The Astrophysical Journal, 864:49 (52pp), 2018 September 1 https://doi.org/10.3847/1538-4357/aacb79
© 2018. The American Astronomical Society. All rights reserved.

The CO Luminosity Density at High-z (COLDz) Survey: A Sensitive, Large-area Blind Search for Low-J CO Emission from Cold Gas in the Early Universe with the Karl G. Jansky Very Large Array

Riccardo Pavesi[1], Chelsea E. Sharon[1,2], Dominik A. Riechers[1], Jacqueline A. Hodge[3,4,5], Roberto Decarli[3,6], Fabian Walter[3], Chris L. Carilli[7,8], Emanuele Daddi[9], Ian Smail[10], Mark Dickinson[11], Rob J. Ivison[12,13], Mark Sargent[14], Elisabete da Cunha[15], Manuel Aravena[16], Jeremy Darling[17], Vernesa Smolčić[18], Nicholas Z. Scoville[19], Peter L. Capak[20], and Jeff Wagg[21]

[1] Department of Astronomy, Cornell University, Space Sciences Building, Ithaca, NY 14853, USA; rp462@cornell.edu
[2] Department of Physics & Astronomy, McMaster University, 1280 Main Street West, Hamilton, ON L85-4M1, Canada
[3] Max-Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany
[4] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
[5] Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands
[6] INAF—Osservatorio di Astrofisica e Scienza dello Spazio, via Gobetti 93/3, I-40129, Bologna, Italy
[7] National Radio Astronomy Observatory, P.O. Box O, Socorro, NM 87801, USA
[8] Cavendish Astrophysics Group, University of Cambridge, Cambridge, CB3 0HE, UK
[9] Laboratoire AIM, CEA/DSM-CNRS-Universite Paris Diderot, Irfu/Service d'Astrophysique, CEA Saclay, Orme des Merisiers, F-91191 Gif-sur-Yvette cedex, France
[10] Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK
[11] Steward Observatory, University of Arizona, 933 N. Cherry Street, Tucson, AZ 85721, USA
[12] European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748, Garching, Germany
[13] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[14] Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Brighton, BN1 9QH, UK
[15] Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia
[16] Núcleo de Astronomía, Facultad de Ingeniería, Universidad Diego Portales, Av. Ejército 441, Santiago, Chile
[17] Center for Astrophysics and Space Astronomy, Department of Astrophysical and Planetary Sciences, University of Colorado, 389 UCB, Boulder, CO 80309-0389, USA
[18] University of Zagreb, Physics Department, Bijenička cesta 32, 10002 Zagreb, Croatia
[19] Astronomy Department, California Institute of Technology, MC 249-17, 1200 East California Boulevard, Pasadena, CA 91125, USA
[20] Spitzer Science Center, California Institute of Technology, MC 220-6, 1200 East California Boulevard, Pasadena, CA 91125, USA
[21] SKA Organization, Lower Withington, Macclesfield, Cheshire SK11 9DL, UK
Received 2018 February 13; revised 2018 May 29; accepted 2018 June 6; published 2018 August 29

Abstract

We describe the CO Luminosity Density at High-z (COLDz) survey, the first spectral line deep field targeting CO(1–0) emission from galaxies at $z = 1.95$–2.85 and CO(2–1) at $z = 4.91$–6.70. The main goal of COLDz is to constrain the cosmic density of molecular gas at the peak epoch of cosmic star formation. By targeting both a wide ($\sim$51 arcmin^2) and a deep ($\sim$9 arcmin^2) area, the survey is designed to robustly constrain the bright end and the characteristic luminosity of the CO(1–0) luminosity function. An extensive analysis of the reliability of our line candidates and new techniques provide detailed completeness and statistical corrections as necessary to determine the best constraints to date on the CO luminosity function. Our blind search for CO(1–0) uniformly selects starbursts and massive main-sequence galaxies based on their cold molecular gas masses. Our search also detects CO(2–1) line emission from optically dark, dusty star-forming galaxies at $z > 5$. We find a range of spatial sizes for the CO-traced gas reservoirs up to $\sim$40 kpc, suggesting that spatially extended cold molecular gas reservoirs may be common in massive, gas-rich galaxies at $z \sim 2$. Through CO line stacking, we constrain the gas mass fraction in previously known typical star-forming galaxies at $z = 2$–3. The stacked CO detection suggests lower molecular gas mass fractions than expected for massive main-sequence galaxies by a factor of $\sim$3–6. We find total CO line brightness at $\sim$34 GHz of $0.45 \pm 0.2\ \mu$K, which constrains future line intensity mapping and CMB experiments.

Key words: galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: ISM – radio lines: galaxies – surveys

1. Introduction

Although the process of galaxy assembly through star formation is believed to have reached a peak rate at redshifts of $z = 2$–3 (i.e., $\sim$10–11 billion yr ago), the fundamental driver of this evolution is still uncertain (Madau & Dickinson 2014). In order to understand the physical origin of the cosmic star formation history (i.e., the rate of star formation taking place per unit comoving volume), we need to quantify the mass of cold, dense gas in galaxies as a function of cosmic time, because this gas phase controls star formation (Kennicutt & Evans 2012). In particular, the evolution of the cold gas mass distribution can provide strong constraints on models of galaxy formation by simultaneously measuring the gas availability and, through a comparison to the star formation distribution function, the global efficiency of the star formation process (see Carilli & Walter 2013 for a review). In this work, we carry out the first fully "blind" deep-field spectral line search for CO(1–0) line emission, arguably the best tracer of the total

molecular gas mass at the peak epoch of cosmic star formation, by taking advantage of the greatly improved capabilities of NSF's Karl G. Jansky Very Large Array (VLA).

To date, observations of the immediate fuel for star formation, i.e., the cold molecular gas, have mostly been limited to follow-up studies of galaxies that were preselected from optical/near-infrared (NIR) deep surveys (and hence based on stellar light) or selected in the submillimeter based on dust-obscured star formation as submillimeter galaxies (SMGs; for reviews, see, e.g., Blain et al. 2002; Casey et al. 2014). In particular, optical/NIR color-selection techniques (e.g., "BzK," "BM/BX"; Daddi et al. 2004; Steidel et al. 2004) have explored significant samples of massive star-forming galaxies at $z \sim 1.5$–2.5 (Daddi et al. 2008, 2010a; Tacconi et al. 2010, 2013), and the submillimeter selection has been particularly effective in identifying the most highly star-forming galaxies at this epoch for CO follow-up (e.g., Bothwell et al. 2013). Although such targeted CO studies are fundamental to explore the properties of known galaxy populations, they need to be complemented by blind CO surveys that do not preselect their targets, which may potentially reveal gas-dominated and/or systems with uncharacteristically low star formation rates missed by other selection techniques.

Targeted CO studies have found more massive gas reservoirs at $z \sim 2$ compared to local galaxies. Cold molecular gas is therefore believed to be the main driver for the high star formation rates of normal galaxies at these redshifts (e.g., Greve et al. 2005; Daddi et al. 2008, 2010b, 2010a; Genzel et al. 2010; Tacconi et al. 2010; Bothwell et al. 2013). Recent studies have claimed tentative evidence for an elevated star formation efficiency, i.e., star formation rate generated per unit mass of molecular gas, at $z \sim 2$ compared to local galaxies (e.g., Genzel et al. 2015; Schinnerer et al. 2016; Scoville et al. 2016, 2017; Tacconi et al. 2018). Such an elevated star formation efficiency could be related to massive, gravitationally unstable gas reservoirs. The interstellar gas content of galaxies therefore appears to be the main driver of the star formation history of the universe during the epoch when galaxies formed at least half of their stellar mass content (e.g., Madau & Dickinson 2014). Although targeted molecular gas studies currently allow us to observe larger galaxy samples more efficiently than blind searches, their preselection could potentially introduce an unknown systematic bias. Critically, such studies may not uniformly sample the galaxy cold molecular gas mass function. The best way to address such potential biases, and thus to complement targeted studies, is through deep-field blind surveys, in which galaxies are directly selected based on their cold gas content. Although some targeted CO(1–0) deep studies have previously been attempted (most notably Aravena et al. 2012 and Rudnick et al. 2017), these studies have typically targeted overdense (proto-)cluster environments. Hence, a blind search approach to sample a representative cosmic volume is needed in order to assess the statistical significance of such previous studies.

The CO(1–0) line emission is one of the most direct tracers of the cold molecular interstellar medium (ISM) in galaxies.[22] Its line luminosity can be used to estimate the cold molecular gas mass by means of a conversion factor ($\alpha_{\rm CO}$; see Bolatto et al. 2013 for a review). Although other tracers of the cold ISM have been utilized to date, including mid-J CO lines and the dust continuum emission, these are less direct tracers because they require additional, uncertain conversion factors (e.g., CO excitation corrections and dust-to-gas ratios). Specifically, while the ground-state CO(1–0) transition traces the bulk gas reservoir, mid-J CO lines such as CO(3–2) and higher-J lines are likely to preferentially trace the fraction of actively star-forming gas. Hence, their brightness requires additional assumptions about line excitation in order to provide a measurement of the total gas mass. Furthermore, different populations of galaxies may be characterized by significantly different CO excitation conditions (e.g., BzK, SMGs, and quasar hosts; Riechers et al. 2006, 2011c, 2011a; Daddi et al. 2010b; Ivison et al. 2011; Bothwell et al. 2013; Carilli & Walter 2013; Narayanan & Krumholz 2014), which also show considerable individual scatter (e.g., Sharon et al. 2016).

Long-wavelength dust continuum emission has been suggested to be a measure of the total gas mass and is utilized to great extent in recent surveys with the Atacama Large (sub-) Millimeter Array (ALMA) to investigate large samples of far-infrared (FIR)-selected galaxies (Eales et al. 2012; Bourne et al. 2013; Groves et al. 2015; Decarli et al. 2016; Scoville et al. 2016, 2017). Nonetheless, there remain substantial uncertainties in the accuracy of the calibration for this method at high redshift, especially below the most luminous, most massive sources.[23] Another caveat to using FIR continuum emission instead of CO comes from the finding that the dust emission measured by ALMA may not always trace the bulk of the gas distribution. This is made clear by the small sizes of the dust-emitting regions compared to the star-forming regions and the gas as traced by CO emission (e.g., Ivison et al. 2011; Riechers et al. 2011b, 2014; Simpson et al. 2015; Hodge et al. 2016; Chen et al. 2017; Miettinen et al. 2017).

Disentangling the causes for the observed increased star formation activity at $z \sim 2$ is not straightforward, since an increased availability of cold gas may be difficult to distinguish from increased star formation efficiency due to the uncertainty in deriving gas masses for representative samples of galaxies. Now, thanks to the unprecedented sensitivity and bandwidth of the VLA and ALMA, CO deep-field studies can be carried out efficiently, and these are ideal to address such potential selection effects. Previous deep-field studies with the Plateau de Bure Interferometer (PdBI; now the NOrthern Extended Millimeter Array, NOEMA) in the HDF-N (Decarli et al. 2014; Walter et al. 2014) and ALMA in the HUDF (ALMA SPECtroscopic Survey in the Hubble Ultra-Deep Field Pilot, or ASPECS-Pilot; Decarli et al. 2016; Walter et al. 2016) have provided the first CO blind searches covering mid-J transitions such as CO(3–2),[24] which are accessible at millimeter wavelengths. These studies have yielded crucial constraints on the molecular gas mass function at $z \sim 1$–3, subject to assumptions on the excitation of the CO line ladder to infer the corresponding molecular gas content.[25] They have found broad agreement with models of the CO luminosity evolution with

[22] In this work, CO always refers to the most abundant isotopologue, ^{12}CO.

