galaxy And Mass Assembly 15h-A field (GAMA15A) (Z17, H19 and I.W. et al., manuscript in preparation). Among them, COSMOS is the unique one showing clear overdense region(s).

The same field was also observed with another narrowband filter HSC-NB97312, the bandpass of which partially overlaps with that of DECam-NB964 (Fig. 1). A hint of overdensity around LAGER-z7OD1 is also visible among the HSC-NB973 selected LAEs12, but not as strong as that seen in DECam-NB964. We stacked the DECam-NB964 image and the HSC-NB973 image to improve the depth of the narrowband image and selected three more members of LAGER-z7OD1. We also plot in Fig. 2a four more members, which were selected as lower-grade candidates (compared with those presented in H19) but were later spectroscopically confirmed (see next paragraph). Note that these seven additional member galaxies are only used to illustrate the spatial profile of
LAGER-z7OD1 (but not to calculate the overdensity), as they were not selected in a uniform and unbiased approach. See Methods for details.

Spectroscopic observations have been conducted to confirm the member LAEs, measure their redshifts and remove potential contaminants that may show continuum or emission lines at blueward of 9,600 Å. Three members have been confirmed in a previous study14. We carried out new spectroscopic follow-ups using the Inamori Magellan Areal Camera and Spectrograph (IMACS) at the 6.5 m Magellan I Baade Telescope (6–8 February 2017 and 21–23 February 2018), and the Low Dispersion Survey Spectrograph 3 (LDSS3) at the 6.5 m Magellan II Clay Telescope (10–11 January and 29–31 December 2019). The average seeing during the observations was ~0.8″. We carefully reduced the observed data and ruled out foreground identifications for member galaxies. Details of the data reduction are presented in Methods and a dedicated spectroscopic paper in preparation (along with identifications of LAEs outside of LAGER-z7OD1 and in other fields). In total, we have obtained spectroscopic confirmations for 16 member LAEs (red solid symbols in Fig. 2a). Lyα lines were not detected in three additional members that were put on masks; however, we are unable to rule them out as their Lyα lines might incidentally overlap with sky lines, or their Lyα line width be too broad to be detected14 (>500 km s⁻¹). The two- and one-dimensional spectra of the confirmed LAEs are presented in Supplementary Fig. 1 and the redshift distribution in Fig. 1.

The scale (66×30×26 cMpc⁻³), overdensity (δ_g = 5.11±2.96) and LOS velocity dispersion of spectroscopic confirmations (~765 km s⁻¹) of our protocluster LAGER-z7OD1 are similar to those of the previously detected protoclusters at redshifts of 5.7–6.6 (refs. 15,16) and simulation predictions15,16. We estimate the total present-day cluster mass M_{protocluster} of LAGER-z7OD1 following the widely used formula15:

\[ M_{protocluster} = (1 + \delta_g) \bar{V} V, \]

where V is the volume of the protocluster, \( \bar{V} \) (3.88×10⁶ M_⊙ cMpc⁻³) is the mean matter density of the Universe and \( \delta_g \) the mean overdensity. \( \delta_g \) is related to the observed galaxy overdensity through \( 1 + b \delta_g = C(1 + \delta) \), where b is the bias parameter and C the correction factor for the redshift space distortion. For \( \delta = 5.11 \) at \( z = 7 \), we find \( C = 0.79 \) and \( \delta_m = 0.87 \). The present-day mass M_{protocluster} of LAGER-z7OD1 is estimated to be 3.70±0.38×10¹⁵ M_⊙, comparable to the mass of the nearby COMA cluster15 (~2×10¹⁵ M_⊙).

See Methods for details.

The three-dimensional distribution of the spectroscopic confirmations is shown in Fig. 3. LAGER-z7OD1 shows an elongated shape and consists of two subprotoclusters. The overdensities of the two subprotoclusters are 6.76±3.77 (left) and 9.34±4.21 (right), respectively, where the boundaries of the two substructures are defined as the blue squares in Fig. 2. If we treat the two substructures as isolated, their present-day masses are expected to be 1.39±0.28×10¹⁵ M_⊙ and 1.60±0.28×10¹⁵ M_⊙, respectively.

