



A [C II] 158 μm emitter associated with an O I absorber at the end of the reionization epoch

Yunjing Wu¹, Zheng Cai^{1,2,3}✉, Marcel Neeleman⁴, Kristian Finlator^{5,6}, Shiwu Zhang¹, J. Xavier Prochaska^{7,8}, Ran Wang⁹, Bjorn H. C. Emonts¹⁰, Xiaohui Fan¹¹, Laura C. Keating¹², Feige Wang¹¹, Jinyi Yang¹¹, Joseph F. Hennawi¹³ and Junxian Wang^{14,15}

The physical and chemical properties of the circumgalactic medium at $z \gtrsim 6$ have been studied successfully through the absorption in the spectra of background quasi-stellar objects^{1–3}. 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^{4–6}. 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. 7–9). Deep, dedicated surveys have searched for Ly α emission associated with strong C IV absorbers at $z \approx 6$, but only a few Ly α -emitter candidates have been detected. Interpreting these detections is moreover ambiguous because Ly α is a resonant line^{10–12}, raising the need for complementary techniques for detecting absorbers' host galaxies. Here we report a [C II] 158 μm emitter detected using the Atacama Large Millimeter Array that is associated with a strong low-ionization absorber, O I, at $z = 5.978$. The projected impact parameter between O I and [C II] emitter is 20.0 kpc. The measured [C II] luminosity is 7.0×10^7 solar luminosities. Further analysis indicates that strong O I absorbers may reside in the circumgalactic medium of massive halos one to two orders of magnitude more massive than expected values^{8,14}.

Metal absorption systems (for example, O I, C IV and Mg II) are powerful probes of the enrichment of the high-redshift intergalactic medium (IGM)^{1–3}. Because the first excitation energies of O I $\lambda 1302$ and neutral hydrogen are almost identical, O I is considered one of the best indicators to trace the metal enrichment and the neutral IGM close to the reionization epoch^{15,16}. Simulations suggest that at $z = 5–6$, feedback from star-forming galaxies can transport metals such as oxygen to 50 proper kpc ($\sim 9''$) (ref. 7–9). 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 models¹⁴. We select one of the strongest ultraviolet O 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 $10^{14.2} \text{ cm}^{-2}$ in the small optical depths regime (equation (8) in ref. 14).

The [C II] moment-0 map shown in Fig. 1a is constructed by collapsing the emission line channels from the [C II] data cube. At the redshift of the O I absorber, an emission line is detected at 4.3σ significance, which we interpret as [C II] emission from a galaxy at the O I absorber's redshift. Here we refer to this [C II] emitter as [C 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^{-6} for this line to be a CO interloper (Methods). Then, we show the spectrum of [C II]2054 in Fig. 1d. The yellow shaded region represents the collapsed channels. The velocity-integrated [C II] flux density is measured to be $0.0758 \pm 0.0177 \text{ Jy km s}^{-1}$, corresponding to a [C II] luminosity of $7.0 \times 10^7 L_{\odot}$, where $L_{\odot}$ 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) survey²¹ over 4.2 arcmin^2 and found that the estimated probability of [C II]2054 caused by noise fluctuation is $\sim 0.6\%$ (Extended Data Fig. 5 and details in Methods). Furthermore, from our moment-0 image (Fig. 1a), the projected impact parameter between [C II]2054 and the O I absorber is $3.5''$, corresponding to 20.0 proper kpc (pkpc) at $z = 5.978$. Throughout this paper we assume a Λ Cold Dark Matter (CDM) cosmology with $\Omega_{\Lambda} = 0.7$, $\Omega_{\text{M}} = 0.3$ and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

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 I absorber at $z = 6$ (Extended Data Figs. 6 and 7 in Methods).

The [C II] 158 μm fine structure line emission is the dominant cooling line in the interstellar medium, and it is almost unaffected by dust attenuation²². The ALMA allows us to use [C II] to probe the physical conditions of the gas in galaxies and to quantify the SFR of the galaxy in the post-reionization epoch²³.

¹Department of Astronomy, Tsinghua University, Beijing, China. ²Center for Astronomy Technology, Tsinghua University, Beijing, China. ³Peng Cheng Laboratory, Shenzhen, China. ⁴Max-Planck-Institut für Astronomie, Heidelberg, Germany. ⁵New Mexico State University, Las Cruces, NM, USA. ⁶Cosmic Dawn Center (DAWN), Niels Bohr Institute, University of Copenhagen / DTU Space, Technical University of Denmark, Lyngby, Denmark. ⁷Department of Astronomy and Astrophysics, UCO/Lick Observatory, University of California, Santa Cruz, CA, USA. ⁸Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), The University of Tokyo, Kashiwa, Japan. ⁹Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing, China. ¹⁰National Radio Astronomy Observatory, Charlottesville, VA, USA. ¹¹Steward Observatory, University of Arizona, Tucson, AZ, USA. ¹²Leibniz-Institut für Astrophysik Potsdam, Potsdam, Germany. ¹³Department of Physics, University of California, Santa Barbara, CA, USA. ¹⁴CAS Key Laboratory for Research in Galaxies and Cosmology, Department of Astronomy, University of Science and Technology of China, Hefei, China. ¹⁵School of Astronomy and Space Science, University of Science and Technology of China, Hefei, China. ✉e-mail: zcaiz@mails.tsinghua.edu.cn

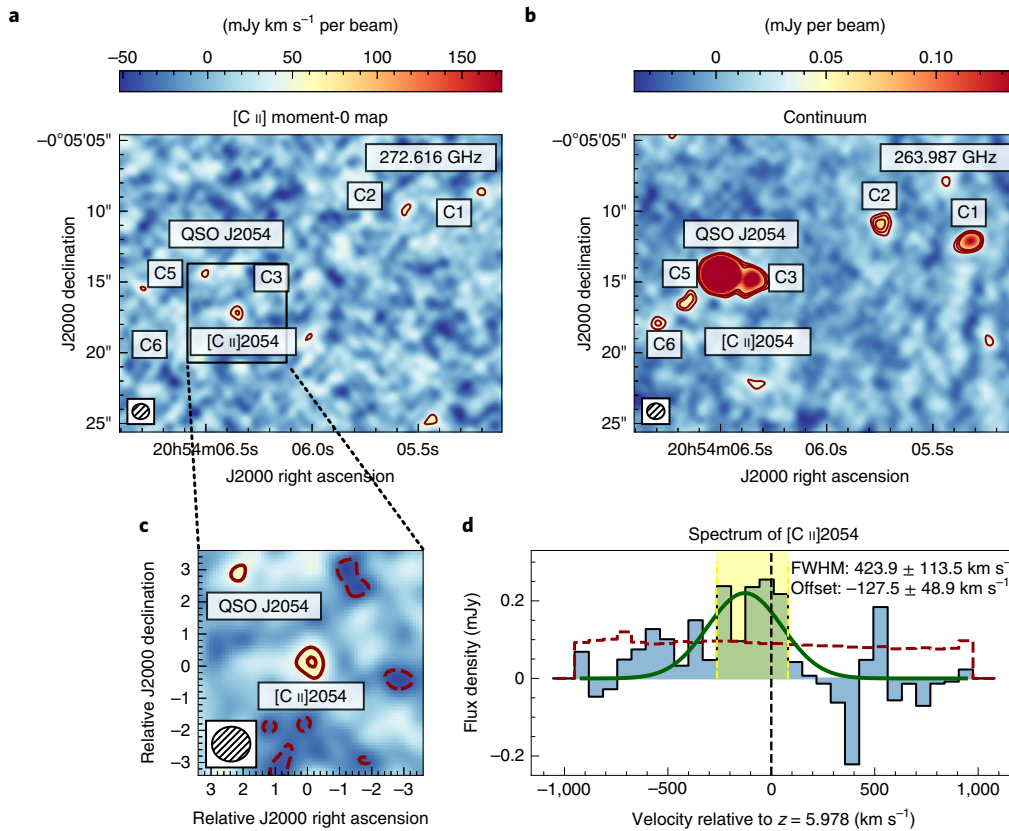


Fig. 1 | ALMA observations. **a**, [C II] moment-0 map. The outer contour is at a 3σ level, with contours in steps of 1σ . The 1σ is $0.0177 \text{ Jy km s}^{-1}$ per beam, calculated using the pixel-to-pixel standard deviation. The QSO appears in the continuum-subtracted image because of a water line of H_2O with the molecular transition of $J=3_{2,2}-3_{1,3}$ at $z=6.04$. **b**, Continuum map. Contours are drawn at 3σ , 4σ and 5σ . The 1σ noise is $12.4 \mu\text{Jy}$ per beam. Six continuum sources within the primary beam can be seen (QSO J2054 is C4). Colour bars in **a** and **b** represent the integrated flux and flux density respectively. **c**, Zoom-in of the [C II] moment-0 map. The sizes of the synthesized beams are plotted at the bottom left of this panel. The dashed contour is at a -2σ level. **d**, Spectrum of [C II]2054. The red dashed line shows the 1σ r.m.s. noise. The [C II] moment-0 map is collapsed based on the emission range shown by the yellow shaded region. The black dashed line represents velocity that is relative to the O I redshift $z=5.978$. The dark green line shows a single-Gaussian model fit to the data. The integral [C II] flux is $0.0758 \pm 0.0177 \text{ Jy km s}^{-1}$. FWHM, full width at half maximum. Source data

The velocity-integrated [C II] luminosity of [C II]2054 is estimated to be $L_{[\text{C II}]} = (7.0 \pm 1.7) \times 10^7 L_{\odot}$ (ref. ²⁴), corresponding to a [C II]-based SFR ($\text{SFR}_{[\text{C II}]}$) of $6.8 \pm 1.7 M_{\odot} \text{ yr}^{-1}$ (ref. ²⁵), where $M_{\odot}$ is the solar mass. No submillimetre continuum is detected at the position of [C II]2054, yielding a 3σ upper limit of $37 \mu\text{Jy}$. Converting this luminosity to a total infrared SFR (SFR_{TIR}) yields the upper limit of $\text{SFR}_{\text{TIR}} < 11 M_{\odot} \text{ yr}^{-1}$ (ref. ²⁶). To estimate the halo mass directly from the [C II] luminosity, we adopt a relation between $L_{[\text{C II}]}$ and halo mass at $z \approx 6$ (ref. ¹³) by assuming that [C II]2054 resides in the galaxy main sequence. Then, the derived halo mass is $4.1 \times 10^{11} M_{\odot}$, with a 1σ 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 1σ 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^{14,18–20}. As mentioned above, the O I absorber has an REW of $0.12 \text{ \AA}$, and the measured projected impact parameter is 20.0 pkpc (Fig. 1c). Broadly, simulations predict that metal line absorbers with an REW of $0.12 \pm 0.05 \text{ \AA}$ arise at impact parameters of $5\text{--}30 \text{ pkpc}$, bracketing our measurement. Specifically, in Fig. 2a we show the comparison of our data with the Illustris simulation¹⁴. 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 II]2054 are able to enrich the IGM at distances of $\geq 20 \text{ pkpc}$. Note that 11.1% of the data points in the simulation have a similar or larger impact parameter than that of [C II]2054. Thus, the projected distance between O I and [C 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 II]2054 based on the SFR. As we mentioned above, [C II]2054 has a halo mass of $4.1 \times 10^{11} M_{\odot}$ with a 1σ uncertainty of 0.5 dex. Different simulations, including Illustris, Sherwood, HVEL and FAST simulations, directly investigated the halo mass distribution for O I absorbers with column densities of $N_{\text{OI}} > 10^{14} \text{ cm}^{-2}$ at $z=6$. They find that O I absorbers with REW in the range $0.12 \pm 0.05 \text{ \AA}$ at $z=6$ are typically embedded in $10^9\text{--}10^{10} M_{\odot}$ halos^{8,14}. 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.