[23] The dust continuum method to determine gas masses may be affected by the metallicity dependence of the dust-to-gas ratio (Sandstrom et al. 2013; Berta et al. 2016), trends in dust temperature with redshift (e.g., Magdis et al. 2012), or galaxy population (e.g., Faisst et al. 2017).

[24] The ASPECS-Pilot survey simultaneously covered the CO(2–1) line in the redshift range $z \sim 1.0$–1.7, the CO(3–2) line at $z \sim 2.0$–3.1, and higher-J CO transitions at higher redshift.

[25] A key challenge in these studies is the uncertainty in assigning candidate emission lines to the correct CO transition, in cases where the redshift of the observed line candidates is not independently known.

redshift by finding an elevated molecular gas cosmic density at $z > 1$ in comparison to $z \sim 0$, but they may suggest a tension with luminosity function models at $z \gtrsim 1$ by finding a larger number of CO line candidates than expected (Decarli et al. 2016).

In order to more statistically characterize the molecular gas mass function in galaxies at $z = 2$–3 and 5–7 than previously possible while avoiding some of the previous selection biases, we have carried out the COLDz survey[26]: a blind search for CO(1–0) and CO(2–1) line emission using the fully upgraded VLA.[27] The main objective of this survey is to constrain the CO(1–0) luminosity function at $z = 2$–3, which provides the most direct census of the cold molecular gas at the peak epoch of cosmic star formation free from excitation bias and based on a direct selection of the cold gas mass in galaxies. As such, the COLDz survey is highly complementary to millimeter-wave surveys like ASPECS and targeted studies. The CO(1–0) intensity mapping technique explored by Keating et al. (2015, 2016) is complementary to our approach. Intensity mapping offers sensitivity to the aggregate line emission signal from galaxies but only measures the second raw moment of the luminosity function (therefore not distinguishing between the characteristic luminosity and volume density). While the intensity mapping technique allows us to cover significantly larger areas of the sky, it does not directly measure gas properties of individual galaxies and is therefore complementary to direct searches such as COLDz.

In a previous paper (Lentati et al. 2015; Paper 0), we described a first interesting example of the galaxies identified in this survey. In this work (Paper I), we describe the survey, present the blind search line catalog, analyze the results of line stacking, and outline the statistical methods employed to characterize our sample. In Paper II, we will present the analysis of the CO luminosity functions and our constraints on the cold gas density of the universe at $z = 2$–7 (Riechers et al. 2018).

In Section 2 of this work, we describe the VLA COLDz observations, the calibration procedure, and the methods to mosaic and produce the signal-to-noise ratio (S/N) cubes. In Section 3, we describe our blind line search through matched filtering in 3D. In Section 4, we present our "secure" and "candidate" CO line detections in both the deeper (in COSMOS) and wider (in GOODS-N) fields. In Section 5, we utilize stacking of galaxies with previously known spectroscopic redshifts to provide strong constraints on their CO luminosity. In Section 6, we derive constraints to the total CO line brightness at ∼34 GHz. In Section 7, we discuss the implications of our results in the context of previous surveys. We conclude with the implications for future surveys with current and planned instrumentation. A more detailed analysis of the line search methods, statistical characterization of the candidate sample properties, and upper limits found for additional galaxy samples are presented in the Appendices.

In this work, we adopt a flat, ΛCDM cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_M = 0.3$ and a Chabrier IMF.

[26] The COLDz survey data, together with complete candidate lists, and analysis routines may be found online at coldz.astro.cornell.edu.

[27] The recently expanded VLA, with its new Ka-band detectors, new 3 bit samplers, simultaneous 8 GHz bandwidth, and improved sensitivity, for the first time enables carrying out this survey study.

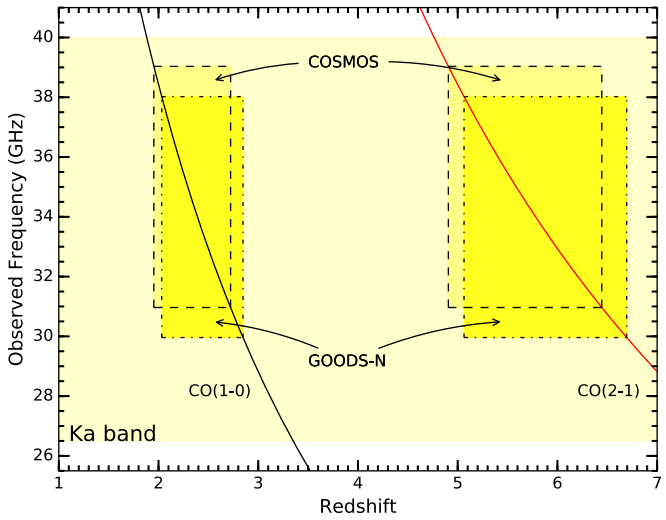

Figure 1. Frequency coverage of the VLA COLDz survey in the Ka band. The frequency range covers CO(1–0) at $z = 1.95$–2.85 and CO(2–1) at $z = 4.91$–6.70.

2. Observations

In order to constrain both the characteristic luminosity, L^*_{CO}, or "knee" of the CO(1–0) luminosity function, and the bright end, we have optimized our observing strategy following the "wedding cake" design to cover a smaller deep area and a shallower wide area. We have used the wide-band capabilities of the upgraded VLA to obtain continuous coverage of 8 GHz in the Ka band (PI: Riechers; IDs 13A-398, 14A-214) in a region of the COSMOS field (centered on the dusty starburst AzTEC-3 at $z = 5.3$ as a line reference source; Capak et al. 2011; Riechers et al. 2014) and in the GOODS-N/CANDELS-Deep field in order to take advantage of the availability of excellent multiwavelength data (Giavalisco et al. 2004; Grogin et al. 2011; Koekemoer et al. 2011).

The COSMOS data form a seven-pointing mosaic (center: R.A. = $10^h0^m20\overset{s}{.}7$, decl. = $2°35'17''$) with continuous frequency coverage between 30.969 and 39.033 GHz. The GOODS-N data form a 57-pointing mosaic (center: R.A. = $12^h36^m59^s$, decl. = $62°13'43\overset{''}{.}5$) with continuous coverage between 29.981 and 38.001 GHz (Figures 1, 2). The total on-source time was approximately 93 hr in the COSMOS field and 122 hr in the GOODS-N field. The frequency range targeted in this project covers CO(1–0) at $z = 1.95$–2.85 and CO(2–1) at $z = 4.91$–6.70, such that the space density of CO(2–1) line emitters is expected to be smaller than that for CO(1–0) (Figure 1; e.g., Popping et al. 2014, 2016). Both the large redshift spacing and the expected redshift evolution of the space density of CO emitters lessens the severity of the redshift ambiguity in our survey compared to previous studies.

At 34 GHz, the VLA primary beam can be described as a circular Gaussian with FWHM ∼ 80″, so our pointing centers were optimized to achieve a sensitivity that is approximately uniform in the central regions of the mosaics by choosing a spacing of 55″ ($<80''/\sqrt{2}$) in a standard hexagonally packed mosaic (Condon et al. 1998). During each observation, we targeted a set of seven pointings in succession, alternating through phase calibration. We performed pointing scans at the beginning of each observation, with additional pointing observations throughout for observations longer than 2 hr. Most of the COSMOS and GOODS-N data were taken in the D

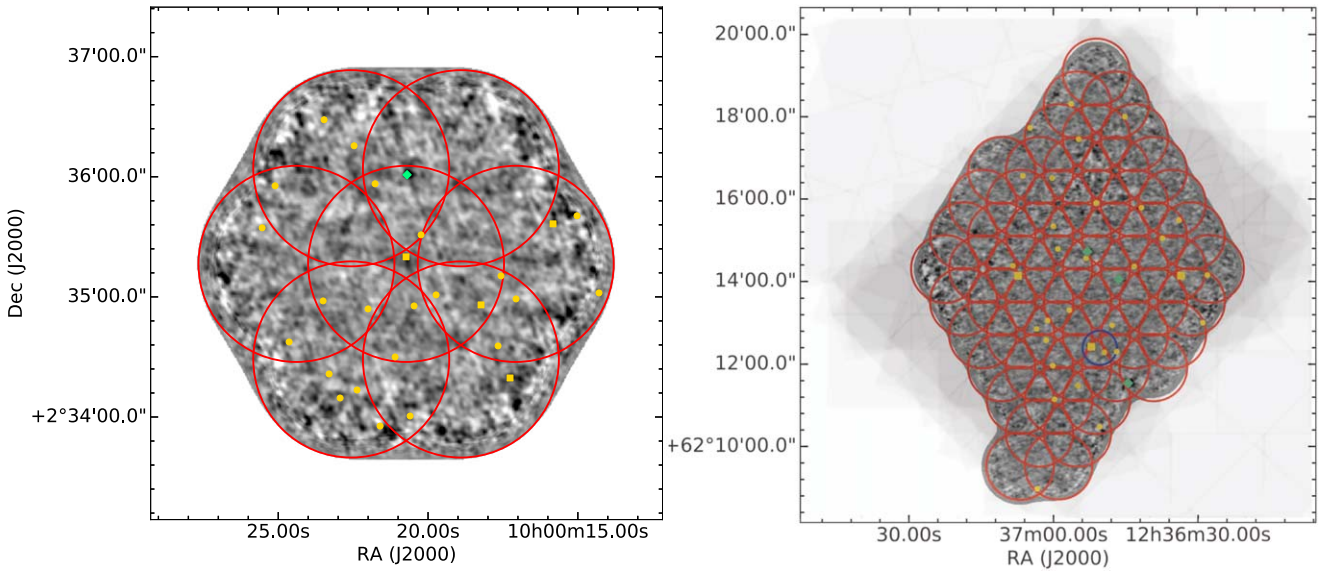

Figure 2. The CO deep-field regions covered by COLDz in the COSMOS (left) and GOODS-N/CANDELS (right) fields. The mosaics are composed of 7 and 57 pointings, respectively (as shown by the red circles). The gray scale corresponds to the frequency-averaged signal data cube. The positions of our line candidates from Tables 3 and 7 are marked by yellow squares and circles, respectively. Green markers indicate the positions of the most significant ∼34 GHz continuum sources. The field covered by the blind search from Decarli et al. (2014) is shown by a blue circle. We covered the majority of the CANDELS-Deep footprint in GOODS-N, shown as the background gray scale of the *HST*/WFC3 F160W exposure map (Grogin et al. 2011) for comparison.

