A [C\text{\textsc{ii}}] 158 \mu m emitter associated with an O\text{\textsc{i}} absorber at the end of the reionization epoch

Yunjing Wu\textsuperscript{1}, Zheng Cai\textsuperscript{1,2,3,4}, Marcel Neeleman\textsuperscript{4}, Kristian Finlator\textsuperscript{5,6}, Shiwu Zhang\textsuperscript{1}, J. Xavier Prochaska\textsuperscript{7,8}, Ran Wang\textsuperscript{9}, Bjorn H. C. Emonts\textsuperscript{10}, Xiaohui Fan\textsuperscript{11}, Laura C. Keating\textsuperscript{12}, Feige Wang\textsuperscript{11}, Jinyi Yang\textsuperscript{11}, Joseph F. Hennawi\textsuperscript{13} and Junxian Wang\textsuperscript{7,8,9}

The physical and chemical properties of the circumgalactic medium at z \geq 6 have been studied successfully through the absorption in the spectra of background quasi-stellar objects\textsuperscript{1-4}. One of the most crucial questions is to investigate the nature and location of the source galaxies that give rise to these early metal absorbers\textsuperscript{5-8}. Theoretical models suggest that momentum-driven outflows from typical star-forming galaxies can eject metals into the circumgalactic medium and the intergalactic medium at z = 5-6 (refs. 9,10). Deep, dedicated surveys have searched for Ly\text{\textsc{a}} emission associated with strong C\text{\textsc{iv}} absorbers at z \approx 6, but only a few Ly\text{\textsc{a}}-emitter candidates have been detected. Interpreting these detections is moreover ambiguous because Ly\text{\textsc{a}} is a resonant line\textsuperscript{11,12}, raising the need for complementary techniques for detecting absorbers’ host galaxies. Here we report a [C\text{\textsc{ii}}] 158 \mu m emitter detected using the Atacama Large Millimeter Array that is associated with a strong low-ionization absorber, O\text{\textsc{i}}, at z = 5.978. The projected impact parameter between O\text{\textsc{i}} and [C\text{\textsc{ii}}] emitter is 20.0 kpc. The measured [C\text{\textsc{ii}}] luminosity is 7.0 \times 10^7 solar luminosities. Further analysis indicates that strong O\text{\textsc{i}} absorbers may reside in the circumgalactic medium of massive halos one to two orders of magnitude more massive than expected values\textsuperscript{5,13,14}.

Metal absorption systems (for example, O\text{\textsc{i}}, C\text{\textsc{iv}} and Mg\text{\textsc{ii}}) are powerful probes of the enrichment of the high-redshift intergalactic medium (IGM)\textsuperscript{15,16}. Because the first excitation energies of O\text{\textsc{i}} \lambda 1302 and neutral hydrogen are almost identical, O\text{\textsc{i}} is considered one of the best indicators to trace the metal enrichment and the neutral IGM\textsuperscript{17,18}. Simulations suggest that at z = 5-6, feedback from star-forming galaxies can transport metals such as oxygen to 50 proper kpc (~9″) (ref. 9). Measurements of the impact parameters and the star formation rates (SFRs) of the source galaxies can directly constrain the efficiency of galactic winds in transporting metals and then test different feedback models\textsuperscript{19}. We select one of the strongest ultraviolet O\text{\textsc{i}} \lambda 1302 absorbers at z = 5.978 towards quasi-stellar object (QSO) J2054-0005 (at z = 6.04) as our preliminary target. The absorber has a rest-frame equivalent width (REW) of 0.12 Å (ref. 17), corresponding to a best-fit column density of 10^{14.2} cm\textsuperscript{-2} in the small optical depths regime (equation (8) in ref. 17).

The [C\text{\textsc{ii}}] moment-0 map shown in Fig. 1a is constructed by collapsing the emission line channels from the [C\text{\textsc{ii}}] data cube. At the redshift of the O\text{\textsc{i}} absorber, an emission line is detected at 4.3σ significance, which we interpret as [C\text{\textsc{ii}}] emission from a galaxy at the O\text{\textsc{i}} absorber’s redshift. Here we refer to this [C\text{\textsc{ii}}] emitter as [C\text{\textsc{ii}}]2054. To check the reliability of this detection, we searched for signals in the whole data cube first. No other sources have a signal either higher than 4σ level or lower than ~4σ (Extended Data Figs. 1 and 2). Note that there is a low probability of 10^{-4} for this line to be a CO interloper (Methods). Then, we show the spectrum of [C\text{\textsc{ii}}]2054 in Fig. 1d. The yellow shaded region represents the collapsed channels. The velocity-integrated [C\text{\textsc{ii}}] flux density is measured to be 0.0758 ± 0.0177 Jy km s\textsuperscript{-1}, corresponding to a [C\text{\textsc{ii}}] luminosity of 7.0 \times 10^7 L\textsubscript{\odot}, where L\textsubscript{\odot} is the solar luminosity. We further determined the reliability of this detection by checking the XX/YY polarization correlation and individual exposure maps (Extended Data Figs. 3 and 4 and details in Methods). Moreover, we applied our target selection criteria and algorithm in the ALMA (Atacama Large Millimeter Array) Spectroscopic Survey in the Hubble Ultra Deep Field (ASPECS) survey\textsuperscript{19} over 4.2 arcmin\textsuperscript{2} and found that the estimated probability of [C\text{\textsc{ii}}]2054 caused by noise fluctuation is ~0.6% (Extended Data Fig. 5 and details in Methods). Furthermore, from our moment-0 image (Fig. 1a), the projected impact parameter between [C\text{\textsc{ii}}]2054 and the O\text{\textsc{i}} absorber is 3.5″, corresponding to 20.0 proper kpc (pkpc) at z = 5.978. Throughout this paper we assume a Cold Dark Matter (CDM) cosmology with Ω\textsubscript{m} = 0.3, Ω\textsubscript{\Lambda} = 0.7, and H\textsubscript{0} = 70 km s\textsuperscript{-1} Mpc\textsuperscript{-1}.

Other than QSO J2054, in the far-infrared dust-continuum image (Fig. 1b), we detect another five continuum targets with signal-to-noise ratio (S/N) above 4. The archival Hubble Space Telescope (HST) broadband data were used to estimate their photometric redshifts. All of these galaxies are securely identified as foreground sources that do not associate with the O\text{\textsc{i}} absorber at z = 6 (Extended Data Figs. 6 and 7 in Methods).

The [C\text{\textsc{ii}}] 158 \mu m fine structure line emission is the dominant cooling line in the interstellar medium, and it is almost unaffected by dust attenuation\textsuperscript{20}. The ALMA allows us to use [C\text{\textsc{ii}}] to probe the physical conditions of the gas in galaxies and to quantify the SFR of the galaxy in the post-reionization epoch\textsuperscript{21}.
The velocity-integrated [C\textsc{ii}] luminosity of [C\textsc{ii}] 2054 is estimated to be $L_{\text{[C\textsc{ii}]}_r} = (7.0 \pm 1.7) \times 10^{10} L_\odot$ (ref. 24), corresponding to a [C\textsc{ii}]-based SFR (SFR\textsubscript{[C\textsc{ii}]}\textsubscript{r}) of $6.8 \pm 1.7 M_\odot$yr\textsuperscript{-1} (ref. 25), where $M_\odot$ is the solar mass. No submillimetre continuum is detected at the position of [C\textsc{ii}] 2054, yielding a 3$\sigma$ upper limit of 37$\mu$Jy. Converting this luminosity to a total infrared SFR (SFR\textsubscript{IR}) yields the upper limit of SFR\textsubscript{IR} $< 11 M_\odot$yr\textsuperscript{-1} (ref. 26). To estimate the halo mass directly from the [C\textsc{ii}] luminosity, we adopt a relation between $L_{\text{[C\textsc{ii}]}_r}$ and halo mass at $z \approx 6$ (ref. 13) by assuming that [C\textsc{ii}] 2054 resides in the galaxy main sequence. Then, the derived halo mass is $4.1 \times 10^{11} M_\odot$, with a 1$\sigma$ scattering of 0.5 dex. We also estimate the halo mass by converting the SFR to the stellar mass and then to the halo mass. The two methods of halo mass estimation yield consistent results within the 1$\sigma$ level (Methods).