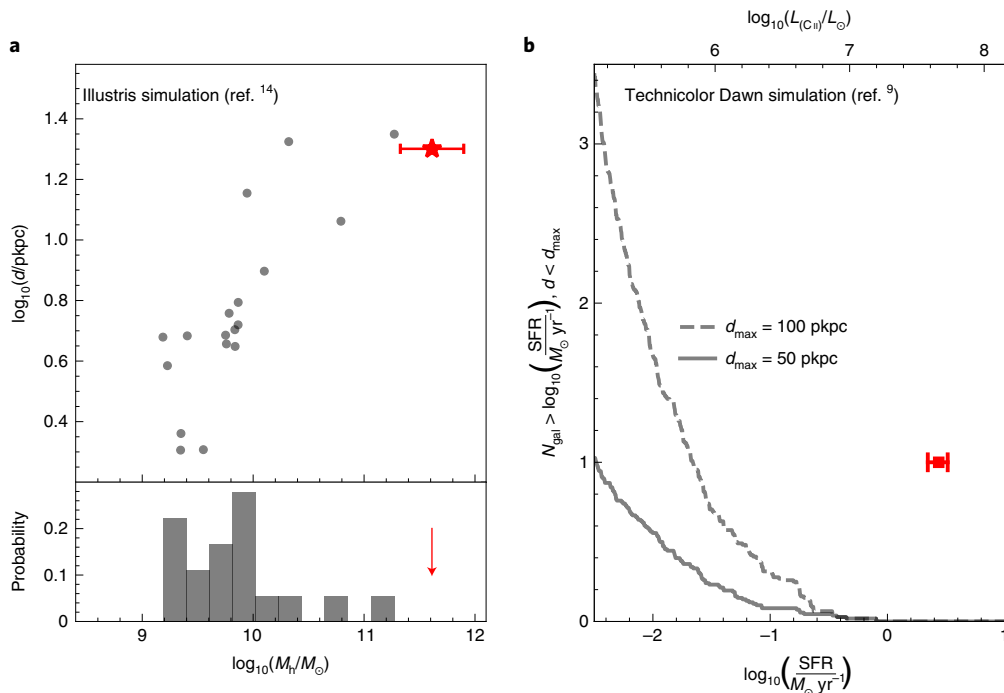


Fig. 2 | Comparison between simulations and observations. **a**, The upper panel shows the relationship between projected impact parameter and halo mass of strong O I absorbers in the Illustris simulation¹⁴. The O I absorbers have an REW of $0.12 \pm 0.05 \text{ \AA}$ (consistent with observations). The red star represents [C 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 II]2054 is one order of magnitude larger than the median value predicted by Illustris simulations. **b**, The number of galaxies that cluster about strong O I absorbers at $z=6$ is predicted to vary with SFR. The solid and dashed curves are biased SFR functions⁹. They are generated by accumulating catalogues of all galaxies that fall within 50 pkpc to 100 pkpc of O I absorbers with $\text{REW} > 0.12 \text{ \AA}$ ($N_{\text{OI}} > 10^{14.2} \text{ cm}^{-2}$). The predicted luminosity of the O I associated [C II] emitter is roughly two orders of magnitude fainter than that of [C II]2054. N_{gal} is the number of galaxies within an impact parameter d_{max} from strong O I absorbers in simulations.

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 epoch⁹. This set of simulations incorporates realistic small-scale ultraviolet background fluctuations and reproduces a broad range of observations of reionization and early galaxy growth⁹. They simultaneously reproduce observations of the galaxy stellar mass function, the Si 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 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 I absorbers at $z=6$ is predicted to vary with [C II]-based SFR. The solid and dashed curves are generated by compiling catalogues of all galaxies that fall within 50 pkpc to 100 pkpc of strong O I absorbers, respectively. We note that these two lines are normalized by the number of O I absorbers identified within the simulation (Methods). The strong O I absorbers have rest-frame equivalent widths of $\text{REW} > 0.12 \text{ \AA}$ ($N_{\text{OI}} > 10^{14.2} \text{ cm}^{-2}$). The SFR on the lower x axis is the 100-Myr average for physically associated galaxies, whereas the upper x axis converts this SFR to $L_{[\text{C II}]}$ (ref. 25). We plot the SFR of [C II]2054 in this panel using a red square with an error bar. Consistent with previous simulations^{7,8}, Technicolor Dawn simulations predict that the characteristic luminosity of physically associated [C II] emitters should be roughly two orders of magnitude lower in $L_{[\text{C II}]}$ than [C II]2054; that is, far fewer than one such galaxy is predicted to lie within the area probed by our ALMA observation.

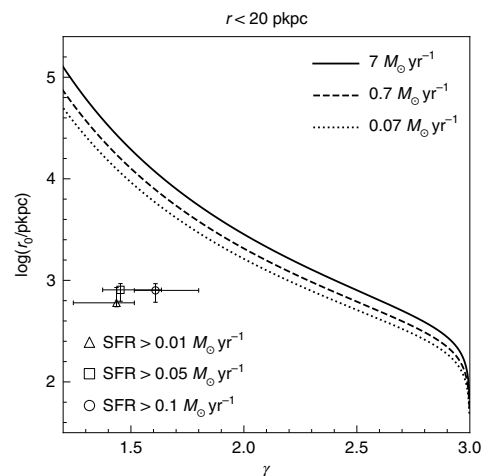


Fig. 3 | Comparison of cross-correlation function between simulations and observations. We assume the galaxy–absorber cross-correlation to be $\xi(r) = (r/r_0)^{-\gamma}$. This function expresses the fractional excess density of galaxies located at a distance r from an absorber in terms of the correlation length r_0 and a power-law slope γ . Symbols indicate cross-correlation functions predicted by Technicolor Dawn for galaxies clustered about synthetic O I absorbers with $\text{REW} \geq 0.12 \text{ \AA}$. 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 $\text{SFR} \geq 7, 0.7$ and $0.07 M_{\odot} \text{ yr}^{-1}$, within 20.0 pkpc of an absorber at $z=5.9$, respectively. Comparison with the symbols show that the predicted r_0 is ~ 10 times too low to explain the observations.

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_{264}^a (mJy)	flux _{err} (mJy)			S/N
C1	0.64	0.06			10.7
C2	0.12	0.02			6.0
C3	0.19	0.012			15.8
J2054	3.37	0.012			280
C5	0.067	0.014			4.8
C6	0.069	0.016			4.3
[C II] moment-0 map					
Source name	flux _{C II} (Jy km s ^{−1})	flux _{err} (Jy km s ^{−1})			S/N
[C II]2054	0.0758	0.0177			4.3
HST observations (AB magnitude)					
Source name ^b	F606W(V)	F814W(I)	F105W(Y)	F125W(J)	F160W(H)
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	25.53 ± 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	— ^c	— ^c	25.05 ± 0.02	24.55 ± 0.01	24.10 ± 0.01
C3-1	27.48 ± 0.09	26.92 ± 0.07	25.58 ± 0.02	25.17 ± 0.01	24.91 ± 0.01
C3-2	24.59 ± 0.01	24.06 ± 0.01	23.709 ± 0.004	23.604 ± 0.003	23.457 ± 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	22.053 ± 0.001	21.059 ± 0.001	20.880 ± 0.001	20.704 ± 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
[C II]2054	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 $0.6''$ diameter and photometry was applied in the PSF-matched multiband images (Methods). ^a Continuum flux density at 264 GHz. ^b HST resolves these sources as multiple targets. We number the resolved sources from 1 to 5. ^c 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 O I 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 [C II] luminosity function at this redshift²⁹, the mean space density of galaxies at $z = 5.9$ with a $SFR \geq 6.8 M_{\odot} \text{ yr}^{-1}$ is $1.4 \times 10^{-3} \text{ cMpc}^{-3}$ (where cMpc is comoving Mpc). A sphere of radius 20 pkpc containing one such galaxy has a probability of 2×10^{-5} , corresponding to a mean overdensity of $\Delta = 5 \times 10^4$ (Methods). Our observations therefore confirm that bright galaxies and strong absorbers correlate strongly^{6,9}. We quantify this correlation using the galaxy-absorber cross-correlation function $\xi(r) = (r/r_0)^{-\gamma}$, 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 O I absorbers with $REW \geq 0.12 \text{ \AA}$. As galaxies with $SFR \geq 6.8 M_{\odot} \text{ yr}^{-1}$ 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 does increase with SFR threshold⁹, the trend is not nearly strong enough to explain the discrepancy.

[C II]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 [C II]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 [C II] emitter associated with a reionization-epoch O I absorber is more than one to two orders of magnitude higher in [C II] 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 [C II]2054. Future surveys will be needed to test conclusively whether bright galaxies are more common around strong O I 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 O I 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 $M_{UV} \lesssim -19.5$, corresponding to a halo mass of $10^{11} M_{\odot}$. Based on the survey depth, the mean galaxy number density³⁰ 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 O I absorbers at $z \gtrsim 6$ 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 O I absorbers.

Methods

Motivation of this observational project. The observational programme was inspired by the successful detection of two host galaxies of damped Ly α absorbers (DLAs) in 2017 (ref. 46). One of the galaxies has the unexpected SFR of $110 \pm 10 M_{\odot} \text{ yr}^{-1}$ (ref. 30). This DLA is also associated with a strong O I absorber. The O I 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^{34,47}. These observations were then designed to constrain the galaxy counterparts associated with O I absorbers at $z \approx 6$. Based on the O I 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 O I absorbers are strong enough to have the associated Si II $\lambda 1260/1304$ 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 cMpc, then this strong O I absorber is selected from a survey of ~ 1 cGpc, indicating that the selected strong O I is rare.