Table 1
COLDz Observations Summary

Field	Pointing	D Configuration	D → DnC Configuration	DnC Configuration	DnC → C Configuration
Baseline range (m)		40–1000	40-2100	40–2100	40–3400
COSMOS	1–7	82 hr	⋯	11 hr	⋯
GOODS-N	1–7	13 hr	⋯	⋯	⋯
GOODS-N	6	⋯	⋯	3 hr	⋯
GOODS-N	8–14	15 hr	⋯	⋯	⋯
GOODS-N	15–21	15 hr	⋯	⋯	⋯
GOODS-N	22–28	14 hr	1.4 hr	⋯	⋯
GOODS-N	29–35	⋯	3 hr	12 hr	⋯
GOODS-N	36–42	⋯	⋯	11 hr	3 hr
GOODS-N	43–49	14 hr	⋯	1.3 hr	⋯
GOODS-N	50–56	10.5 hr	⋯	3.5 hr	⋯
GOODS-N	57	2 hr	⋯	⋯	⋯

Note. We list the total on-source time in different array configurations for all pointings in each group combined.

configuration of the VLA. Some of the observations, especially for the GOODS-N pointings, were fully or partially carried out in the DnC configuration, in reconfiguration from D to DnC (D → DnC), and in reconfiguration from DnC to C (DnC → C). Pointings are named sequentially from GN1 to GN57 (groups of GN1–7, GN8–14, etc. were observed together; Table 1).

The total area imaged, down to a sensitivity of ∼30% of the peak, in COSMOS is 8.9 arcmin^2 at 31 GHz and 7.0 arcmin^2 at 39 GHz. In GOODS-N, the total area is 50.9 arcmin^2 at 30 GHz and 46.4 arcmin^2 at 38 GHz. The correlator was set up in 3 bit mode at 2 MHz spectral resolution (corresponding to ∼18 km s^{-1} at 34 GHz) to simultaneously cover the full 8 GHz bandwidth for each polarization (Figure 1). Tuning frequency shifts between tracks, and sometimes in the same track, were used to mitigate the edge channel noise increase in order to achieve a uniform depth across the frequency range (Table 2).

Table 2
Lines, Redshift Ranges, and Volumes Covered by COLDz

Transition	ν_0 (GHz)	$z_{\rm min}$	$z_{\rm max}$	$\langle z \rangle$	Volume (Mpc^3)
		COSMOS			
CO(1–0)	115.271	1.953	2.723	2.354	20,189
CO(2–1)	230.538	4.906	6.445	5.684	30,398
		GOODS-N			
CO(1–0)	115.271	2.032	2.847	2.443	131,042
CO(2–1)	230.538	5.064	6.695	5.861	193,286

Note. The comoving volume is calculated to the edges of the mosaic and does not account for varying sensitivity across the mosaic, which is accounted for by the subsequent completeness correction. The average redshift is cosmic volume weighted.

2.1. COSMOS Observations

The data set in the COSMOS field consists of 46 dynamically scheduled observations between 2013 January 26 and 2013 May 14, each about 3 hr in duration. Flux calibration was performed with reference to 3C286, and J1041+0610 was observed for phase and amplitude calibration. Three frequency tunings offset in steps of 12 MHz were adopted to cover the gaps between spectral windows and obtain uninterrupted bandwidth.

2.2. GOODS-N Observations

The GOODS-N data set consists of 90 observations between 2013 January 27 and 2014 September 27, each about 2 hr in duration. Pointing 6, which covers the 3 mm PdBI pointing of the CO deep field in Decarli et al. (2014) in the HDF-N, was observed both as part of the 1–7 pointing set and in two additional targeted observations to achieve better sensitivity. Pointing GN57 was observed for 3 hr (127 minutes on source) in the D-array configuration on 2015 December 18 in order to follow up the most significant negative line feature in GN1–56 (see Appendix D for details). We used J1302+5748 for phase calibration, and the flux was calibrated by either observing 3C286 (in seven observations) or in reference to the phase calibrator (in the remaining observations). An average phase calibrator flux at 34 GHz of $S = 0.343$ Jy and spectral index of -0.2 was assumed in the observations in 2013, and $S = 0.21$ Jy and spectral index of -0.6 was assumed in 2014, as regularly measured in the tracks where a primary flux calibrator was observed. Based on track-to-track variations of the calibrator flux, we estimate an ∼20% total flux calibration uncertainty. The spectral setup employed uses two dithered sets of spectral windows with a relative shift of 16 MHz in order to fully cover the 8 GHz bandwidth available without gaps.

2.3. Data Processing

Data calibration was performed in CASA version 4.1 using the VLA data reduction pipeline (v.1.2.0). CASA version 4.5 was used to recalculate visibility weights using the improved version of `statwt` that excludes flagged channels when calculating weights and for imaging and mosaicking (McMullin et al. 2007). The pipeline radio-frequency interference (RFI) flagging, which uses CASA `rflag` to identify transient lines, was switched off, as recommended by the developers, since it can potentially remove narrow spectral lines and there is little RFI in the Ka band (with the exception of the 31.487–31.489 GHz range, which we flag prior to running the CASA pipeline). The pipeline was further modified to only flag the first and last channel of each spectral window (instead of three channels), regardless of proximity to baseband edges, to minimize the gap between sub-bands. We find that the bandpass is sufficiently flat that this choice gives the best trade-off between sensitivity in the end channels and additional noise, although some noise increase at the band edges is visible in Figure 3. After executing the pipeline, we visually inspected the visibilities in the calibrator fields to identify any necessary additional flagging. We then re-executed the pipeline to obtain a final calibration. In addition, for most GOODS-N observations, we modify the pipeline to flux-calibrate in reference to the gain calibrator (whenever a primary flux calibrator was not observed).

We identified a small number of noisy spectral channels in our observations that are not removed by the calibration pipeline. The noisy channels were initially discovered as narrow spikes of a small number of channels in amplitude-versus-frequency plots of visibilities from the science target fields, and they are mostly associated with single antennas. Being very narrow in frequency (one or two 2 MHz channels), the noise spikes are not significantly reduced by the statistical weights obtained from `statwt`, which minimizes the effects of all other noise features, since the weights are computed per spectral window. Including one of these noisy channels for the affected antenna during the imaging of a single pointing of the mosaic from a single observation track increases the rms noise by ∼20% in that frequency channel. Selecting channels whose standard deviation exceeds the mean standard deviation in that spectral window for that antenna by 3σ is a sufficient criterion to exclude most of the problematic noise spikes (these are only of order ∼0.2% of all channel–antenna combinations). This method is partially redundant to the algorithms in `rflag` (which we did not execute as part of the pipeline), but it reduces the risk of removing real spectral lines, since the noisy channels are selected for individual antennas. We also found that many noisy channels in the same antenna repeat over time during an observation and would therefore be more problematic if left in the data cube. We find a concentration of noise spikes in roughly four peaks over the frequency range that correlate with peaks in the weighted calibrated amplitudes as a function of frequency. We consider this to be indicative of random electronic problems that manifest as increased noise and thus are more prevalent in certain hardware components of the correlator than others. The presence of four peaks is likely associated with the underlying basebands, since there appears to be one peak in each baseband, but no precise correlation of the noise peak frequencies to the baseband edges could be identified. The feature is stochastic and does not appear to preferentially affect any particular subset of antennas. These noise spikes are at least twice as narrow as the narrowest blindly selected line candidates (which are rare among all candidates), and therefore residual anomalous noise spikes are believed not to measurably affect our line search.

Calibrated data from each pointing were imaged separately without any `CLEAN` cycles, because the fields do not contain strong continuum or line sources (see Section 4). We imaged the total intensity (sum of the two polarizations) using natural weighting and choosing a pixel size of $0\farcs5$ consistently in the two fields. The smallest adopted channel width is 4 MHz, equivalent to 35 km s^{-1} midband, which is less than the typical line width from galaxies. With this choice, our data cubes have ∼2000 channels after averaging polarizations. A crucial aspect of the imaging procedure, necessary for blind line searches, is to avoid any frequency regridding by interpolation; therefore, we image using the nearest channel, rather than interpolating. Interpolation would introduce correlations between the noise of adjacent channels, undermining the statistical basis of the search for spectral lines and producing a significant number of spurious noise lines. Disabling frequency interpolation when imaging visibilities may introduce a very small (less than half a channel) frequency error that we consider negligible because 4 MHz channels (35 km s^{-1}) are smaller than the typical line width.

The geometric average synthesized beam size for COSMOS ranges between $2\farcs2$ and $2\farcs8$ as a function of frequency, and the

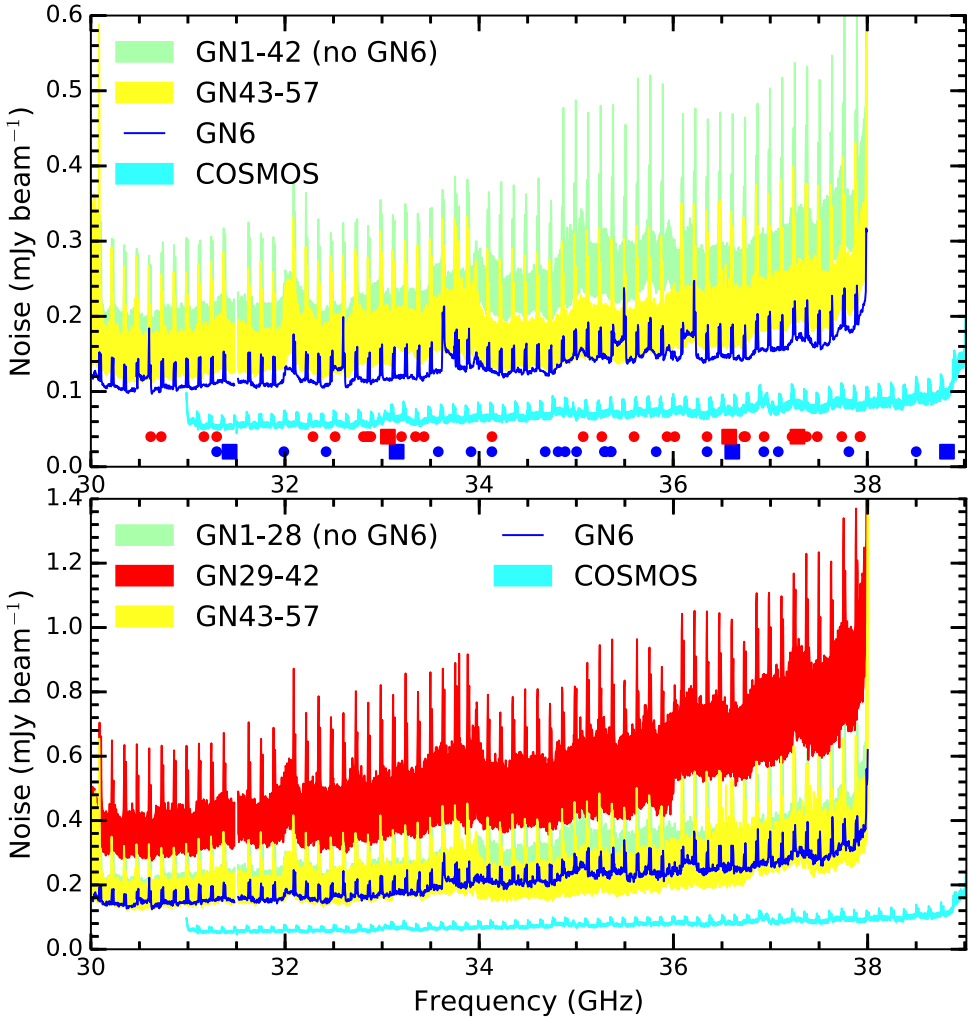

Figure 3. Measured rms noise per pointing in 4 MHz channels as a function of frequency at the native spatial resolution (top) and after smoothing to a common beam size (bottom). The bands are for groups of pointings that were observed in similar conditions and thus have similar noise characteristics. The GOODS-N pointings GN29–GN42, which were observed at higher resolution (predominantly in the DnC configuration), suffer a significant noise increase in the bottom panel due to spatial smoothing. The frequencies of our line candidates (blue in COSMOS, red in GOODS-N) from Tables 3 and 7 are marked by squares and circles, respectively, in the top panel.

