LETTER

https://doi.org/10.1038/s41586-018-0025-2

A massive core for a cluster of galaxies at a redshift of 4.3

T. B. Miller[1,2]*, S. C. Chapman[1,3,4], M. Aravena[5], M. L. N. Ashby[6], C. C. Hayward[6,7], J. D. Vieira[8], A. Weiß[9], A. Babul[10], M. Béthermin[11], C. M. Bradford[12,13], M. Brodwin[14], J. E. Carlstrom[15,16,17,18], Chian-Chou Chen[19], D. J. M. Cunningham[1,20], C. De Breuck[19], A. H. Gonzalez[21], T. R. Greve[22], J. Harnett[23], Y. Hezaveh[24], K. Lacaille[1,25], K. C. Litke[26], J. Ma[21], M. Malkan[27], D. P. Marrone[26], W. Morningstar[24], E. J. Murphy[28], D. Narayanan[21], E. Pass[1,29], R. Perry[1], K. A. Phadke[8], D. Rennehan[10], K. M. Rotermund[1], J. Simpson[30,31], J. S. Spilker[26], J. Sreevani[8], A. A. Stark[6], M. L. Strandet[9,32] & A. L. Strom[33]

Massive galaxy clusters have been found that date to times as early as three billion years after the Big Bang, containing stars that formed at even earlier epochs[1–3]. The high-redshift progenitors of these galaxy clusters—termed 'protoclusters'—can be identified in cosmological simulations that have the highest overdensities (greater-than-average densities) of dark matter[4–6]. Protoclusters are expected to contain extremely massive galaxies that can be observed as luminous starbursts[7]. However, recent detections of possible protoclusters hosting such starbursts[8–11] do not support the kind of rapid cluster-core formation expected from simulations[12]: the structures observed contain only a handful of starbursting galaxies spread throughout a broad region, with poor evidence for eventual collapse into a protocluster. Here we report observations of carbon monoxide and ionized carbon emission from the source SPT2349-56. We find that this source consists of at least 14 gas-rich galaxies, all lying at redshifts of 4.31. We demonstrate that each of these galaxies is forming stars between 50 and 1,000 times more quickly than our own Milky Way, and that all are located within a projected region that is only around 130 kiloparsecs in diameter. This galaxy surface density is more than ten times the average blank-field value (integrated over all redshifts), and more than 1,000 times the average field volume density. The velocity dispersion (approximately 410 kilometres per second) of these galaxies and the enormous gas and star-formation densities suggest that this system represents the core of a cluster of galaxies that was already at an advanced stage of formation when the Universe was only 1.4 billion years old. A comparison with other known protoclusters at high redshifts shows that SPT2349-56 could be building one of the most massive structures in the Universe today.

In a multiband survey over 2,500 deg^2 of sky, the South Pole Telescope (SPT) discovered a population of rare ($n \approx 0.04\,\mathrm{deg}^{-2}$), extremely bright (with observed flux densities (S) at 1.4 mm of more than 20 mJy), millimetre-selected sources[13,14]. Subsequent Atacama Large Millimeter/submillimeter Array (ALMA) imaging at wavelengths of 870 μm showed that more than 90% of these SPT-selected sources are single high-redshift submillimetre galaxies (SMGs)[15] with intrinsic flux densities of $S_{870\mu m} = 5$–10 mJy, gravitationally lensed by factors of 5–20 (ref. [16]), and with a median redshift of $z \approx 4$ (ref. [17]). However, about 10% of these sources show no evidence of lensing and may instead be intrinsically very luminous galaxies or even groups of many rapidly star-forming galaxies. We made observations at 870 μm using a low-resolution bolometer camera on the APEX telescope (LABOCA), revealing the brightest such source in the SPT 2,500 deg^2 survey—SPT2349-56 ($S_{1.4mm} = 23.3$ mJy)—to consist of two elongated sources with a combined flux density of $S_{870\mu m} \approx 110$ mJy (Fig. 1), with the brighter, southern source comprising about 77 mJy of this flux density. A redshift survey by ALMA[17] further resolved SPT2349-56 into a pair of bright 3-mm sources associated with the southern LABOCA source, with both members of the pair lying at $z = 4.3$.

To better understand the nature of this structure, we undertook deep ALMA spectral imaging of the brighter southern peak of the extended LABOCA source. We used a 358-GHz map containing the redshifted $[\mathrm{C\,II}]_{1900.5\mathrm{GHz}}$ line to search for line-emitting galaxies. We carried out a blind spectral-line survey (see Methods) of the data cube, revealing 14 line emitters at redshifts of about 4.31 with high significance (with a signal-to-noise ratio of more than 7). We detected 12 of these emitters individually in the 1.1-mm continuum map at a significance of more than 5σ, with 1.1-mm flux densities ranging from 0.2 mJy to 5 mJy (Fig. 1). The remaining two line emitters (M and N) are both detected at lower significance in the 1.1-mm continuum map, but have robust counterparts detected by the Infrared Array Camera (IRAC) on board the Spitzer Space Telescope (Extended Data Table 1 and Extended Data Fig. 1). Eight of these sources are also detected (at significances of more than 5σ) in the carbon monoxide (4–3) line. The ALMA spectra are shown in Fig. 1.

Measurements of both the continuum and the spectral lines of the 14 galaxies allow us to estimate their star-formation rates (SFRs) and gas masses (Table 1 and Extended Data Table 1). The physical properties of these sources indicate that this protocluster already harbours massive galaxies that are rapidly forming stars from an abundant gas supply. The two brightest sources, A and B, have SFRs in excess of 1,000 Solar masses per year ($M_\odot\,\mathrm{yr}^{-1}$) within their resolved, roughly 3-kpc radii. The total SFR of the 14 sources is $6{,}000 \pm 600\,M_\odot\,\mathrm{yr}^{-1}$. Multicolour imaging with the Spectral and Photometric Imaging Receiver (SPIRE) instrument on board the Herschel Space Observatory (at wavelengths of 250 μm, 350 μm and 500 μm), in addition to the 870-μm LABOCA

[1]Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada. [2]Department of Astronomy, Yale University, New Haven, CT, USA. [3]Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada. [4]National Research Council, Herzberg Astronomy and Astrophysics, Victoria, British Columbia, Canada. [5]Núcleo de Astronomía, Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Santiago, Chile. [6]Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA. [7]Center for Computational Astrophysics, Flatiron Institute, New York, NY, USA. [8]Department of Astronomy, University of Illinois, Urbana, IL, USA. [9]Max-Planck-Institut für Radioastronomie, Bonn, Germany. [10]Department of Physics and Astronomy, University of Victoria, Victoria, British Columbia, Canada. [11]Aix-Marseille Université, CNRS, LAM, Laboratoire d'Astrophysique de Marseille, Marseille, France. [12]California Institute of Technology, Pasadena, CA, USA. [13]Jet Propulsion Laboratory, Pasadena, CA, USA. [14]Department of Physics and Astronomy, University of Missouri, Kansas City, MO, USA. [15]Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL, USA. [16]Department of Physics, University of Chicago, Chicago, IL, USA. [17]Enrico Fermi Institute, University of Chicago, Chicago, IL, USA. [18]Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL, USA. [19]European Southern Observatory, Garching, Germany. [20]Department of Astronomy and Physics, Saint Mary's University, Halifax, Nova Scotia, Canada. [21]Department of Astronomy, University of Florida, Gainesville, FL, USA. [22]Department of Physics and Astronomy, University College London, London, UK. [23]School of Physics, University of Sydney, Sydney, New South Wales, Australia. [24]Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, Stanford, CA, USA. [25]Department of Physics and Astronomy, McMaster University, Hamilton, Ontario, Canada. [26]Steward Observatory, University of Arizona, Tucson, AZ, USA. [27]Department of Physics and Astronomy, University of California, Los Angeles, CA, USA. [28]National Radio Astronomy Observatory, Charlottesville, VA, USA. [29]Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada. [30]Institute for Astronomy, Royal Observatory, University of Edinburgh, Edinburgh, UK. [31]Centre for Extragalactic Astronomy, Department of Physics, Durham University, Durham, UK. [32]International Max Planck Research School (IMPRS) for Astronomy and Astrophysics, Bonn, Germany. [33]Observatories of The Carnegie Institution for Science, Pasadena, CA, USA. *e-mail: tim.miller@yale.edu

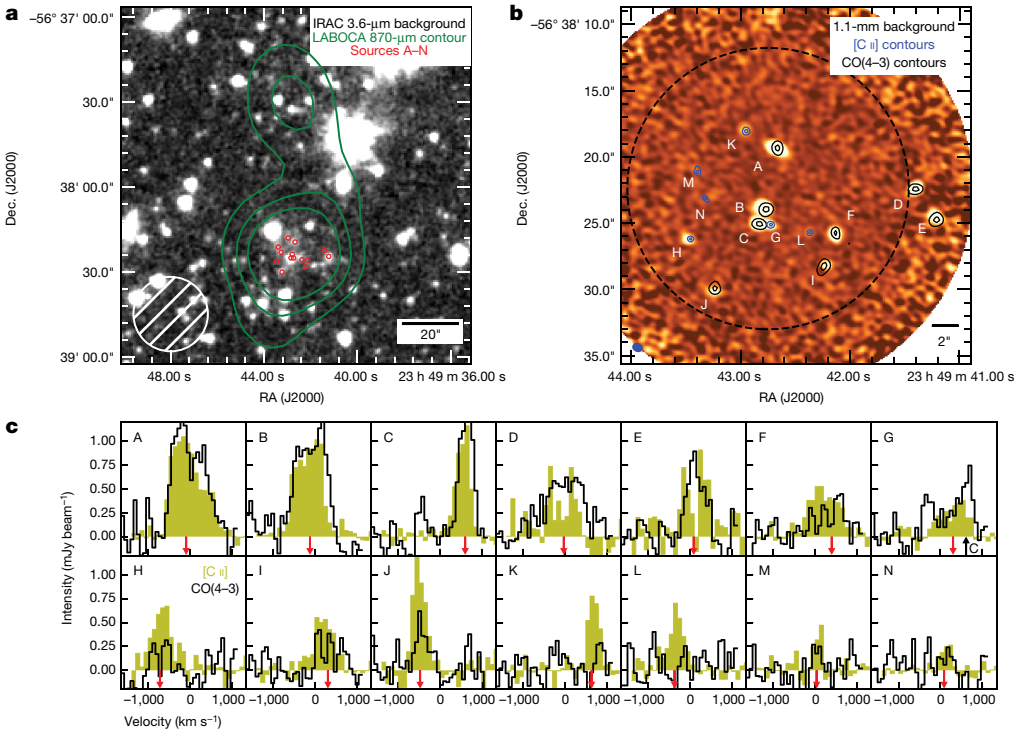


Fig. 1 | The SPT2349-56 field and spectra of the constituent galaxies. **a**, LABOCA 870-μm contours of SPT2349-56, overlaid on the IRAC 3.6-μm image; the 26″ beam at 870 μm is shown in white. Contours represent signal-to-noise rations of 3,7 and 9, moving inwards. The small red circles show the locations of the 14 protocluster sources. **b**, ALMA band-7 imaging (276 GHz, 1.1 mm), showing the 14 confirmed protocluster sources, labelled A to N. Black and blue contours denote 75% (outer contour) and 90% (inner contour) of the peak flux for each source, based on the CO(4–3) and [C II] lines, respectively. The dashed black line shows where the primary beam is at 50% of its maximum. The filled blue ellipse shows the 0.4″ naturally weighted synthesized beam. **c**, CO(4–3) spectra (black lines) and [C II] spectra (yellow bars) for all 14 sources, centred at the biweight cluster redshift $z = 4.304$. The [C II] spectra are scaled down in flux by a factor of ten, for clarity of presentation. The red arrows show the velocity offsets determined by fitting a Gaussian profile to the CO(4–3) spectra for all sources except for G, H, K, L, M and N, for which we used [C II] (because these sources are not detected in the CO(4–3) spectra).

map, suggest that the northern LABOCA structure also lies at $z = 4.3$ (see Methods). The sources detected in the ALMA 870-μm imaging therefore comprise just 50% of the total flux density of the southern LABOCA source, and 36% of the total LABOCA flux density, suggesting that the roughly 500-kpc extent of the protocluster contains a total star-formation rate of 16,500 $M_\odot$ yr^{-1}. Modelling the spectral energy distribution on the basis of this combined submillimetre photometry yields an infrared luminosity (at wavelengths from 8 μm to 1,100 μm) of $(8.0 \pm 1.0) \times 10^{13}$ times the Solar luminosity ($L_\odot$). The gas masses of the 14 protocluster galaxies—estimated from the CO(4–3) line, or from the [C II] line if undetected in CO(4–3) (see Methods)—range from $1 \times 10^{10} M_\odot$ to $1 \times 10^{11} M_\odot$, with a total gas mass of roughly 6×10^{11} $(X_{CO}/0.8) M_\odot$ (where X_{CO} is the conversion factor from CO(1–0) luminosity to total gas mass). A follow-up survey of colder molecular gas in the CO(2–1) line with the Australia Telescope Compact Array (ATCA, a radio telescope) detects the bulk of this large gas repository, especially in the central region near sources B, C and G, and confirms that the assumed line-intensity ratio, CO(4–3) to CO(1–0), used in the Methods when calculating the total gas mass, is consistent with the average measurements from ATCA.

Table 1 | Derived physical properties of SPT2349-56 protocluster members

Source	ΔV [km s^{-1}]†	SFR [$M_\odot$ yr^{-1}]	M_{gas} [$10^{10} M_\odot$]
A	−90±35	1,170±390	12.0±2.1
B	−124±31	1,227±409	11.2±2.0
C	603±12	907±302	6.7±1.2
D	−33±40	530±182	8.4±1.5
E	84±21	497±179	4.8±0.9
F	395±82	505±169	3.4±0.7
G	308±42	409±137	2.9±1.3‡
H	−719±28	310±105	4.4±2.0‡
I	310±78	268±91	2.2±0.5
J	−481±35	243±85	2.2±0.5
K	631±12	208±71	3.1±1.4‡
L	−379±18	122±43	3.3±1.5‡
M	34±21	75±34	1.2±0.6†
N	90±25	64±29	1.0±0.5†

†Velocity offsets were measured relative to the mean redshift, $z = 4.304$.
‡The [C II] line was used to derive M_{gas} in these cases, as CO(4–3) was not detected.