ALMA observations and data reduction. There are a few advantages in using ALMA to probe the source galaxies associated with an O I absorber at $z \approx 6$. First, the sub-mm [C II] 158 μm line is one of the strongest far-infrared lines from galaxies^{22,32}. Second, the background QSO is less dominant in emission in the submillimetre regime. Third, at this redshift, the [C II] emission is shifted within ALMA Band 6, which has a high throughput and is thus efficient for emission line detections. Finally, cosmological simulations suggest that O I absorbers have a relatively small halo with predicted impact parameters less than $10''$, which is perfectly within the ALMA field of view¹⁴.

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 [C II] emission at the redshift of the O I absorber and QSO, respectively. The remaining two SPWs were used to obtain a continuum image of the field. The on-source time was 2.3 h. The data were reduced using the CASA 5.4.0 pipeline³³ following the standard calibration procedures. To obtain the [C II] 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 ‘tclean’ with natural weighting to reach the highest sensitivity. The [C II] intensity map (that is, moment-0 map) was generated from the [C II] emission-line channels based on the task ‘immoments’ within CASA. We define the emission-line channels as flux continuously greater than 1σ of the noise. Then, the [C II] emission full width half maximum (FWHM) of $\sim 400 \text{ km s}^{-1}$ is consistent with a galaxy with an SFR of $\sim 10 M_{\odot} \text{ yr}^{-1}$. The flux scale was calibrated by observing J2148+0657, whereas the phase stability was checked by observing J2101+0341. Absolute flux uncertainties are expected to be less than 10%.

The final data cubes and continuum images were obtained using natural weighting to maximize sensitivity. This resulted in a beam size of $1.2'' \times 1.1''$ and a 1σ background standard deviation of $12.4 \mu\text{Jy}$ per beam for the continuum image. A moment-0 map image was also generated centred on the redshifted [C II] emission line. For the moment-0 map, the emission line channels were collapsed

resulting in a beam size of $1.2'' \times 1.1''$ and a 1σ sensitivity of $0.0177 \text{ Jy km s}^{-1}$ per beam. For each source, 2σ contour areas are regarded as emission regions. We use these emission regions as the photometric aperture for resolved sources, whereas for unresolved sources the peak surface brightness represents the total flux. [C II]2054 is unresolved with a deconvolved size of $0.95'' \times 0.74''$. The photometry results are shown in Table 1. To further determine the credibility of the detection of [C II]2054, we checked signals on the binned channel maps, XX and YY polarization-correlation maps and three individual exposures. Both XX and YY polarization-correlation maps and all three individual exposures show positive signals. The details are given in the next section.

Details of reliability of [C II]2054 detection. We did three tests to study the robustness of the detection of [C II]2054. First, we searched for signals in the whole data cube. Each moment-0 image has a width of 344 km s^{-1} in velocity, ranging over the whole channel from -641.3 km s^{-1} to $+734.4 \text{ km s}^{-1}$ around the redshift of [C II]2054. No other sources had either a signal higher than the 4σ level or negative signal lower than -4σ . 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 4.3σ 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 [C II] 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’ results. The measured integral fluxes are $0.0768 \pm 0.0247 \text{ Jy km s}^{-1}$ and $0.0667 \pm 0.0254 \text{ Jy km s}^{-1}$. The S/Ns of [C II]2054 in the XX and YY maps are 3.1 and 2.6, respectively, close to the expected value of $\frac{4.3}{\sqrt{2}}$.

Next, we split different exposures based on adjusting the parameter ‘vis’ in ‘tclean’. The intensity maps are also obtained from the same emission-line channels. Results are shown in Extended Data Fig. 4. The measured integral fluxes are $0.0935 \pm 0.0319 \text{ Jy km s}^{-1}$, $0.0786 \pm 0.0335 \text{ Jy km s}^{-1}$ and $0.0490 \pm 0.0269 \text{ Jy km s}^{-1}$. The S/Ns of [C II]2054 in the three individual exposures are 2.9, 2.3 and 1.8 respectively, close to the expected value of $\frac{4.3}{\sqrt{3}}$. Thus, none of the evidence supports that [C II]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^* galaxies at the $\sim 4\sigma$ level. These galaxies have a stellar mass of $\sim 10^9 M_{\odot}$ and a typical halo mass on the order of $\sim 10^{11} M_{\odot}$. We expect to marginally detect the L^* galaxies at the $\sim 4\sigma$ level. Then, according to this expectation, we design the r.m.s. of ALMA observations. Thus, we design our selection criteria based on the marginal detection of a typical L^* galaxy at $z \approx 6$. Three selection criteria are employed as follows.

First, the [C II] emitter should have a velocity offset within $\pm 150 \text{ km s}^{-1}$ from the O I absorber. This is because high- z star-forming galaxies with stellar mass of $\sim 10^9 M_{\odot}$ exhibit outflow velocities of $\lesssim 150 \text{ km s}^{-1}$ (refs. 6,14,52) at $z = 2-6$. Note that our detection limit corresponds to a derived stellar mass of $5 \times 10^8 M_{\odot}$, 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 sightline⁵³. 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 O I absorbers have to be smaller than $4.0''$ at $z \approx 6$. 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 $z \approx 6$, simulations suggest that strong O I absorbers with REW equal or greater than $0.1 \text{ \AA}$ have a projected impact parameter within $\leq R_{\text{vir}}$ (the virial radius) of the dark-matter halo according to hydrodynamical simulations (see, for example, refs. 7–9,14). In addition, according to all these simulations, the actual projected impact parameter is within $R \leq 0.7 \times R_{\text{vir}}$. Furthermore, our designed halo mass limit of $10^{11} M_{\odot}$ (note that our detected [C II]2054 has a derived halo mass of $4 \times 10^{11} M_{\odot}$) corresponds to an R_{vir} of 20–30 kpc at $z \approx 6$. Thus, when we search for $\sim L^*$ source galaxies in the field, we set the impact parameter limit of the detection of an $\sim L^*$ galaxy to $4.0''$. Note that this is also a safe constraint because the projection effect may make the actual separation greater than the limit.

Third, according to the [C II] line width analysis, galaxies with a [C II] luminosity of $L_{[\text{C II}]} \approx 10^8 L_{\odot}$ (stellar mass of $10^9 M_{\odot}$) have line widths of [C II] greater than 250 km s^{-1} . This is according to the FWHM measurement from various previous studies^{51,53,54}. Thus, we search for galaxies with an FWHM of [C II] greater than 250 km s^{-1} . Note that the line width criteria may 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^{31,55}. 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 v.1.1.0 in every map. After proceeding with a source-target-finding algorithm, to get an estimate of the probability of finding peaks having $\geq 4\sigma$ 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 GHz that corresponds to a total velocity range of $9 \times 10^3 \text{ km s}^{-1}$. Two of the four SPWs were set to observe the [C II] emission at the redshift of O I 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 [C II]2054). Excluding two lines (water emission and QSO [C II] emission) caused by the background QSO, the number of sources having $S/N \geq 4\sigma$ is 12. For each SPW, we have an effective search area of $\sim 23''$, corresponding to a primary beam limit of 0.01 at the edge of the field of view. Note that our selection criteria have a search area of $\pi \times 4''^2 \approx 50.2 \text{ arcsec}^2$, a velocity range within $\pm 150 \text{ km s}^{-1}$ and a linewidth of $\geq 250 \text{ km s}^{-1}$. Scaling to our selection criteria, we calculate the expected number of sources having $S/N \geq 4\sigma$ as

$$n = 12 \times \frac{\pi \times 4''^2 \text{ arcsec}^2}{\pi \times 23''^2 \text{ arcsec}^2} \times \frac{300 \text{ km s}^{-1}}{9 \times 10^3 \text{ km s}^{-1}} \times \frac{6}{12} = 0.60\%, \quad (1)$$

where we find that six of 12 emitters have a line width of $\geq 250 \text{ km s}^{-1}$.

Therefore, using our own survey of four SPWs in Band 6, we find the probability of randomly finding $\geq 4\sigma$ source is also 0.6% based on our target selection criteria and assuming a Poisson distribution. Note that the probability calculated from our own survey could be regarded as the most reliable estimate of the probability of $\geq 4\sigma$ sources, as the noise properties are identical between [C II]2054 and other $\geq 4\sigma$ sources.

Second, we estimate the probability of finding $\geq 4\sigma$ sources in random-field observations by a source-matching method. Aravena et al.⁵⁶ searched for [C II] emitters from the ASPECS Pilot data cubes. In their survey, they searched for line emitters over a $\sim 1 \text{ arcmin}^2$ sky area and a ~ 60 GHz frequency range, with a lower frequency limit of 212 GHz and a higher frequency limit of 272 GHz. This frequency range corresponds to a wavelength range of $3.1 \times 10^6 \text{ \AA}$ and a velocity range of $\Delta v = 9 \times 10^4 \text{ km s}^{-1}$. Then, they matched the line-emitter catalogue and optical dropout galaxies by constraining the offset within $1''$ of the whole frequency range. They found that 12 of 50 optical dropout galaxies have $\geq 4\sigma$ line-emitter counterparts. Using results from ref.⁵⁶, the expected number of found targets having $\geq 4\sigma$ is

$$n = \frac{12}{50} \times \frac{\pi \times 4''^2 \text{ arcsec}^2}{\pi \times 1''^2 \text{ arcsec}^2} \times \frac{300 \text{ km s}^{-1}}{9 \times 10^4 \text{ km s}^{-1}} \times \frac{6}{12} = 0.64\%, \quad (2)$$

where $\frac{6}{12}$ represents the fraction of sources with $\text{FWHM} \geq 250 \text{ km s}^{-1}$ in the total number of sources with $\geq 4\sigma$ (Fig. 8 in ref.⁵⁶). Therefore, using the survey of ref.⁵⁶, we find that the probability of finding a $\geq 4\sigma$ source is 0.64% using our target selection criteria (assuming a Poisson distribution).

Third, we estimate the probability of finding $\geq 4\sigma$ sources in a blind line survey. We first checked this probability in the newly released deep ASPECS survey over a large field of 4.2 arcmin^2 (refs.^{21,57}). Using the source detection algorithm^{31,55}, we searched for line emitters over a 4.2 arcmin^2 sky area and a ~ 60 GHz frequency range. The velocity range is larger than that of our selection criteria by a factor of 300. We identified 1,209 targets in the whole 4.2 arcmin^2 field (Extended Data Fig. 5). After scaling to our search region ($\pi \times 4''^2 \approx 50.2 \text{ arcsec}^2$), a velocity range within $\pm 150 \text{ km s}^{-1}$ and a linewidth of $\geq 250 \text{ km s}^{-1}$, the expected number of such randomly found targets is

$$n = 1,209 \times \frac{300 \text{ km s}^{-1}}{9 \times 10^4 \text{ km s}^{-1}} \times \frac{\pi \times 4''^2 \text{ arcsec}^2}{4.2 \times 3,600 \text{ arcsec}^2} \times \frac{719}{1,209} = 0.79\%, \quad (3)$$

where $\frac{719}{1,209}$ represents the fraction of sources with $\text{FWHM} \geq 250 \text{ km s}^{-1}$ in the total number of sources with $\geq 4\sigma$. Therefore, using a large survey over an area of 4.2 arcmin^2 , we find that the probability of randomly finding a $\geq 4\sigma$ source is 0.79% using our target selection criteria (assuming a Poisson distribution).