beam axis ratio is in the range of 0.8–1, while for GOODS-N, the synthesized beam size differs more significantly from pointing to pointing. In particular, the main difference is between the subset of "high-resolution pointings" (GN20–GN42) and the rest (GN1–GN19 and GN43–GN57). The geometric average beam size ranges between approximately $1\farcs3$ and $2\farcs0$ for the high-resolution pointings and $2\farcs1$ and $3\farcs1$ for the rest. The beam axis ratio is in the range of 0.6–0.9. The individual pointing cubes are subsequently smoothed to a common beam size of $3\farcs38 \times 2\farcs91$ for COSMOS and $4\farcs1 \times 3\farcs2$ for GOODS-N using the CASA task `imsmooth`. This compromise in resolution and S/N is necessary in order to mosaic all pointings together and search for line emission in a uniform manner. This is called the Smoothed-mosaic. Separately, we have also mosaicked the pointings with their native resolution (after removing the beam information from the headers), and this Natural-mosaic (where the resolution is set by natural weighting) was used to exclusively search for spatially unresolved sources for which the spatial size information is not important. Figure 2 shows the spatial coverage provided by the individual pointings in our two mosaicked fields.

The CASA function `linearmosaic` was used to mosaic the images of our COSMOS data together. We wrote a custom script to optimally mosaic the images of the GOODS-N data using Equation (1), which takes into account the different noise levels in different pointings per channel in order to compute optimal weights for mosaicking:

$$I = \frac{\sum_p I_p A(\boldsymbol{x} - \boldsymbol{x}_p) \big/ \sigma_p^2}{\sum_p A(\boldsymbol{x} - \boldsymbol{x}_p)^2 \big/ \sigma_p^2}, \quad (1)$$

where A is the primary beam function, $\boldsymbol{x}_p$ is the pointing center position, I_p represents the specific intensity data from pointing p, and σ_p is the noise level in pointing p (computed on a per-channel basis).

All COSMOS pointings were always observed in every execution and for comparable amounts of time. The GOODS-N pointings were observed in blocks of seven over the course of several months due to scheduling constraints. Therefore, they have slightly different noise levels, partly due to the upgraded 3 bit samplers in the later (2014) observations. Furthermore, some GOODS-N data were not taken in the D configuration but rather in a combination of the DnC configuration in transition

between the D and the DnC and in transition between the DnC and the C configuration. Therefore, when smoothed to a common beam, these pointings have higher noise, because the information on the longest baselines is effectively discarded. For these reasons, the noise is significantly spatially varying in the smoothed version of the GOODS-N mosaic, which we take into account when analyzing the data (Figure 3 shows the noise before and after smoothing). All pointings suffer a noise increase due to smoothing, because the targeted beam for the smoothing process has to be larger than every beam in any pointing at any frequency and also includes those beams that have different position angles. In the case of the COSMOS mosaic, the CASA function `linearmosaic` produces the mosaic edge at 30% of peak level sensitivity (per-channel). For consistency, we therefore apply the same criterion in our GOODS-N mosaics. In order to do this, we define a mask that produces the mosaic edge at 30% of peak sensitivity in the Natural-mosaic and utilize the same mask for the Smoothed-mosaic for consistency.

2.4. Constructing the S/N Cubes

In order to search for emission lines in our data, we produce an S/N cube by calculating a noise value for each pixel and in each frequency channel of the mosaics. Spatial variations in the noise are introduced by mosaicking pointings with different noise levels and primary beam corrections. The noise in the resulting mosaic can be calculated assuming statistical independence of the noise in different pointings and can therefore be calculated by summing their standard deviations in quadrature, with weights given by Equation (1),

$$\sigma(x) = \frac{1}{\sqrt{\sum_p A(\boldsymbol{x} - \boldsymbol{x}_p)^2 / \sigma_p^2}}, \quad (2)$$

where σ_p is the measured noise in the individual pointing images, A is the primary beam function, and $\boldsymbol{x}_p$ is the pointing center position. In the special case of pointings with approximately equal noise (as in our COSMOS data), we can use a simplified expression, where the denominator is simply the square root of the sensitivity map output from CASA's `linearmosaic` function:

$$\sigma(x) = \frac{\sigma}{\sqrt{\sum_p A(\boldsymbol{x} - \boldsymbol{x}_p)^2}}. \quad (3)$$

The frequency variation of the noise is accounted for by measuring the noise in each frequency channel in the individual pointings. In COSMOS, where the noise variations from pointing to pointing can be neglected, we calculate the S/N by multiplying the signal cube by the square root of the sensitivity map (which gives spatially uniform noise) and then dividing each channel map by its standard deviation to normalize the pixel value distribution. In GOODS-N, we measure the noise in each pointing and apply Equation (2) to compute noise and S/N cubes.

3. Line Search Methods

The main objective of this survey is to carry out a blind search for CO(1–0) and CO(2–1) emission lines in the COLDz data set. No other bright lines are expected to contaminate the 30–39 GHz frequency range (Figure 1). The line brightness sensitivity is approximately equal for the low- and high-redshift

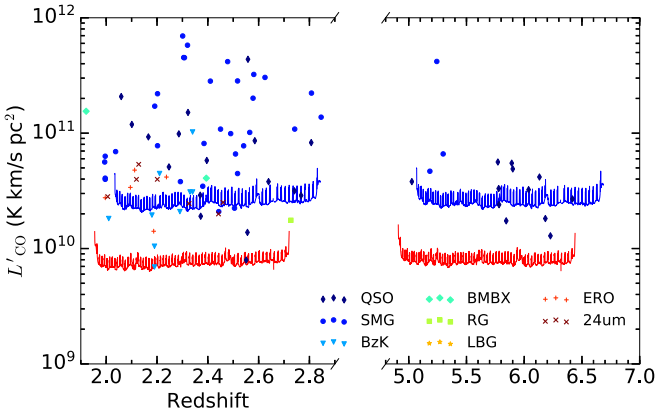

Figure 4. Line detection sensitivity limit reached by our observations (GOODS-N in blue, COSMOS in red). We assume a line FWHM of 200 km s^{-1} and a 5σ limit on the average pre-smoothing noise limit for direct comparison to Decarli et al. (2016). For comparison, we overplot all $z > 1$ CO detections to date from the compilation by Carilli & Walter (2013), as updated by Sharon et al. (2016). Colors mark different source types (quasars, SMGs, 24 μm–selected galaxies, Lyman-break galaxies, color-selected galaxies, and radio galaxies). At the sensitivity in the COSMOS field, we would be able to detect all previously detected CO emitters at high redshift in this compilation.

bins (corresponding to CO(1–0) and CO(2–1), respectively; Figure 4). We therefore expect a higher source density in the low-redshift bin due to the expected evolution of the cosmic gas density (Popping et al. 2014, 2016). Hence, we will assume that all detected features correspond to CO(1–0) unless data at other wavelengths suggest that they belong to the higher-redshift bin. In order to detect emission lines in our data cubes, we have implemented a previously published method (SPREAD; Decarli et al. 2014) and developed three new methods to explore the differences between different detection algorithms (see Appendix A for details).

The objective of a line search algorithm is to systematically assess the significance (expressed as S/N) of candidate emission lines in the data. The relevant information available to us is the strength of the signal, the number of independent samples that make up the line, and the spatial and frequency structure (for which we have priors based on previous samples of CO detections at high redshift). In particular, we expect most CO sources to be either unresolved or resolved over a few beams at most at the $\sim 3''$ resolution of our mosaics (which corresponds to $\sim$25 kpc at $z \sim 2.5$ and $\sim$17 kpc at $z \sim 6$), and we expect the line FWHM to be in the range of 50 to 1000 km s^{-1} (Carilli & Walter 2013).

Our line search method of choice, matched filtering in 3D (MF3D), expands on the commonly used spectral matched filtering (MF1D; e.g., AIPS `serch`). Matched filtering corresponds to convolving the data with a filter, or template, that is "matched" to the sources of interest in order to attenuate the noise and concentrate the full S/N of the source in the peak pixel. A detailed description of the MF3D method is presented in Appendix A.

We also implement and test some of the previously used methods on our data, in particular SPREAD and matched filtering in the spectral domain, i.e., in 1D. The main limitation of SPREAD is that it does not employ the full spatial information available but only utilizes signal strength. While matched filtering in 1D (MF1D) is arguably the optimal search method for completely unresolved sources (for which the spectrum at the peak spatial pixel contains the full information), it still

requires a prescription for identifying pixels belonging to the same source, and it needs to be generalized to account for the possibility that some sources may be slightly extended. Besides accounting for extended sources, the MF3D also captures the spreading of the S/N over different spatial positions in different frequency channels, which is at least in part a consequence of moderate S/N. For this reason, it is natural to use the spatial information by using templates that include a spatial profile. Therefore, we extended the method to matched filtering with 3D templates. A description of the detailed implementation of all line search methods and a more detailed comparison is presented in Appendix A.

4. Results of the Line Search

The MF3D procedure provides output including the maximal S/N for each line candidate, the position in the cube where that maximal S/N is achieved, the number of templates for which the candidate has $>4\sigma$ significance, and the template size (spatial and frequency width) where the highest S/N is achieved. We run the line search down to a low S/N threshold of 4σ. The number of identified features is very large, due to the large number of statistical elements in our data cubes. Specifically, we estimate approximately 2.8×10^6 and 1.7×10^7 independent elements for the COSMOS and GOODS-N fields, respectively, by dividing the mosaic area by the beam area and dividing by a line FWHM of 200 km s^{-1}. However, we caution that naively estimating the extent of the noise tails from these numbers does not provide a good estimate, as previously described by Vio & Andreani (2016) and Vio et al. (2017; also see Appendix F.2 for more details).

We mask radio continuum sources in our fields, which contaminate the line candidates: one in the COSMOS field at 10:00:20.67+02:36:01.5 with a flux of 0.024 mJy beam^{-1} and three in the GOODS-N field at 12:36:44.42+62:11:33.5 with a flux of 0.3 mJy beam^{-1}, 12:36:52.92+62:14:44.5 with a flux of 0.17 mJy beam^{-1}, and 12:36:46.34+62:14:04.46 with a flux of 0.07 mJy beam^{-1} (J. Hodge et al. 2018, in preparation). Even though the continuum fluxes of these sources only have low significance in the individual channels ($<0.3\sigma$ and $<2\sigma$ per 4 MHz channel for the brightest source in COSMOS and GOODS-N, respectively), we remove any candidate within 2.″5 of the spatial positions of these sources because they are likely spurious and caused by noise superposed to the continuum signal. Specifically, once we remove the continuum flux from their spectra, the significance of these line candidates becomes lower than $\sim 4.5\sigma$, indicating that they likely correspond to noise peaks.