The detected ALMA sources also enable an initial estimate of the mass of the protocluster. We determine the mean redshift using the biweight estimator[18] to be $\langle z \rangle_{bi} = 4.3040^{+0.0020}_{-0.0019}$. The velocity dispersion of the galaxy distribution is $\sigma_{bi} = 408^{+82}_{-56}$ km s^{-1} according to the biweight method[18], which is the standard approach for galaxy samples of this size. Other common methods (gapper[18] and Gaussian fit) agree to within 3% and provide similar errors. Under the assumption that SPT2349-56 is approximately virialized, the mass-dispersion relation for galaxy clusters[19] indicates a dynamical mass of $M_{dyn} = (1.16 \pm 0.70) \times 10^{13} M_\odot$, which is an upper limit if the system has not yet virialized. Given the possible selection effect of requiring a bright source (with $S_{1.4mm}$ values of more than 15 mJy) within the 1′ SPT beam for detection, we also further consider the possibility that our structure may represent an end-on filament being projected into a compact but unbound configuration, rather than a single gravitationally bound halo. Our analysis in the Methods suggests that this is not as likely as a relatively bound system in a massive halo, given the velocity dispersion measured as a function of position, and other supporting arguments. However, we cannot rule this possibility out completely, and further analysis and observations of the larger angular scale of the structure will be required to more fully understand the nature of this system.

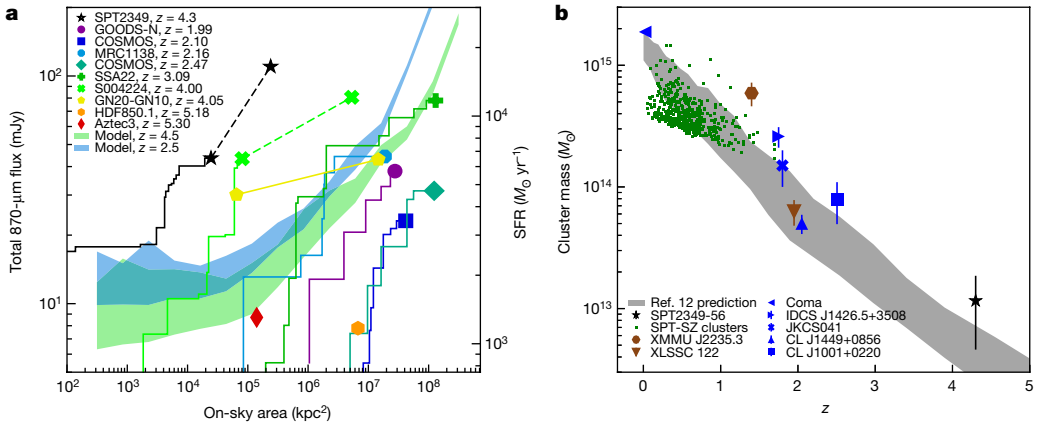


Fig. 2 | Comparison of SPT2349-56 to other cluster and protocluster systems. a, The cumulative 870-μm flux density is plotted against on-sky area for SPT2349-56 and for other SMG-rich overdensities at high redshifts (see Methods). The solid black line shows the ALMA-identified sources in SPT2349-56, while the dashed line includes the wider-field LABOCA-detected structure. The blue and green shaded regions denote the maximum flux density versus area curves obtained in 100 Monte Carlo realizations of a theoretical model for submillimetre-luminous protoclusters at $z = 4.5$ and $z = 2.5$, based on an N-body simulation (see Methods). Most of the literature SMG overdensities are consistent with the model expectations, whereas SPT2349-56 lies vastly above the region spanned by the model. A recently discovered $z = 4$ protocluster[23], SMM J004224 (light green cross), is quite a unique system, but more than ten times less dense than—and probably only about 50% the total luminosity of—SPT2349-56. **b**, Cluster mass versus redshift is shown for SPT2349-56 and for other massive galaxy clusters (from the literature) that have detected intracluster media and well defined masses. The colour scheme highlights the different methods used for selecting massive clusters (brown, X-ray; blue, optical; green, Sunyaev–Zel'dovich effect). Error bars represent 1σ standard deviation. We also show the mean protocluster most-massive-progenitor mass versus redshift relation predicted by N-body simulations[12]. The location of SPT2349-56 in this plane suggests a very massive descendant (with a halo mass of more than $10^{15} M_{\odot}$ at $z = 0$), although we caution that the complex growth histories of dark-matter halos make it difficult to reliably predict the $z = 0$ halo mass from the halo mass at a given epoch[20].

If the total halo mass represented by these 14 SMGs is indeed roughly $10^{13} M_{\odot}$, then the protocluster could conceivably be the progenitor of a galaxy cluster that is larger than $10^{15} M_{\odot}$, comparable to the Coma cluster at $z = 0$ (Fig. 2), as we deduce from simulations that track the hierarchical halo growth of an approximately $10^{15} M_{\odot}$ galaxy cluster from early epochs[12]. The location of SPT2349-56 in this particular plane suggests a very massive descendant, but we caution that N-body simulations indicate that it is difficult to reliably predict $z = 0$ halo masses from the halo mass at a given epoch, owing to the large halo-to-halo variations in dark-matter halo-growth histories[20].

To study the relative overdensity and concentration of SPT2349-56, it is desirable to compare it with other active protoclusters at high redshifts. SPT2349-56 is highly overdense, as it harbours ten SMGs with $S_{1.1\mathrm{mm}}$ values of more than about 0.5 (a level at which we are 'complete', meaning that we are not missing any sources and that there is uniform sensitivity across our search area) located within a circle of diameter 19″ (130 kpc), corresponding to a number density of $N(S_{1.1\mathrm{mm}} > 0.5\ \mathrm{mJy}) \approx 2 \times 10^4\ \mathrm{deg}^{-2}$. By comparison, the average number of field sources with $S_{1.1\mathrm{mm}}$ values of more than 0.5 mJy within this area across all redshifts is less than one[21]; thus, the SPT2349-56 field is overdense by more than a factor of ten. When we account for the fact that all sources in the SPT2349-56 field are at the same redshift, the volume density is more than 1,000 times the field density, assuming a redshift binning of $\Delta z = 0.1$ and the redshift distribution for SMGs[22]. In Fig. 2, we plot 'curves of growth' of the total 870-μm flux density against on-sky area for SPT2349-56 and other SMG-rich protoclusters (see Methods for details of the comparison sample). For SPT2349-56, we plot both the total flux density of the 14 confirmed protocluster members detected with ALMA, and the total flux density of the extended LABOCA structure. The curve of growth for SPT2349-56 rises much more steeply than those of the other high-redshift protoclusters, demonstrating its extreme density. For SPT2349-56, the on-sky area encompassing the accumulated 870-μm flux density (and thus approximately the total SFR) is as much as three orders of magnitude lower than for other protoclusters at $z > 2$. Thus SPT2349-56 clearly stands out as the densest collection of SMGs: although some other protoclusters contain as many SMGs, they extend over much larger areas on the sky, with separations often exceeding 10 arcmin (22 co-moving megaparsecs (cMpc) to 15 cMpc at redshifts of 4.3 to 2). This comparison shows that SPT2349-56 has probably been observed during a much more advanced stage of cluster formation than other high-redshift protoclusters—as a cluster core that is in the process of assembly, rather than an extended structure that may not even collapse to form a cluster by the present day[12].

Also shown in Fig. 2 is the maximal curve of growth predicted by a theoretical model for submillimetre-luminous protocluster regions at redshifts of about 4.5 (see Methods). Except for SPT2349-56 and the recent Herschel discovery SMM J004224 (ref. [23]), the comparison high-redshift protoclusters exhibit $S_{870\mu\mathrm{m}}$ curves of growth that are fairly consistent with the model expectations. The model prediction for the region spanned by SPT2349-56 is roughly 10% of the observed total flux density of the 14 ALMA sources. The underprediction is more severe if we consider the extended LABOCA source: only about 5% of the observed flux density is recovered. This discrepancy may suggest that environmental effects (such as enhanced galaxy interactions or gas accretion in high-density environments) that are not included in this theoretical model are responsible for the extremely high SFR density exhibited by SPT2349-56. An alternative theoretical approach—'zoom' hydrodynamical simulations of protoclusters[24]—can potentially capture such environmental effects. But so far such simulations have been unable to reproduce the extremely high SFR inferred for SPT2349-56: of the 24 protocluster simulations presented by these authors[24], the maximum total SFR attained was about $1{,}700\ M_{\odot}\ \mathrm{yr}^{-1}$, an order of magnitude less than that of SPT2349-56. However, the volume of the N-body simulation from which the 24 halos were selected was $1\ h^{-3}\ \mathrm{cGpc}^3$, which may be too small to contain an object as rare as SPT2349-56. Nevertheless, the existence of SPT2349-56, which contains an unprecedented concentration of rapidly star-forming SMGs from a time when the Universe was only 1.4 billion years old, poses a formidable challenge to theoretical models seeking to explain the origin and evolution of galaxy (proto)clusters.

As outlined above, SPT2349-56 may represent a much more advanced stage of cluster formation than the typical $z > 4$ protoclusters identified to date. Given that the cores of present-day galaxy clusters are characterized by massive elliptical galaxies with old-to-intermediate-age stellar populations[25], and that SMGs are thought to be the high-redshift progenitors of present-day ellipticals[22], it is likely that at least

some of the 14 SMGs located at the same redshift within a region of less than 130 kpc in diameter will soon merge to form a massive elliptical galaxy at the core of a lower-redshift galaxy cluster.

Theoretical studies have shown that at redshifts of more than 4, the progenitors of galaxy clusters should span more than 5 cMpc (refs. [12,26]), corresponding to an angular scale of as much as a degree; we are thus possibly observing only a small part of a much larger structure. For SPT2349-56, it is unknown whether the overdensity extends over such a large scale, as more detailed observations are required to characterize the field surrounding SPT2349-56. We have shown that the extended LABOCA-detected complex has submillimetre colours similar to the core region identified by our ALMA observations, and thus it is likely that the entire complex is found at redshifts of about 4.3. We have also identified five additional bright SPIRE sources in the surrounding roughly 18×18 cMpc field, with similar red colours lying several arcmin from the core structure (see Methods). These are candidates for further protocluster members located in an extended, collapsing structure, similar to the comparison SMG overdensities shown in Fig. 2. If all of these sources are confirmed to lie at $z = 4.31$, this would approximately double the far-infrared luminosity of the cluster, making it by far the most active system known in the Universe. Given that SPT2349-56 was selected from a blind millimetre survey of 2,500 deg^2 (approximately one-sixteenth of the sky), it is unlikely there are more than about 16 such structures across the entire sky. A full analysis of other unlensed sources from the SPT survey, to identify possible systems similar to SPT2349-56, will place stronger constraints on early structure formation in the Universe.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0025-2.

Received: 19 September 2017; Accepted: 24 January 2018;
Published online 25 April 2018.

1. Wang, T. et al. Discovery of a galaxy cluster with a violently starbursting core at $z = 2.506$. *Astrophys. J.* **828**, 56 (2016).
2. Mantz, A. B. et al. The XXL Survey: XVII. X-ray and Sunyaev-Zel'dovich properties of the redshift 2.0 galaxy cluster XLSSC 122. *Astron. Astrophys.* https://doi.org/10.1051/0004-6361/201630096 (2017).
3. Stanford, S. A. et al. IDCS J1426.5+3508: discovery of a massive, infrared-selected galaxy cluster at $z = 1.75$. *Astrophys. J.* **753**, 164 (2012).
4. Springel, V. et al. Simulating the joint evolution of quasars, galaxies and their large-scale distribution. *Nature* **435**, 629–636 (2005).
5. Overzier, R. A. et al. ΛCDM predictions for galaxy protoclusters—I. The relation between galaxies, protoclusters and quasars at $z \sim 6$. *Mon. Not. R. Astron. Soc.* **394**, 577–594 (2009).
6. Chiang, Y.-K., Overzier, R. A., Gebhardt, K. & Henriques, B. Galaxy protoclusters as drivers of cosmic star formation history in the first 2 Gyr. *Astrophys. J. Lett.* **844**, L23 (2017).
7. Miley, G. & De Breuck, C. Distant radio galaxies and their environments. *Astron. Astrophys. Rev.* **15**, 67–144 (2008).
8. Casey, C. M. et al. A massive, distant proto-cluster at $z = 2.47$ caught in a phase of rapid formation? *Astrophys. J. Lett.* **808**, L33 (2015).
9. Chapman, S. C. et al. Do submillimeter galaxies really trace the most massive dark-matter halos? Discovery of a high-z cluster in a highly active phase of evolution. *Astrophys. J.* **691**, 560–568 (2009).
10. Tamura, Y. et al. Spatial correlation between submillimetre and Lyman-α galaxies in the SSA22 protocluster. *Nature* **459**, 61–63 (2009).
11. Ma, J. et al. Stellar masses and star formation rates of lensed, dusty, star-forming galaxies from the SPT survey. *Astrophys. J.* **812**, 88 (2015).
12. Chiang, Y.-K., Overzier, R. & Gebhardt, K. Ancient light from young cosmic cities: physical and observational signatures of galaxy proto-clusters. *Astrophys. J.* **779**, 127 (2013).
13. Vieira, J. D. et al. Extragalactic millimeter-wave sources in the South Pole Telescope survey data: source counts, catalog, and statistics for an 87 square-degree field. *Astrophys. J.* **719**, 763–783 (2010).
14. Mocanu, L. M. et al. Extragalactic millimeter-wave point source catalog, number counts and statistics from 771 square degrees of the SPT-SZ survey. *Astrophys. J.* **779**, 61 (2013).
15. Vieira, J. D. et al. Dusty starburst galaxies in the early Universe as revealed by gravitational lensing. *Nature* **495**, 344–347 (2013).
16. Spilker, J. et al. ALMA imaging and gravitational lens models of South Pole Telescope-selected dusty, star-forming galaxies at high redshifts. *Astrophys. J.* **826**, 112 (2016).
17. Strandet, M. L. et al. The redshift distribution of dusty star forming galaxies from the SPT survey. *Astrophys. J.* **822**, 80 (2016).
18. Beers, T. C. et al. Measures of location and scale for velocities in clusters of galaxies—a robust approach. *Astron. J.* **100**, 32 (1990).
19. Evrard, A. E. et al. Virial scaling of massive dark matter halos: why clusters prefer a high normalization cosmology. *Astrophys. J.* **672**, 122–137 (2008).
20. Cole, S., Helly, J., Frenk, C. S. & Parkinson, H. The statistical properties of Lambda cold dark matter halo formation. *Mon. Not. R. Astron. Soc.* **383**, 546–556 (2008).
21. Ono, Y., Ouchi, M., Kurono, Y. & Momose, R. Faint submillimeter galaxies revealed by multifield deep ALMA observation: number counts, spatial clustering, and a dark submillimeter line emitter. *Astrophys. J.* **795**, 5 (2014).
22. Simpson, J. M. et al. An ALMA survey of submillimeter galaxies in the extended Chandra deep field south: the redshift distribution and evolution of submillimeter galaxies. *Astrophys. J.* **788**, 125 (2014).
23. Oteo, I. et al. An extreme proto-cluster of luminous dusty starbursts in the early Universe. Preprint at https://export.arxiv.org/pdf/1709.02809 (2017).
24. Granato, G. L. et al. The early phases of galaxy clusters formation in IR: coupling hydrodynamical simulations with GRASIL-3D. *Mon. Not. R. Astron. Soc.* **450**, 1320–1332 (2015).
25. Renzini, A. Stellar population diagnostics of elliptical galaxy formation. *Annu. Rev. Astron. Astrophys.* **44**, 141–192 (2006).
26. Oñorbe, J. et al. How to zoom: bias, contamination and Lagrange volumes in multimass cosmological simulations. *Mon. Not. R. Astron. Soc.* **437**, 1894–1908 (2013).