We further explore whether the protocluster LAGER-z7OD1 would collapse into a single cluster. Similar to previous studies15,16, we estimate the linear overdensity of LAGER-z7OD1 to be \( \delta_1 = 0.54 \) at \( z = 7 \) (equation 18 of ref. 16). As the growth of linear perturbation \( \delta_1 \) is proportional to \( t^{\epsilon_1} \), \( \delta_1 \) will be larger than the threshold \( \delta_1 > 1.69 \) at \( z = 2 \), where \( \delta_1 = 1.69 \) is the critical value of linear overdensity of a spherical perturbation at the time it collapses11. Thus, we expect LAGER-z7OD1 collapses into a cluster at lower redshift. The discovery of LAGER-z7OD1 indicates that the formation of such large-scale structure had already begun by redshift 7.0, making it an ideal laboratory for understanding galaxy formation and large-scale structure formation.

During the epoch of reionization (EoR), the hard ultraviolet photons that escaped from a galaxy could ionize the intergalactic medium (IGM) and generate a HII region. The HII regions could gradually grow and merge with adjacent ones into sufficiently large ionized bubbles10,11,22,23, which can reduce resonant scattering of Lyα, significantly increase the IGM transmission, and thus enhance the number density of galaxies. On the basis of the relation between the bubble size and Lyα luminosity in a simulation of reionization work11,12, we show the predicted bubbles in Fig. 3 and the bubble sizes in Supplementary Table 1. The summed volume of the ionized bubbles of all 21 LAEs is 6.38×10⁶ cMpc³, with the four most luminous ones (with \( L_{Ly\alpha} > 2×10^{40} \) erg s⁻¹, that is, LAE-1, -2, -3 and -15) contributing 60.3% of the total ionized volume. This total ionized volume is even slightly larger than the volume of LAGER-z7OD1 (5.15×10⁶ cMpc³). This demonstrates substantial overlaps between individual bubbles, indicating that the individual bubbles are in the act of merging into a one or two giant bubbles (Fig. 3). As a comparison, the total predicted volume of all the 49 uniformly selected LAEs in COSMOS field is 12.71×10⁶ cMpc³, corresponding to 11.1% of the total volume surveyed by DECam-NB964. See Methods for details. Such predicted giant bubbles are large enough to be resolved by future 21 cm programmes, for example, SKA1-Low with resolution of ~7.3 arcmin at \( z = 7.3 \) (ref. 23), corresponding to ~19 cMpc.

The merged bubble (with a predicted size of ≥30 cMpc) could significantly increase the IGM transmission, and thus enhance the Lyα visibility of member LAEs15. Note Z17 and H19 have revealed a bright-end excess in the Lyα luminosity function in the COSMOS field, also suggesting the existence of big ionized bubbles at \( z ≈ 7 \) that reduce the opacity of neutral IGM around the luminous LAEs. Meanwhile, if the Lyα transmission through the IGM has been significantly boosted in most LAEs in LAGER-z7OD1, it may lead to larger Lyα equivalent widths than seen in field LAEs5,28. It is yet unclear whether the intrinsic Lyα equivalent widths of LAEs in the protocluster. However, the expected larger Lyα equivalent widths is not seen, compared with the field LAEs in COSMOS (see Methods and Extended Data Fig. 1 for details), though the large uncertainties in the equivalent-width measurements and the small sample size prevent us from reaching a robust conclusion. One possibility is that high-redshift protoclusters are highly biased regions and might contain LAEs with physical properties deviating substantially from the field LAEs13,18. It is yet unclear whether the intrinsic Lyα escape (before IGM scattering) in clustered LAEs is the same as that in field LAEs.

Moreover, the expected excess of close companions due to potentially enhanced Lyα transmission, in or behind the large bubbles of the luminous LAEs (\( L_{Ly\alpha} > 2×10^{40} \) erg s⁻¹, that is, LAE-1, -2, -3 and -15), is not seen (Figs. 2a and 3). This is likely in part due to the possibility
that while the biased dark matter halos can increase the galaxy merger/interaction, and thus enhance the star formation in the overdense region, the feedback from the ultraviolet background may suppress the star formation in the nearby fainter galaxies.