To constrain models of IGM metal enrichment at the end of reionization, we compare our observed impact parameter against predictions from the Illustris, Sherwood, HVEL and FAST cosmological hydrodynamic simulations\textsuperscript{14,18–20}. As mentioned above, the O\textsc{i} absorber has an REW of 0.12 Å, and the measured projected impact parameter is 20.0 pkpc (Fig. 1c). Broadly, simulations predict that metal line absorbers with an REW of 0.12 ± 0.05 Å arise at impact parameters of 5–30 pkpc, bracketing our measurement. Specifically, in Fig. 2a we show the comparison of our data with the Illustris simulation\textsuperscript{14}. The comparisons with other simulations are shown in Extended Data Fig. 8 and in Methods. In Fig. 2a, the red star shows our observations, whereas other points are from the Illustris simulation. The simulation is able to predict our observed impact parameter, and has demonstrated that galaxies similar to [C\textsc{ii}] 2054 are able to enrich the IGM at distances of 20 pkpc. Note that 11.1% of the data points in the simulation have a larger impact parameter than that of [C\textsc{ii}] 2054. Thus, the projected distance between O\textsc{i} and [C\textsc{ii}] 2054 is larger than the median value expected from the hydrodynamical simulations.

For a direct comparison of the models, we need to derive the halo mass of [C\textsc{ii}] 2054 based on the SFR. As we mentioned above, [C\textsc{ii}] 2054 has a halo mass of $4.1 \times 10^{11} M_\odot$ with a 1$\sigma$ uncertainty of 0.5 dex. Different simulations, including Illustris, Sherwood, HVEL and FAST simulations, directly investigated the halo mass distribution for O\textsc{i} absorbers with column densities of $N_{\text{O}} > 10^{12}$ cm\textsuperscript{-2} at $z = 6$. They find that O\textsc{i} absorbers with REW in the range $0.12 \pm 0.05$ Å at $z = 6$ are typically embedded in $10^{10}$–$10^{11}$ $M_\odot$ halos\textsuperscript{6,14}. This value is one to two orders of magnitude smaller than that implied by our observations. Thus, our results suggest that low-mass systems in these simulations are likely to be overly efficient at producing metal absorbers.

A caveat to this result stems from the fact that the analysis of these simulations identified each metal absorber with its nearest halo. In reality, the nearest halo may not necessarily be the true source of an absorber because a halo that is slightly more distant but significantly more massive may contribute more of its metals.
Letters Nature Astronomy

We note that these two lines are normalized by the number of O \textsc{i} absorbers identified with [C \textsc{ii}]-based SFR. The solid and dashed curves are generated by accumulating catalogues of all galaxies that fall within 50 pkpc to 100 pkpc of O \textsc{i} absorbers. We assume the galaxy–absorber cross-correlation to be $\xi(r) = (r/r_0)^{-\gamma}$. This function expresses the fractional excess number density of galaxies located at a distance $r$ from an absorber in terms of the correlation length $r_0$ and a power-law slope $\gamma$. Symbols indicate cross-correlation functions predicted by Technicolor Dawn for galaxies clustered about synthetic O \textsc{i} absorbers with REW $\geq$ 0.12 Å. Error bars are added by assuming Poisson fluctuations in the number of simulated host galaxies. The solid, dashed and dotted lines show predictions of an average of one galaxy of SFR $\geq$ 7 M$_\odot$ yr$^{-1}$, 0.7 M$_\odot$ yr$^{-1}$, and 0.07 M$_\odot$ yr$^{-1}$, within 20.0 pkpc of an absorber at $z = 5.9$, respectively. Comparison with the symbols show that the predicted $r_0$ is too low to explain the observations.

This consideration indicates the need for an alternative framework for quantifying the galaxy–absorber relationship that can circumvent this ambiguity.

Recently, a set of Technicolor Dawn simulations have been developed for a detailed study of the metal enrichment of galaxies in the reionization epoch. This set of simulations incorporates realistic small-scale ultraviolet background fluctuations and reproduces a broad range of observations of reionization and early galaxy growth. They simultaneously reproduce observations of the galaxy stellar mass function, the Si \textsc{iv} column density distribution and the ultraviolet background amplitude at $z > 5$. Hence, the galaxy populations and the metallicity of their circumgalactic medium are realistic.

Our discovery of [C\textsc{ii}]2054 is not expected from the Technicolor Dawn simulations. In Fig. 2b, we show how the number of galaxies that cluster about strong O\textsc{i} absorbers at $z = 6$ is predicted to vary with SFR. The solid and dashed curves are biased SFR functions. They are generated by accumulating catalogues of all galaxies that fall within 50 pkpc to 100 pkpc of O\textsc{i} absorbers with REW $> 0.12$ Å ($N_{\text{O}} > 10^{44.4}$ cm$^{-2}$). The predicted luminosity of the O\textsc{i} associated [C\textsc{ii}] emitter is roughly two orders of magnitude fainter than that of [C\textsc{ii}]2054. $N_{\text{Gal}}$ is the number of galaxies within an impact parameter $d_{\text{max}}$ from strong O\textsc{i} absorbers in simulations.

**Fig. 2 | Comparison between simulations and observations.** a. The upper panel shows the relationship between projected impact parameter and halo mass of strong O\textsc{i} absorbers in the Illustris simulation. The O\textsc{i} absorbers have an REW of 0.12 ± 0.05 Å (consistent with observations). The red star represents [C\textsc{ii}]2054. The error bar shows the 1σ uncertainty of the estimated halo mass. The lower panel shows the halo mass distribution. The red arrow shows that the host halo mass of [C\textsc{ii}]2054 is one order of magnitude larger than the median value predicted by Illustris simulations. $d$ is the projected impact parameter between the metal-line absorber and its host galaxy. b. The number of galaxies that cluster about strong O\textsc{i} absorbers at $z = 6$ is predicted to vary with SFR. The solid and dashed curves are biased SFR functions. They are generated by accumulating catalogues of all galaxies that fall within 50 pkpc to 100 pkpc of O\textsc{i} absorbers with REW $> 0.12$ Å ($N_{\text{O}} > 10^{44.4}$ cm$^{-2}$). The predicted luminosity of the O\textsc{i} associated [C\textsc{ii}] emitter is roughly two orders of magnitude fainter than that of [C\textsc{ii}]2054. $N_{\text{Gal}}$ is the number of galaxies within an impact parameter $d_{\text{max}}$ from strong O\textsc{i} absorbers in simulations.
Table 1 | Photometric results of sources in the J2054 field. In our ALMA observations, for each source, 2σ contour areas are regarded as emission regions