Acknowledgements This paper makes use of the following ALMA data (http://www.almaobservatory.org/en/home/): ADS/JAO.ALMA#2016.1.00236.T and ADS/JAO.ALMA#2015.1.01543.T. ALMA is a partnership of the European Southern Observatory (ESO, representing its member states), the National Science Foundation (NSF, USA) and the National Institute of Natural Sciences (NINS, Japan), together with the National Research Council (NRC, Canada) and the National Security Council (NSC) and the Academia Sinica Institute of Astronomy and Astrophysics (ASIAA, Taiwan), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, Associated Universities Inc. (AUI)/National Radio Astronomy Observatory (NRAO) and the National Astronomical Observatory of Japan (NAOJ). This work is also based in part on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. The SPT is supported by the NSF through grant PLR-1248097, with partial support through grant PHY-1125897, the Kavli Foundation and the Gordon and Betty Moore Foundation grant GBMF 947. This publication is based on data acquired with the Atacama Pathfinder Experiment (APEX) under programmes E-299.A-5045A-2017 and ID M-091.F-0031-2013. APEX is a collaboration between the Max-Planck-Institut für Radioastronomie, the ESO, and the Onsala Space Observatory. Supporting observations were obtained at the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the NSF (USA), the NRC (Canada), Comisión Nacional de Investigación Científica y Tecnológica (CONICYT, Chile), Ministerio de Ciencia, Tecnologa e Innovacion Productiva (Argentina), and Ministerio da Ciencia, Tecnologia e Inovacao (Brazil). The Australia Telescope Compact Array (ATCA) is part of the Australia Telescope National Facility which is funded by the Australian Government for operation as a National Facility managed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO). D.P.M., J.S.S., J.D.V., K.C.L. and J.S. acknowledge support from the US NSF under grant AST-1312950. S.C.C., T.B.M. and A.B. acknowledge support from the National Sciences and Engineering Research Council (NSERC). S.C.C. and T.B.M. acknowledge the Canada Foundation for Innovation (CFI) and the Killam trust. M.A. acknowledges partial support from the Fondo Nacional de Desarrollo Científica y Tecnológico (FONDECYT, Chile) through grant 114009. The Flatiron Institute is supported by the Simons Foundation. J.D.V. acknowledges support from an A.P. Sloan Foundation Fellowship. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

Reviewer information *Nature* thanks P. Capak and C. Papovich for their contribution to the peer review of this work.

Author contributions T.B.M. led the data analysis and assembled the paper. S.C.C. designed the study, proposed the ALMA observations, re-imaged the data, and analysed the data products. C.C.H. developed the theoretical model and advised on the literature comparison. M.A. led the ATCA follow-up and the blind emission-line studies. A.W. procured and analysed the deep LABOCA imaging. M.Br. provided the cluster mass and evolution context and discussion. J.S.S. reimaged the calibrated data. K.A.P. performed the spectral energy distribution (SED) fitting. T.B.M, S.C.C., M.A., K.A.P. and A.W. made the figures. S.C.C., T.B.M., M.A., C.C.H., J.D.V. and A.W. wrote the manuscript. All authors discussed the results and provided comments on the paper. The authors are ordered alphabetically after A.W.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0025-2.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to T.B.M.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Observations. *SPT and SPIRE discovery.* The SPT[27] possesses a unique combination of sensitivity, selection wavelengths (3 mm, 2 mm and 1.4 mm) and beam size that potentially make it ideal for finding the active core regions of galaxy clusters forming at the earliest epochs. Finding very distant ($z > 4$), gravitationally lensed millimetre sources in the SPT survey is quite straightforward, as the contrast to such distant bright sources is high relative to the weak (generally undetected) galactic foregrounds (Extended Data Fig. 2). However, searching for the rare SMGs in the SPT 2,500 deg^2 survey that are unlensed—and therefore are candidates for active groups and protoclusters like SPT2349-56—involves sifting through the many gravitationally lensed sources, and typically involves multistage follow-up efforts using various facilities: a single-dish mapping instrument such as APEX-LABOCA to better localize the emission within the roughly 1′ SPT beam; deep optical imaging to search for bright lensing galaxies; and high-resolution ALMA mapping. The spatially extended sources in SPT2349-56 found with LABOCA span more than an arcmin. With the upcoming deep surveys using the next-generation SPT-3G receiver, this 'extended-beam' thermal-source structure may present a unique signature of many early forming protoclusters, affording the first complete census of the early epochs of structure formation.

A shallow, wide-field image taken by SPIRE (on board the Herschel observatory) over a 100 deg^2 subregion of the SPT-SZ survey[28] reveals the red colours of SPT2349-56, and also that SPT2349-56 appears to reside in something of a void in the $z \approx 1$ foreground that dominates the SPIRE galaxy population. However, the high redshift of SPT2349-56 means that it is not much brighter than many other SPIRE sources in this field, and, aside from its colours, SPT2349-56 does not stand out substantially from the field despite its extreme properties. SPT2349-56 is not detected in the all-sky Planck survey[29], the lower sensitivity of Planck compared with SPT being exacerbated by beam dilution in the 3′ Planck beam.

Obtaining the redshift for SPT2349-56 was beyond the scope of the original SPT–SubMillimeter Galaxy (SMG) redshift survey, owing to the faintness of the unlensed components relative to the typical bright, gravitationally lensed SMGs found in the bulk of the SPT-SMG sample. In ALMA channels Cy 0 and 1, SPT2349-56 was included in the 3-mm spectral-scan redshift survey[15,30], but no lines were detected in the short, roughly 1-min integrations with about 16 ALMA antennae. In Cy 3, a deeper follow-up 3-mm spectral scan was able to tentatively identify two CO lines and a double-source structure with a likely redshift of $z = 4.30$, confirmed by APEX/FLASH C + detection[17].

APEX–LABOCA discovery. We obtained 870-μm imaging of SPT2349-56 using LABOCA on the APEX telescope. A shallow image with a 1.6-hour integration time was observed on 27 Sep 2010, reaching a noise level of about 5 mJy per beam (root mean square, r.m.s.). In August 2017 we obtained a deeper image (18.8-hour integration time; project ID E-299.A-5045A-2017; principal investigator S.C.C.), reaching a minimum noise level of 1.3 mJy per beam, and less than 2.0 mJy per beam r.m.s. for 75.3 arcmin^2 and less than 1.5 mJy per beam r.m.s. for 32.4 arcmin^2 (shown in Fig. 1 and Extended Data Fig. 3). All observations were carried out using standard raster-spiral observations[31] under good weather conditions (precipitable water vapour was 0.6 mm and 0.8 mm for the 2010 and 2017 observing campaigns, respectively). Calibration was achieved through observations of Uranus, Neptune and secondary calibrators and was found to be accurate within 8.5% r.m.s. The atmospheric attenuation was determined via skydips every 2 h as well as from independent data from the APEX radiometer, which measures the line-of-sight water vapour column every minute. The data were reduced and imaged using the BoA reduction package[32]. LABOCA's central frequency and beam size are 345 GHz and 19.2″, resolving the SPT 1.4-mm elongated source into two bright LABOCA sources.

Both LABOCA observations yield consistent calibration results, with a peak intensity at 21″ resolution of 50 mJy per beam for the brighter, southern component (right ascension (RA) 23 h 49 min 42.70 s; declination (dec.) −56° 38′ 23.4″). In addition the LABOCA map reveals a second source to the north at RA 23 h 49 min 42.86 s, dec. −56° 37′ 31.02″, with a peak flux density at 21″ resolution of 17 mJy per beam. Both sources are clearly extended even at LABOCA's relatively coarse spatial resolution, with an observed source size of 28″ × 25″ and 31″ × 24″ for the sourthern and northern sources, respectively. These components are connected by a faint bridge emission. The total 870-μm flux density of the SPT2349-56 system is 110.0 ± 9.5 mJy, of which about 77 mJy are associated with the southern component, 25 mJy with the northern component, and 7 mJy with the connection between the components (using the subregions shown in Extended Data Fig. 3). One additional submillimetre source is detected at $> 5\sigma$ in the LABOCA image to the east of the primary source, but has blue colours inconsistent with $z \approx 4$, and is not likely to be a member of the extended protocluster.

ALMA imaging and spectroscopy. Observations made using ALMA band 3 targeted the CO(4–3) line in SPT2349-56, centred in the lowest frequency of the spectral windows adopted (86–88 GHz), taken under a cycle 3 program (2015.1.01543.T; principal investigator K.L.). Data were taken on 24 June 2016 with a 47-min integration time. The array used 36 antennas with baselines ranging from 15 m to 704 m, and provided a naturally weighted synthesized beam size of about 1″. The asteroid Pallas and source J2343-5626 were used to calibrate the flux and phase respectively. Data were processed using the standard ALMA pipeline with natural beam weighting.

ALMA band-7 imaging (276 GHz) was obtained under a cycle 4 program (2016.0.00236.T; principal investigator S.C.C.), targeting the peak of the brightest LABOCA source. Observations were obtained on 14 December 2016 in a 40–2 array configuration with baseline lengths of 15–459 m, giving a naturally weighted synthesized beam size of about 1″. There were 40 antennas available, with a total on-source integration time of 22 min. The asteroid Ceres and J2357-5311 were used as flux and phase calibrators respectively. The [C II] line ($\nu_{\rm rest} = 1{,}900.5$ GHz) was observed as part of the same ALMA project on 23 March 2017, tuning in band 7 to the redshifted line at $\nu_{\rm obs} = 358.3$ GHz in the upper sideband covering 356–360 GHz. These observations used the 40–2 array configuration with baselines of 16–459 m, giving a naturally weighted synthesized beam size of about 0.5″. An on-source integration time of 14 min was obtained, and J2357-5311 was used as both flux and phase calibrator. The data were reprocessed using the Common Astronomy Software Applications (CASA) and the standard ALMA-supplied calibration, using natural beam weighting to maximize sensitivity.

One-dimensional spectra were extracted from the centroid of the line emission for each source and binned into 75 km s^{-1} channels. Spectra are presented in Fig. 1, and are smoothed using a Gaussian filter with a full width at half-maximum (FWHM) of 100 km s^{-1} for presentation. A Gaussian line profile is fit using a least-squares method, providing errors to the velocity offsets from $z = 4.304$ in Table 1 and line widths in Extended Data Table 2. The continuum level is left as a free parameter in the fitting function, and is then subtracted to derive line fluxes and for presentation.

Blind search for [C II]. We performed a blind search for [C II] line emission in the ALMA band 7 data cube towards SPT2349-56. For this, we followed the procedure used to detect line emitters in the ASPECS survey[33]. We use a data cube channelized at 100 km s^{-1}, without primary beam correction and continuum subtraction. We used the Astronomical Image Processing System (AIPS) task SERCH. This task convolves the data cube along the frequency axis with a Gaussian kernel defined by different input linewidths, subtracts the surrounding continuum, and reports all channels and pixels that have a signal-to-noise ratio (SNR) over a specified limit. The SNR is defined as the maximum significance level achieved after convolving over the Gaussian kernels. We used a set of different Gaussian kernels, from 200 km s^{-1} to 600 km s^{-1}, and searched for all line peaks with SNRs of more than 4.0.

Once all peaks were identified, we used the IDL routine CLUMPFIND[34] to isolate individual candidates. A full list of 68 positive line peaks with SNRs of more than 4.0 was thus obtained. We quantified the reliability of our line search on the basis of the number of negative peaks in our ALMA cube, using the same line procedure. We found 43 negative peaks with SNRs of less than 5.8, and none at a higher SNR. This means that all positive line candidates with SNRs greater than 6.0 are probably real (100% purity). Out of the 14 [C II] line candidates detected, all have SNRs of more than 6.3, and 12 are associated with continuum detections in the ALMA data.

ATCA CO(2–1) line detection: observations. We used ATCA in its H168 hybrid array configuration to observe the CO(2–1) emission line ($\nu_{\rm rest} = 230.5380$ GHz) of SPT2349-56 (with a primary beam size of 53″). The observations were performed as part of project C2818 during 2, 3 and 11 October 2016 under good weather conditions (atmospheric seeing values 90–400 m), and with five working antennae.

We used the ATCA 7-mm receivers, with the compact array broadband backend configured in the wide-bandwidth mode[35]. This leads to a total bandwidth of 2 GHz per correlator window and a spectral resolution of 1 MHz per channel (6.9 km s^{-1} per channel). The spectral windows were centred at observing frequencies of 43.5 GHz and 45.0 GHz, and aimed at observing the CO line and continuum emission, respectively.

Gain and pointing calibrations were performed every 10 min and 1 h, respectively. The bright sources 1921-293, 1934-638 and 2355-534 were used as bandpass, flux and gain calibrators, respectively. We expect the flux calibration to be accurate to within 15%, given the comparison of the Uranus and 1934-638 fluxes. The software package MIRIAD[36] and CASA[37] were used for editing, calibration and imaging.