The discovery of the protocluster LAGER-z7OD1 provides an excellent opportunity to probe the rise and merging of ionized bubbles around the midpoint of the EoR. Future deep and multi-tiband (Hubble Space Telescope, James Webb Space Telescope, Atacama Large Millimeter/submillimeter Array and so on) observations could reveal the detailed reionization processes, for example, through searching for undetected Lyα fainter galaxies partially responsible for the ionization budget, better constraining the Lyα line equivalent widths and the Lyα profiles, measuring the Lyα velocity offsets relative to their system redshifts, and mapping the star formations histories of the galaxies.

Methods
Throughout this study, we adopt the recent Planck cosmological parameters: Ωm = 0.3111, ΩΛ = 0.6889 and H0 = 67.66 km s−1 Mpc−1, where Ωm and ΩΛ are the densities of total matter and dark energy and H0 is the Hubble constant.

LAGER survey. LAEs are promising probes for characterizing the cosmic reionization. We are carrying out a large-area narrowband imaging survey, LAGER, to search for the LAEs at z ≈ 7.0. Using the DECam installed on the CTIO. We designed and procured a narrowband filter (DECam-NB964) for the LAGER survey, with a central wavelength of ~9,642 Å and full-width at half-maximum of ~92 Å to avoid the atmospheric absorption and strong OH emission lines. The filter DECam-NB964 was installed on the DECam system in December 2015. Owing to the large field of view (FoV) (~3 deg2) and red-sensitive camera, LAGER is one of the most efficient surveys in searching for LAEs in the EoR. See the filter design paper for more details. We adopt a ‘wedding cake’ observing strategy with two deep fields aiming to discover faint LAEs and several shallower fields aiming to discover numerous luminous LAEs. Up to now, candidate LAEs have been selected in four fields, that is, COSMOS, CDFS, WIDE12 and GAMA15A.

LAGER-z7OD1 member galaxies. Member galaxies from H19. The 14 LAEs from H19 were selected by the narrowband technique. This technique is widely used in the literature and has been proven effective at searching for LAEs. Briefly, the selection criteria in H19 include: (1) the signal-to-noise ratio (S/N) of the DECam-NB964 signal is larger than 5; (2) DECam-NB964 excess over the underlying broadband to ensure the rest frame equivalent width of Lyα is larger than 10 Å; (3) non-detection in bluer broadbands (we adopt the recently release HSC SSP ultradeep broadband images). We visually inspected each LAE selected in four fields, that is, COSMOS, CDFS, WIDE12 and GAMA15A.

Additional members. The same field was also observed with another narrowband filter HSC-NB973, the bandpass of which partially overlaps with that of
OVERDENSITY IN LAGER-7OD1

Letters Nature Astronomy

The resulting one-dimensional spectra (both IMACS and LDSS3) have a spectral resolution of ~6 Å. We carefully examine the two-dimensional spectra of all spectroscopic targets and identify 13 sources as LAEs based on single-line detections. Including the three previous confirmations48, we now have spectroscopic confirmations for 16 member LAEs in LAGER-7OD1 (red solid symbols in Fig. 2a). The spectra of these 16 LAEs are presented in Supplementary Fig. 1.

Single-line detections (no other lines, no continuum) might still be confirmed by foreground strong emission line galaxies, for example, Hα, [OIII] and [OII]. Due to the limited spectral quality (and partial overlap with sky lines for some of them), we are unable to secure the characteristic asymmetric line profile26 (with a red wing) of high-z Ly lines for many sources. Meanwhile, while some lines are too narrow to be [OII] doublet, for some broader ones [OII] cannot be completely ruled out based on the line profile alone11. However, the contamination rate is expected to be low. For example, a recent spectroscopic survey of high-redshift LAEs at z = 5.7 reported a low contamination rate of <10% in their spectroscopic detections15. Even if we consider a contamination rate of 10%, our single-line identifications would be reliable for most sources. More critically and fortunately, in COSMOS, ultradeep broadband images are available to rule out almost all low-z interlopers of emission line galaxies, and we expect that the sample of H19 includes only ~0.14 ([OII]), 0.52 ([OIII]) and 0.16 (Hα) low-z emission line galaxies over the whole COSMOS field6.