The probability of [C II]2054 as a CO interloper. CO (3–2), CO (4–3), CO (5–4) and CO (6–5) transitions can possibly fall into our observed spectral window, corresponding to redshifts of 0.3, 0.7, 1.1 and 1.5, respectively. Given that the measured integral flux is $0.0758 \pm 0.0177 \text{ Jy km s}^{-1}$, this corresponds to possible CO line luminosities of 1.8×10^6 , 1.7×10^7 , 5.4×10^7 and $1.2 \times 10^8 \text{ Jy km s}^{-1} \text{ Mpc}^2$, respectively. After integrating the luminosity function, the mean densities of CO (3–2), CO (4–3), CO (5–4) and CO (6–5) are 2×10^{-11} , 4×10^{-11} , 3×10^{-11} and $3 \times 10^{-11} \text{ kpc}^{-3}$, respectively. The survey volume is assumed to be a sphere with a radius of $3.5''$, corresponding to projected distances of 14.4, 24.9, 28.7 and 29.6 pkpc , respectively. After taking all of these into account, the probabilities of finding CO in this field are all close to $\sim 10^{-6}$. This indicates that the probability of a CO interloper is extremely low.

HST broadband imaging. For the field surrounding QSO J2054-0005, the HST archival database has two optical observations, F606W(V) and F814W(I)⁵⁴, and three near-infrared broadband images, F105W(Y), F125W(J) and F160W(H). These broad filter images are used to estimate the photometric redshifts of targets in this field through the use of the Lyman break dropout technique⁵⁵. Galaxies at $z \approx 6$ 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 'EPSFBuilder' in Astropy v.4.2.1. These stars were matched by using the Gaia archival catalogue. After matching PSFs to that of F160W, the measured FWHMs of the PSF in all five bands are $0.23''$ (consistent with the FWHM of the PSF in F160W). These results are also consistent with the PSF-matching results in the CANDELS Multi-wavelength Catalog⁵⁶. 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.

[C II] luminosity, SFR and halo mass estimation. The velocity-integrated [C II] flux density is measured to be $0.0758 \pm 0.0177 \text{ Jy km s}^{-1}$, corresponding to a luminosity of $L_{[\text{C II}]} = (7.0 \pm 1.7) \times 10^7 L_{\odot}$. According to ref.²⁴, $L_{[\text{C II}]} / L_{\odot} = 1.04 \times 10^{-3} \Delta v \nu_0 (1+z)^{-1} D_L^2$, where $\nu_0 = 1,900.5369 \text{ GHz}$ is the rest-frame [C II] line frequency, Δv is the integrated line flux in Jy km s^{-1} and D_L is the luminosity distance in Mpc. The [C II]-based SFR, $\text{SFR}_{[\text{C II}]}$, is approximately $6.8 \pm 1.7 M_{\odot} \text{ yr}^{-1}$ (with 0.3 dex uncertainty)²⁵. Comparing with the ALPINE survey²⁹, this [C II] emitter is located at the faint end. Nevertheless, in a random survey, the probability of finding such a galaxy in this survey volume is 2×10^{-5} . Also, although the predicted SFRs of O I associated galaxies are typically lower than $\sim 0.1 M_{\odot} \text{ yr}^{-1}$ in these cosmological simulations, we detect a bright galaxy with an SFR of $6.8 M_{\odot} \text{ yr}^{-1}$. Therefore, for galaxies associated with strong O I absorber, this detection is exceptional.

Because the star formation rate and stellar mass information are not shown in ref.¹⁴, a comparison between the halo masses is conducted for a direct comparison with cosmological simulations. We derive the halo mass, M_{h} , directly from a [C II] luminosity– M_{h} relation proposed by simulating galaxies with an SFR of $0.01 M_{\odot} \text{ yr}^{-1}$ to $330 M_{\odot} \text{ yr}^{-1}$ at $z \approx 6$ (ref.¹³). Galaxies with an SFR similar to that of [C II]2054 can be reproduced in this simulation. Using the relation between [C II] luminosity and halo mass in their work (equation (8) in ref.¹⁴), we derive a halo mass of [C II]2054 is $4.1 \times 10^{11} M_{\odot}$ with a 1σ scattering of 0.5 dex.

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 [C II]-derived total SFR to the stellar mass^{27,37,38}; the predicted stellar mass of [C II]2054 is then $5.8 \times 10^8 M_{\odot}$ with 0.3 dex uncertainty. We further calculate the dynamical mass and the molecular gas fractions of $f_{\text{molgas}} = \frac{M_{\text{molgas}}}{M_{\text{molgas}} + M_{\text{star}}}$ (where M_{star} represents the stellar mass of a galaxy) to confirm the estimation of stellar mass. From the spectrum of [C II]2054, the [C II] velocity dispersion ($\sigma_{[\text{C II}]}$) is 187 km s^{-1} , corresponding to a circular velocity of $v_{\text{c}} = 1.763 \sigma_{[\text{C II}]} / \sin(i)$. Thus, considering the inclination, the lower limit of v_{c} is 329.7 km s^{-1} . Furthermore, according to the mentioned unresolved size of [C II]2054, the circularized effective radius is $r_{\text{e}} = 4.8 \text{ kpc}$ (ref.³⁹). The disk-like gas dynamical mass is estimated from equation (4) in ref.⁴⁰, $\frac{M_{\text{dyn}}}{M_{\odot}} = 1.16 \times 10^5 \left(\frac{v_{\text{c}}}{\text{km s}^{-1}} \right) \left(\frac{2r_{\text{e}}}{\text{kpc}} \right)$. Therefore, the derived dynamical mass is $3.7 \times 10^8 M_{\odot}$, consistent with our derived stellar mass³⁸. Then, using the relation between [C II] luminosity and molecular gas mass⁴¹, the derived molecular gas mass ($M_{\text{molgas}}^{[\text{C II}]}$) is $2.0 \times 10^9 M_{\odot}$. Thus, the molecular gas fraction is 0.78, consistent with the results from the ALPINE survey⁴⁰. Then, we can estimate the halo mass from the stellar mass based on equation (1) of ref.²⁸. The predicted halo mass is about $1.1 \times 10^{11} M_{\odot}$. Note that the conversion between the stellar mass and halo mass has a scatter of 0.3 dex. In this method, the halo mass is estimated by using three scaling relations: [C II] luminosity–SFR, SFR–stellar mass and stellar mass–halo mass. All of these derivations have scatters 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.

[C II]2054 continuum properties. No continuum was detected at the position of [C II]2054, resulting in a 3σ upper limit of 37 mJy per beam. To convert this continuum flux to a total infrared luminosity (TIR; $8\text{--}1,000 \text{ }\mu\text{m}$), we estimate the TIR luminosity by fitting a modified blackbody (that is, $S_{\nu} \approx \nu^{\beta} / (\exp(h\nu/kT_{\text{dust}}) - 1)$) to this single upper limit. We assume that [C II]2054 has the same continuum properties as a typical Lyman break galaxy at $z \approx 6$, and therefore we adopt a dust temperature of $T_{\text{dust}} = 38 \text{ K}$ and emissivity index of $\beta = 2.0$ (ref.⁴²). This translates to a 3σ upper limit of $L_{\text{TIR}} = 7.3 \times 10^{10} L_{\odot}$. Assuming a greybody spectrum^{26,43} and the dust temperature and emissivity index mentioned previously, we obtain a TIR-based SFR of $\text{SFR}_{\text{TIR}} < 11 M_{\odot} \text{ yr}^{-1}$. [C II]2054 also was not detected in any of the HST broadband imaging using SExtractor⁴⁴ (Extended Data Fig. 6). From the HST imaging, we determine a 3σ upper limit of the ultraviolet continuum at the position of [C II]2054 of $3.413 \times 10^{44} \text{ erg s}^{-1}$ (F105W band). This corresponding to an upper limit of the ultraviolet-based SFR of $\text{SFR}_{\text{UV}} < 15 M_{\odot} \text{ yr}^{-1}$.

Continuum sources. There are five HST broadband observations in this field (F606W(V), F814W(I), F105W(Y), F125W(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 approach⁴⁵. Model spectra of different galaxies at different redshifts are generated from the models presented in the galaxy template model (BC03) (ref. 46). For the galaxy models, we take the metallicity to be $Z = 0.02 Z_{\odot}$ (where $Z_{\odot}$ 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 $1 M_{\odot} \text{ yr}^{-1}$ with an exponentially declining star formation history with timescales of $\tau_{\text{SFR}} = 20 \text{ Gyr}$. To convert these galaxy templates to colours, we first apply a dust attenuation⁴⁷ with $E(B-V) = 0.0-0.4$. Then, we take into account the effects of IGM attenuation⁴⁸. 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 $z > 4$:

$$F606W - F814W > 2 \quad (4)$$

$$F814W - F105W < 3 \quad (5)$$

$$F606W - F814W > 1.5 \times (F814W - F105W) + 0.5 \quad (6)$$

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 detected in the F606W filter, which could indicate that they are V-dropout galaxy candidates. In the right panel of Extended Data Fig. 7, we further study their colour using the F814W, F105W and F125W filters and use the following criteria to select $z \approx 6$ galaxies:

$$F814W - F105W > 0.2 \quad (7)$$

$$F105W - F125W < 0.4 \quad (8)$$

$$F814W - F105W > 1.6 \times (F105W - F125W) + 0.36 \quad (9)$$

With these criteria, C1-1 and C5-3 are both ruled out as galaxies at $z \approx 6$. We conclude that all of the continuum sources except C2 are regarded as low- z foreground galaxies. The redshift of C2 is difficult to constrain without the F606W and F814W imaging.