The calibrated visibilities were inverted with the CASA task CLEAN using natural weighting. No cleaning was applied, given the relatively low significance of the CO-line detection in individual channels. The final data cube, averaged along the spectral axis, yields an r.m.s. of 0.23 mJy per beam for each 100 km s^{-1} channel, with a synthesized beam size of 5.6″ × 4.5″ (position angle 70.4°) at 43.5 GHz.

ATCA CO(2–1) line detection: results. One source was formally detected at the centre, corresponding to C II/continuum sources B + C + G. The central source (C) is unresolved at the resolution of the ATCA observations. The other two sources

are marginally detected to the west and north of the central source, coinciding with the location of C II/continuum sources D + E and A + K, respectively. We extracted spectra at these locations and obtained integrated line intensities by fitting Gaussian profiles to the identified line emission (Extended Data Fig. 4).

We computed CO luminosities using the integrated line intensities. We computed gas masses by assuming a ultraluminous infrared galaxy (ULRIG) X_{CO} factor of 0.8 ($M_\odot$ (K km s^{-1} pc^2)$^{-1}$), also assuming that the CO gas is in local thermodynamic equilibrium: thus $L'_{(CO2-1)} \approx L'_{(CO1-0)}$ (ref. [38]). The results of the CO-line observations are summarized in Extended Data Table 2 Collapsing the line-free spectral window along the spectral axis over the 2-GHz bandwidth leads to a non-detection of the continuum emission down to 80 μJy per beam (3σ standard deviation).

These results confirm the finding from CO(4–3) line that the main reservoir (72%) of molecular gas resides in the B + C + G system, with a smaller fraction hosted in the west and north locations.

Spitzer imaging. This field was twice observed at 3.6 μm and 4.5 μm with the IRAC[39] on board the Spitzer Space Telescope[40]. It was first observed in August 2009 as part of a large program to obtain follow-up imaging of a large sample of SPT-selected SMG sources (project 60194; principal investigator J.D.V.). The observing scheme for this project involved obtaining 36 dithered 100-s integrations at 3.6 μm and, separately, a much shallower 12 × 30-s integration at 4.5 μm. Later, in cycle 8, the field was covered serendipitously as part of the Spitzer–SPT deep-field survey (project 80032; principal investigator M.A. Stanford[40]). This project surveyed 92 deg^2 uniformly in both IRAC passbands, to a depth of 4 × 30 s. Using established techniques, we combined all exposures covering the SPT target from projects 60194 and 80032 at 3.6 μm and 4.6 μm, to obtain the best possible SNR in our final mosaics, which were pixellated to 0.5″. Nine of the 14 sources identified by ALMA are detected in the IRAC bands at greater than 3σ in at least one of the 3.6-μm or 4.5-μm channels (Extended Data Fig. 1).

Analysis of the surrounding field with SPIRE and LABOCA imaging. In Extended Data Fig. 5, our deep SPIRE RGB image is shown with LABOCA contours overlaid. A source sample was culled from the 250-μm-selected catalogue (there were 135 sources with SNR (250 μm) values of greater than 3 in an area of 52 arcmin^2), where the source peaks are best defined. To account for the large beam-size difference with SPIRE (ranging from 36″ at 500 μm to 18″ at 250 μm), we used a deblending code, with the 250-μm positions as spatial priors, which provides the standard parameters as well as the covariance matrices highlighting the degeneracies (almost none at 250 μm, but statistically significant at 500 μm). This code, FASTPHOT[41], takes into account these degeneracies to estimate the flux measurement errors.

Colour–colour and colour–flux diagrams are shown in Extended Data Fig. 5. The colour–colour diagram shows a 250-μm-selected sample with an SNR(250 μm) of more than 3, and is dominated by the $z \approx 1$ cosmic infrared background (blue and green colours) in the foreground of SPT2349-56. The colour–flux diagram shows an additional SNR(500 μm) of more than 3, cut to highlight just the well detected 500-μm source subsample. These diagrams highlight the extreme and red properties of SPT2349-56, but make clear that one of the three 250-μm peaks within the SPT2349-56 LABOCA structure is very likely to be a foreground galaxy (the green symbol highlighted in the figure shows very blue colours). Nevertheless, a full ALMA mapping of the structure is warranted given the uncertainties involved in the SPIRE deconvolution procedure.

Five red sources consistent with $z \approx 4$ ($S_{500\mu m} > S_{350\mu m} > S_{250\mu m}$; Extended Data Table 3) are found in the surrounding roughly 10′ × 10′ field, and are candidates for additional protocluster members in an extended, collapsing structure. If all these sources are bona fide $z = 4.3$ sources, this would substantially increase the total 870-μm flux density (and thus the far-infrared luminosity) of the cluster beyond the 110 mJy found in the central structure, making it by far the most active system known in the Universe (see Fig. 2).The deep LABOCA map marginally detects the closest of the five red SPIRE sources at about 3σ, consistent with expectations given the SPIRE flux densities. Full analysis of these surrounding SMGs will require follow-up efforts.

Properties, comparisons and simulations. *Derivation of physical properties.* We briefly describe here our procedures for calculating various physical quantities from the observables. To derive the star-formation rate, we measured the 870-μm flux density directly in the lower sideband (line-free bands) of our ALMA band-7 observations from cycle 4, finding consistent measurements with those in previous shallower observations[16]. We adopt an SFR-to-$S_{870\mu m}$ ratio of $150 \pm 50\, M_\odot$ yr^{-1} mJy^{-1}, which is typical for SMGs[42]. The uncertainty in this ratio derives from variations in the dust temperature distribution among the SMG population, which are driven mainly by differences in the ratio of the luminosity absorbed by dust to the total dust mass[43]. This, combined with the measurement error, dictates the error on the SFR in Table 1.

We calculated the gas mass from the CO(4–3) line luminosity, which we converted to CO(1–0) luminosity by using a ratio between the brightness temperatures of these lines, $r_{4,1} = 0.41 \pm 0.07$, that was found from the average of a sample of unlensed SMGs with multiple CO-line transitions[44]. We used a conservative conversion factor, $\alpha_{CO} = 0.8 \frac{M_\odot}{\text{K km s}^{-1}\text{ pc}^2}$, and multiplied by 1.36 to account for the addition of helium. When CO(4–3) was not detected to a large extent, we used our [C II] line luminosity and the average CO(4–3)/[C II] ratio for our detected sample (denoted with double daggers in Table 1).

Spectral energy distribution of SPT2349-56. The SPT, LABOCA and SPIRE measurements resolve the SPT2349-56 structure to varying degrees, but none can isolate the core region resolved by our present ALMA observations with any confidence. We thus assembled a photometric catalogue of the total SPT2349-56 flux density from 250 μm to 850 μm, and modelled the resulting total spectral energy distribution (SED) to estimate some global properties of the system. Although the SPT does measure the flux at 1.4 mm, 2.0 mm and 3.0 mm (being 27.9 ± 4.6 mJy, 5.2 ± 1.1 mJy and 0.5 ± 0.1 mJy, respectively), we did not include these points in the SED fit, because the measurements are uncertain owing to the elongated structure of SPT2349-56 and difficulties with the filtering used to make the map. At IRAC wavelengths in the mid-infrared, we identified nine SMGs detected at more than 3σ, and included the sum in the SED fit. We deferred further analysis to follow-up work with forthcoming deeper Spitzer–IRAC data and follow-up optical and near-infrared photometry.

We used Code Investigating GALaxy Emission (CIGALE)[45,46] for SED fitting of the combined photometry of the source. The SED modelling assumes a single-component star-formation history and solar metallicity[47]. A Chabrier[48] initial mass function is assumed. The resulting best-fitting SED is shown in Extended Data Fig. 6. The infrared luminosity (at 8–1,100 μm) is $(8.0 \pm 1.0) \times 10^{13} L_\odot$ (where $L_\odot$ is the luminosity of the Sun).

Protocluster comparison sample. To place SPT2349-56 in context and to compare it with other systems claimed to be protoclusters, we identified from the literature various SMG-rich overdensities at redshifts between 2 and 5. Although a direct comparison of the number counts (number per deg^2) of SMG-overdense systems can be performed, it involves making somewhat arbitrary choices of enclosed areas and redshift boundaries. We opted in Fig. 2 to show instead a curve of growth analysis of the 870-μm flux density. We considered only galaxies that have been confirmed to be protocluster members via spectroscopic redshifts. The data were drawn from a recent compilation[49] and original references therein.

The Great Observatories Origins Deep Survey North (GOODS-N) overdensity at $z = 1.99$ (refs. [9,50,51]) spans a roughly 10′ by 10′ field in the Hubble Deep Field North, and contains nine SMGs within a redshift range (Δz) of 0.008. The probability of this large an overdensity being drawn from the field distribution by chance is less than 0.01%. Interestingly, only a modest overdensity of Lyman-break galaxies is found in this GOODS-N structure.

The Cosmic Evolution Survey (COSMOS) $z = 2.5$ SMG overdensity[8] is similar to the GOODS-N structure in terms of the numbers and luminosities of the component SMGs, the angular size of the system, and the modest overdensity of LBGs associated with it.

The overdensity in the survey MRC1138 was originally discovered as an overdensity of Lyα and Hα emitters[52]. Follow-up observations[53,54] revealed the presence of five SMGs, in addition to an active galactic nucleus (AGN) known as the Spiderweb Galaxy. This is a radio-loud AGN that resides in a large Lyα halo.

The SSA22 protocluster was one of the first discovered by observing an overdensity of Lyman-break galaxies[55]. It is an extremely extended structure located at $z = 3.09$, with Lyα emitters spanning more than 50 cMpc (ref. [56]). Submillimetre observations of the field have revealed a population of at least eight SMGs[10,50,57–59].

The COSMOS $z = 2.1$ protocluster[60] lacks enough deep 850-μm data to characterize the Herschel–SPIRE sources identified in the structure. We estimated 870-μm flux densities by taking the published infrared luminosities (L_{IR}; integrated over 3–1,100 μm) and used the SED of the galaxy collision Arp 220 to estimate $S_{870\mu m}$, finding that $L_{IR} = 2 \times 10^{12} L_\odot$ corresponds to $S_{870\mu m} = 1$ mJy at $z \approx 2$. For the SSA22 protocluster, we used the measured 870-μm flux density when available and otherwise estimate it from the 1.1-mm flux using a standard conversion at $z \sim 3$ of $S_{870\mu m} = 2 \times S_{1.1mm}$. To create the curves of growth for Fig. 2, we defined the centre of each protocluster by computing the median right ascension and declination of all submillimetre sources. We checked that randomly adjusting the centres of the curve-of-growth tracks by about 1′ did not boost the curves by more than 10%, showing that the curves of growth for the literature SMG overdensities are insensitive to the adopted centre.

Recently, there have also been detections of SMG overdensities at $z \approx 4$. The first, GN20, at $z = 4.05$, was discovered through the serendipitous detection of CO(4–3) emission from two SMGs[61], with two further SMGs detected subsequently[62]. An excess of *B*-band dropouts was also observed in this structure, several of these dropouts being confirmed spectroscopically to lie at $z \approx 4.05$. The second SMG overdensity at a redshift greater then 4, HDF850.1, contains a single SMG, a quasar and 11 spectroscopically confirmed galaxies. This SMG has a confirmed redshift of $z = 5.18$ (ref. [63]). The AzTEC-3 overdensity is centred on a single SMG at $z = 5.3$,

with 12 spectroscopically confirmed optical galaxies at the same redshift. This is a relatively dense structure, with most of the galaxies residing within a circle of about 1′ in diameter. The most luminous example at $z \approx 4$, SMM J004224, was found from the Herschel surveys[23], with several additional 870-μm sources in the surrounding field. In Fig. 2 we also plot the flux from the satellites that have observed SPIRE colours consistent with $z = 4$ (ref. [64]).

Overdensities of SMGs and optical galaxies have also been found around high-redshift radio galaxies (HzRGs)[65], continuing to confirm HzRGs as useful beacons of structure forming in the early Universe. However, none of these systems comes close to the level of overdensity found in SPT2349-56, and they also suffer from the bias inherent in targeting these sources—namely, that one or more protocluster members has to be radioluminous.

There have also been discoveries of compact binary hyperluminous infrared galaxy (HyLIRG) systems, the most luminous of which is the $z = 2.4$ source HATLAS J084933 (ref. [66]), with others approaching this luminosity[67,68]. In each of these systems, the dynamics and SFRs are dominated by two SMGs, but there is no strong evidence of any surrounding protocluster in the form of an excess of galaxies selected optically or in the submillimetre. In one case[68], there is evidence for a relative void around the structure. These systems may simply be instances of very rare events in fairly typical (but still massive) halos[69], analogous to hyperluminous quasars[70].

Theoretical studies of *N*-body simulations have shown that the progenitors of $z = 0$ dark-matter halos with masses greater than $10^{15} M_{\odot}$ should extend to effective radii greater than about 5 cMpc at redshifts higher than 2 (refs. [12,26]). Given that the overdensities listed above are mostly concentrated in small areas, it is difficult to asses their exact evolution or to compare them easily with simulated structures. Interpreting a small overdense region at high redshift as a 'protocluster core' is certainly prone to misinterpretation, and small overdense regions at high redshift can evolve into halos spanning a range of masses at the current epoch[12]. Chiang et al.[12] suggest investigating if an overdensity extends to larger scales (more than 20 cMpc), in order to better determine whether it will form a cluster with a mass greater than $10^{15} M_{\odot}$. But this is difficult at high redshift, because the excess of galaxies will be less pronounced on larger scales, and it is challenging to detect high-redshift, low-luminosity galaxies.

To assess the evolution of SPT2349-56, we also compared it with a selection of the highest-mass galaxy clusters at lower redshifts (less than around 2.5). The clusters XMMU J2235.3 (ref. [71]) and XLSSC 122 (ref. [2]) were both discovered in X-ray surveys with masses derived from the X-ray light profile. Meanwhile IDCS J1426.5+3508 (refs. [3,72]), JKCS041 (ref. [73]), CL J1449+0856 (refs. [74,75]) and CL J1001+0220 (ref. [1]) were discovered as overdensities of massive red galaxies and have corresponding X-ray detections. Masses for all were estimated using these X-ray detections and are consistent with dynamical masses estimated using the velocity dispersion. We also show the sample of clusters discovered in the 2,500 deg^2 SPT survey using the Sunyaev-Zel'dovich effect, with mass estimates based on the velocity dispersions[76].