LAGER-8, 12 and 14 were also spectroscopically observed, but not yet confirmed. The non-detections of the Ly in their spectra do not necessarily rule out these candidates. We observe the Poisson errors in the number densities of LAEs in DECam-NB964. Among the 8 HSC-NB973 selected LAEs in the area, four (LAE-1, -2, -11 and -15) were detected in DECam-NB964 and presented by H19. An additional source (LAE-20) shows tentative signal (S/N = 4.7) in the DECam-NB964 image (see also next paragraph), while the remaining 3 (LAE-22, -23 and -24; blue dashed circles in Fig. 2a; namely HSC-LAE-24, -6 and -16, respectively, in ref. 15) are invisible in DECam-NB964. The latter three had not been spectroscopically observed, and are candidate LAEs probably at slightly higher redshifts beyond the probe of NB964 (Fig. 1). At the current stage, we do not consider these 3 LAEs as member galaxies of LAGER-7OD1 as they may locate at slightly but sufficiently higher redshifts than that of the structure.

We further stack the DECam-NB964 and HSC-NB973 images to search for fainter LAEs located within the common volume sampled by two filters, and include four more LAE candidates (LAE-5, -6, -9 and -10), which were selected as LAE-5, -6, -9 and -10, DECam-NB964 (Fig. 1). At the current stage, we do not consider these 3 LAEs as member galaxies of LAGER-7OD1 as they may locate at slightly but sufficiently higher redshifts than that of the structure.

Oversized field measurements (see also next paragraph) suggest that the large-scale overdensity selected in DECam-NB964 is also present in LAGER-7OD1. A hint of overdensity around LAGER-7OD1 is also present in DECam-NB964 and LAGER-7OD1 is δ = 5.11 ± 0.08. The LAE sample suffers incompleteness during the detection and selection procedures (see H19 for details). However, as the narrow- and broadband images utilized for LAE detection and selection have rather uniform depths throughout the COSMOS field, the incompleteness is constant over the field, and thus, cancels out in the calculation of the overdensity. Note we use the mean overdensity in Supplementary Table 1 for the spectroscopically confirmed LAEs in LAGER-7OD1. To account for this effect, we distribute randomly, excluding such contaminants (even if possible) from the overdensity calculation.

As aforementioned, the COSMOS field is unique among four LAGER fields, showing clear overdense region(s). The average LAE number density in the COSMOS field could be biased by cosmic variance, and such effect may also affect the calculation of the overdensity. We compare the luminosity function of LAEs in the COSMOS field with those in other three LAGER fields and find the luminosity function to agree within 1σ. Present-day mass of LAGER-7OD1 is estimated to be 3.70 × 10^13 M_solar (where the errors are derived through Poisson errors). For δ = 5.1 at z = 7, we find C = 0.79 and δ_0 = 0.87. Thus, the present-day mass M_7OD1 of LAGER-7OD1 is estimated to be 3.70 × 10^13 M_solar. Decreasing the volume by 30% would yield a 17% lower M_7OD1 and the COSMOS field, thus even if a small fraction of the 49 candidates selected over the whole field are indeed contaminants, we would expect no more than one of them within the area of LAGER-7OD1 (assuming the contaminants randomly distribute over the field). Therefore, we opt to keep all three of them as valid candidates. We further note that even if we were able to secure all such contaminations over the whole field, excluding such contaminations from the calculation would yield even higher overdensity.