| ALMA continuum image |  |  |  |
|---------------------|---|---|---|
| Source name | S_{\text{int}} (mJy) | flux_{\text{CII}} (mJy) | S/N |
| C1 | 0.64 | 0.06 | 10.7 |
| C2 | 0.12 | 0.02 | 6.0 |
| C3 | 0.19 | 0.02 | 15.8 |
| J2054 | 3.37 | 0.02 | 280 |
| C5 | 0.067 | 0.014 | 4.8 |
| C6 | 0.069 | 0.016 | 4.3 |

| [CII] moment-0 map |  |  |  |
|---------------------|---|---|---|
| Source name | flux_{\text{CII}} (Jy km s^{-1}) | flux_{\text{CII}} (Jy km s^{-1}) | S/N |
| [CII] J2054 | 0.0758 | 0.0177 | 4.3 |

| HST observations (AB magnitude) |  |  |  |
|---------------------|---|---|---|
| Source name | F606W(V) | F814W(H) | F105W(Y) | F125W(J) |
| C1-1 | 2σ ≥ 29.45 | 26.86 ± 0.06 | 26.33 ± 0.05 | 25.70 ± 0.02 | 25.09 ± 0.02 |
| C1-2 | 28.29 ± 0.19 | 27.18 ± 0.08 | 26.53 ± 0.06 | 25.64 ± 0.02 | 25.04 ± 0.02 |
| C1-3 | 26.54 ± 0.04 | 26.27 ± 0.04 | 25.98 ± 0.04 | 25.33 ± 0.02 | 25.42 ± 0.02 |
| C1-4 | 27.00 ± 0.06 | 26.09 ± 0.03 | 25.21 ± 0.02 | 24.71 ± 0.01 | 24.33 ± 0.01 |
| C1-5 | 26.68 ± 0.04 | 26.19 ± 0.03 | 25.37 ± 0.02 | 24.59 ± 0.01 | 24.14 ± 0.01 |
| C2 | 27.48 ± 0.09 | 26.92 ± 0.07 | 25.58 ± 0.02 | 25.17 ± 0.01 | 24.91 ± 0.01 |
| C3-1 | 24.59 ± 0.01 | 24.06 ± 0.01 | 23.70 ± 0.004 | 23.60 ± 0.003 | 23.45 ± 0.003 |
| C3-3 | 26.74 ± 0.05 | 26.51 ± 0.05 | 25.59 ± 0.02 | 25.12 ± 0.01 | 24.87 ± 0.01 |
| QSO J2054 | 26.54 ± 0.04 | 22.05 ± 0.001 | 21.059 ± 0.001 | 20.88 ± 0.001 | 20.70 ± 0.001 |
| C5-1 | 27.77 ± 0.12 | 27.15 ± 0.08 | 26.36 ± 0.05 | 26.18 ± 0.03 | 26.15 ± 0.04 |
| C5-2 | 26.60 ± 0.04 | 26.44 ± 0.04 | 26.05 ± 0.04 | 25.93 ± 0.02 | 25.64 ± 0.03 |
| C5-3 | 2σ ≥ 29.45 | 27.38 ± 0.10 | 27.29 ± 0.12 | 26.44 ± 0.04 | 25.90 ± 0.03 |
| C6 | 26.22 ± 0.03 | 25.64 ± 0.02 | 24.42 ± 0.01 | 23.91 ± 0.01 | 23.53 ± 0.01 |
| [CII] J2054 | 2σ ≥ 29.45 | 2σ ≥ 29.21 | 2σ ≥ 28.97 | 2σ ≥ 29.38 | 2σ ≥ 29.22 |

We use these emission regions as the photometric aperture for resolved sources, whereas for unresolved sources the peak surface brightness represents the total flux. For HST observations, we used a uniform aperture with a 0.6" diameter and photometry was applied in the PSF-matched multiband images (Methods). *Continuum flux density at 264 GHz.* HST resolves these sources as multiple targets. We number the resolved sources from 1 to 5. C2 is located in the gap of two charge-coupled device chips of HST and therefore does not have these magnitudes.

Although Fig. 2 appears to suggest that our observations can be accommodated within the Illustris physical model even though it is largely unexpected in Technicolor Dawn, we note that the latter simulation’s resolution is a factor of ~5 higher whereas its cosmological volume is relatively small to contain a representative population of bright galaxies (ultraviolet-based absolute magnitude of M_{UV} < -20), which makes the two simulations complementary. Given that faint galaxies with L = 0.01 L* (where L* represents the characteristic galaxy luminosity at this redshift) are one order of magnitude more prevalent than massive galaxies with L* and that they are able to host strong OI absorbers in both models, both models predict the observational identification of a massive host to be unlikely. We further quantify the discrepancy between simulations and observations by re-casting the observations as a constraint on the galaxy–absorber cross-correlation function. Based on the [CII] luminosity function at this redshift, the mean space density of galaxies at z = 5.9 with a SFR ≥ 6.8 M_{\odot} yr^{-1} is 1.4 × 10^{-3} cMpc^{-3} (where cMpc is comoving Mpc). A sphere of radius 20 pkpc containing one such galaxy has a probability of 2 × 10^{-3}, corresponding to a mean overdensity of Δ = 5 × 10^{4} (Methods). Our observations therefore confirm that bright galaxies and strong absorbers correlate strongly.

We quantify this correlation using the galaxy–absorber cross-correlation function ξ(r) = (r/r_0)^{-γ}, which expresses the fractional excess number density of galaxies located at a distance r from an absorber in terms of the correlation length r_0 and a power-law slope γ (see detailed calculations of r_0 and γ in Methods).

In Fig. 3, we use solid curves to show combinations of r_0 and γ that predict an average of one galaxy within 20.0 pkpc of an absorber at z = 5.9. Symbols indicate cross-correlation functions predicted directly by Technicolor Dawn for galaxies clustered around synthetic OI absorbers with REW ≥ 0.12 Å. As galaxies with SFR ≥ 6.8 M_{\odot} yr^{-1} are too rare to arise in the current simulation volume, we consider two lower SFR thresholds (Fig. 3, legend). In all cases, the predicted r_0 is ~10 times too low to explain the observation. Moreover, although r_0 increases with SFR threshold, the trend is not nearly strong enough to explain the discrepancy.

[CII] J2054 could be explained if its SFR were overestimated because faint galaxies are more abundant than bright ones. For example, it could be undergoing a temporary starburst. Alternatively, a strong intervening gravitational lens could boost its flux. To evaluate whether these possibilities would be sufficient to explain our observations, we use dashed and dotted curves to plot combinations of r_0 and γ that would explain [CII] J2054 if its true SFR were 10 and 100 times lower, respectively. Even in these cases, the simulation underpredicts r_0 by a factor of 10.