Correlation length and power-law index calculation. To calculate the correlation length, we adopt the method displayed in ref. 9, $\xi_{g-\text{abs}}(r) = \frac{1}{n_0} \frac{\Delta N(r)}{\Delta V} - 1$, where $\Delta N(r)$ represents the mean number of galaxies within a finite spherical shell of volume ΔV located a distance r from an absorber, and n_0 is the mean number density of galaxies averaged over the entire simulation volume. This equation can be rearranged as $n_0 [1 + \xi_{g-\text{abs}}(r)] = \frac{\Delta N(r)}{\Delta V}$. To quantify this relation more physically, we assume that this correlation obeys a power law by introducing the correlation length (r_0) and a power-law index (γ), that is, $\xi_{g-\text{abs}}(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$. After reducing $\xi_{g-\text{abs}}$ in these two equations, we obtain $n_0 \left[1 + \left(\frac{r}{r_0}\right)^{-\gamma}\right] = \frac{\Delta N(r)}{\Delta V}$. By integrating this equation over r , we get $4\pi \int_0^r n_0 \left[1 + \left(\frac{r}{r_0}\right)^{-\gamma}\right] r^2 dr = N$. We detect only one galaxy in the survey volume. Thus, $N = 1$ and $4\pi \int_0^r n_0 \left[1 + \left(\frac{r}{r_0}\right)^{-\gamma}\right] r^2 dr = 1$. Finally, we obtain $\log(r_0) = \frac{\log[A(\gamma)]}{\gamma} + \log(r)$, where $A = \left(\frac{1}{4\pi n_0} - \frac{r^3}{3}\right) \times \frac{3-\gamma}{r^3}$.

To compare with observations, the number of galaxies that are brighter than [CII]2054 within the observed projection impact parameter ($r = 20 \text{ pkpc}$) are calculated. The mean number density (n_0) of [CII] emitters at this redshift was calculated by integrating the [CII] luminosity function²⁹. Thus, the mean density of [CII] emitters at $z \approx 6$ is $1.4 \times 10^{-3} \text{ cMpc}^{-3}$. Assuming a spherical volume with a radius of 20 pkpc, this volume containing one such bright galaxy has a mean overdensity of $\Delta = 5 \times 10^4$. Thus, with a fixed r , the r_0 and γ relation can be obtained and shown as in Fig. 3.

Comparison between observations and different simulations. We compare the measured impact parameter and the derived halo mass of [CII]2054 and that of the other three simulations, that is, Sherwood, HVEL and FAST simulations. The comparisons are shown in Extended Data Fig. 8. These absorbers are selected with

an REW of $0.12 \pm 0.05 \text{ \AA}$. We find that, as the mass of host halo increases, absorbers can be ejected to increasing distances from the halo. These absorbers are located closest to the host halos with typical masses of $10^{10} M_{\odot}$. However, the derived halo mass of [CII]2054 is one order of magnitude larger than the median value predicted by all of these simulations. Our results suggest that low-mass systems in these simulations are overly efficient at producing metal absorbers. Furthermore, our observations strongly constrain all of these simulations.

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^{-5} , 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/aq/> 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

- Codoreanu, A. et al. The comoving mass density of Mg II from $z \sim 2$ to 5.5. *Mon. Not. R. Astron. Soc.* **472**, 1023–1051 (2017).
- Becker, G. D. et al. The evolution of O I over $3.2 < z < 6.5$: reionization of the circumgalactic medium. *Astrophys. J.* **883**, 163 (2019).
- Cooper, T. J. et al. Heavy element absorption systems at $5.0 < z < 6.8$: metal-poor neutral gas and a diminishing signature of highly ionized circumgalactic matter. *Astrophys. J.* **882**, 77 (2019).
- 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).
- Bielby, R. M. et al. Into the Ly α jungle: exploring the circumgalactic medium of galaxies at $z \sim 4-5$ with MUSE. *Mon. Not. R. Astron. Soc.* **493**, 5336–5356 (2020).
- Díaz, 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).
- 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).
- Finlator, K. et al. The host haloes of O I absorbers in the reionization epoch. *Mon. Not. R. Astron. Soc.* **436**, 1818–1835 (2013).
- Finlator, K., Doughty, C., Cai, Z. & Díaz, G. The faint host galaxies of C IV absorbers at $z > 5$. *Mon. Not. R. Astron. Soc.* **493**, 3223–3237 (2020).
- Dijkstra, M., Haiman, Z. & Spaans, M. Ly α radiation from collapsing protogalaxies. I. Characteristics of the emergent spectrum. *Astrophys. J.* **649**, 14–36 (2006).
- Zheng, Z., Cen, R., Trac, H. & Miralda-Escudé, J. Radiative transfer modeling of Ly α emitters. I. Statistics of spectra and luminosity. *Astrophys. J.* **716**, 574–598 (2010).
- Hayes, M. et al. Escape of about five per cent of Lyman- α photons from high-redshift star-forming galaxies. *Nature* **464**, 562–565 (2010).
- Leung, T. K. D. et al. Predictions of the $L_{[\text{CII}]}$ -SFR and [CII] luminosity function at the epoch of reionization. *Astrophys. J.* **905**, 102 (2020).
- Keating, L. C., Puchwein, E., Haehnelt, M. G., Bird, S. & Bolton, J. S. Testing the effect of galactic feedback on the IGM at $z \sim 6$ with metal-line absorbers. *Mon. Not. R. Astron. Soc.* **461**, 606–626 (2016).
- Oh, S. P. Probing the dark ages with metal absorption lines. *Mon. Not. R. Astron. Soc.* **336**, 1021–1029 (2002).
- Furlanetto, S. R. & Loeb, A. Metal absorption lines as probes of the intergalactic medium prior to the reionization epoch. *Astrophys. J.* **588**, 18–34 (2003).
- Becker, G. D., Sargent, W. L. W., Rauch, M. & Calverley, A. P. High-redshift metals. II. Probing reionization galaxies with low-ionization absorption lines at redshift six. *Astrophys. J.* **735**, 93 (2011).
- Vogelsberger, M. et al. Properties of galaxies reproduced by a hydrodynamic simulation. *Nature* **509**, 177–182 (2014).

19. Bolton, J. S. et al. The Sherwood simulation suite: overview and data comparisons with the Lyman α forest at redshifts $2 \leq z \leq 5$. *Mon. Not. R. Astron. Soc.* **464**, 897–914 (2017).
20. Bird, S. et al. Damped Lyman α absorbers as a probe of stellar feedback. *Mon. Not. R. Astron. Soc.* **445**, 2313–2324 (2014).
21. Uzgil, B. D. et al. The ALMA Spectroscopic Survey in the HUDF: a search for [CII] emitters at $6 \leq z \leq 8$. *Astrophys. J.* **912**, 67 (2021).
22. Carilli, C. L. & Walter, F. Cool gas in high-redshift galaxies. *Annu. Rev. Astron. Astrophys.* **51**, 105–161 (2013).
23. Lagache, G., Cousin, M. & Chatzikos, M. The [CII] 158 μ m line emission in high-redshift galaxies. *Astron. Astrophys.* **609**, A130 (2018).
24. Wang, R. et al. Star formation and gas kinematics of quasar host galaxies at $z \sim 6$: new insights from ALMA. *Astrophys. J.* **773**, 44 (2013).
25. Schaerer, D. et al. The ALPINE-ALMA [CII] survey. Little to no evolution in the [CII]-SFR relation over the last 13 Gyr. *Astron. Astrophys.* **643**, A3 (2020).
26. Murphy, E. J. et al. Calibrating extinction-free star formation rate diagnostics with 33 GHz free-free emission in NGC 6946. *Astrophys. J.* **737**, 67 (2011).
27. Salmon, B. et al. The relation between star formation rate and stellar mass for galaxies at $3.5 \leq z \leq 6.5$ in CANDELS. *Astrophys. J.* **799**, 183 (2015).
28. Ma, X. et al. Simulating galaxies in the reionization era with FIRE-2: galaxy scaling relations, stellar mass functions, and luminosity functions. *Mon. Not. R. Astron. Soc.* **478**, 1694–1715 (2018).
29. Yan, L. et al. The ALPINE-ALMA [C II] survey: [C II] 158 μ m emission line luminosity functions at $z \sim 4$ –6. *Astrophys. J.* **905**, 147 (2020).
30. Smit, R. et al. The star formation rate function for redshift $z \sim 4$ –7 galaxies: evidence for a uniform buildup of star-forming galaxies during the first 3 Gyr of cosmic time. *Astrophys. J.* **756**, 14 (2012).
31. Walter, F. et al. ALMA Spectroscopic Survey in the Hubble Ultra Deep Field: survey description. *Astrophys. J.* **833**, 67 (2016).
32. De Looze, I., Baes, M., Bendo, G. J., Cortese, L. & Fritz, J. The reliability of [CII] as an indicator of the star formation rate. *Mon. Not. R. Astron. Soc.* **416**, 2712–2724 (2011).
33. 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. 376, 127 (Astronomical Society of the Pacific, 2007).
34. Marshall, M. A. et al. Limits to rest-frame ultraviolet emission from far-infrared-luminous $z \sim 6$ quasar hosts. *Astrophys. J.* **900**, 21 (2020).
35. Shapley, A. E., Steidel, C. C., Pettini, M. & Adelberger, K. L. Rest-frame ultraviolet spectra of $z \sim 3$ Lyman break galaxies. *Astrophys. J.* **588**, 65–89 (2003).
36. Stefanon, M. et al. CANDELS multi-wavelength catalogs: source identification and photometry in the CANDELS Extended Groth Strip. *Astrophys. J. Suppl. Ser.* **229**, 32 (2017).
37. 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 $z \sim 0$ –6. *Astrophys. J. Suppl. Ser.* **214**, 15 (2014).
38. Capak, P. L. et al. Galaxies at redshifts 5 to 6 with systematically low dust content and high [CII] emission. *Nature* **522**, 455–458 (2015).
39. 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).
40. Dessauges-Zavadsky, M. et al. The ALPINE-ALMA [CII] survey. Molecular gas budget in the early Universe as traced by [CII]. *Astron. Astrophys.* **643**, A5 (2020).
41. Zanella, A. et al. The [CII] emission as a molecular gas mass tracer in galaxies at low and high redshifts. *Mon. Not. R. Astron. Soc.* **481**, 1976–1999 (2018).
42. Faisst, A. L. et al. ALMA characterizes the dust temperature of $z \sim 5.5$ star-forming galaxies. *Mon. Not. R. Astron. Soc.* **498**, 4192–4204 (2020).
43. Izumi, T. et al. Subaru high- z exploration of low-luminosity quasars (SHELLQs). III. Star formation properties of the host galaxies at $z \geq 6$ studied with ALMA. *Publ. Astron. Soc. Jpn* **70**, 36 (2018).
44. Bertin, E. & Arnouts, S. SExtractor: software for source extraction. *Astron. Astrophys. Suppl. Ser.* **117**, 393–404 (1996).
45. Ono, Y. et al. Great Optically Luminous Dropout Research Using Subaru HSC (GOLDRUSH). I. UV luminosity functions at $z \sim 4$ –7 derived with the half-million dropouts on the 100 deg² sky. *Publ. Astron. Soc. Jpn* **70**, S10 (2018).
46. Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).
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. Neeleman, M. et al. [CII] 158- μ m 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 Ly α galaxies at $z > 2$ – III. The star formation rates of neutral gas reservoirs at $z \sim 2.7$. *Mon. Not. R. Astron. Soc.* **446**, 3178–3198 (2015).
51. Loiacono, F. et al. The ALPINE-ALMA [C II] survey. Luminosity function of serendipitous [C II] line emitters at $z \sim 5$. *Astron. Astrophys.* **646**, A76 (2021).
52. Steidel, C. C. et al. The structure and kinematics of the circumgalactic medium from far-ultraviolet spectra of $z \sim 2$ –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 [CII] 158 μ m 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: search for [CII] 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 Y.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-HF2-51448.001-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 J2054 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 J.X.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., J.F.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>.