Bubbles and galaxy formation. Previous studies have presented a semi-numerical simulation to investigate the relation between the bubble size and Ly luminosity of high-redshift LAEs in the EoR. In the simulation, the star formation rate was assumed to be proportional to the growth rate of the dark matter halo, the escape fraction of the ionizing photons was assumed to be 0.2, and the Ly line emissivity was calculated based on the star formation history of the galaxy using the population synthesis code STARBUST9927. Finally, the evolution of ionized bubble and Ly luminosity (derived from the ionizing photons that do not escape) were obtained after calculating the radiative transfer in the IGM. On the basis of the relation between the bubble size and Ly luminosity (at z = 8, Fig. 15 in ref. 17), we show the predicted bubble size at z = 8 in Fig. 1 and predict a bubble size in Supplementary Table 1 for the spectroscopically confirmed LAEs in LAGER-7OD1.

Note that the derived bubble sizes are model dependent. Reference 27 has assumed a constant mean IGM density outside an H I bubble with a clumping factor C = 3 considered to account for the inhomogeneity of the IGM. Increasing the clumpiness C would not significantly decrease the bubble size28. Reference 13 has pointed out that the higher IGM density near the virial radius may reduce the predicted bubble sizes. The predicted bubble size is sensitive to the Lyman continuum escape fraction, which was assumed to be a constant of 0.2 in ref. 15. But note the observational results at z > 4 suggest that only a small fraction of galaxies
has a high escape fraction of > 10% (refs. 13–15), and the ionizing continuum escape fraction could be mass dependent16–18. Moreover, in the model of ref. 19, the Lyα escape fraction was assumed to be a constant of 0.6, and both the bubble size and Lyα luminosity tightly correlate with galaxy stellar mass. However, it is known that high-redshift LAEs on average are low mass and young galaxies, that is, the Lyα escape fraction is mass and stellar age dependent (for example, refs. 16–18). Consequently, our LAEs could be less massive and younger than the model predictions of ref. 19, and thus would be expected to have considerably smaller bubble sizes.

Nevertheless, considering the mean neutral hydrogen fraction at z = 7.0 (x_HI = 0.2–0.4, H19) and the significant overdensity of LAGER-zOD1, it is reasonable to believe that the IGM in LAGER-zOD1 was close to fully ionized at z ≈ 7.0. However, it is yet uncertain whether the member LAEs we detected alone could produce such a giant bubble, as their predicted bubble sizes are remarkably model dependent. In the cases aforementioned, more undetected Lyα fainter galaxies (with lower star formation rates and/or Lyα escape fraction) could have contributed to the reionization around LAGER-zOD1.

**Equivalent width/colour distribution.** The Lyα equivalent widths of narrowband selected LAEs can be well represented by the colour between the narrowband and the underlying broadband. For LAEs with spectroscopic redshifts, one could derive more precise Lyα equivalent-width measurements, after correcting for the wavelength dependence (non-boxcar shape) of the narrowband transmission and the redshift dependence of continuum contribution to narrowband photometry (see Fig. 2 of H19). For LAEs with particular field LAEs that have spectroscopic redshifts, here, we simply use the HSC–DECam-NB964 colour as an indicator of Lyα equivalent width and compare the colour distribution of LAEs inside the LAGER-zOD1 with those field LAEs (Extended Data Fig. 1). A Kolmogorov–Smirnov test shows no statistical difference between two samples.

### Data availability

The candidate selection is based on the following images in COSMOS field: DECam-NB964 (NORMLab Prop. ID: 2016A-0386, 2017B-0336; CTAC Prop. ID: 2016A-0610), HSC SSP programme and HSC-NB973 (Prop. ID: S16B-0011), which are available at https://archivel.nd.mri.nao.ac.jp/ and https://hsc-release.mtk.nao.ac.jp/doc/index.php/chorus, respectively. The spectroscopic datasets and the datasets generated or analysed during this study are available from the corresponding authors upon reasonable request. The LAE catalogue for LAGER-zOD1 used in this study is available in the Supplementary Information.

### Code availability

Codes used in this study are not publicly released yet but are available from the corresponding authors on reasonable request.