In summary, our ALMA detection of a [CII] emitter associated with a reionization-epoch OI absorber is more than one to two orders of magnitude higher in [CII] luminosity than expected.
from hydrodynamical simulations. Casting this observation as a constraint on the galaxy–absorber cross-correlation function suggests that the correlation length predicted by Technicolor Dawn is ~10 times too low. Our observations suggest that simulations may incorrectly identify faint galaxies as the principal repositories of observed low-ionization metal ions. This could indicate that the predicted environments of brighter galaxies are overly ionized or that simulated faint galaxies are too efficient at creating and/or expelling metals. Conversely, it remains possible that our ALMA observations lead us to overestimate the underlying SFR and halo mass of [CII]2054. Future surveys will be needed to test conclusively whether bright galaxies are more common around strong OI absorbers than expected from current theoretical models, as currently suggested by our data. Meanwhile, testing the strong prediction that many faint galaxies should exist in the vicinity of our OI absorber will require a larger survey using the James Webb Space Telescope (JWST). The three-hour JWST grism observations allow us to detect galaxies with $M_{\text{IC}} \leq -19.5$, corresponding to a halo mass of $10^{11} M_\odot$. Based on the survey depth, the mean galaxy number density and the galaxy–absorber cross-correlation function based on our pilot ALMA programme, we expect that JWST can uncover ~1 host galaxy per QSO field. If we observe the field of 10 existing OI absorbers at $z \geq 6$ with a total integration of 30 h, we will construct a sample larger by factor of 10 and ultimately test whether massive galaxies are commonly associated with OI absorbers.

**Methods**

**Motivation of this observational project.** The observational programme was inspired by the successful detection of two host galaxies of damped Ly\textalpha absorbers (DLAs) in 2017 (ref. 17). One of the galaxies has the unexpected SFR of $\sim 10^1 M_\odot \text{yr}^{-1}$ (ref. 18). This DLA is also associated with a strong OI absorber. The OI absorber chosen in this work has an equivalent width among the highest at $z > 5$ and is probably associated with a DLA or sub-DLA (ref. 19). These observations were then designed to constrain the galaxy counterparts associated with OI absorbers at $z \geq 6$. Based on the OI absorber survey, six of 17 sightlines contain strong absorbers that can be detected in the QSO spectra. There are 10 absorbers in the six sightlines. Furthermore, two of the 10 OI absorbers are strong enough to have the associated Si II $\lambda 1526/1304$ and C II $\lambda 1334$ absorptions, indicating that they could also be associated with DLAs or sub-DLAs located in the H I Gunn–Peterson trough. Assuming each QSO sightline can probe 500 Mpc, then this strong OI absorber is selected from a survey of 1 Gpc, indicating that the selected strong OI is rare.

**ALMA observations and data reduction.** There are a few advantages in using ALMA to probe the source galaxies associated with an OI absorber at $z \geq 6$. First, the sub-mm [CII] 158 \mu m line is one of the strongest far-infrared lines from galaxies (ref. 20). Second, the background QSO is less dominant in emission in the submillimetre regime. Third, at this redshift, the [CII] emission is shifted within the sub-mm [CII] 158 \mu m line. No other sources had either a signal higher than the 4$\sigma$ level. The S/Ns of [CII]2054 in the three individual exposures are 2.9, 2.3 and 1.8 respectively, close to the expected value of $\sqrt{4}$ Thus, none of the evidence supports that [CII]2054 is a false positive.

**Target selection criteria and source robustness.** According to our ALMA proposal, our detection limit is down to the typical $L_*$ at the ~4$\sigma$ level. The observed [CII]2054 mass of $10^7 M_\odot$ and a typical $L_*$ mass on the order of $10^7 M_\odot$ to $10^8 M_\odot$ is consistent with a galaxy at $z \geq 6$. The ALMA observations are consistent with the ALMA Band 6, which has a high throughput and is thus efficient for emission line detections. Finally, cosmological simulations suggest that OI absorbers have a relatively small halo with predicted impact parameters less than 10 h$^{-1}$ kpc, which is perfectly within the ALMA field of view.

The ALMA observations used four 1.875 GHz spectral windows (SPWs) and were obtained in the time division mode. Each SPW was divided into 43 channels. Two of the SPWs were centred on the [CII] emission at the redshift of the OI absorber and QSO, respectively. The remaining two SPWs were used to obtain a continuum image of the field. The on-source time was 2.5 h. The data were reduced using the CASA 5.4.0 pipeline following the standard calibration procedures. To obtain the [CII] emission-line data cube, we subtracted the continuum via the CASA task ‘uvcontsub’ on the line-free frequencies. After obtaining the calibrated and continuum-subtracted data set, we generated the image cubes using the task ‘clean’ with natural weighting to reach the highest sensitivity. The [CII] intensity map (that is, moment-0 map) was generated from the [CII] emission-line channels based on the task ‘immoments’ within CASA. We define the emission-line channels as those with a signal-to-noise ratio greater than 1.0$\sigma$.

**Details of reliability of [CII]2054 detection.** We did three tests to study the robustness of the detection of [CII]2054. First, we searched for signals in the whole cube. Each moment-0 image has a width of 344 km s$^{-1}$ in velocity, ranging from $-173$ km s$^{-1}$ to $+173$ km s$^{-1}$ around the redshift of [CII]2054. No other sources had either a signal higher than the 4$\sigma$ level or negative signal lower than ~4$\sigma$. Results are shown in Extended Data Figs. 1 and 2. Note that the QSO also appears in this continuum-subtracted image because of a serendipitous water line at this frequency. The probability of finding a 4$\sigma$ source caused by random fluctuation is small.

Then, we split the XX and YY correlation based on setting the parameter ‘stokes’ in ‘tclean’ to ‘XX’ and ‘YY’, respectively. The intensity maps were generated from the same [CII] emission line channels as mentioned above. The results are shown in Extended Data Fig. 3. The upper panels show the XX intensity map and the corresponding one-dimensional spectrum, and the lower panels show the YY intensity map and the corresponding one-dimensional spectrum. The measured integral fluxes are 0.0768 $\pm 0.0335$ Jy km s$^{-1}$ and 0.0490 $\pm 0.0306$ Jy km s$^{-1}$. The S/Ns of [CII]2054 in the XX and YY maps are 3.1 and 2.6, respectively, close to the expected value of $\sqrt{2}$. Thus, none of the evidence supports that [CII]2054 is a false positive.

**Target selection criteria and source robustness.** According to our ALMA proposal, our detection limit is down to the typical $L_*$ at the ~4$\sigma$ level. The observed [CII]2054 mass of $10^7 M_\odot$ and a typical $L_*$ mass on the order of $10^7 M_\odot$ to $10^8 M_\odot$ is consistent with a galaxy at $z \geq 6$. Three selection criteria are employed as follows.

First, the [CII] emitter should have a velocity offset within $\pm 150$ km s$^{-1}$ from the OI absorber. This is because high-$z$ star-forming galaxies with stellar mass of $\sim 10^{10} M_\odot$ exhibit outflow velocities of $\sim 150$ km s$^{-1}$ (refs. 31,32) at $z \geq 2$. Note that our detection limit corresponds to a derived stellar mass of $\geq 10^8 M_\odot$, resulting in a stellar mass that is lower than or comparable to that of the comparison sample. Furthermore, the projected separation could make the line-of-sight velocity component even lower compared to the outflow velocity measured along the galaxy sightline. Therefore, it is safe to constrain the outflow velocity, that is, the velocity offset between the galaxy and absorber, to between $-150$ km s$^{-1}$ and $+150$ km s$^{-1}$.