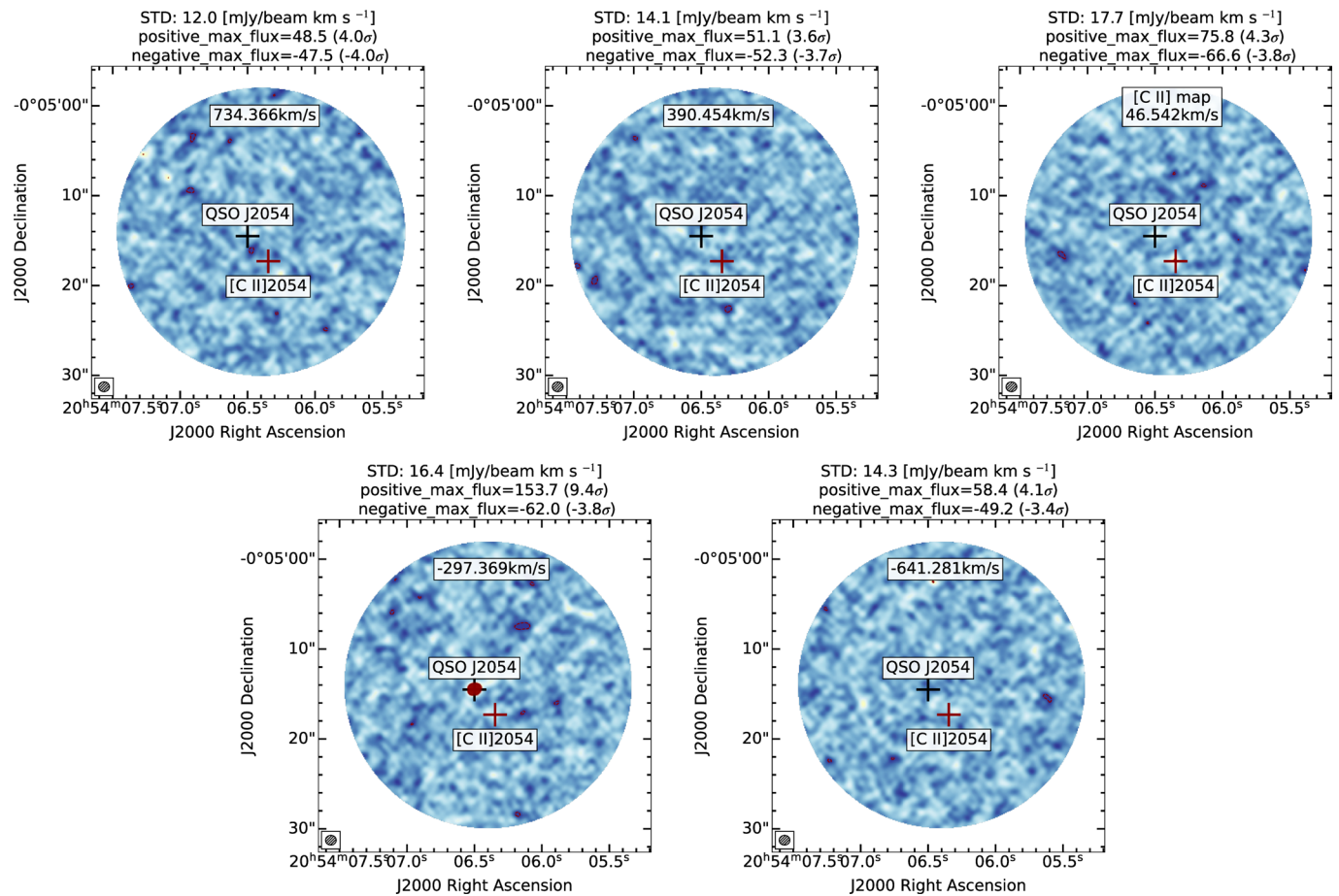
Correspondence and requests for materials should be addressed to Zheng Cai.

Peer review information *Nature Astronomy* thanks Matthieu Béthermin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

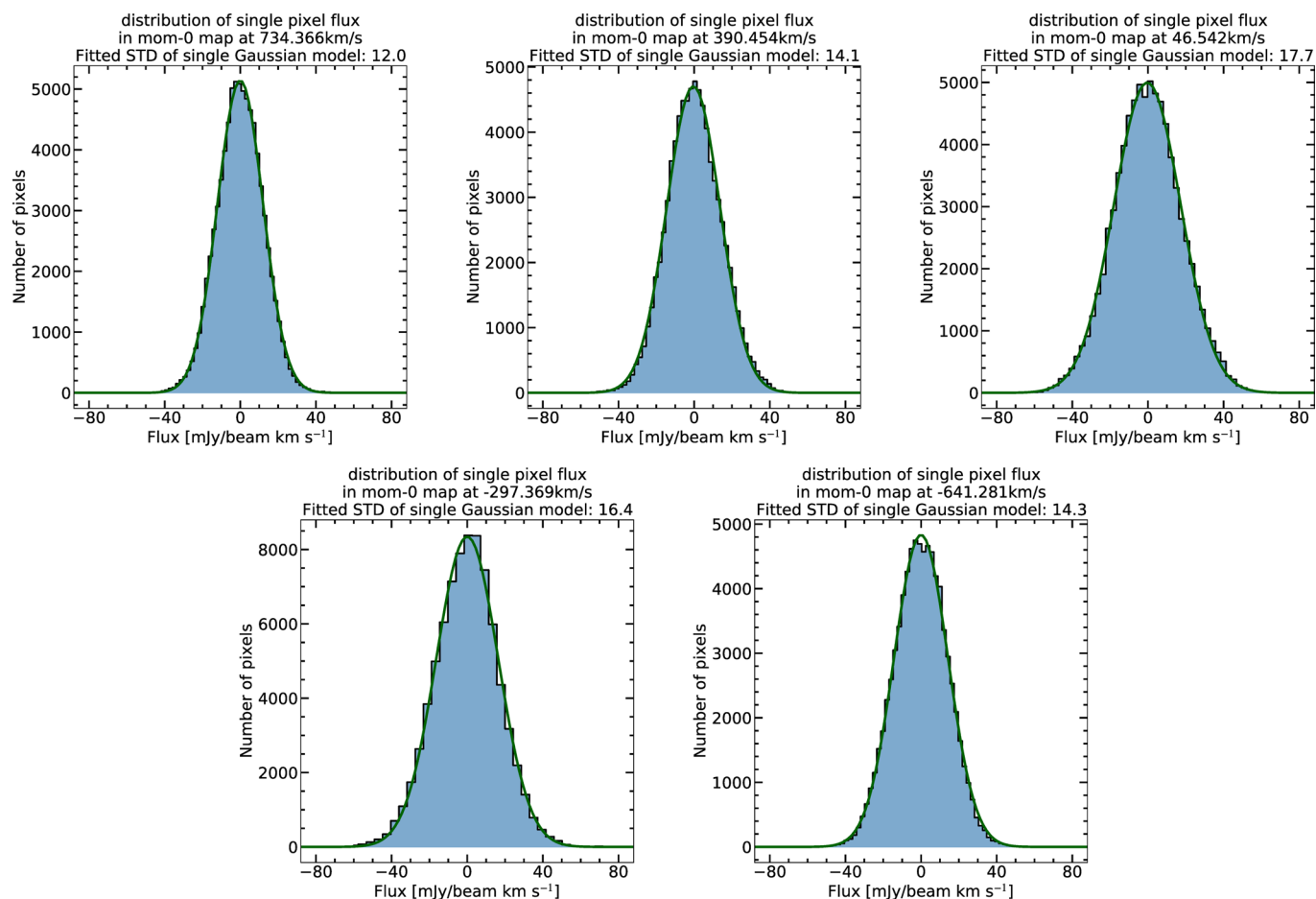
Reprints and permissions information is available at www.nature.com/reprints.

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

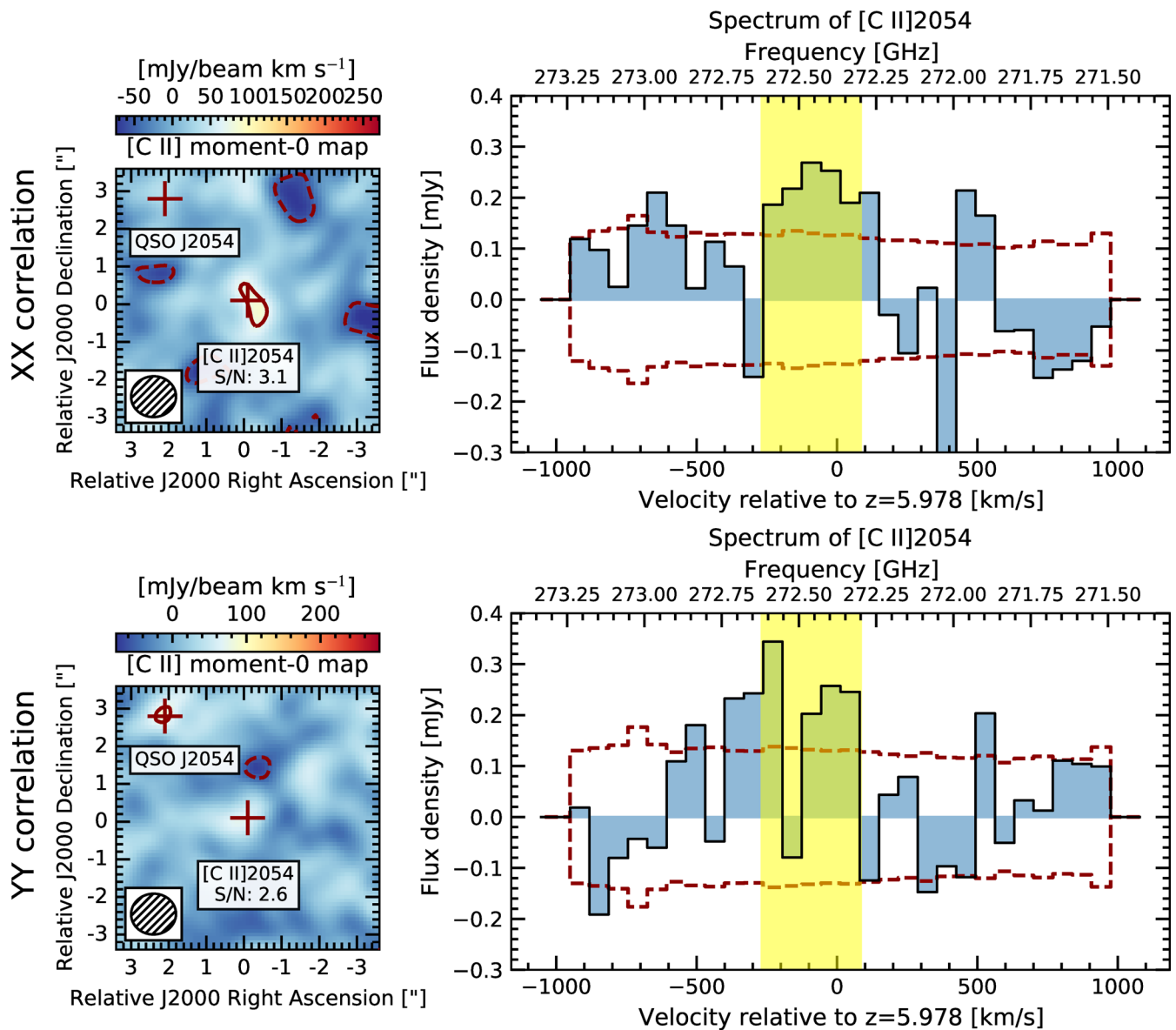
© The Author(s), under exclusive licence to Springer Nature Limited 2021