Geometry and dynamics. In Extended Data Fig. 7 we investigate the geometry and kinematics of the SPT2349-56 system. By analysing the velocity distribution relative to the mean redshift of 4.304, ΔV, versus projected radius for our sample of 14 galaxies, we can see that—even with our most conservative mass estimate—at least 12 of the 14 galaxies appear to be bound. We show the escape velocity as a function of radius for a point source and a Navarro–Frenk–White (NFW) profile with a mass of $1.16 \times 10^{13} M_{\odot}$. The NFW profile assumes a virial radius of 200 kpc and a concentration of 5—values typical of massive halos at this epoch, found in the *N*-body simulations described in the section on 'Simulations' below. Even if the projected radius were to account for only a fraction of the true physical distance (that is, separated along the line of sight), it is likely that most of the galaxies with low relative velocity (ΔV less than 500 km s^{-1}) will still be bound and eventually collapse into a single galaxy. The cumulative distribution of relative velocities is also shown, alongside that of a Gaussian distribution with $\sigma = 408$ km s^{-1}. The observed distribution is smooth and well fit by the Gaussian. The relative kinematics of the 14 galaxies is consistent with most, if not all, members being mutually bound with SPT2349-56, following a Gaussian ΔV distribution characteristic of a virialized system.

An alternative interpretation of this system is that we are observing individual galaxies, or separate less-massive halos along a filament aligned with our line of sight. The probability of this occurring might be boosted by our selection technique, given that the beam size of SPT is approximately 1′. We are preferentially sensitive to structure on this scale, and require several of the most luminous galaxies found in the Universe together in the SPT beam to exceed our detection threshold.

To explore this possibility, we also show in Extended Data Fig. 7 the most extreme interpretation of SPT2349-56, whereby none of the velocity offsets is peculiar, and the physical distribution of the SPT2349-56 galaxies is represented by the relative velocities being entirely due to cosmic expansion rather than peculiar motions. In this case, the galaxies of course are maximally separated along a (proper distance) 3-Mpc filament (compared with their 130-kpc maximum tangential extent). However, this cannot be the true distribution, and a more realistic interpretation would entail some 20–30% of velocities being Hubble flow (600×900 pc extent), and the remainder representing peculiar motions. (Note that this would be true for any massive halo at this epoch, given the expected velocity dispersion of about 400 km s^{-1} and Hubble constant of 470 km s^{-1} Mpc^{-1}.) This explanation, if substantiated, would alter our interpretation of the system, specifically decreasing the mass estimate as the velocity offsets are not due to peculiar motions and the system is not virialized.

However, given the observed kinematics and spatial configuration of the galaxies discussed above, we argue that SPT2349-56 is less likely to consist of multiple groups, widely separated on a line-of-sight filament. To back up this claim, we performed a systemic search for similar filamentary structures in the *N*-body simulations described in 'Simulations' below. We first searched for groups at redshifts greater than 4 with more than eight galaxies within a projected radius of more than 150 kpc but can be extended up to 5 Mpc along the line of sight. There are three such systems in our 1-cGpc^3 simulation, and they are also displayed in Extended Data Fig. 7, showing their geometry. Each of these three analogous systems in the simulation has a total halo mass of about $10^{13} M_{\odot}$. None is extended much past 500 kpc, suggesting that, at redshifts of around 4, there are no substantially extended filaments hosting massive galaxies in the simulation that are comparable to even the gas masses of our SPT2349-56 ALMA galaxies. These simulated structures are not filaments; they are close to collapsed structures that are slightly cigar-shaped, and although SPT2349-56 could in principle be distributed like this, it does not fundamentally change our discussion here. (We also note that we are plotting two different things in Extended Data Fig. 7: velocity offsets for SPT2349-56, and the actual geometry of the simulated galaxies.) These results further suggest that most of the velocity offsets in the SPT2349-56 galaxies are truly due mostly to peculiar motion rather than line-of-sight projection. At least one of the simulation systems found appears to be a chance projection of multiple groups, but it has a characteristically different ΔV distribution. The system has multiple smaller groups with $\sigma \approx 200$ km s^{-1}, reflecting their smaller masses, separated in velocity space by 500 km s^{-1}.

We also note that a filamentary structure feeding a halo of approximate mass $10^{13} M_{\odot}$ should have a characteristic width of 200 kpc to 400 kpc (refs. [77,78]), compared with the 80 kpc (FWHM) width we find for our 14 ALMA galaxies. This is suggestive that, even with a direct line-of-sight view down a filament, the configuration is far more concentrated than one would infer based on the typical size of filament. We then ask what sort of environment a roughly 80-kpc-wide filament (consistent with our SPT2349-56 galaxies) would connect, and infer a typical halo mass of around $2 \times 10^{12} M_{\odot}$ (ref. [77]). Given that the SPT2349-56 galaxies contain at least this much mass entirely in their cold gas (as we have adopted the lowest plausible $\alpha_{\rm CO} = 0.8$ conversion factor), and are probably hosted by haloes at least twice as large as this, it becomes somewhat contradictory that they could be found along a filament that is connecting a much less massive halo.

A further important piece of evidence against these 14 galaxies being a filament is that we find no additional sources in the surrounding spectral windows in either side band, in the band-7 or the band-3 ALMA data (Extended Data Fig. 7d). Although all 14 sources are clustered within 1,500 km s^{-1} of each other, our full frequency coverage extends over relative velocities of 6,500 km^{-1} (for [C II]) and 27,000 km s^{-1} (for CO(4–3)), or many tens of Mpc line-of-sight distance. If SPT2349-56 were actually an elongated filament, one might expect to see a distribution of sources over a much larger fraction of our observed frequency bandwidth. Although we can not rule out the possibility that we are observing multiple galaxies along our line of sight, the observed geometry and kinematics suggest that SPT2349-56 consists of a single group of mutually gravitationally bound galaxies.

Another way to assess the configuration of this system is by using the Millennium Simulation[4] database, with the implementation of the galaxy-formation recipe[79]. We searched the model output at $z \approx 4$ for galaxies with a total baryonic mass in excess of around $10^{11} M_{\odot}$, consistent with the combined mass of gas and stars estimated for the brightest ALMA-resolved galaxies in SPT2349-56. We used this total baryonic mass cut as it meant we were less sensitive to the details of the early star-formation histories of galaxies in the model. We found only one galaxy in the 3.2×10^8 cMpc^3 simulation volume at $z = 4$ with a total baryonic mass above $10^{11} M_{\odot}$. Half of the baryonic mass in this model galaxy is found in stars, and it has a $3 \times 10^8 M_{\odot}$ black hole. Moreover, this galaxy is the central galaxy of a $1 \times 10^{13} M_{\odot}$ halo—the optimal environment for finding merging galaxies, according to simulations[80]. Searching the environment of this system, we found another four massive galaxies are distributed across a region with a roughly 0.5-cMpc diameter around the central galaxy, with baryonic masses of more than $7 \times 10^{10} M_{\odot}$ (more than 5% of the primary). By redshift zero, these galaxies are all predicted to reside in a $1 \times 10^{15} M_{\odot}$ halo, consistent with our other assessments of the outcome of this system.

Simulations. To further place SPT2349-56 in context, we compared it with the predictions of a theoretical model for SMG overdensities[69,81]. We used the MultiDark[82] *N*-body simulation, which is one of the largest (2.91 Gpc^3) available *N*-body simulations that still resolves SMG-like halos (with halo masses of more than roughly $10^{12}M_{\odot}$). We analysed the $z=4.68$ and $z=4.1$ snapshots, which are the available snapshots that are closest in redshift to SPT2349-56. We created halo catalogues using the Rockstar halo finder[83], and assigned stellar masses to dark-matter halos using a relation derived based on a sub-halo abundance-matching relation[84]. To assign SFRs, it is assumed that the distribution of specific SFRs (SFRs per unit stellar mass, hereafter SSFR) is the sum of two Gaussians, corresponding to quiescently star-forming and starburst galaxies[85]. The median SSFR value is based on the abundance-matching-derived relation[84], and the starburst fraction and the widths of the Gaussian distributions are set on the basis of observations of massive, high-redshift star-forming galaxies similar to the members of SPT2349-56 (ref. [85]). The dust mass, $M_{\rm d}$, is estimated from the stellar mass by using empirical gas-fraction and metallicity relations[86]. Once SFR, stellar mass (M_*), and $M_{\rm d}$ values are assigned to each halo, $S_{870\mu m}$ is calculated using the following fitting function, which was derived from results obtained by performing dust radiative transfer calculations on hydrodynamical simulations of both isolated and interacting galaxies[81,87]:

$$S_{870\rm mm}=0.81\ \rm mJy\left(\frac{SFR}{100\ M_{\odot}yr^{-1}}\right)^{0.43}\left(\frac{M_{\rm d}}{10^8M_{\odot}}\right)^{0.54}$$

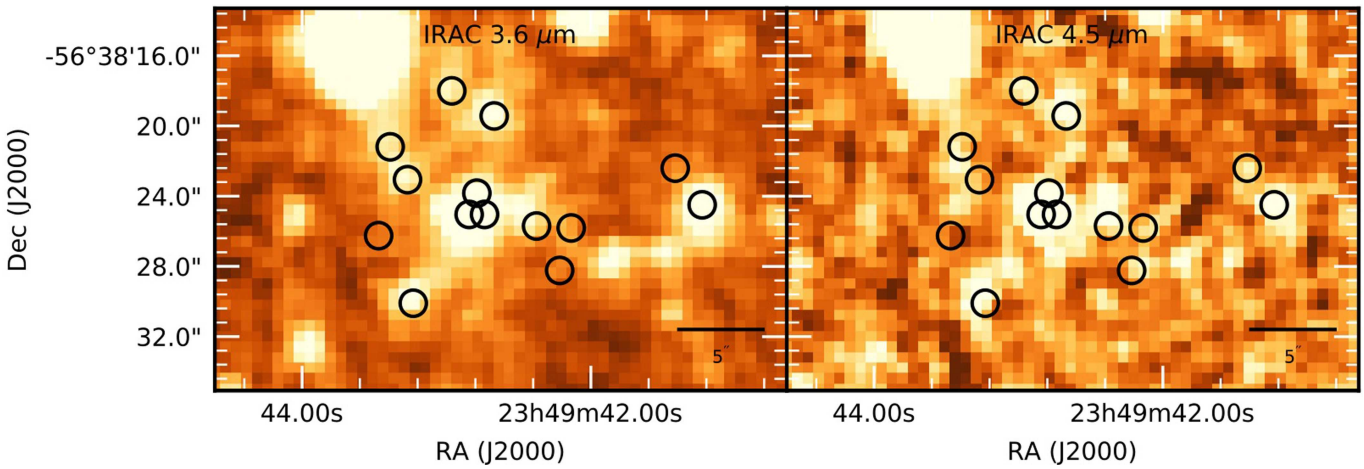
where $S_{870\mu m}$ is the 870-μm flux density, SFR is the star-formation rate, and $M_{\rm d}$ is the dust mass. A scatter of 0.13 dex is included when applying the relation.

Once $S_{870\mu m}$ was assigned to each halo, we searched the entire simulation volume for the most luminous regions. We began at each independent halo and calculated the total $S_{870\mu m}$ of all halos within a given projected radius, *r*, and line-of-sigh distance along a given axis of this halo. We used a line-of-sight distance of 1 Mpc for the $z=4$ snapshot and 2 Mpc for the $z=2.5$ snapshot, reflecting the Hubble constant at each epoch and the expected velocity dispersion of roughly 400 $km\,s^{-1}$. For each value of *r*, we recorded the largest total $S_{870\mu m}$ obtained (across all halos). We performed 100 Monte Carlo iterations for each snapshot; in each iteration, galaxy properties were re-assigned, drawing from the distributions described above. The shaded region in Fig. 2 shows the entire region spanned by the 100 realizations of the maximum $S_{870\mu m}$ versus area curves. To compare SPT2349-56 with lower-redshift protoclusters, we performed a similar analysis on a snapshot at $z=2.49$ with 20 Monte Carlo iterations.

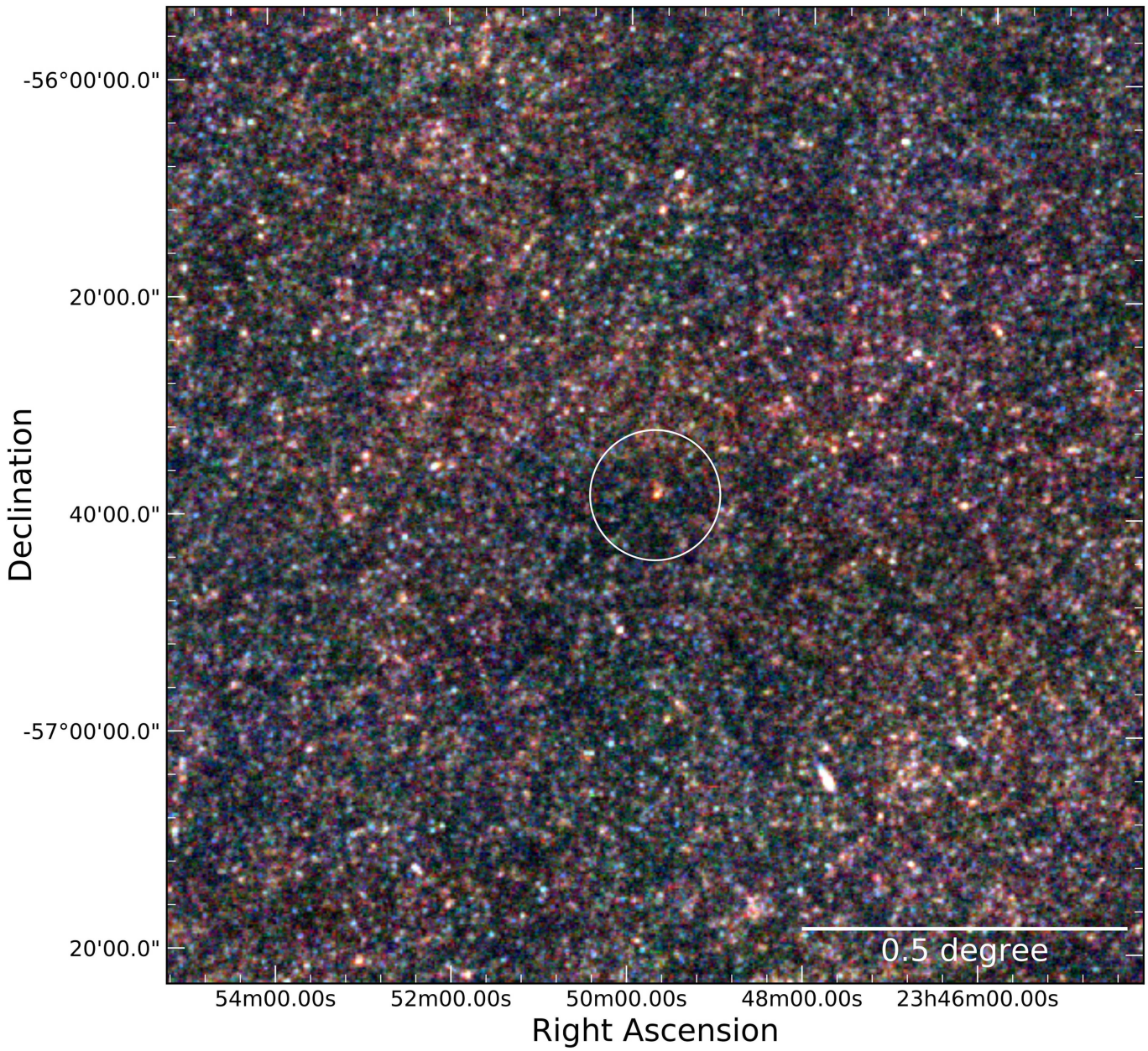
Data availability. Data can be made available upon reasonable request to T.B.M.