Second, the projected impact parameter between the source galaxy and the strong OI absorbers have to be smaller than $4.0$ h$^{-1}$ kpc at $z \geq 6$. This is because strongly metal-enriched gas should be present in the circumgalactic medium and be located within the dark-matter halo of the host galaxy. At $z \geq 6$, simulations suggest that strong OI absorbers with REW equal or greater than 0.1 A have a projected impact parameter within $\lessapprox R_\text{vir}$ (the virial radius) of the dark-matter halo according to hydrodynamical simulations (see, for example, refs. 33,34) in addition, according to all these simulations, the actual projected impact parameter is within $R \lessapprox 0.7 R_\text{vir}$. Furthermore, our designed halo mass limit of $10^9 M_\odot$ (note that our detected [CII]2054 has a derived halo mass of $4 \times 10^9 M_\odot$, corresponding to an $R_\text{vir}$ of $20\sim 30$ kpc at $z \leq 6$). Thus, when we search for $L_*$ source galaxies in the field, we set the impact parameter limit of the detection of an $L_*$ galaxy to 4.0$\sigma$. Note that this is also a safe constant limit because the projection effect may make the actual separation greater than the limit.

Third, according to the [CII] line width analysis, galaxies with a [CII] luminosity of $L_{\text{CII}} \sim 10^8 L_\odot$ (stellar mass of $10^8 M_\odot$ have line widths of [CII] greater than $250$ km s$^{-1}$. This is according to the FWHM measurement from various previous studies (ref. 19). The consistent source selection criteria for galaxies with an FWHM of [CII] greater than $250$ km s$^{-1}$ Note that the line width criteria can be helpful to further eliminate the false-positive sources generated by noise. Correlating noise may more frequently generate a narrow line, because the peak of the noise could influence adjacent channels (see also Extended Data Fig. 5).

With these three criteria, we searched for emitters from the continuum-subtracted data cube using the algorithm FINDCLUMPS (ref. 35). First, we perform floating averages of a given number of channels with different window sizes (for example, 3, 4- and 5-channel windows) at different frequencies. After obtaining...
these maps, we estimate the r.m.s. values based on the pixel-to-pixel standard deviation in every averaged map. Then, we search for peaks exceeding a certain S/N threshold using the Python task DAOFIND v1.1.0 in every map. After proceeding with a source-centred-search algorithm, to get an estimate of the probability of finding peaks having $\geq \sigma$ significance we adopt the following three approaches.

First, we check the probability of $\geq 4\sigma$ in our own spectral windows (SPWs). The four SPWs all reside in Band 6, with central frequencies of 272.3, 270.5, 256.0 and 257.8 GHz, respectively. Each has a frequency bandwidth of $-2\Delta\nu$ that corresponds to a total velocity range of $9 \times 10^4$ km s$^{-1}$. Two of the four SPWs were set to observe the [CII] emission at the redshift of OI absorber and the background QSO, respectively. The remaining two SPWs were used to obtain a continuum image of the field. We performed the line-search algorithm on the four SPWs and found 14 sources in the whole field with $S/N \geq 4\sigma$ (including [CII]2054). Corresponding to a primary beam limit of 0.01 at two lines (water emission and QSO [CII] emission) caused by the background CO line luminosities of $1.8 \times 10^{23} \mathrm{cm}^{-2} \cdot \mathrm{s}^{-1}$, CO (3–2), CO (4–3), CO (5–4) and CO (6–5) are $2 \times 10^{23} \mathrm{cm}^{-2} \cdot \mathrm{s}^{-1}$, which correspond to the peak flux densities of $3.1 \times 10^{-16} \mu\mathrm{Jy}$ per beam. To estimate the photometric redshifts of targets in this field through the use of the Lyman break dropout technique, we fit the HST broadband imaging. For the field surrounding QSO J2054-0005, the HST archival database has two optical observations, F606W (V) and F814W (I),39 and three near-infrared broadband images, F105W (J), F125W (H) and F160W (H). These broad filter images were used to estimate the photometric redshifts of targets in this field.

HST broadband imaging. For the field surrounding QSO J2054-0005, the HST archival database has two optical observations, F606W (V) and F814W (I), and three near-infrared broadband images, F105W (J), F125W (H) and F160W (H). These broad filter images were used to estimate the photometric redshifts of targets in this field through the use of the Lyman break dropout technique. Galaxies at $z = 6$ are expected to drop out of the F606W filter.

For photometry of the galaxies, we matched the point-spread function (PSF) first. We generated PSFs in different filters by stacking stars in the field by using an aperture radius of 0.3 petalight years. Stars were matched in the PSF generation to an archival catalogue. After matching PSFs to that of F1060W, the measured FWHMs of the PSF in all five bands are $0.23''$ (consist with the FWHM of the PSF in F160W). These results are also consistent with the PSF-matching results in the CANDELS Multi-wavelength Catalogue. Then, we chose the diameter of the photometric apertures as $0.6''$, corresponding to 2.5 times the PSF-matched FWHM. Table 1 shows the HST broadband photometric results of all the ALMA-detected continuum sources. The HST observations are shown in Extended Data Fig. 6.

[CII] luminosity, SFR and halo mass estimation. The velocity-integrated [CII] flux density is measured to be $0.0758 \pm 0.0177 \mu\mathrm{Jy}$ per beam, corresponding to a luminosity of $L_{\mathrm{CII}}(0.1 \times 10^{-20} \mu\mathrm{Jy}) = 0.000101 \times 10^{20} \mathrm{erg} \cdot \mathrm{s}^{-1} \times (1 + z)^2$. Thus, the [CII] luminosity is $3.413 \times 10^{23} \mathrm{erg} \cdot \mathrm{s}^{-1} \times (1 + z)^2$, which is the luminosity distance in Mpc. The [CII]-based SFR, $\dot{M}_{\text{SFR}}$, is approximately $6.7 \times 10^{8} \, M_{\odot} \, \text{yr}^{-1}$ (with 0.3 dex uncertainty). Comparing with the ALPINE survey,27 this [CII] emitter is located at the faint end. Nevertheless, in a random survey, the probability of finding such a galaxy in this survey volume is $2 \times 10^{-4}$. Also, albeit only 25% of the predicted SFRs of OI associated galaxies are typically larger than $0.1 \times 10^{10} \, M_{\odot} \, \text{yr}^{-1}$ in these cosmological simulations, we detect a bright galaxy with an SFR of $6.8 \times 10^{8} \, M_{\odot} \, \text{yr}^{-1}$. Therefore, for galaxies associated with strong OI absorber, this detection is exceptional.

Because the star formation rate and stellar mass information are not shown in ref. 26, a comparison between the halo masses is conducted for a direct comparison with PSFs cosmological simulations. We derive the halo mass, $M_{\text{halo}}$, directly from a [CII] luminosity–$M_{\text{halo}}$ relation proposed by simulating galaxies with an SFR of $0.1 \times 10^{10} \, M_{\odot} \, \text{yr}^{-1}$ to $3.0 \times 10^{10} \, M_{\odot} \, \text{yr}^{-1}$ at $z \approx 6$ (ref. 24). Galaxies with an SFR similar to that of [CII]2054 can be reproduced in this simulation. Using the relation between [CII] luminosity and halo mass in their work (equation (8) in ref. 24), we derive a halo mass of $M_{\text{halo}} = 4.1 \times 10^{11} \, M_{\odot}$ with a 1σ scattering.