In Table 3, we present the list of the secure line emitters in COSMOS and GOODS-N that were independently spectroscopically confirmed. While we are confident that our highest-S/N ($>6.4\sigma$) candidates correspond to real CO emission lines because they all have identified multiwavelength counterparts, we also define a longer list of line candidates that have significantly lower purity ($\sim$5%–40%) as a statistical sample in Appendix E, as described below. Although only a fraction of these tabulated sources are real emission lines, they provide statistical information once we account for their fractional purity, and therefore they may be used to constrain the CO luminosity function. While a fraction of these lower-significance candidates may be expected to correspond to real CO emission, we advise caution in interpreting these lower-significance candidates on a per-source basis until they are independently confirmed.

In order to determine the reliability of the line candidates presented in Appendix E, we compare the S/N distribution to that for "negative" line candidates following the standard practice (e.g., Decarli et al. 2014; Walter et al. 2016) that relies on the symmetry of interferometric noise. We provide a detailed description of our candidate purity estimation in Appendix F, but we point out that an excess of positive candidates over the negatives for S/N above a threshold is an indication that at least a fraction of those positive candidates may correspond to real sources, rather than being due to noise. By adopting this criterion, we determined the S/N thresholds for our candidate lists consistently for both fields by cutting at the S/N level that includes as many negative line candidates as unconfirmed positive candidates. Thus, we exclude from the count the high-S/N confirmed sources (four in COSMOS and two in GOODS-N; Table 3), and we require that the number of unconfirmed sources is greater than the number of negative lines down to the same S/N threshold, thereby constituting an excess. This procedure determines S/N thresholds on the candidate catalog of 5.25σ for the smaller COSMOS field and 5.5σ for the wider GOODS-N field Appendix E. The threshold is chosen to be higher in the wide GOODS-N mosaic because the larger number of statistical elements produces more pronounced noise tails.

4.1. Measuring Line Candidate Properties

After selecting the blind search line candidates, we separately measure their line properties using a standard method described in the following. The statistical corrections were computed adopting identical methods in the artificial source analysis (Appendix F.3).

In order to extract the spectrum of the line candidates, we fit a 2D Gaussian to the velocity-integrated line maps and extract the flux in elliptical apertures with sizes equal to the FWHM of the fitted Gaussians. For the integrated line maps, we use a velocity range equal to the FWHM of the template that maximizes the S/N. This procedure is expected to provide the highest S/N of the extracted flux. In the infinite S/N case, this aperture choice includes half of the total flux, and we therefore correct the extracted flux scale of the spectrum by a factor of two. We then fit a Gaussian line profile to the aperture spectrum and measure its peak flux and velocity width, from which we derive the integrated fluxes reported in Tables 3 and 7. We also measure peak fluxes for the candidates, which are expected to best represent the correct flux for unresolved sources. For the peak fluxes, we extract the spectrum at the highest pixel in the integrated line map. We find that the peak fluxes are compatible with aperture fluxes for point-like sources, and so we choose to adopt the aperture fluxes because they measure the full flux of extended sources at the expense of slightly larger uncertainties. We calculate the positional and size uncertainty of the 2D Gaussian fitting using the CASA task `imfit` applied to the same integrated line maps described above. The positional uncertainty is relevant when establishing counterpart associations (as detailed in Appendix E for the full candidate list). It is dominated by the detection S/N and the spatial size of the synthesized beam or extended emission.

In the COSMOS field, we can measure aperture fluxes in the Natural-mosaic to make full use of the highest S/N (the fluxes are typically within 20% of the values measured in the Smoothed-mosaic). Specifically, the seven pointings of the

Table 3
Catalog of the Secure Line Candidates Identified in Our Analysis that Have Been Independently Confirmed (See Table 7 for the Remainder of the Full Statistical Sample)

ID	R.A. (J2000.0)	Decl. (J2000.0)	Frequency (GHz)	Redshift	Flux (Jy km s^{-1})	FWHM (km s^{-1})	S/N	Opt/NIR c.part?	Comments
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
COSMOS									
COLDz.COS.0	10:00:20.70	+02:35:20.5	36.609 ± 0.002	5.2974 ± 0.0003[a]	0.17 ± 0.02	390 ± 40	14.7	Y	AzTEC-3
COLDz.COS.1	10:00:15.80	+02:35:37.0	31.430 ± 0.003	2.6675 ± 0.0004	0.11 ± 0.03	430 ± 80	10.6	Y	…
COLDz.COS.2	10:00:18.20	+02:34:56.5	33.151 ± 0.006	2.4771 ± 0.0006	0.13 ± 0.03	830 ± 130	9.6	Y	Source reported by Lentati et al. (2015)
COLDz.COS.3	10:00:17.23	+02:34:19.5	38.822 ± 0.003	1.9692 ± 0.0002	0.37 ± 0.10	240 ± 50	9.2	Y	Extended
GOODS-N									
COLDz.GN.0	12:36:33.45	+62:14:08.85	36.578 ± 0.005	5.3026 ± 0.0009[a]	0.344 ± 0.074	610 ± 100	8.56	Y	GN10
COLDz.GN.3	12:37:07.37	+62:14:08.98	33.051 ± 0.006	2.4877 ± 0.0006	0.34 ± 0.12	580 ± 120	6.14	Y	GN19
COLDz.GN.31[b]	12:36:52.07	+62:12:26.49	37.283 ± 0.007	5.1833 ± 0.0008[a]	0.148 ± 0.057	490 ± 140	5.33	Y	HDF850.1

Notes. Columns are (1) line ID; (2)–(3) R.A. and decl. (J2000); (4) central line frequency and uncertainty based on Gaussian fitting; (5) CO(1–0) redshift and uncertainty, unless otherwise noted; (6) velocity-integrated line flux and uncertainty; (7) line FWHM as derived from a Gaussian fit; (8) S/N measured by MF3D; (9) presence of a spatially coincident optical/NIR counterpart; (10) comments.
[a] CO(2–1) redshift.
[b] Source is below the formal catalog threshold adopted here and therefore not part of the statistical sample.

mosaic have an approximately equal beam size. This allows us to calculate an average beam size for each channel and hence to correctly measure aperture fluxes. These are the fluxes we report in Tables 3 and 7 for the most significant candidates, which we also use for the luminosity function.

In the GOODS-N field, on the other hand, we are limited to measuring aperture fluxes for resolved objects in the Smoothed-mosaic because of the strong beam size variations across the mosaic that make it impossible to precisely define a beam in the Natural-mosaic. Nonetheless, since most of the candidates are unresolved in the original data (show highest S/N in the Natural-mosaic), in those cases, we report the peak fluxes measured in the Natural-mosaic without concern for missing any flux and without being affected by the beam size variations.

In the GOODS-N field, there is another beam size effect that needs to be taken into account even in the Smoothed-mosaic. The measured beam size is actually larger than the formal $4\farcs1 \times 3\farcs2$ size that was targeted with the CASA task `imsmooth` and is slightly pointing-dependent, as explained in Appendix C. The measured beam area is ∼1.4 times larger in the D-array-only pointings and ∼1.7 times larger in the higher-resolution pointings than the target size for the smoothing procedure because of the precise *uv*-plane coverage and the effect of tapering. Therefore, we measure the correct beam size after smoothing by Gaussian-fitting to the smoothed dirty beam in each pointing for each channel. We correct the aperture flux for each candidate line detection in the Smoothed-mosaic by calculating an effective beam area given by a weighted average of the beams of the overlapping pointings weighted by the square of the primary beams (the same weighted average that determines the flux in the mosaic). We calculate aperture fluxes in this way in the Smoothed-mosaic and confirm that the peak pixel flux in unresolved sources matches this corrected aperture flux within the uncertainties.

The measured CO line fluxes are affected by the effect of a warmer cosmic microwave background (CMB) at the redshift of our sources, which is a uniform background (hence invisible to an interferometer) at the small scales of galaxy sizes (da Cunha et al. 2013a). While we do not expect corrections for our $z = 2$–3 sources to be significant (∼20%–25%), a larger correction (up to a factor of ∼2) may be required if the gas kinetic temperature were lower than expected. On the other hand, the CO(2–1) line luminosity from the $z > 5$ sources may be underestimated by up to a factor of ∼2–5 (da Cunha et al. 2013a). We do not apply any of these corrections to the measured line flux values reported here. These effects will be further discussed in Paper II, in the context of the CO luminosity function.

4.2. Individual Candidates

We have identified 26 line candidates in the COSMOS field down to an S/N threshold of 5.25 and 31 candidates in the GOODS-N field down to an S/N threshold of 5.5 (Tables 3 and 7). The top four sources in COSMOS and two among the highest-S/N sources in GOODS-N have been independently confirmed through additional CO transitions (Daddi et al. 2009; Riechers et al. 2010a, 2011b, 2018; Ivison et al. 2011; R. Pavesi et al. 2018, in preparation). Furthermore, we include COLDz.GN.31 in this set of independently confirmed sources (Table 3), although it is slightly below the formal 5.5σ cutoff, because it corresponds to CO(2–1) line emission from HDF850.1 (Walter et al. 2012). This line source does not contribute to our evaluation of the CO(2–1) luminosity function because it does not satisfy the significance threshold to be included in the statistical sample (Paper II). For reference, we here briefly describe these individual secure candidates, and we show their CO line maps and spectra in Figures 5 and 6. Line maps and spectra of the complete statistical sample are presented in Appendix E for reference.

We detect four previously known dust-obscured massive starbursting galaxies and three secure sources in the COSMOS field that lie within the scatter of the high-mass end of the main sequence at $z \sim 2$ (Lentati et al. 2015; R. Pavesi et al. 2018, in preparation) These galaxies may be representative of a galaxy population that has not been well studied to date, due to our novel selection technique.

COLDz.COS.0.—We identify the brightest candidate in the COSMOS field with CO(2–1) from the $z = 5.3$ SMG AzTEC-3, detected at an S/N of 15 and chosen to be near the center of our survey region. This galaxy is known to reside in a massive protocluster (Riechers et al. 2010a, 2014; Capak et al. 2011). The line flux is compatible with the previously measured value of 0.23 ± 0.03 Jy km s^{-1} (Riechers et al. 2010a) within the relative flux calibration uncertainty. This source is also detected at 3 GHz with a flux of 20 ± 3 μJy (Smolčić et al. 2017) and by SCUBA-2 at 850 μm as part of the S2COSMOS survey with a significance of 9.3σ and a flux of $8.1^{+1.1}_{-1.3}$ mJy (J. M. Simpson et al. 2018, in preparation).

COLDz.COS.1.—This high-S/N detection is matched in position (offset $0\farcs3 \pm 0\farcs3$) and CO(1–0) redshift to a source with photometric redshift ($z_{\rm phot} = 2.6$–2.9) in the COSMOS2015 catalog (Laigle et al. 2016). We have confirmed its redshift with ALMA through a detection of the CO(3–2) line (R. Pavesi et al. 2018, in preparation). This source is also detected at 3 GHz with a flux of 15 ± 2 μJy (Smolčić et al. 2017) and at 850 μm with a significance of 6.0σ and a flux of $4.9^{+1.1}_{-1.2}$ mJy (J. M. Simpson et al. 2018, in preparation).