27. Carlstrom, J. E. et al. The 10 meter South Pole Telescope. *Publ. Astron. Soc. Pacific* **123**, 568 (2011).
28. Ashby, M. L. N. et al. The Spitzer South Pole Telescope deep field: survey dDesign and infrared array camera catalogs. *Astrophys. J. Suppl. Ser.* **209**, 22 (2013).
29. Planck Collaboration. Planck 2013 results. XXII. Constraints on inflation. *Astron. Astrophys.* **571**, A22 (2014).
30. Weiß, A. et al. ALMA redshifts of millimeter-selected galaxies from the SPT survey: the redshift distribution of dusty star-forming galaxies. *Astrophys. J.* **767**, 88 (2013).
31. Siringo, G. et al. The Large APEX BOlometer CAmera LABOCA. *Astron. Astrophys.* **497**, 945–962 (2009).
32. Weiß, A. et al. LABOCA observations of nearby, active galaxies. *Astron. Astrophys.* **490**, 77–86 (2008).
33. 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).
34. Williams, J. P. et al. Clumpfind: determining structure in molecular clouds (Astrophysics Source Code Library ascl:1107.014, 2011).
35. Wilson, W. E. et al. The Australia Telescope Compact Array broad-band backend: description and first results. *Mon. Not. R. Astron. Soc.* **416**, 832–856 (2011).
36. Sault, R. J., Teuben, P. J. & Wright, M. C. H. A retrospective view of Miriad. *Astron. Data Analysis Software Systems IV* **77**, 433 (1995).
37. Mcmullin, J. P., Waters, B., Schiebel, D., Young, W. & Golap, K. CASA architecture and applications. *Astron. Data Analysis Software Systems XVI* **376**, 127 (2007).
38. Solomon, P. M. et al. The molecular interstellar medium in ultraluminous infrared galaxies. *Astrophys. J.* **478**, 144–161 (1997).
39. Fazio, G. G. et al. The Infrared Array Camera (IRAC) for the Spitzer Space Telescope. *Astrophys. J. Suppl. Ser.* **154**, 10–17 (2004).
40. Werner, M. W. et al. The Spitzer Space Telescope Mission. *Astrophys. J. Suppl. Ser.* **154**, 1–9 (2004).
41. Béthermin, M., Dole, H., Cousin, M. & Bavouzet, N. Submillimeter number counts at 250, 350 and 500 microns in BLAST data. *Astron. Astrophys.* **516**, A43 (2010).
42. Barger, A. J. et al. Is there a maximum star formation rate in high-redshift galaxies? *Astrophys. J.* **784**, 9 (2014).
43. Safarzadeh, M., Hayward, C. C., Ferguson, H. C. & Somerville, R. S. What shapes the far-infrared spectral energy distributions of galaxies? *Astrophys. J.* **818**, 62 (2016).
44. Bothwell, M. S. et al. A survey of molecular gas in luminous sub-millimetre galaxies. *Mon. Not. R. Astron. Soc.* **429**, 3047–3067 (2013).
45. Burgarella, D. et al. Star formation and dust attenuation properties in galaxies from a statistical ultraviolet-to-far-infrared analysis. *Mon. Not. R. Astron. Soc.* **360**, 1413–1425 (2005).
46. Noll, S. et al. Analysis of galaxy spectral energy distributions from far-UV to far-IR with CIGALE: studying a SINGS test sample. *Astron. Astrophys.* **507**, 1793–1813 (2009).
47. Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).
48. Chabrier, G. The galactic disk mass function: reconciliation of the Hubble Space Telescope and nearby determinations. *Astrophys. J.* **586**, L133–L136 (2003).
49. Casey, C. M. The ubiquity of coeval starbursts in massive galaxy cluster progenitors. *Astrophys. J.* **824**, 36 (2016).
50. Chapman, S. C., Blain, A. W., Smail, I. & Ivison, R. J. A redshift survey of the submillimeter galaxy population. *Astrophys. J.* **622**, 772–796 (2005).
51. Blain, A. W., Chapman, S. C., Smail, I. & Ivison, R. Clustering of submillimetre-selected galaxies. *Astrophys. J.* **611**, 725–731 (2004).
52. Kurk, J. D. et al. A search for clusters at high redshift. I. Candidate Lyalpha emitters near 1138-262 at $z=2.2$. *Astron. Astrophys.* **358**, L1–L4 (2000).
53. Kuiper, E. et al. A SINFONI view of flies in the Spiderweb: a galaxy cluster in the making. *Mon. Not. R. Astron. Soc.* **415**, 2245–2256 (2011).
54. Dannerbauer, H. et al. An excess of dusty starbursts related to the Spiderweb galaxy. *Astron. Astrophys.* **570**, A55 (2014).
55. Steidel, C. et al. A large structure of galaxies at redshift $z\sim3$ and its cosmological implications. *Astrophys. J.* **492**, 428–438 (1998).
56. Hayashino, T. et al. Large-scale structure of emission-line galaxies at $z=3.1$. *Astron. J.* **128**, 2073–2079 (2004).
57. Chapman, S. C. et al. Submillimeter imaging of a protocluster region at $z=3.09$. *Astrophys. J.* **548**, L17–L21 (2001).
58. Geach, J. E. et al. A submillimetre survey of Lyman α haloes in the SA 22 protocluster at $z=3.1$. *Mon. Not. R. Astron. Soc.* **363**, 1398–1408 (2005).
59. Umehata, H. et al. ALMA deep field in SSA22: a concentration of dusty starbursts in a $z=3.09$ protocluster core. *Astrophys. J. Lett.* **815**, L8 (2015).
60. Hung, C.-L. et al. Large scale structure around a $z=2.1$ cluster. *Astrophys. J.* **826**, 130 (2016).
61. Daddi, E. et al. Two bright submillimeter galaxies in a $z=4.05$ protocluster in GOODS-North, and accurate radio-infrared photometric redshifts. *Astrophys. J.* **694**, 1517–1538 (2009).
62. Daddi, E. et al. A CO emission line from the optical and near-IR undetected submillimeter galaxy GN10. *Astrophys. J.* **695**, L176–L180 (2009).
63. Walter, F. et al. The intense starburst HDF 850.1 in a galaxy overdensity at $z\sim5.2$ in the Hubble Deep Field. *Nature* **486**, 233–236 (2012).
64. Lewis, A. J. R. et al. Ultra-red galaxies signpost candidate proto-clusters at high redshift. Preprint at https://arxiv.org/pdf/1711.08803v1.pdf (2017).
65. Noirot, G. et al. HST grism confirmation of two $z\sim2$ structures from the clusters around radio-loud AGN (CARLA) survey. *Astrophys. J.* **830**, 90 (2016).
66. Ivison, R. J. et al. Herschel-ATLAS: a binary HyLIRG pinpointing a cluster of starbursting proto-ellipticals. *Astrophys. J.* **772**, 137 (2013).
67. Fu, H. et al. The rapid assembly of an elliptical galaxy of 400 billion solar masses at a redshift of 2.3. *Nature* **498**, 338–341 (2013).
68. Chapman, S. C. et al. A millimetre-wave redshift search for the unlensed HyLIRG, HS1700.850.1. *Mon. Not. R. Astron. Soc.* **453**, 951–959 (2015).
69. Miller, T. B., Hayward, C. C., Chapman, S. C. & Behroozi, P. S. The bias of the submillimetre galaxy population: SMGs are poor tracers of the most-massive structures in the $z\sim2$ Universe. *Mon. Not. R. Astron. Soc.* **452**, 878–883 (2015).
70. Trainor, R. F. & Steidel, C. C. The halo masses and galaxy environments of hyperluminous QSO's at $z\sim2.7$ in the Keck Baryonic Structure Survey. *Astrophys. J.* **752**, 39 (2012).
71. Rosati, P. et al. Multi-wavelength study of XMMU J2235.3-2557: the most massive galaxy cluster at $z>1$. *Astron. Astrophys.* **508**, 583–591 (2009).
72. Brodwin, M. et al. IDCS J1426.5+3508: the most massive galaxy cluster at $z>1.5$. *Astrophys. J.* **817**, 122 (2016).
73. Andreon, S. et al. JKCS 041: a Coma cluster progenitor at $z=1.803$. *Astron. Astrophys.* **565**, A120 (2014).
74. Gobat, R. et al. A mature cluster with X-ray emission at $z=2.07$. *Astron. Astrophys.* **526**, A133 (2011).
75. Gobat, R. et al. WFC3 GRISM confirmation of the distant cluster Cl J1449+0856 at $<z>=2.00$: quiescent and star-forming galaxy populations. *Astrophys. J.* **776**, 9 (2013).
76. Bocquet, S. et al. Mass calibration and cosmological analysis of the SPT-SZ galaxy cluster sample using velocity dispersion and X-ray measurements. *Astrophys. J.* **799**, 214 (2014).
77. Mandelker, N., van Dokkum, P. G., Brodie, J. P., van den Bosch, F. C. & Ceverino, D. Cold filamentary accretion and the formation of metal poor globular clusters and halo stars. Preprint at https://arxiv.org/abs/1711.09108v1 (2017).
78. Cautun, M., van de Weygaert, R., Jones, B. J. T. & Frenk, C. S. Evolution of the cosmic web. *Mon. Not. R. Astron. Soc.* **441**, 2923–2973 (2014).
79. Bower, R. G. et al. Breaking the hierarchy of galaxy formation. *Mon. Not. R. Astron. Soc.* **370**, 645–655 (2006).
80. Hopkins, P. F., Hernquist, L., Cox, T. J. & Kereš, D. A cosmological framework for the co-evolution of quasars, supermassive black holes, and elliptical galaxies. I. Galaxy mergers and quasar activity. *Astrophy. J. Suppl. Ser.* **175**, 356–389 (2008).

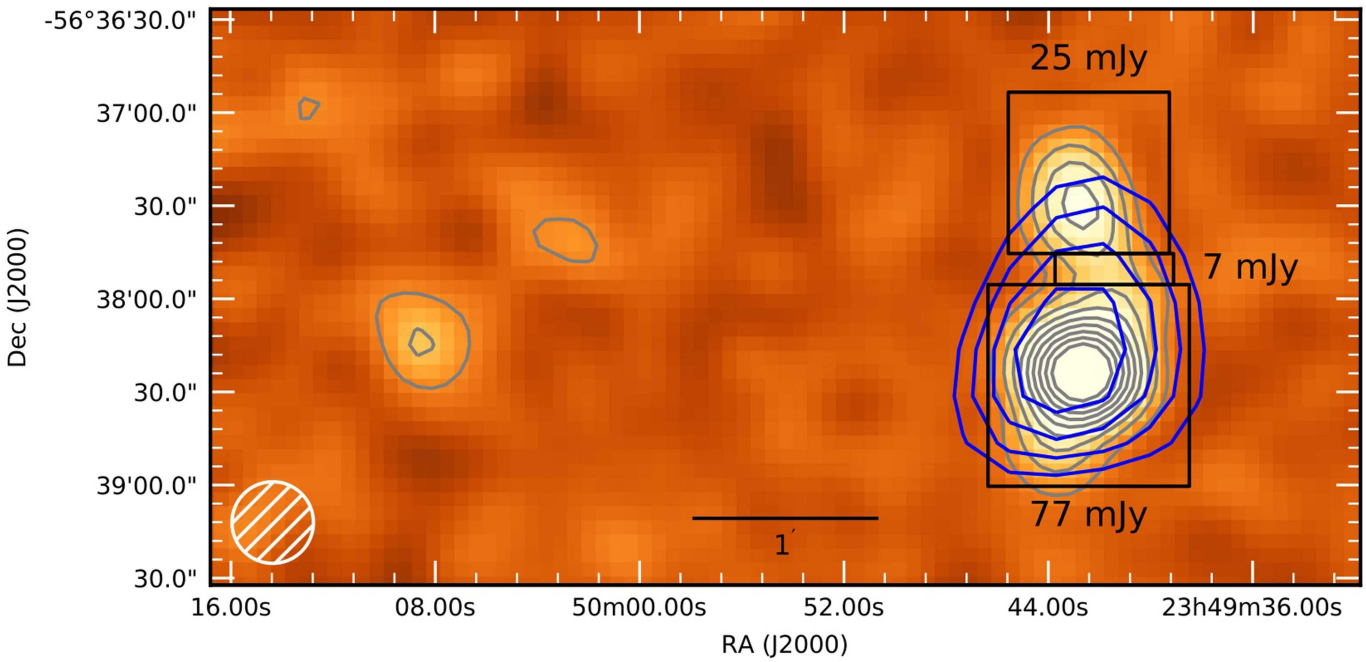
81. Hayward, C. C. et al. Spatially unassociated galaxies contribute significantly to the blended submillimetre galaxy population: predictions for follow-up observations of ALMA sources. *Mon. Not. R. Astron. Soc.* **434**, 2572–2581 (2013).
82. Prada, F., Klypin, A. A., Cuesta, A. J., Betancort-Rijo, J. E. & Primack, J. Halo concentrations in the standard Λ cold dark matter cosmology. *Mon. Not. R. Astron. Soc.* **423**, 3018–3030 (2012).
83. Behroozi, P. S., Wechsler, R. H. & Wu, H.-Y. The rockstar phase-space temporal halo finder and the velocity offsets of cluster cores. *Astrophys. J.* **762**, 109 (2011).
84. Behroozi, P. S., Wechsler, R. H. & Conroy, C. The average star formation histories of galaxies in dark matter halos from $z = 0$–8. *Astrophys. J.* **770**, 57 (2013).
85. Sargent, M. T. et al. The contribution of starbursts and normal galaxies to infrared luminosity functions at $z < 2$. *Astrophys. J.* **747**, L31 (2012).
86. Hayward, C. C. et al. Submillimetre galaxies in a hierarchical universe: number counts, redshift distribution and implications for the IMF. *Mon. Not. R. Astron. Soc.* **428**, 2529–2547 (2013).
87. Hayward, C. C. et al. What does a submillimeter galaxy selection actually select? The dependence of submillimeter flux density on star formation rate and dust mass. *Astrophys. J.* **743**, 159 (2011).