To further confirm the halo mass estimation, we use another method of deriving the halo mass from the stellar mass. To do this, we first need to derive the stellar mass. We convert the [CII]–derived total SFR to the stellar mass $M_{\star}$ of $3.0 \times 10^{10} \, M_{\odot}$, the predicted stellar mass of [CII]2054 is then $5.8 \times 10^{10} \, M_{\odot}$ with 0.3 dex uncertainty. We further calculate the dynamical mass and the molecular gas fractions of $f_{\text{molgas}} = M_{\text{molgas}} / M_{\text{halo}}$ (where $M_{\text{halo}}$ represents the stellar mass of a galaxy) to confirm the estimation of stellar mass. From the spectrum of [CII]2054, the [CII] velocity dispersion ($\sigma_{\text{CII}}$) is $187 \, \text{km} \cdot \text{s}^{-1}$, corresponding to a circular velocity of $v_{\text{c}} = 1.763 \times 10^3 \, \text{km} \cdot \text{s}^{-1}$, then, using the mentioned unresolved size of $0.1 \times 10^{10} \, M_{\odot} \, \text{yr}^{-1}$ at $z \approx 6$ (ref. 24). Comparing with the ALPINE survey,27 we find that the derived dynamical mass is $3.7 \times 10^{10} \, M_{\odot}$ consistent with our derived stellar mass. Then, using the relation between [CII] luminosity and molecular gas mass $M_{\text{molgas}} / M_{\text{halo}}$, we predict the molecular gas mass ($M_{\text{molgas}}/M_{\text{halo}}$) is $2.0 \times 10^{-2}$, thus, the molecular gas fraction is $0.78$, consistent with the results from the ALPINE survey. Therefore, we can estimate the halo mass from the stellar mass based on equation (1) of ref. 27. The predicted halo mass is about $1.1 \times 10^{10} \, M_{\odot}$. Note that the conversion between the stellar mass and halo mass has a scatter of 0.3 dex. Thus, the combined uncertainty is close to three times the 0.3 dex scatter, and this estimated halo mass is consistent with the value we derived in the last paragraph within the 1σ level.

[CII]2054 continuum properties. No continuum was detected at the position of [CII]2054, resulting in a 3σ upper limit of 37 jy per beam. To convert this continuum flux to a total infrared luminosity (TIR; 8–1000 μm), we estimate the TIR luminosity by fitting a modified blackbody (that is, $S_{\nu} \approx \text{exp}(\nu/T_{\text{dust}}) - 1$) to this single upper limit. We assume that [CII]2054 has the same continuum properties as a typical Lyman break galaxy at $z \approx 6$ and therefore we derive an infrared SFR of $10^{11} \, M_{\odot} \, \text{yr}^{-1}$. [CII]2054 also was not detected in any of the HST broadband imaging using SEExtractor40 (HST Data Release 6). For the broadband imaging, we used a large survey over an area of $4.2 \times 10^{10} \, \text{K} \cdot \text{m} \cdot \text{K}^{-1} \cdot \text{cm}^{-2}$, this corresponds to possible CO line luminosities of $1.8 \times 10^{20} \, \text{K} \cdot \text{m} \cdot \text{K}^{-1} \cdot \text{cm}^{-2} \times 10^{10} \, \text{K} \cdot \text{m} \cdot \text{K}^{-1} \cdot \text{cm}^{-2}$, respectively. After integrating the luminosity function, the mean densities of CO (3–2), CO (4–3), CO (5–4) and CO (6–5) are $2 \times 10^{20} \, \text{cm}^{-3} \cdot \text{cm}^{-3} \cdot \text{cm}^{-3} \cdot \text{cm}^{-3}$, respectively. The survey volume is assumed to be a sphere with a radius of $3.5 \, \text{arcmin}$, corresponding to projected distances of 14.4, 24.9, 28.7 and 29.6 kpc, respectively. After taking all of these into account, the probabilities of finding CO in this field are all close to $10^{-4}$. This indicates that the probability of a CO interloper is extremely low.
Continuum sources. There are five HST broadband observations in this field (F606W(V), F814W(I), F105W(Y), F215W(J) and F160W(H)); the photometric results for each source are tabulated in Table 1. To constrain the photometric redshifts of these continuum sources, we use colour–colour diagrams as shown in Extended Data Fig. 7, and compare the colours of the galaxies with those of simulated galaxies. In the HST image (Extended Data Fig. 6), C1, C3 and C5 are resolved as multiple targets. Therefore, we plot the colours of these multiple sources individually.

For the model galaxies, we use a general star-forming-galaxies approach47. Model spectra of different galaxies at different redshifts are generated from the models presented in the galaxy template model (BC03) (ref. 48). For the galaxy models, we take the metallicity to be Z = 0.02Z⊙ (where Z⊙ is the solar metallicity) with an age of 25 Myr, and we assume a Salpeter initial mass function. We further assume a constant SFR of 1 M⊙yr−1 with an exponentially declining star formation history with timescales of τr = 20 Gyr. To convert these galaxy templates to colours, we first apply a dust attenuation49 with E(B−V) = 0.0–0.4. Then, we take into account the effects of IGM attenuation46. Next, we convolve the spectral energy distribution with the transmission curves of the five HST broadband filters to get the colours of the simulated galaxies. In the left panel of Extended Data Fig. 7, the dark blue dots are the low-redshift template galaxies with redshift of z = 2.0–4.0. The red points represent star-forming galaxies at z = 5.5–6.5. We use the following criteria to select galaxies at z > 4:

$$F606W - F814W > 2$$

$$F814W - F105W < 3$$

$$F606W - F814W > 1.5 \times (F814W - F105W) + 0.5$$

Based on these criteria, we rule out almost all of the continuum sources as high-redshift candidates, except C1-1 and C5-3. In this panel, C1-1 and C5-3 are not resolved as multiple targets. Therefore, we plot the colours of these multiple sources individually.

We detect only one galaxy in the survey volume. Thus, the probability of randomly finding a galaxy in our survey volume will be 4 × 10−5, still much smaller than unity.

The results of the second QSO field. The results reported here are only a part of our ALMA observational programme. Actually, observation of three fields have been proposed and two of the three fields have been finished. The analysis of the second field was finished recently, but the third field has not been observed yet. Preliminary analysis shows no detection in the second QSO field. Thus, if the second field is included, then our survey volume is larger by a factor of 2. The probability of randomly finding a galaxy in our survey volume will be 4 × 10−4, which is still much smaller than unity.

Data availability
Both ALMA and HST data sets used in this work are publicly available. The data reported in this paper are available through the ALMA archive at https://almascience.eso.org/ajp with project code 2017.1.01088.S and through the HST archive at https://archive.stsci.edu/hst/ with project codes 15410 and 12974. Other data are available from the corresponding author upon reasonable request.

Code availability
The ALMA data were reduced using the CASA pipeline version 5.4.0, available at https://casa.nrao.edu/casa_obtaining.shtml.