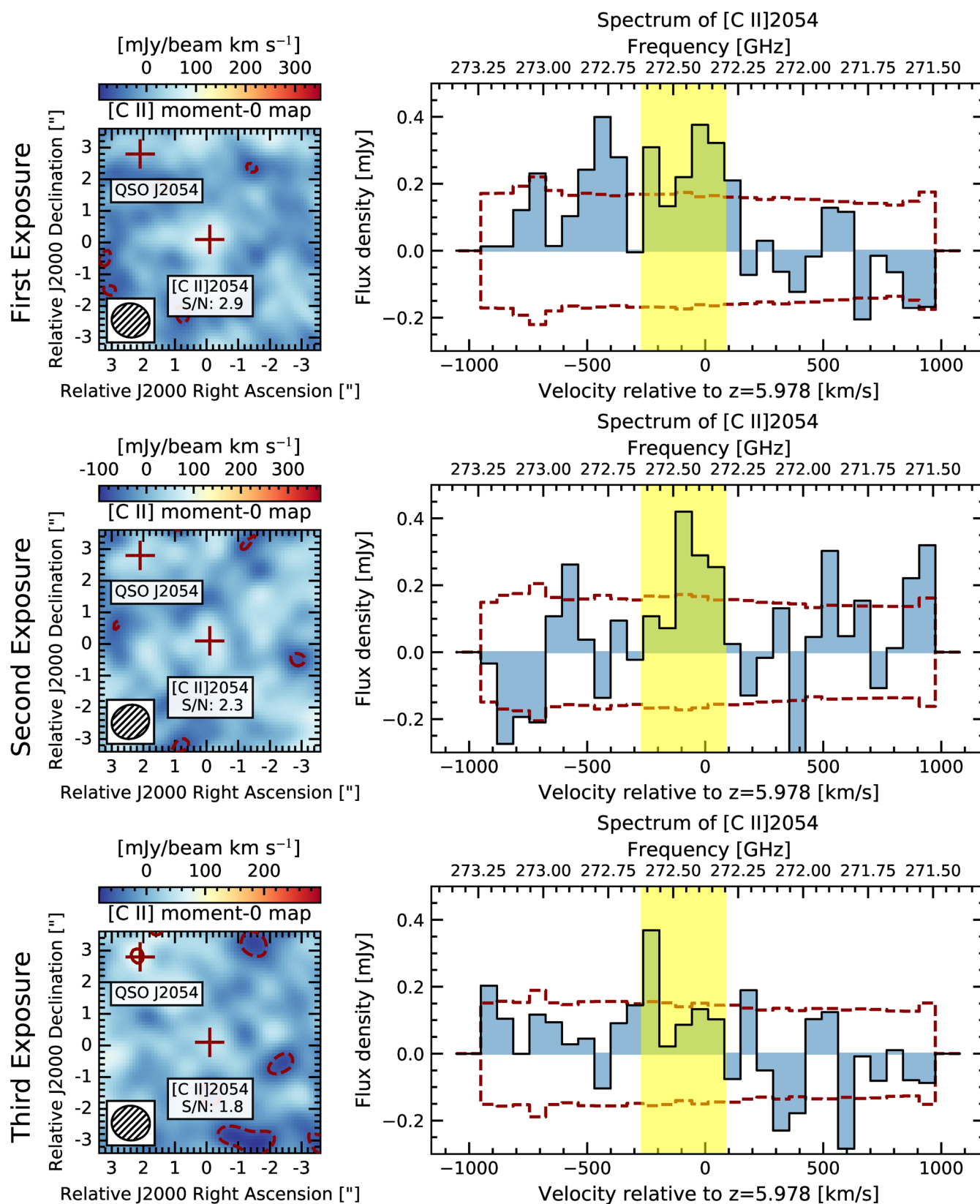
Extended Data Fig. 1 | Channel maps in the whole observed field of view. These channel maps are arranged from -641.281 to + 734.366 km/s. The channel width is set as the same as [C II] intensity map as 344 km/s for each map. The black and dark red cross indicate the position of QSO J2054 and [C II]2054, respectively. The sizes of the synthesized beams are demonstrated in the bottom-left of these panel.



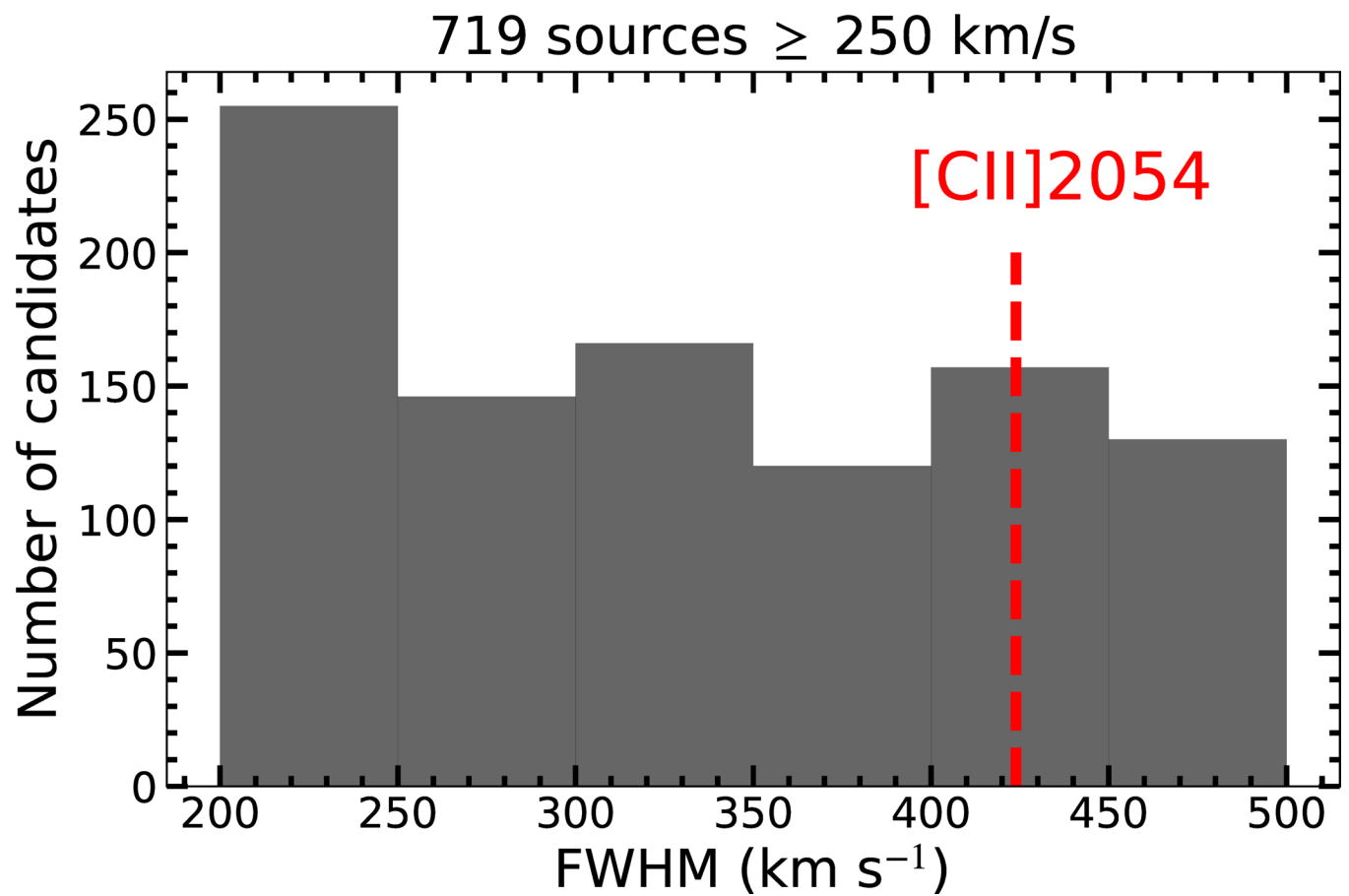
Extended Data Fig. 2 | The detailed pixel flux distributions of different Channel maps. The green lines show the best-fit single-Gaussian models. We regard the single-Gaussian fitted standard deviation (STD) as the noise in these intensity maps.