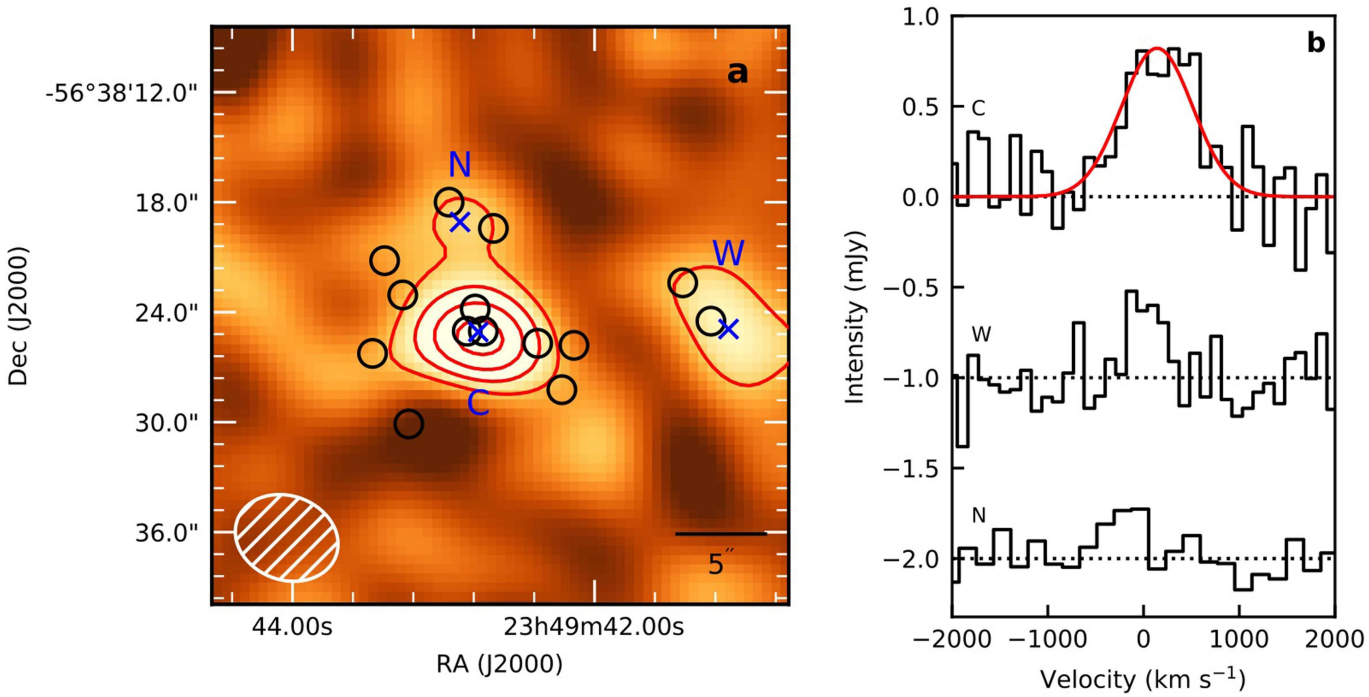
Extended Data Fig. 1 | IRAC observations of SPT2349-56. Circles show the locations of the 14 sources detected in ALMA band 7 (see Methods). Nine of the 14 ALMA sources are detected in the IRAC bands with at least 3σ confidence, including the two faintest [C II] sources from the blind line survey.



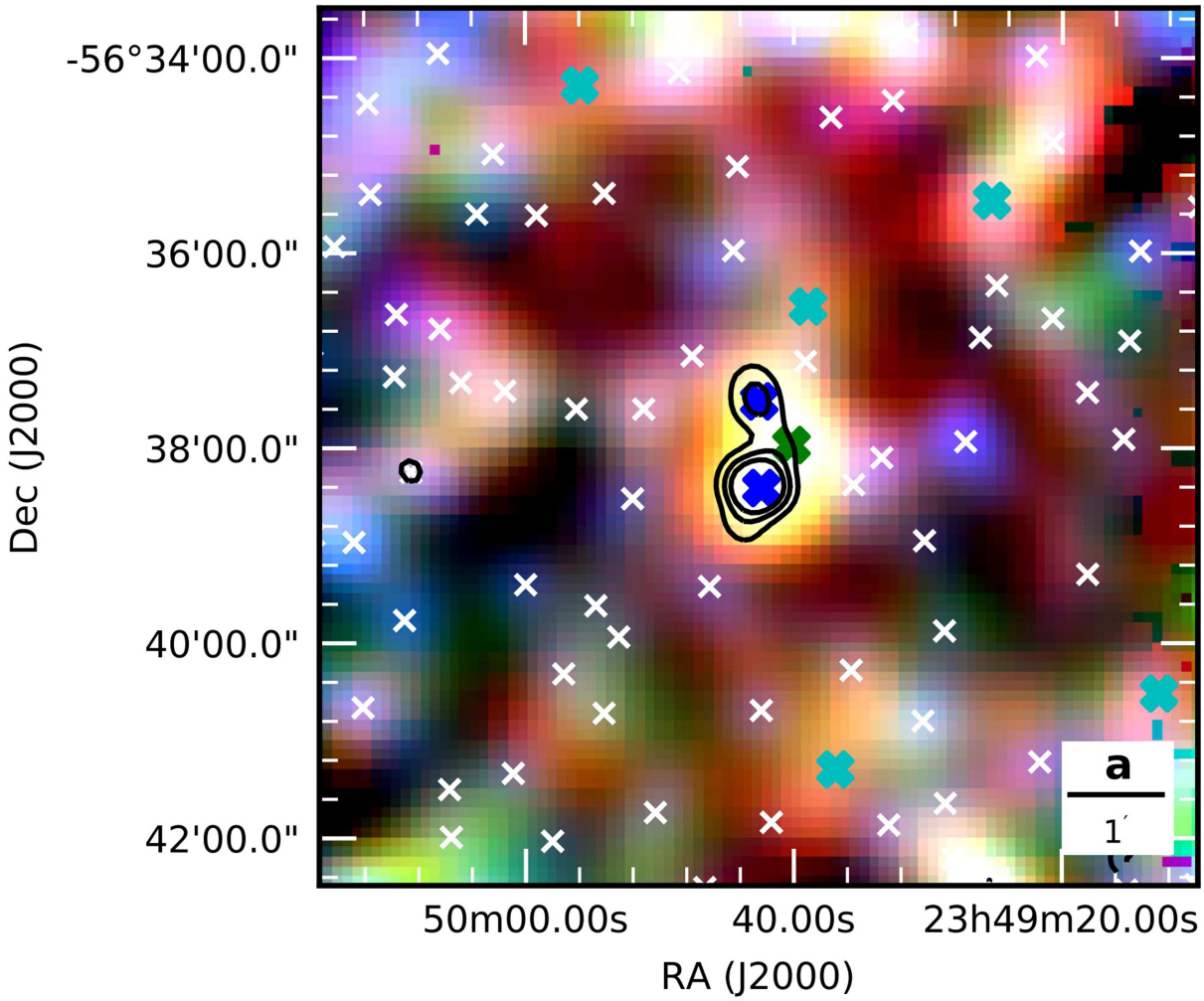
Extended Data Fig. 2 | Herschel–SPIRE image. An RGB scale is used to represent sources selected at wavelengths of 500 μm (red), 350 μm (green) and 250 μm (blue), with the red SPT2349-56 extended complex clearly visible in a relative void in the foreground $z \approx 1$ cosmic infrared background (blue to green coloured galaxies).

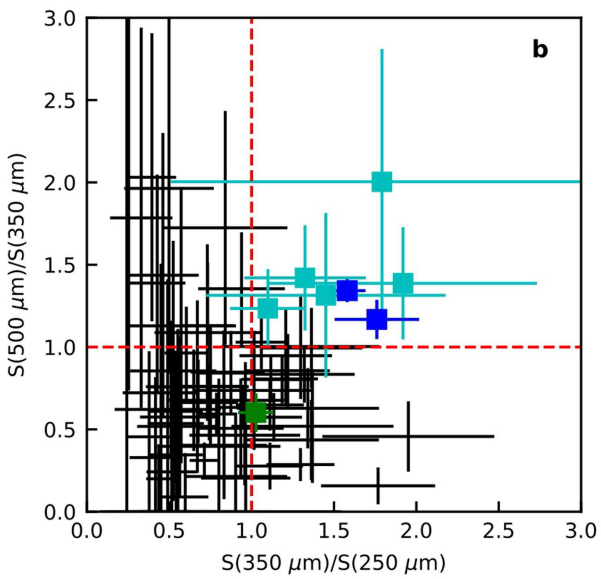


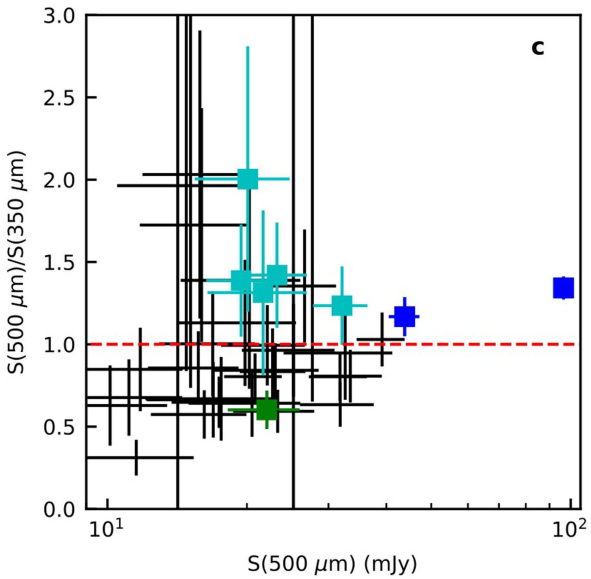
Extended Data Fig. 3 | Wide-field 870-μm image and photometry. A wide-field LABOCA image (21″ beam size; white circle) of SPT2349-56. The image r.m.s. noise is 1.3 mJy at the centre of region shown to the right, increasing to 2 mJy at the edges of this region. The total flux density recovered is 110.0 ± 9.5 mJy. Subregions are drawn with black outlines, showing three different regions and their recovered flux densities. Grey contours start at 3σ and increase in steps of 3σ. SPT 1.4-mm contours are also shown (blue), revealing that even with the 1′ beam of SPT, SPT2349-56 is resolved. One additional submillimetre source is detected at $> 5\sigma$ in the LABOCA image to the east (left) of the primary source, though Herschel–SPIRE photometry indicates that it is unlikely to be at $z \approx 4$.



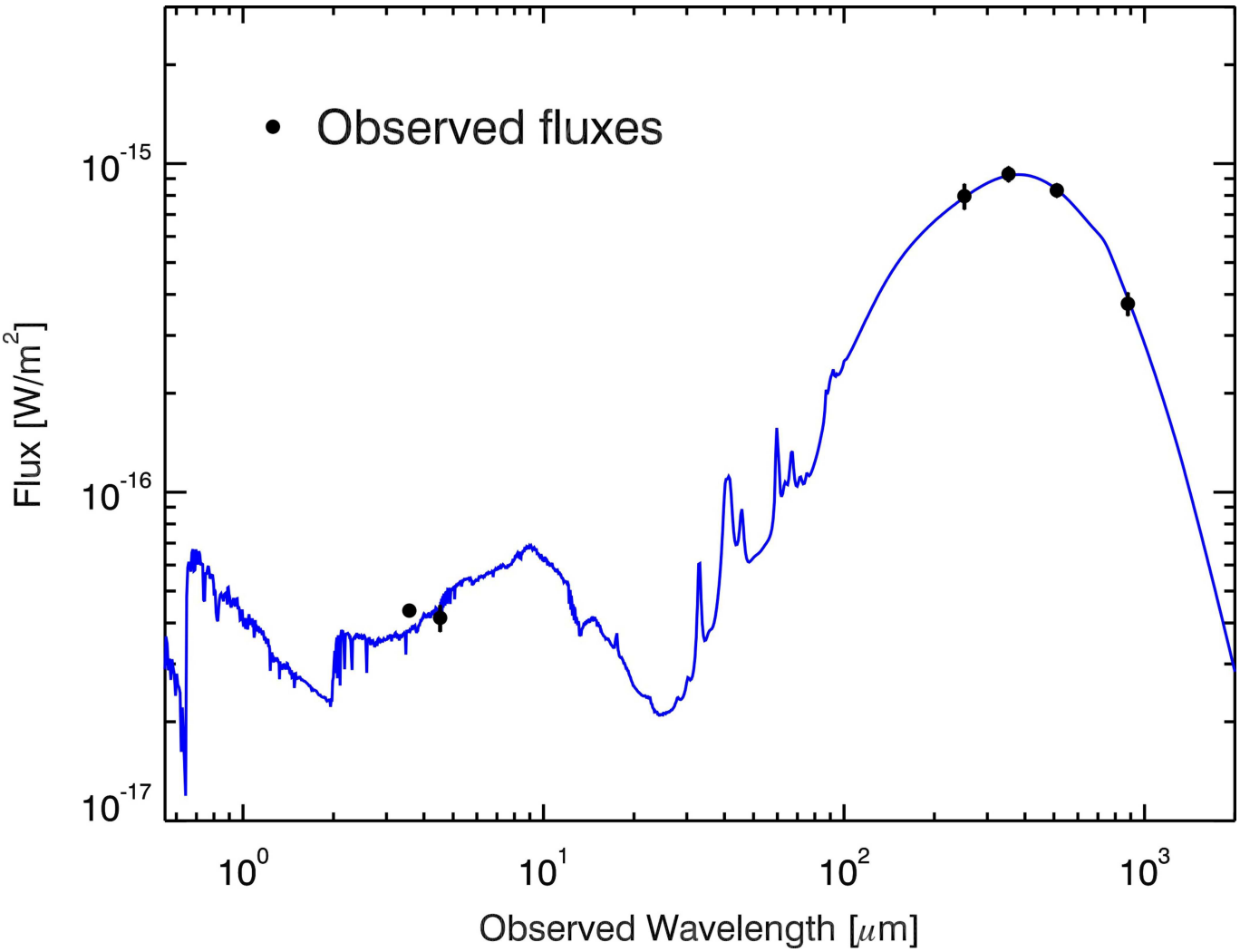
Extended Data Fig. 4 | CO(2–1) observations of SPT2349-56. a, The colour map and red contours trace the CO(2–1) line integrated over the central 830 km s^{-1}, with the contours spaced by 2σ, starting at 2σ. The grey contours show the 1.1-mm ALMA continuum detections. N, north; W, west; C, centre. Black circles show the location of the 14 [C II]/continuum sources identified with ALMA. **b**, One-dimensional spectra extracted at the positions indicated with blue crosses in panel **a**.