Received: 8 December 2020; Accepted: 29 July 2021; Published online: 27 September 2021

References
1. Codoreanu, A. et al. The comoving mass density of MgII from z = 2 to 5.5. Mon. Not. R. Astron. Soc. 472, 1023–1031 (2017).
2. Becker, G. D. et al. The evolution of OI over 3.2 < z < 6.5: reionization of the circumgalactic medium. Mon. Not. R. Astron. Soc. 483, 163 (2019).
3. Cooper, T. J. et al. Heavy element absorption systems at z = 5.0–c < 6.8: metal-poor neutral gas and a diminishing signature of highly ionized circumgalactic matter. Astrophys. J. 882, 77 (2019).
4. Cai, Z., Fan, X., Dave, R., Finlator, K. & Oppenheimer, B. Probing the metal enrichment of the intergalactic medium at z = 5–6 using the Hubble Space Telescope. Astrophys. J. Lett. 849, L18 (2017).
5. Biehl, R. M. et al. Into the Lyα jungle: exploring the circumgalactic medium of galaxies at z = 4–5 with MUSE. Mon. Not. R. Astron. Soc. 493, 5336–5356 (2020).
6. Diaz, C. G. et al. Faint LAEs near z > 4.7: C IV absorbers revealed by MUSE. Mon. Not. R. Astron. Soc. 502, 2645–2663 (2021).
7. Oppenheimer, B. D., Davé, R. & Finlator, K. Tracing the re-ionization-epoch intergalactic medium with metal absorption lines. Mon. Not. R. Astron. Soc. 396, 729–758 (2009).
8. Finlator, K. et al. The host haloes of OI absorbers in the reionization epoch. Mon. Not. R. Astron. Soc. 436, 1818–1835 (2013).
9. Finlator, K., Doughty, C., Cai, Z. & Diaz, G. The faint host galaxies of CIV absorbers at z > 5. Mon. Not. R. Astron. Soc. 493, 3223–3237 (2020).
10. Dijkstra, M., Haiman, Z. & Spans, M. Lyr radiation from collapsing protogalaxies. I. Characteristics of the emergent spectrum. Astrophys. J. 649, 14–36 (2006).
11. Zheng, Z., Cen, R., Trac, H. & Miralda-Escudé, J. Radiative transfer modeling of Lyman alpha forest. I. Statistics of spectra and luminosity. Astrophys. J. 716, 574–598 (2010).
12. Hayes, M. et al. Escape of five per cent of Lyman-α photons from high-redshift star-forming galaxies. Nature 464, 562–565 (2010).
13. Leung, T. K. D. et al. Predictions of the Lyα SFR and [C II] luminosity function at the epoch of reionization. Astrophys. J. 905, 102 (2020).
14. Keating, L. C., Puchwein, E., Haehnelt, M. G., Bird, S. & Bolton, J. S. Testing the effect of galactic feedback on the IGM at z ~ 6 with metal-line absorbers. Mon. Not. R. Astron. Soc. 461, 606–626 (2016).
15. Oh, S. P. Probing the dark ages with metal absorption lines. Mon. Not. R. Astron. Soc. 336, 1021–1029 (2002).
16. Furlanetto, S. R. & Loeb, A. Metal absorption lines as probes of the intergalactic medium. Mon. Not. R. Astron. Soc. 309, 71–78 (1999).
29. Marshall, M. A. et al. Limits to rest-frame ultraviolet emission from high-redshift galaxies. Astron. J. 756, 14 (2012).
30. Walter, F. et al. Calibrating extinction-free star formation rate diagnostics in high-redshift galaxies. Astron. J. 130, 117 (2000).
31. Walter, F. et al. ALMA Spectroscopic Survey in the Hubble Ultra Deep Field: ALMA Spectroscopic Survey in the HUDF: a search for [C II] emitters at z ~ 6. Astrophys. J. 959, 147 (2020).
32. Smee, J. L. F. et al. The star formation rate function for redshift z ~ 4–7 galaxies: evidence for a uniform buildup of star-forming galaxies during the first 3 Gyr of cosmic time. Astrophys. J. 879, 322 (2018).
33. Walter, F. et al. ALMA Spectroscopic Survey in the Hubble Ultra Deep Field: survey description. Astrophys. J. 883, 67 (2017).
34. De Looze, I., Baes, M., Bendo, G. J., Cortese, L. & Fritz, J. The reliability of [C II] as an indicator of the star formation rate. Mon. Not. R. Astron. Soc. 416, 2712–2724 (2011).
35. McMullin, J. P., Waters, B., Schiebel, D., Young, W. & Golap, K. CASA architecture and applications. In Proc. 16th Annual Conference on Astronomical Data Analysis Software and Systems (eds Shaw, R. A. et al.) Vol. 32, 214–217 (2016).
36. Shapley, A. E., Steidel, C. C., Pettini, M. & Adelberger, K. L. Rest-frame ultraviolet spectra of z ~ 3 Lyman break galaxies. Astrophys. J. 588, 65–89 (2003).
37. Stefanon, M. et al. CANDLES multi-wavelength catalogs: source identification and photometry in the CANDLES Extended Groth Strip. Astrophys. J. Suppl. Ser. 229, 32 (2017).
38. Speagle, J. S., Steinhardt, C. L., Capak, P. L. & Silverman, J. D. A highly consistent framework for the evolution of the star-forming ‘main sequence’ from z ~ 0–6. Astrophys. J. Suppl. Ser. 214, 15 (2014).
39. Capak, P. L. et al. Galaxies at redshifts 5 to 6 with systematically low dust content and high [C II] emission. Nature 522, 455–458 (2015).
40. Fujimoto, S. et al. The ALPINE-ALMA [CII] survey: size of individual star-forming galaxies at z ~ 4–6 and their extended halo structure. Astrophys. J. 900, 1 (2020).
41. Dessauges-Zavadsky, M. et al. The ALPINE-ALMA [CII] survey. Molecular gas budget in the early Universe as traced by [C II]. Astron. Astrophys. 643, A5 (2020).
42. Zanella, A. et al. The [C II] emission as a molecular gas mass tracer in galaxies at low and high redshifts. Mon. Not. R. Astron. Soc. 481, 1976–1999 (2018).
43. Faist, A. L. et al. ALMA characterizes the dust temperature of z ~ 5.5 star-forming galaxies. Mon. Not. R. Astron. Soc. 498, 4192–4204 (2020).
44. Izumi, T. et al. Subaru high-z exploration of low-luminosity quasars (SHELLQs). III. Star formation properties of the host galaxies at z ≥ 6 studied with ALMA. Publ. Astron. Soc. Jpn 70, 36 (2018).
45. Bertin, E. & Arnouts, S. SExtractor: software for source extraction. Astron. Astrophys. Suppl. Ser. 117, 393–404 (1996).
46. Ono, Y. et al. Great Optically Luminous Dropout Research Using Subaru HSC (GOLDURUSH). I. UV luminosity functions at z ~ 4–7 derived with the half-million dropouts on the 100 deg² sky. Publ. Astron. Soc. Jpn 70, S10 (2018).
47. Calzetti, D. et al. The dust content and opacity of actively star-forming galaxies. Astrophys. J. 533, 682–695 (2000).
48. Inoue, A. K., Shimizu, I., Iwata, I. & Tanaka, M. An updated analytic model for attenuation by the intergalactic medium. Mon. Not. R. Astron. Soc. 442, 1805–1820 (2014).
49. Neelam, M. et al. [CII] 158-μm emission from the host galaxies of damped Lyman-alpha systems. Science 355, 1285–1288 (2017).
50. Fumagalli, M., O’Meara, J. M., Prochaska, J. X., Rafelski, M. & Kanekar, N. Directly imaging damped LyMAN α galaxies at z > 2 – III. The star formation rates of neutral gas reservoirs at z ~ 2.7. Mon. Not. R. Astron. Soc. 446, 3178–3198 (2015).
51. Loiacono, F. et al. The ALPINE-ALMA [CII] survey. Luminosity function of serendipitous [CII] 158-μm emitters at z ~ 5. Astron. Astrophys. 646, A76 (2021).
52. Steidel, C. C. et al. The structure and kinematics of the circumgalactic medium: evidence from far-ultraviolet spectra of z ~ 2–3 galaxies. Astrophys. J. 717, 289–322 (2010).
53. Béthermin, M. et al. The ALPINE-ALMA [CII] survey: data processing, catalogs, and statistical source properties. Astron. Astrophys. 643, A2 (2020).
54. Fujimoto, S. et al. First identification of 10 kpc [C II] 158-μm halos around star-forming galaxies at z ~ 5–7. Astrophys. J. 887, 107 (2019).
55. Decarli, R. et al. The ALMA Spectroscopic Survey in the Hubble Ultra Deep Field: multiband constraints on line-luminosity functions and the cosmic density of molecular gas. Astrophys. J. 902, 110 (2020).
56. Aravena, M. et al. The ALMA Spectroscopic Survey in the Hubble Ultra Deep Field: selection criteria for [C II] line and dust emission in 6 < z < 8 galaxies. Astrophys. J. 833, 71 (2016).
57. González-López, J. et al. The Atacama Large Millimeter/submillimeter Array Spectroscopic Survey in the Hubble Ultra Deep Field: CO emission lines and 3 mm continuum sources. Astrophys. J. 882, 139 (2019).