Received: 6 August 2020; Accepted: 9 December 2020; Published online: 25 January 2021

### References

1. Ouchi, M. et al. The discovery of primeval large-scale structures with forming clusters at redshift 6. *Astrophys. J.* **620**, L1–L4 (2005).

2. Wang, J. X., Malhotra, S. & Rhoads, J. E. An overdensity of Lyα emitters at redshift z = 5.7 near the Hubble Ultra Deep Field. *Astrophys. J.* **622**, L77–L80 (2005).

3. Malhotra, S. et al. An overdensity of galaxies at z = 5.9 ± 0.2 in the Hubble Ultra Deep Field confirmed using the ACS grism. *Astrophys. J.* **626**, 666–679 (2005).

4. Jiang, L. et al. A giant protocluster of galaxies at redshift 5.7. *Nat. Astron.* **2**, 962–966 (2018).

5. Harikane, Y. et al. SILVERRUSH. VIII. Spectroscopic identifications of early-large-scale structures with protoclusters over 200 Mpc at z = 6–7: strong associations of dusty star-forming galaxies. *Astrophys. J.* **883**, 142 (2019).

6. Calvi, R. et al. MOS spectroscopy of protocluster candidate galaxies at z = 6.5. *Mon. Not. R. Astron. Soc.* **489**, 3294–3306 (2019).

7. Robertson, B. E., Ellis, R. S., Furlanetto, S. R. & Dunlop, J. S. Cosmic reionization and early star-forming galaxies: a joint analysis of new constraints from Planck and the Hubble Space Telescope. *Astrophys. J.* **802**, L19 (2015).

8. Kulkarni, G. et al. Large Ly α opacity fluctuations and low CMB r in models of late reionization with large islands of neutral hydrogen extending to z = 5.5. *Mon. Not. R. Astron. Soc.* **485**, L24–L28 (2019).

9. Castellano, M. et al. Spectroscopic investigation of a reionized galaxy overdensity at z = 7. *Astrophys. J.* **863**, L3 (2018).

10. Tilvi, V. et al. Onset of cosmic reionization: evidence of an ionized bubble merely 680 Myr after the Big Bang. *Astrophys. J.* **891**, L10 (2020).

11. Zheng, Z.-Y. et al. First results from the Lyman Alpha Galaxies in the Epoch of Reionization (LAGER) survey: cosmological reionization at z = 7. *Astrophys. J.* **842**, L22 (2017).
43. Zheng, Z.-Y. et al. Design for the first narrowband filter for the Dark Energy Camera: optimizing the LAGER survey for $z \sim 7$ galaxies. *Publ. Astron. Soc. Pac.* 131, 074302 (2019).

44. Aihara, H. et al. First data release of the Hyper Suprime-Cam Subaru Strategic Program. *Publ. Astron. Soc. Jpn* 70, 58 (2018).

45. Dressler, A. et al. IMACS: The Inamori–Magellan Areal Camera and Spectrograph on Magellan-Baade. *Publ. Astron. Soc. Pac.* 123, 288–332 (2011).

46. Oemler, A., Clardy, K., Kelson, D., Walth, G. & Villanueva, E. COSMOS: Carnegie Observatories System for Multiobject Spectroscopy (ASCL, 2017); http://ascl.net/1705.001

47. Kashikawa, N. et al. The end of the reionization epoch probed by Ly$\alpha$ emitters at $z \sim 6.5$ in the Subaru Deep Field. *Astrophys. J.* 648, 7–22 (2006).

48. Ning, Y. et al. The Magellan M2FS spectroscopic survey of high-redshift galaxies: a sample of 260 Ly$\alpha$ emitters at redshift $z \approx 5.7$. *Astrophys. J.* 903, 4 (2020).

49. Steidel, C. C. et al. A large structure of galaxies at redshift $z \approx 3$ and its cosmological implications. *Astrophys. J.* 492, 428–438 (1998).

50. Ouchi, M. et al. Systematic Identification of LAEs for Visible Exploration and Reionization Research Using Subaru HSC (SILVERRUSH). I. Program strategy and clustering properties of ~2000 Ly$\alpha$ emitters at $z = 6–7$ over the 0.3–0.5 Gpc$^2$ survey area. *Publ. Astron. Soc. Jpn* 70, S13 (2018).