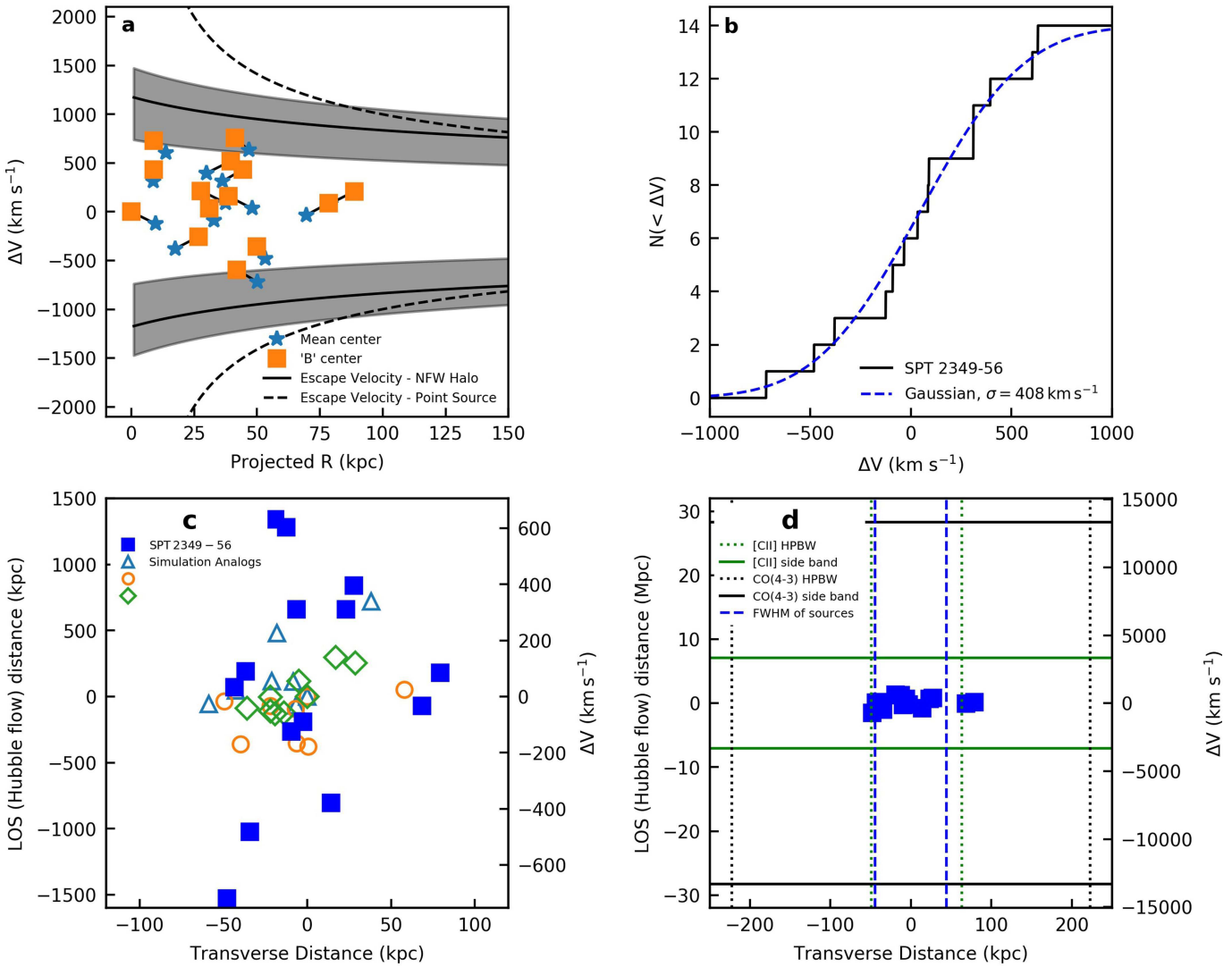




Extended Data Fig. 5 | SPIRE RGB image and source colours in the field surrounding SPT2349-56. a, A deep SPIRE false-colour image is shown with LABOCA contours overlaid. Locations of the 250-μm peaks used for analysis are marked with crosses (the faintest are not visible because of the contrast adopted in the image). The turquoise, blue, green and white crosses relate to the turquoise, blue, green and black lines used in panels **b** and **c**. **b**, **c**, Colour–colour (**b**) and colour–flux (**c**) diagrams for the 250-μm sources. Error bars represent 1σ standard deviation. The colour–colour diagram shows sources with SNR(250 μm) ratios of more than 3 and is dominated by the $z \approx 1$ cosmic infrared background in the foreground of SPT2349-56 (sources with colours ranging from blue to green). The colour–flux diagram applies an additional cut for SNR(500 μm) ratios of more than 3. The colour–colour and colour–flux diagrams show that one of three peaks associated with SPT2349-56 is probably a lower-redshift interloper (green symbol), but also that there are five additional sources (blue symbols) in the surrounding region with colours ($S_{500\mu m} > S_{350\mu m} > S_{250\mu m}$) that are suggestive of $z = 4.3$.



Extended Data Fig. 6 | Spectral energy distribution of SPT2349-56. The SED of the extended SPT2349-56 source is shown, including the summed deconvolved Herschel–SPIRE flux densities, the total 870-μm LABOCA flux density, and the summed IRAC flux densities. Error bars represent 1σ measurement errors. We do not include the SPT 1.4-mm, 2.0-mm and 3.0-mm points because the source is elongated and flux measurements are difficult with the filtering used to make the map. Fitting the SED yields an infrared luminosity of $(8.0 \pm 1.0) \times 10^{13} L_{\odot}$.



Extended Data Fig. 7 | Geometry and dynamics of the SPT2349-56 system. a, Velocity offsets of the 14 sources versus projected (physical) distance, compared with the escape velocity of a $1.16 \times 10^{13} M_{\odot}$ NFW halo (virial radius ≈ 200 kpc; concentration = 5). The grey shaded region shows our estimated halo-mass uncertainty; also shown is the escape velocity assuming a point mass halo of this same mass. Centres for the distribution of the 14 galaxies are shown at the mean of the distribution, and centred on the 'B' galaxy. All galaxies are bound for all but the lowest range of NFW halo masses. **b**, The cumulative velocity distribution of the SPT2349-56 galaxies, compared with a Gaussian distribution with our estimated dispersion (σ, 408 km s^{-1}), is at least consistent with expectations for a relatively bound system. **c**, The physical distribution of the SPT2349-56 galaxies (blue squares), assuming that their redshifts are due to cosmic expansion rather than peculiar motions. This gives an extreme (but unlikely) possibility that the SPT2349-56 galaxies are stretched out along a filament compared with their 130-kpc maximum tangential extent, but this requires that none of the velocity offsets is a peculiar motion. The open symbols show analogues of SPT2349-56 found when we searched specifically for maximally extended filamentary structures in our *N*-body simulations. These simulated structures are not filaments; they are instead rather like collapsed structures that are slightly cigar-shaped. The SPT2349-56 galaxies could in principle be distributed like this, but it does not fundamentally change our discussion here. We also note that we are plotting two different things here: velocity offsets for SPT2349-56, and actual geometry (three-dimensional positions) for the simulation galaxies. Even though our search allowed for structures extending by about 5 Mpc along the line of sight (LOS), we found none that stretches beyond 1 Mpc. **d**, As for panel **c**, except that the full extent of our ALMA band 3 and 7 observations is shown. No structures are observed in the sidebands surrounding the 14 observed sources.

Extended Data Table 1 | Observed properties of SPT2349-56 protocluster members

Source	RA (J2000) [h:m:s]	Dec (J2000) [d:m:s]	$S_{1090\,\mu m}$ [mJy]	$S_{870\,\mu m}$ [mJy]	CO(4-3) $\int$ S dv [Jy km s^{-1}]	CO(4-3) σ_V [km s^{-1}]	[CII] $\int$ S dv [Jy km s^{-1}]	[CII] σ_V [km s^{-1}]
A	23:49:42.67	−56:38:19.3	4.63 ± 0.04	7.8 ± 0.1	0.99 ± 0.03	376 ± 46	8.81 ± 0.26	354 ± 30
B	23:49:42.79	−56:38:24.0	4.35 ± 0.04	8.2 ± 0.1	0.92 ± 0.03	341 ± 38	7.53 ± 0.22	314 ± 28
C	23:49:42.84	−56:38:25.1	2.69 ± 0.04	6.0 ± 0.1	0.55 ± 0.02	154 ± 13	4.43 ± 0.17	160 ± 10
D	23:49:41.42	−56:38:22.6	2.20 ± 0.08	3.5 ± 0.3	0.69 ± 0.04	485 ± 64	3.62 ± 0.78	346 ± 129
E	23:49:41.23	−56:38:24.4	2.12 ± 0.11	3.3 ± 0.4	0.39 ± 0.02	199 ± 23	3.47 ± 1.24	310 ± 137
F	23:49:42.14	−56:38:25.8	1.69 ± 0.05	3.4 ± 0.1	0.28 ± 0.03	396 ± 103	4.28 ± 0.35	353 ± 35
G	23:49:42.74	−56:38:25.1	1.11 ± 0.04	2.7 ± 0.1	-	-	2.45 ± 0.23	305 ± 50
H	23:49:43.46	−56:38:26.2	0.85 ± 0.05	2.1 ± 0.1	-	-	3.63 ± 0.30	236 ± 31
I	23:49:42.22	−56:38:28.3	0.78 ± 0.05	1.8 ± 0.1	0.18 ± 0.03	277 ± 90	3.18 ± 0.32	236 ± 24
J	23:49:43.22	−56:38:30.1	0.61 ± 0.06	1.6 ± 0.2	0.19 ± 0.02	151 ± 38	3.79 ± 0.29	138 ± 15
K	23:49:42.96	−56:38:17.9	0.34 ± 0.04	1.4 ± 0.1	-	-	2.54 ± 0.17	129 ± 12
L	23:49:42.38	−56:38:25.8	0.23 ± 0.04	0.8 ± 0.1	-	-	2.78 ± 0.20	176 ± 20
M	23:49:43.39	−56:38:21.1	0.21 ± 0.05	0.5 ± 0.2	-	-	1.04 ± 0.14	87 ± 23
N	23:49:43.27	−56:38:22.9	0.18 ± 0.04	0.4 ± 0.1	-	-	0.86 ± 0.16	128 ± 26

Columns 6 and 8 show the integrated line intensity for the CO(4–3) and [C II] lines. Columns 7 and 9 show the line width for the CO(4–3) and [C II] lines.

Extended Data Table 2 | Properties of the three ATCA CO(2–1) sources

ATCA source	ALMA ID	$\int$ S dv [Jy km s^{-1}]	σ_V [km s^{-1}]	L′(CO 2-1) 10^{11} [K km s^{-1} pc^2]	$M_{\rm gas}$ [10^{11} M$_\odot$]
Central (C)	B, C, G	0.69 ± 0.076	372 ± 47	1.22 ± 0.12	1.33 ± 0.15
West (W)	D, E	0.16 ± 0.04	166 ± 47	0.29 ± 0.07	0.32 ± 0.08
North (N)	A, K	0.085 ± 0.0028	175 ± 68	0.15 ± 0.05	0.16 ± 0.05

Extended Data Table 3 | Observed properties of all red ($S_{500\mu m} > S_{350\mu m} > S_{250\mu m}$) SPIRE sources in the field surrounding SPT2349-56

RA (J2000) [h:m:s]	Dec (J2000) [d:m:s]	$S_{250\,\mu m}$ [mJy]	$S_{350\,\mu m}$ [mJy]	$S_{500\,\mu m}$ [mJy]	$S_{850\,\mu m}$ [mJy]	d† [arcmin]
23:49:42	−56:38:25	45 ± 3	71 ± 3	96 ± 3	77.0 ± 2.9	-
23:49:43	−56:37:31	21 ± 3	37 ± 3	43 ± 3	25.0 ± 2.8	0.9
23:49:25	−56:35:27	23 ± 4	26 ± 4	32 ± 4	2.9 ± 1.7	5.2
23:49:39	−56:36:33	12 ± 3	16 ± 3	23 ± 4	3.9 ± 1.3	2.1
23:49:36	−56:41:17	7 ± 3	14 ± 3	19 ± 3	3.2 ± 1.6	3.2
23:49:55	−56:34:17	6 ± 4	10 ± 3	20 ± 5	4.8 ± 1.8	5.3
23:49:12	−56:40:31	11 ± 5	16 ± 5	22 ± 5	6.8 ± 2.6	7.7

The LABOCA sources corresponding to SPT2349-56 are listed in the first two rows; the red SPIRE sources in the surrounding field follow. All sources listed are highlighted in Extended Data Fig. 5.
†Distance from central SPT2349-56 source.