Acknowledgements
Z.C. and Z.W. are supported by the National Key R&D Program of China (grant no. 2018YFA0404503) and the National Science Foundation of China (grant no. 12073014). M.N. acknowledges support from European Research Council advanced grant no. 740246 (Cosmic_Gas). F.W. is thankful for support provided by NASA through the NASA Hubble Fellowship grant no. HST-HF2-51448.01-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. J.Y. is a Strittmatter Fellow in Steward Observatory, University of Arizona. The F125W and F160W HST observations in the QSO 2054 field were conducted in the HST program (Proposal ID: 15064). We appreciate the principal investigator Caitlin Casey and the team of this program for collecting these data in Cycle 25.

Author contributions
Y.W. and Z.C. led the data reduction, pipeline development, analysis and manuscript writing. Z.C., M.N., K.F. and I.J.P. conceived the project and led the telescope proposal. Z.C. is the principal investigator of the ALMA program (Program ID: 2017.1.01088.S). M.N., K.F. and S.Z. all participated in the analysis and data reduction. R.W. and B.H.C.E. helped with checking of the ALMA data reduction and analysis. X.F., L.C.K., F.W., J.Y. and J.E.H. and J.W. all helped significantly with the interpretation and commented on the ALMA proposal and the paper. 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/s41550-021-01471-4. Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41550-021-01471-4. Correspondence and requests for materials should be addressed to Zheng Cai.

Peer review information Nature Astronomy thanks Matthieu Béthermin 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.
Extended Data Fig. 1 | Channel maps in the whole observed field of view. These channel maps are arranged from -641.281 to +734.366 km/s. The channel width is set as the same as [CII] intensity map as 344 km/s for each map. The black and dark red cross indicate the position of QSO J2054 and [CII]2054, respectively. The sizes of the synthesized beams are demonstrated in the bottom-left of these panel.
Extended Data Fig. 2 | The detailed pixel flux distributions of different Channel maps. The green lines show the best-fit single-Gaussian models. We regard the single-Gaussian fitted standard deviation (STD) as the noise in these intensity maps.
Extended Data Fig. 3 | Different polarization-correlation maps. Upper panels. The moment-0 and spectrum under XX polarization. The [C II] moment-0 maps is collapsed based on the same emission range as Fig. 1 and shown by the yellow shaded region. The outer contour is at 3σ level, with contours in steps of 1σ. The 1σ rms is $2.47 \times 10^{-2}$ mJy beam$^{-1}$ km s$^{-1}$. Dashed lines represent -2σ-level contours. The integral [C II] flux is $0.0768 \pm 0.0247$ mJy km s$^{-1}$. The synthesized beams are shown in the bottom-left of each mom-0 map. Lower panels. The results under YY polarization. The integral [C II] flux is $0.0667 \pm 0.0254$ mJy km s$^{-1}$, while the 1σ is $2.54 \times 10^{-2}$ Jy beam$^{-1}$ km s$^{-1}$.
Extended Data Fig. 4 | Multi-exposure observations. **Left panel:** The intensity maps of three individual exposures. The integral [CII] fluxes are 0.0935, 0.0786 and 0.0490 Jy beam$^{-1}$ km s$^{-1}$. Meanwhile the 1-$\sigma$ standard deviation are 0.0319, 0.0335, and 0.0269 Jy beam$^{-1}$ km s$^{-1}$, respectively. **Right panel:** The corresponding 1-D spectra of [CII]2054 in the different individual exposure shown in the left. The beam sizes are shown in the bottom-left of each mom-0 map. The yellow shaded region shows the emission range that is used to generate the [CII]moment-0 maps.
Extended Data Fig. 5 | The number of candidates in the deep ASPECS Band-6 datacube. We found 719 sources having FWHM \( \geq 250 \text{ km s}^{-1} \) in a 4.2 arcmin\(^2\) area of the ASPECS survey. In the figure, sources with Full Width Half Maximum (FWHM) between 200 and 500 km s\(^{-1}\) are present. The vertical dashed line represents the FWHM of the single Gaussian fitting of the [CII]2054.
Extended Data Fig. 6 | High-resolution HST broad-band images for five different filters. These images are sorted by the filter central wavelength. From left to right, these images are F606W, F814W, F105W, F125W, and F160W, respectively. Further, different rows represent different continuum sources (as defined in Table 1). The HST photometry is based on apertures with a 0.6 arcsec diameter, as shown by the green and darkgreen circles. The orange (2-5 $\sigma$) and darkred contours (2-4 $\sigma$) represent the 264GHz-continuum and [CII]-emission regions in the ALMA observations, respectively. We find no continuum emission in the [CII]-emission region of [CII]2054 in all five HST images.
Extended Data Fig. 7 | Color-color diagrams used to select galaxies associated with the OI absorber. The color properties of these sources are calculated under the AB magnitudes. The plotted error bars show the propagated uncertainties based on the $1\sigma$ errors in magnitudes. The continuum sources are plotted in blue circles. The red and dark blue dots represent the color properties of simulated star-forming galaxies at $z > 5.5$ and $z < 4$, respectively. High-redshift selection criteria are based on the distribution of these template galaxies and shown in the grey-shaded regions. **Left:** $I - Y$ vs. $V - I$ two-color diagram. In the left panel, we rule out most continuum sources as high-redshift galaxies, except for C1-1 and C5-3. These two galaxies have the same color properties as the simulated high redshift galaxies, and are plotted as a purple and magenta dot. **Right:** $Y - J$ vs. $I - Y$ two-color diagram. In the right panel, we rule out C1-1 and C5-3 as high-redshift candidates. Note QSO J2054-0005 is plotted as a blue-violet star.
Extended Data Fig. 8 | The relationship between the projected impact parameters and halo masses of strong OI absorbers in different simulations.

The halo mass of the [CII]2054 is converted directly from the [CII] luminosity to the halo mass relation (Leung et al. 2020). The error bars represent 1-σ uncertainties of the estimated halo mass. In the three top panels, grey dots represent OI absorbers with the REW of 0.12 ± 0.05 Å (consistent with observations). Meanwhile, Red star represents [CII]2054. Bottom panels show the halo mass distribution of different simulations. In the bottom panels, Red arrow shows that the host halo mass of [CII]2054 is one order of magnitude larger than the median value predicted by all of these simulations.