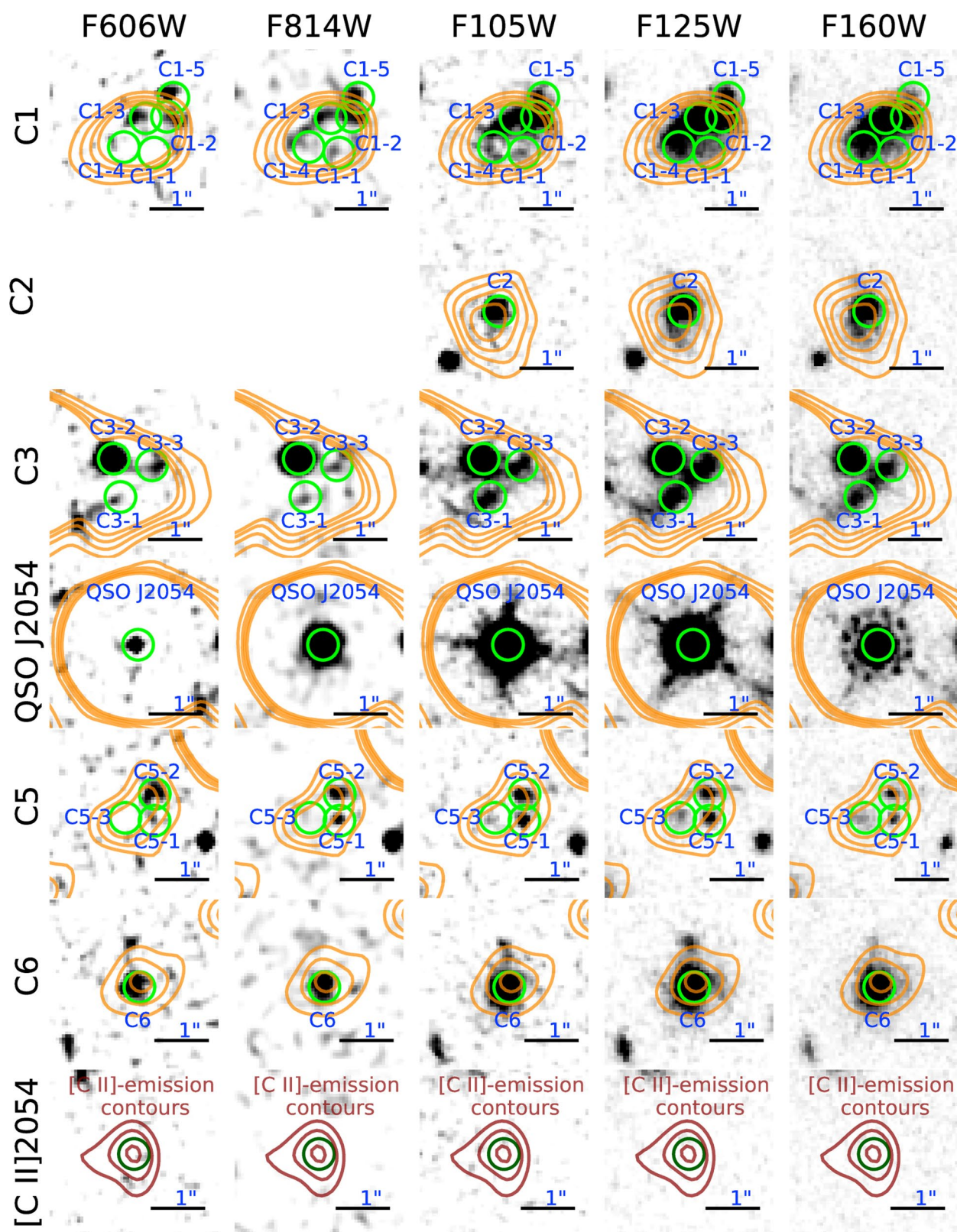
Extended Data Fig. 3 | Different polarization-correlation maps. Upper panels, The moment-0 and spectrum under XX polarization. The [CII] moment-0 maps is collapsed based on the same emission range as Fig. 1 and shown by the yellow shaded region. The outer contour is at 3σ level, with contours in steps of $1-\sigma$. The $1-\sigma$ rms is 2.47×10^{-2} Jy beam⁻¹ km s⁻¹. Dashed lines represent -2σ -level contours. The integral [CII] flux is 0.0768 ± 0.0247 Jy 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 [CII] flux is 0.0667 ± 0.0254 Jy km s⁻¹, while the $1-\sigma$ is 2.54×10^{-2} Jy beam⁻¹ km s⁻¹.



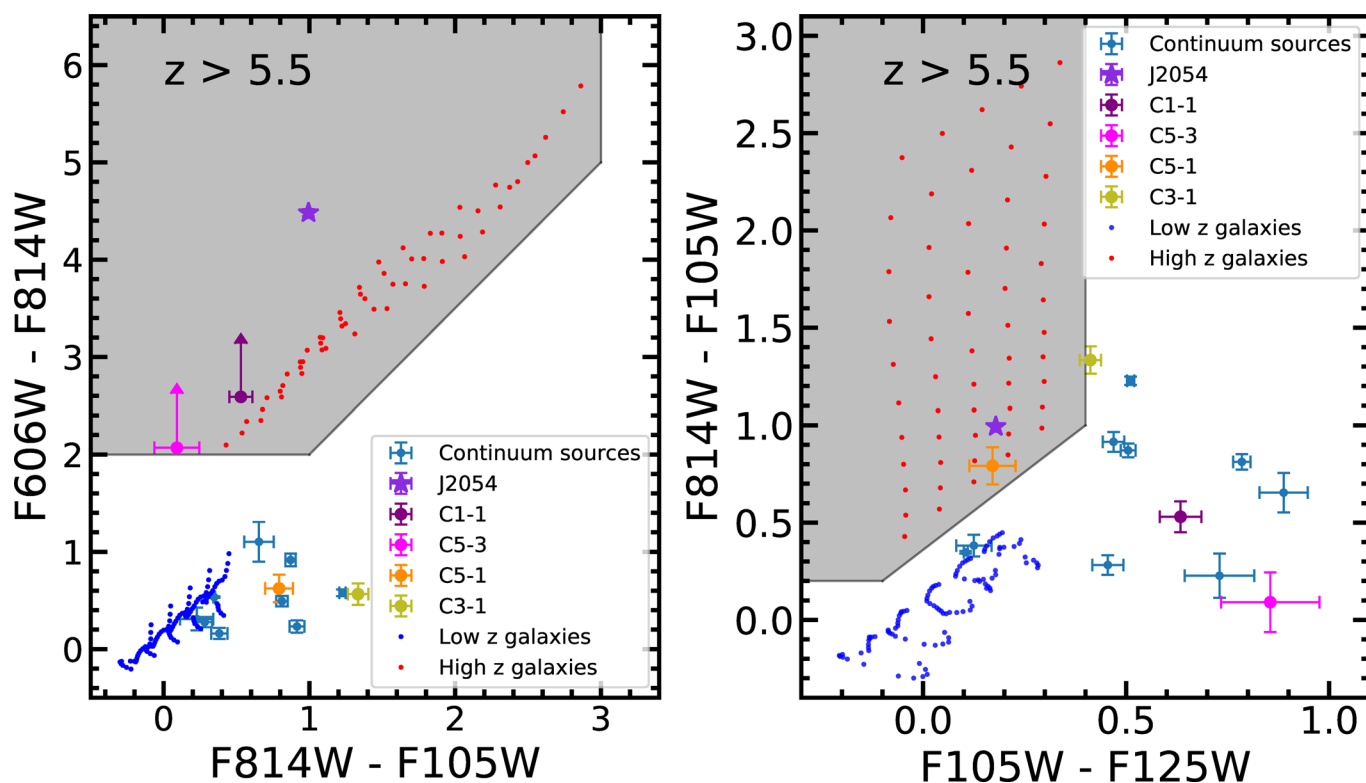
Extended Data Fig. 4 | Multi-exposure observations. Left panel: The intensity maps of three individual exposures. The integral [C II] fluxes are 0.0935, 0.0786 and 0.0490 Jy beam⁻¹ km s⁻¹. Meanwhile the 1- σ standard deviation are 0.0319, 0.0335, and 0.0269 Jy beam⁻¹ km s⁻¹, respectively. Right panel: The corresponding 1-D spectra of [C II]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 [C II]moment-0 maps.



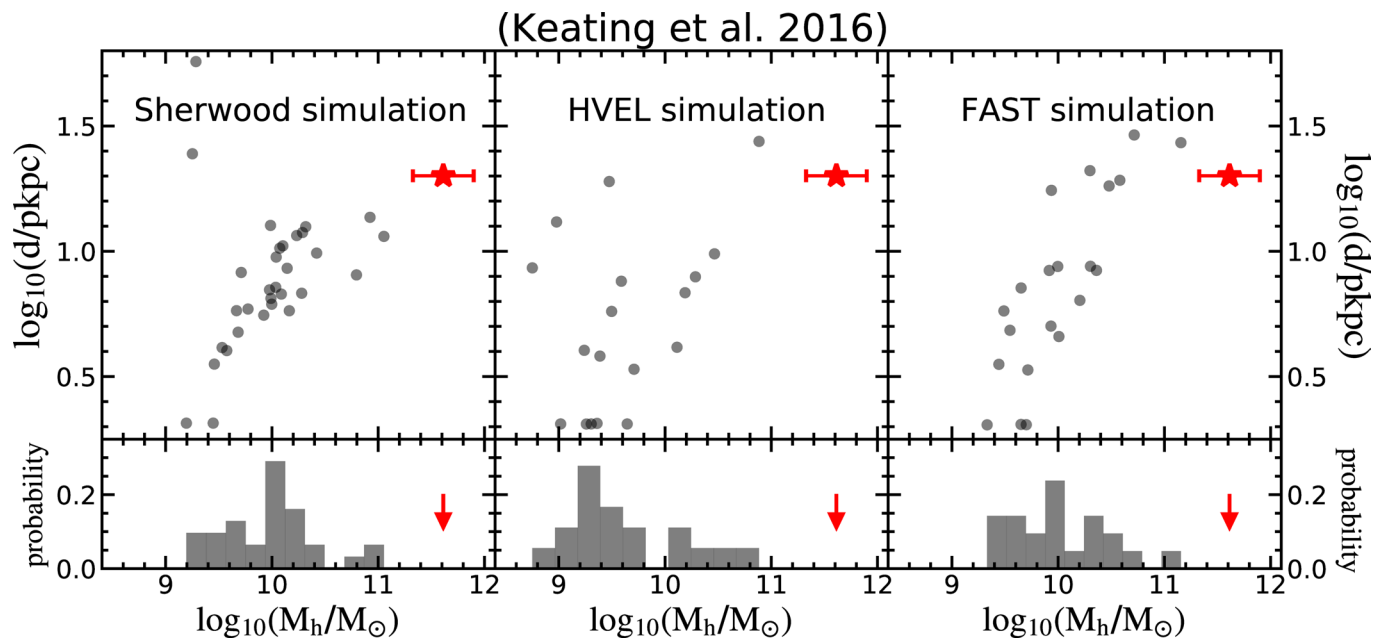
Extended Data Fig. 5 | The number of candidates in the deep ASPECS Band-6 datacube. We found 719 sources having $\text{FWHM} \geq 250 \text{ km s}^{-1}$ in a 4.2 arcmin^2 area of the ASPECS survey. In the figure, sources with Full Width Half Maximum (FWHM) between 200 and 500 km s^{-1} are present. The 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 1σ 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 $1\text{-}\sigma$ uncertainties of the estimated halo mass. In the three top panels, grey dots represent OI absorbers with the REW of $0.12 \pm 0.05 \text{ \AA}$ (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.