COLDz.COS.2.—This high-S/N detection is matched in position (offset $0\farcs3 \pm 0\farcs3$) to a source in the COSMOS2015 catalog (Laigle et al. 2016). We have confirmed its redshift with ALMA through a detection of the CO(3–2) line (R. Pavesi et al. 2018, in preparation), and some of its properties were previously presented in Lentati et al. (2015). The photometric redshift in the COSMOS2015 catalog is highly uncertain and not compatible with the CO redshift of 2.477 within 1σ ($z_{\rm phot} = 2.9$–4.4). This source is also detected at 3 GHz with a flux of 19 ± 3 μJy (Smolčić et al. 2017) and at 850 μm with a significance of 5.9σ and a flux of $4.0^{+0.9}_{-1.0}$ mJy (J. M. Simpson et al. 2018, in preparation).

COLDz.COS.3.—This high-S/N detection is a significantly spatially extended CO source with a deconvolved size of $(4\farcs0 \pm 1\farcs1) \times (1\farcs8 \pm 1\farcs2)$. It is matched in position to two galaxies in the COSMOS2015 catalog (Laigle et al. 2016; offsets of $0\farcs14 \pm 0\farcs3$ and $1\farcs8 \pm 0\farcs3$). We have confirmed its CO(1–0) redshift with ALMA through a detection of the CO(4–3) line (R. Pavesi et al. 2018, in preparation). The cataloged photo-z for both galaxies ($z_{\rm phot} = 1.8$–1.9) is not compatible with the CO redshift of 1.97 within 1σ. This source is also detected at 3 GHz with a flux of 27 ± 3 μJy (Smolčić et al. 2017). The S2COSMOS survey shows a weak signal at 850 μm with a significance of 3.7σ. The formal 4σ limit on the deboosted flux is <4.0 mJy, and the tentative detection suggests a potential source at a flux level of ∼2–3 mJy (J. M. Simpson et al. 2018, in preparation).

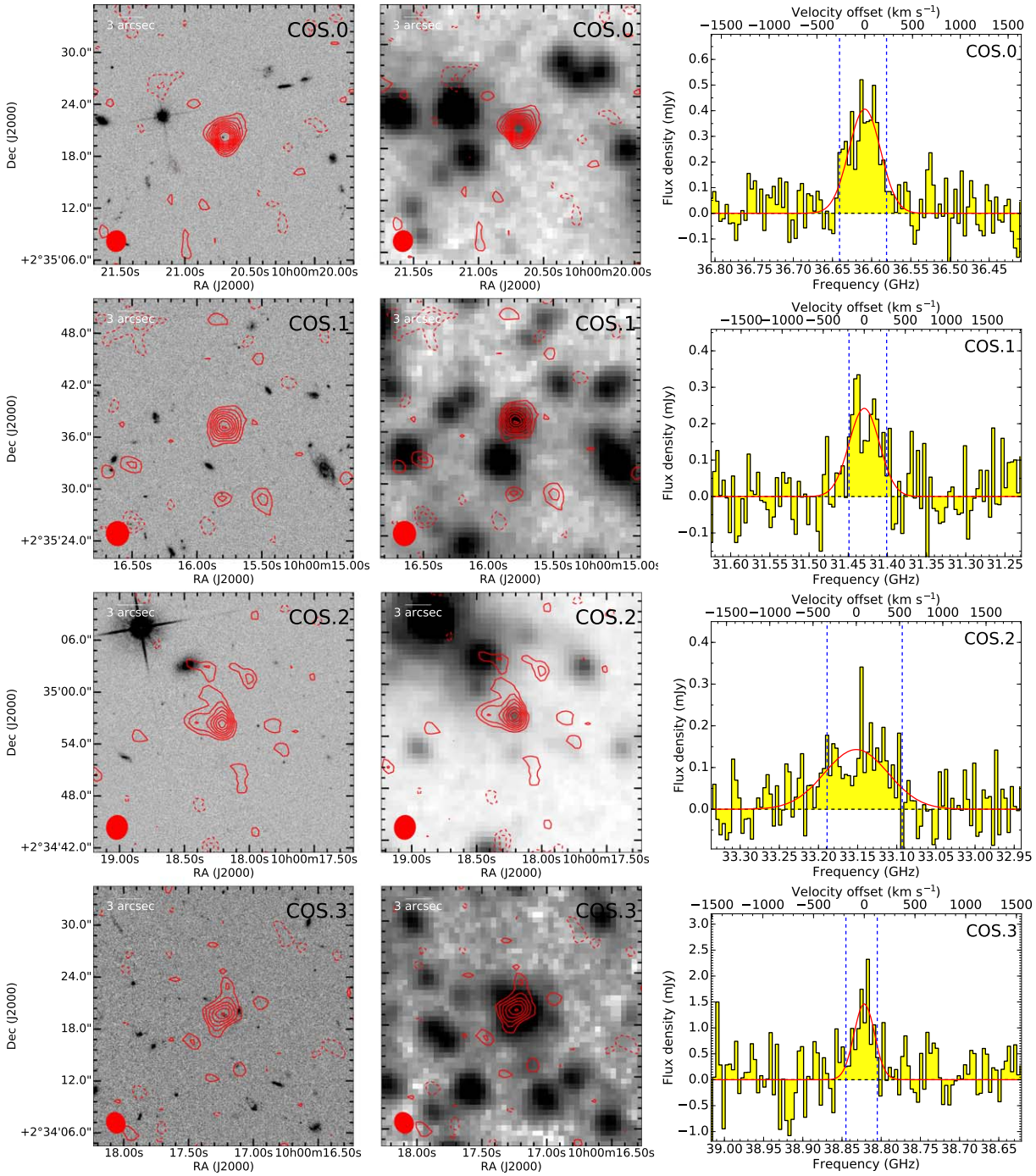

Figure 5. Independently confirmed candidates from our blind line search in the COSMOS field. CLEANed integrated line emission (contours) is shown overlaid on *HST I*-band (left) and IRAC 3.6 μm (middle) images from SPLASH (gray scale; Steinhardt et al. 2014). Contours are shown in steps of 1σ, starting at $\pm 2\sigma$. COS.0 corresponds to CO(2–1) emission from AzTEC-3. Right: extracted line candidate aperture spectra (histograms) and Gaussian fits (red curves) to the line features. The observer-frame frequency resolution of 4 MHz corresponds to $\sim$35 km s^{-1} midband. The velocity range that was used for the overlays is indicated by the dashed blue lines.

COLDz.GN.0.—We identify the brightest candidate in the GOODS-N field with CO(2–1) line emission from GN10, a massive, bright, dust-obscured starbursting galaxy (Pope et al. 2006; Dannerbauer et al. 2008; Daddi et al. 2009). We find a CO redshift of $z = 5.3$, showing that the previous redshift determination ($z = 4.04$) was incorrect. Its properties are described in Riechers et al. (2018). This source is also detected at 1.4 GHz with a flux of 36 ± 4 μJy (Morrison et al. 2010) and by SCUBA-2 at 850 μm in the SCUBA-2 Cosmology Legacy Survey (S2CLS) with a significance of 9.2σ and a flux of 7.5 ± 1.5 mJy (Geach et al. 2017).

COLDz.GN.3.—We identify this source with CO(1–0) line emission from GN19, a merger of two massive, bright, dust-obscured starbursting galaxies at $z = 2.49$ found by Pope et al. (2006) and characterized in detail by Tacconi et al. (2006, 2008), Riechers et al. (2011b), and Ivison et al. (2011). It is detected by the 5.5 GHz eMERGE survey with a flux of 9.6 ± 1.7 μJy (Guidetti et al. 2017). Its line flux is

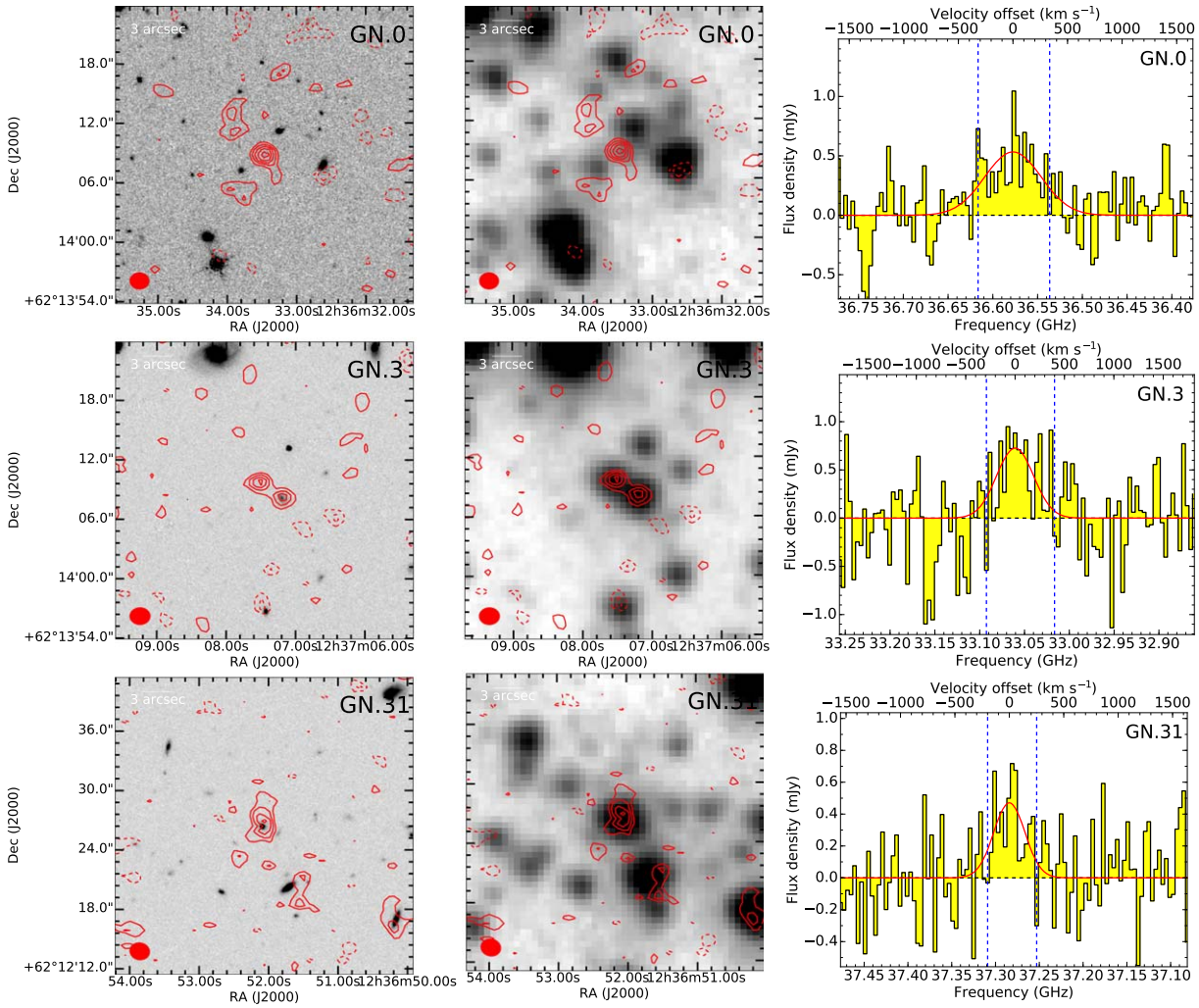

Figure 6. Independently confirmed candidates from our blind line search in the GOODS-N field. Integrated line emission (contours) is shown overlaid on *HST H*-band (left) and IRAC 3.6 μm (middle; gray scale) images. The *HST* and *Spitzer* images were obtained from the CANDELS database. The contours are shown in steps of 1σ, starting at $\pm 2\sigma$. COLDz.GN.0 (GN10) and COLDz.GN.3 (GN19) were separately CLEANed because the high S/N allows us to meaningfully deconvolve the emission. GN.31 corresponds to CO(2–1) emission from HDF850.1. Right: extracted line candidate single-pixel/aperture spectra (histograms; for unresolved/resolved emission) and Gaussian fits to the line features (red curves). The observer-frame frequency resolution of 4 MHz corresponds to $\sim$35 km s^{-1} midband. The velocity range used for the overlays is indicated by the dashed blue lines.

compatible with the previously measured total flux of 0.33 ± 0.04 Jy km s^{-1} from Riechers et al. (2011b). This source is also detected at 1.4 GHz with a flux of 28 ± 4 and 33 ± 4 μJy for the W and E components, respectively (Morrison et al. 2010), and at 850 μm with a significance of 7.9σ and a flux of 6.5 ± 1.1 mJy (Geach et al. 2017).