51. Leitherer, C. et al. StARBurst99: synthesis models for galaxies with active star formation. *Astrophys. J. Suppl. Ser.* 123, 3–40 (1999).

52. Cen, R. & Haiman, Z. Quasar Strömgren spheres before cosmological reionization. *Astrophys. J.* 542, L75–L78 (2000).

53. Grazian, A. et al. Lyman continuum escape fraction of faint galaxies at $z = 6–7$ over the 0.7–2.2 Gpc$^2$ survey area. *Mon. Not. R. Astron. Soc.* 725, 1011–1031 (2010).

54. Finkelstein, S. L. et al. Conditions for reionizing the Universe with a low galaxy ionizing photon escape fraction. *Astrophys. J.* 879, 36 (2019).

55. Ma, X. et al. No missing photons for reionization: moderate ionizing photon escape fractions from the FIRE-2 simulations. *Mon. Not. R. Astron. Soc.* 498, 2001–2017 (2020).

56. Finkelstein, S. L., Rhoads, J. E., Malhotra, S. & Groggin, N. Lyman alpha galaxies: primitive, dusty, or evolved? *Astrophys. J.* 691, 465–481 (2009).

57. Zheng, Z.-Y. et al. A physical model for the evolving ultraviolet luminosity function of high redshift galaxies and their contribution to the cosmic reionization. *Astrophys. J.* 785, 65 (2014).

Acknowledgements

We thank Z.-Y. Cai, Z. Cai and L. Jiang for informative discussions. The work is supported by the National Science Foundation of China (grant numbers 11421303, 11803093, 11773051 and 12022303), the CAS Frontier Science Key Research Program (QYZDJ-SSW-SLH066) and the CAS Pioneer Hundred Talents Program. US investigators on this work have been supported by the US National Science Foundation through NSF grant AST-1518057 and by NASA through WFIRST Science Investigation Team contract NNG16PJ33C. This project used the data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES), and public archival data from the DES. This research is based on observations at Cerro Tololo Inter-American Observatory at NSF's NOIRLab (NOIRLab Prop. ID: 2016A-0386, 2017B-0330, Principal Investigator: S. Malhotra; CNTAC Prop. ID: 2016A-0610, Principal Investigator: L. Infante), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. This work includes data collected at the 6.5 m Magellan Telescope located at Las Campanas Observatory, Chile. We thank the scientists and telescope operators at Magellan Telescope for their help. This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at http://dm.lsst.org. This work is based (in part) on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by the Subaru Telescope and Astronomy Data Center at the National Astronomical Observatory of Japan. Facilities: Magellan Baade (IMACS), Magellan Clay (LDSS3), Blanco (DECam) and Subaru (HSC).

Author contributions

W.H. and J.W. designed the layout of this paper. W.H. reduced the data, performed scientific analysis and wrote the manuscript. J.W. co-led the scientific interpretation and manuscript writing. L.I. led the observing proposals, which yielded new spectroscopic identifications presented in this work. W.H., L.I., H.Y., J.G.-L. and G.P. conducted these observations. All authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41550-020-01291-y. Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41550-020-01291-y. Correspondence and requests for materials should be addressed to W.H. or J.W. Peer review information *Nature Astronomy* thanks José Rodríguez Espinosa and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021, corrected publication 2021
Extended Data Fig. 1 | The HSC-γ – DECam-NB964 (and Lyα EW) distribution of LAEs in the LAGER COSMOS field. The LAEs inside the LAGER-z7OD1 are plotted in blue and those field LAEs in orange. For sources without detection in HSC-γ we simply adopt the 2σ lower limits to their HSC-γ magnitudes. Most sources with color > 2 in the plot are non-detected in HSC-γ. The vertical lines plot the median colors (1.84 and 1.86) respectively. The tick mark of Lyα EW is derived from the color assuming a redshift of 6.931 (corresponding to the center of NB964 transmission curve).