COLDz.GN.31.—We also detect CO(2–1) line emission from the bright, dust-obscured starbursting galaxy HDF850.1 ($z = 5.183$) with a moderate significance of S/N = 5.3. We include this line detection here given the known match, but we do not include it in the statistical analysis because it does not reach the significance threshold for detection by the blind line search. The measured flux is compatible with the previously reported flux of 0.17 ± 0.04 Jy km s^{-1} (Walter et al. 2012). It is detected by the 5.5 GHz eMERGE survey with a flux of 14 ± 3 μJy (Guidetti et al. 2017) but is not detected at 1.4 GHz (Morrison et al. 2010). This source is also detected at 850 μm with a significance of 7.1σ and a flux of 5.9 ± 1.3 mJy (Geach et al. 2017).

The other line candidates identified by our blind line search with moderate significance are to date not independently confirmed (Appendix E). Thus, we only use their properties in a statistical sense in the following to place more detailed constraints on the CO luminosity function. We point out that three out of the seven secure, confirmed sources in our blind search belong to the high-redshift bin and therefore suggest caution in interpreting the indicated CO(1–0) redshift, especially for those line candidates without strong counterparts. We describe the complete candidate sample in Appendix E, where we also discuss potential counterpart associations. In Appendix F, we develop novel statistical techniques to evaluate the purity and completeness of this statistical sample that yield the best constraints to the CO(1–0) luminosity function at $z \sim 2$–3 to date (Paper II).

4.3. Statistical Counterpart Matching

All S/N > 6.4 candidates in COSMOS and GOODS-N have optical, NIR, and/or radio/(sub)millimeter counterparts (in

addition to GN19 and HDF850.1). At lower S/N, it becomes more difficult to establish definitive counterparts due to the modest precision of photometric redshifts and potential (apparent or real) spatial offsets of the emission. Our purity analysis (Appendix F) suggests that the contamination from noise is considerable. As an example, for the candidates shown below S/N = 6, we may expect only one or two out of 10 to be real CO line emitters due to the large sizes of the data cubes. Therefore, we consider the lack of counterparts as a possible indication that a line candidate may be due to noise. On the other hand, the very objective of a blind search for CO-emitting galaxies is to address a potential bias against optical/NIR-faint galaxies. Possible explanations for the lack of counterparts are (1) the stellar light could be too dust-obscured to be visible in the rest-frame optical/NIR; (2) the CO line may correspond to the $J = 2$–1 transition, placing the galaxy at $z > 5$, such that counterparts may only exist below the detection limit; and (3) a CO-bright emitter may be gas-rich but have a low star formation rate and/or stellar mass, which would make it optically "dark."

4.3.1. Optical-NIR Counterparts

We here consider the uncertain line candidates near and below the S/N threshold only. If we match all $5 < \mathrm{S/N} < 6$ candidates in COSMOS (60 in total) to the COSMOS2015 photometric catalog (Laigle et al. 2016) by requiring a spatial separation of $<2''$ and a z_{CO10} or z_{CO21} within the 68th percentile range of the photometric redshifts, we find 10 matches. This is $\sim 2.7\sigma$ higher than the number of matches found for random displacements of the positions of our candidates (randomly expecting $\sim 4.7 \pm 2.0$ associations). We therefore conclude that some ($\sim$3–7) of the 10 associations (out of these top 60 candidates) are likely to be real physical counterparts to real CO line emitters, in agreement with our typical purity estimate of order $\sim$10% for the statistical sample in this S/N interval (Appendix F). Consistently, we also find a 1.8σ excess of positional matches within $<2''$ for this extended candidate sample, 20 matches with a 13.8 ± 3.4 false-positive rate, by spatially associating to the *Spitzer*/IRAC-based catalog by the deep SEDS survey (Ashby et al. 2013). This confirms that at least a fraction of our line candidates in the COSMOS field at these lower S/N levels may have real counterpart associations, to be confirmed by future spectroscopic observations.

We repeat the same procedure in GOODS-N for the candidates with $\mathrm{S/N} > 5.4$, excluding the independently confirmed ones (51 in total). We employ the best redshifts available from Skelton et al. (2014) and Momcheva et al. (2016) using the same selection criteria with a separation requirement of $<2''$. The grism spectroscopy does not significantly impact our matched counts, as almost all of the potential counterparts are too faint and only have photometric redshifts. We only find a slight excess relative to chance associations ($\sim 1.1\sigma$) by finding nine associations at an expected chance rate of 6.3 ± 2.5. The latest "super-deblended" GOODS-N catalog from Liu et al. (2018) does not yield any additional associations besides the secure sources corresponding to GN10 and GN19. In addition, we search for positional matches within $<2''$ for this extended candidate sample by searching for spatial associations in the Ashby et al. (2013) *Spitzer*/IRAC-based catalog from the deep SEDS survey. We do not find any excess of matches over the expected false-positive rate.

The counterpart association signal in GOODS-N does not constitute a significant excess, perhaps due to contamination by chance associations with low-redshift galaxies. However, at least approximately 6–10 line candidates out of the top $\sim$200 have a very close *Spitzer*/IRAC counterpart ($<1''$) and a photometric redshift estimate that is compatible with the CO(1–0) line candidate, as would be expected for real counterpart matches.

In the following, we evaluate the implications of a lack of 3.6 μm counterparts for some of our lower-S/N CO line candidates. The deep *Spitzer*/IRAC images in Figure 5 and Appendix E are derived from the SPLASH observations (Steinhardt et al. 2014), while the *Spitzer*/IRAC images in GOODS-N were obtained as part of the legacy GOODS program (Giavalisco et al. 2004). Due to the moderate resolution of *Spitzer* observations, these images are sometimes contaminated by lower-redshift galaxies or stars, reducing our ability to detect counterparts at higher redshift; hence, in those cases, the following limits may not apply. In order to asses the implications of a counterpart nondetection in the IRAC 3.6 μm images, we use template spectral energy distributions (SEDs) for star-forming galaxies from Bruzual & Charlot (2003), redshift them to $z \sim 2.3$, and convolve them with the IRAC 3.6 μm filter curve using MAGPHYS to estimate the stellar mass limits placed by a lack of detection in COSMOS or GOODS-N (da Cunha et al. 2008, 2015). The expected mass-to-light ratio at this wavelength depends on the stellar population ages and star formation histories, as well as on the degree of dust extinction. The following estimates are thus only indicative. We estimate that the lack of IRAC 3.6 μm counterparts at the $\sim$0.2 and $\sim$0.06 μJy limits ($\sim 3\sigma$; Ashby et al. 2013) of the COSMOS and GOODS-N data correspond to approximate stellar mass upper limits of $\sim 6 \times 10^9$ and $\sim 2 \times 10^9\, M_\odot$, respectively, at $z \sim 2.3$ for a representative $A_V \sim 2.5$.[28] These limits suggest that a lack of infrared counterparts implies either a very low stellar mass or a high degree of dust obscuration. The stellar mass limits would be significantly higher for a line candidate associated with CO(2–1) emission at $z > 5$. Indeed, repeating the same calculations for $z \sim 5.8$, we obtain significantly less constraining stellar mass limits of $\sim 1.3 \times 10^{11}$ and $\sim 4 \times 10^{10}\, M_\odot$ for a representative $A_V \sim 2.5$ in COSMOS and GOODS-N, respectively.

4.3.2. Radio Counterparts

We also searched for counterpart matches in the deep COSMOS 3 GHz continuum catalog (Smolčić et al. 2017), only finding associations for COLDz.COS0, COS1, COS2, and COS3 by using a $3''$ search radius. Of the 18 sources from the Smolčić et al. (2017) catalog located within the boundaries of our mosaic, our secure sources represent the only ones with a redshift estimate (photometric or spectroscopic, when available) falling within our survey volume. All remaining sources from the Smolčić et al. (2017) catalog within our survey area lie in the range $z = 0.1$–1.6. We performed an equivalent search in the catalog from the eMERGE 5.5 GHz survey of the GOODS-N field (Guidetti et al. 2017), finding a single association for GN19 by using a $3''$ search radius. We also searched the VLA 1.4 GHz catalog of the GOODS-N field from Morrison et al. (2010) with the same criteria, finding two

[28] For reference, the limits would be $\sim 1.6 \times 10^9$ and $5 \times 10^8\, M_\odot$ for $A_V < 0.5$ and $\sim 2 \times 10^{10}$ and $7 \times 10^9\, M_\odot$ at $A_V \sim 5$.

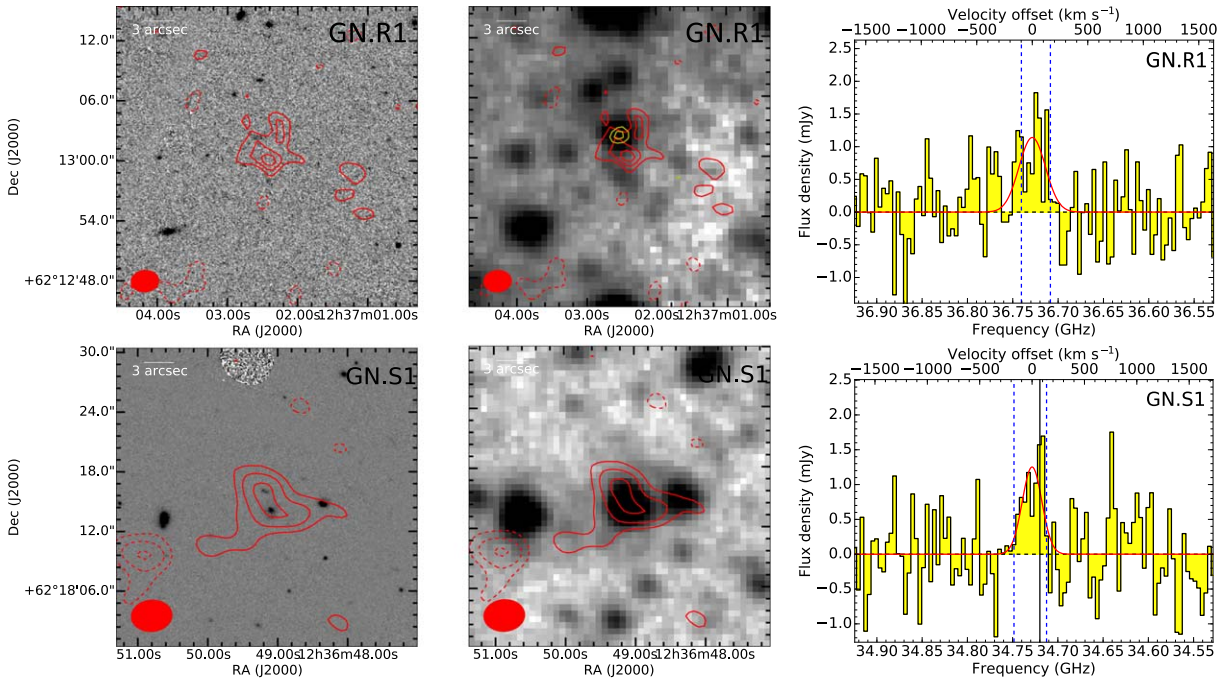

Figure 7. Candidates with radio continuum counterparts and optical spectroscopic redshift below the catalog threshold (i.e., not part of our statistical sample). Left: line map overlays (red contours) over *HST H*-band (left) and IRAC 3.6 μm (middle; gray scale) images. Red contours show the CO line in steps of 1σ, starting at $\pm 2\sigma$. Yellow contours show the radio continuum emission in steps of 2σ, starting at $\pm 3\sigma$ (Morrison et al. 2010). Right: extracted line candidate aperture spectra (histograms) and Gaussian fits (red curves) to the line features. The frequency resolution is the same as in Figure 5. The velocity range used for the overlays is indicated by the dashed blue lines. The solid black line shows the CO(1–0) frequency corresponding to the optical spectroscopic redshift of $z_{\rm spec} = 2.320$.

matches for GN19 and a radio counterpart for GN10. We also used these radio catalogs to search for counterpart associations with CO candidates to a lower significance of 5σ. We found one candidate at an S/N = 5.13 that satisfies the requirement of close association with a radio source (within $3''$) and with a CO redshift that is compatible with the 1σ interval for the photometric redshift listed by the optical-NIR photometric catalogs. This candidate, named COLDz.GN.R1 in the following, is at J2000 12:37:02.53+62:13:02.1 and has an offset of $1\farcs7 \pm 0\farcs6$ from the radio source. We show this candidate in Figure 7. The photometric redshift estimate is 4.73–5.30, which is compatible with the CO(2–1) redshift of $z_{21} = 5.277 \pm 0.001$ implied by the COLDz data. We measure a CO(2–1) line luminosity of $(8 \pm 3) \times 10^{10}$ K km s^{-1}pc^2, which implies a gas mass of $(2.9 \pm 1.1) \times 10^{11}\,M_\odot$ for a standard $\alpha_{\rm CO} = 3.6\,M_\odot$ (K km s^{-1} pc^2)$^{-1}$ and $r_{21} = 1$. The Skelton et al. (2014) catalog reports a stellar mass of $2.8 \times 10^{11}\,M_\odot$, which suggests a molecular gas mass fraction of $\sim$1. The radio continuum fluxes are $S_{1.4\rm GHz} = (26 \pm 4)\,\mu$Jy and $S_{5.5\rm GHz} = (20 \pm 4)\,\mu$Jy (Morrison et al. 2010; Guidetti et al. 2017, respectively). This suggests a star formation rate of $\sim$200–400 $M_\odot$ yr^{-1} when applying the radio–FIR correlation (Delhaize et al. 2017). Although this line candidate has a higher probability of corresponding to real emission than implied by its S/N, we do not include it in the statistical analysis to preserve the unbiased (i.e., CO S/N-limited) nature of our selection.

The deep radio catalogs by Smolčić et al. (2017) in COSMOS and by Morrison et al. (2010) in GOODS-N have a 5σ sensitivity limit of $\sim$11 μJy at 3 GHz and 20 μJy at 1.4 GHz, respectively, which can be converted to upper limits on the $L_{\rm FIR}$ for radio counterparts to our line candidates through the radio–IR correlation. By adopting the relationship from Delhaize et al. (2017), we deduce a detection limit of $L_{\rm FIR} < (4\text{–}7) \times 10^{11}\,L_\odot$ in the $z = 1.953$–2.847 redshift range. On the other hand, recent results have suggested that the radio–FIR correlation in disk-dominated star-forming galaxies may not show a redshift evolution as used by Delhaize et al. (2017) (Molnár et al. 2018). If true, this would suggest less-constraining limits of $L_{\rm FIR} < (1\text{–}3) \times 10^{12}\,L_\odot$. The 5σ sensitivity limit of the eMERGE catalog at 5.5 GHz is approximately 15 μJy and corresponds to limits of $L_{\rm FIR} < (8\text{–}14) \times 10^{11}\,L_\odot$ according to Delhaize et al. (2017) and $L_{\rm FIR} < (2.5\text{–}6) \times 10^{12}\,L_\odot$ according to Molnár et al. (2018). These limits may be constraining, because our measured $L'_{\rm CO}$ would imply median $L_{\rm FIR} \sim 10^{12}\,L_\odot$ and $\sim 4 \times 10^{12}\,L_\odot$ based on the star formation law (Daddi et al. 2010b; Genzel et al. 2010) for our unconfirmed line candidates in the COSMOS and GOODS-N fields, respectively. Possible reasons for the lack of radio counterparts may be due to fainter radio fluxes in our sample than expected from the radio–IR correlation or lower star formation rates than expected based on the gas masses, or candidates may correspond to CO(2–1) emission at $z > 5$. Alternatively, line candidates may not be real and may be due to noise. The possibility of gas-rich, low star formation rate galaxies would be particularly interesting because surveys like the one reported here may be the only way to uncover such a hidden population.

5. Identification and Stacking of Galaxies with Previous Spectroscopic Redshifts

5.1. Identification and Stacking of Previous Mid-J Blind CO Surveys

We searched the GOODS-N data set for low-*J* CO counterparts to the candidate mid-*J* CO detections from our

Table 4
Low-*J* CO Counterpart Search for HDF-N PdBI Blind Mid-*J* CO Candidates

ID	Preferred Redshift	PdBI Preferred Mid-*J* Line	PdBI Covered Mid-*J* Line	This Survey Low-*J* Line	Low-*J* Flux or 3σ Limit (Jy km s^{-1})	$L'_{\rm CO}$ Constraint
ID.01	1.88	2–1	5–4	2–1	<0.05	$r_{52} > 1.6$
ID.02	1.81	2–1	5–4	2–1	<0.02	$r_{52} > 2.6$
ID.03	1.78 (secure)	2–1	5–4	2–1	<0.05	$r_{52} > 1.8$
ID.04	1.71	2–1	5–4	2–1	<0.03	$r_{52} > 0.9$
ID.05	2.85	3–2	5–4	2–1	<0.06	$r_{52} > 1.2$
ID.08 (HDF850.1)	5.19 (secure)	5–4	5–4*	2–1	0.17 ± 0.06	$r_{52} = 0.40 \pm 0.16$
ID.10	2.33	3–2	3–2*/6–5	1–0/2–1	<0.03	r_{31} and $r_{62} > 0.7$
ID.11	2.19	3–2	3–2*/6–5	1–0/2–1	<0.04	r_{31} and $r_{62} > 0.9$
		3–2	7–6	2–1	<0.03	$r_{72} > 1.0$
ID.12	2.19	3–2	3–2*/6–5	1–0/2–1	<0.05	r_{31} and $r_{62} > 0.6$
		3–2	7–6	2–1	<0.03	$r_{72} > 0.7$
ID.13	2.18	3–2	3–2*/6–5	1–0/2–1	<0.04	r_{31} and $r_{62} > 0.6$
		3–2	7–6	2–1	<0.03	$r_{72} > 0.7$
ID.14	2.15	3–2	3–2*/6–5	1–0/2–1	<0.04	r_{31} or $r_{62} > 0.8$
		3–2	7–6	2–1	<0.03	$r_{72} > 0.7$
ID.15	2.15	3–2	3–2*/6–5	1–0/2–1	<0.04	r_{31} and $r_{62} > 1.2$
		3–2	7–6	2–1	<0.03	$r_{72} > 1.2$
ID.17 (HDF850.1)	5.19 (secure)	6–5	6–5*	2–1	0.17 ± 0.06	$r_{62} = 0.24 \pm 0.10$
ID.18	2.07	3–2	3–2*/6–5	1–0/2–1	<0.05	r_{31} and $r_{62} > 1.3$
		3–2	7–6	2–1	<0.04	$r_{72} > 1.3$
ID.19	2.05 (secure)	3–2	3–2*/6–5	1–0/2–1	<0.07	r_{31} and $r_{62} > 0.7$
		3–2	7–6	2–1	<0.04	$r_{72} > 0.9$
ID.20	2.05	3–2	3–2*/6–5	1–0/2–1	<0.05	r_{31} and $r_{62} > 0.8$
		3–2	7–6	2–1	<0.03	$r_{72} > 1.1$
ID.21	3.04	4–3	7–6	2–1	<0.04	$r_{72} > 0.8$

Note. Preferred redshifts are quoted from Decarli et al. (2014). Although we systematically constrain every possible *J* line assignment that would place a CO line in our data, asterisks mark those cases where the assignment is preferred by Decarli et al. (2014). Here ID.03 has a secure redshift identification that places it outside our redshift coverage, and ID.19 has a secure redshift that places it within our coverage.

previous CO blind survey in the HDF-N with the PdBI (Decarli et al. 2014). We find a single match in our candidate list, which corresponds to CO(2–1) line emission from HDF850.1. We systematically searched for every possible mid-*J*/low-*J* CO line combination that would place a low-*J* CO line in our surveyed volume (Table 4). Several of these possible mid-*J*/low-*J* CO line combinations are not the preferred line identifications by Decarli et al. (2014). Therefore, our nondetections are consistent with their preferred redshift in those cases constraining or ruling out several alternative redshift solutions allowed by the PdBI data alone. In order to search for lower-significance candidate lines, we extract spectra at the mid-*J* candidate positions and evaluate the significance of any features or place 3σ upper limits to the line fluxes. By assuming the same line FWHM as the candidate mid-*J* CO lines, we then derive limits on the line brightness temperature ratios (Table 4). We evaluate the S/N by spectral (1D) match-filtering of individual spectra extracted in the central pixel and within five frequency channels around the expected position of the lines. We do not find any significant detections above a 3σ threshold. Some of our upper limits imply super-unity line brightness temperature ratios for the mid-*J* CO candidates. While a lower-*J* level can be less populated than a higher-*J* level, or low optical depths can cause such high line ratios, the physical conditions that give rise to such ratios are rare. In cases where super-unity line ratios are found for redshifts (i.e., mid-*J*/low-*J* CO line combinations) disfavored by Decarli et al. (2014), our data provide supporting evidence for the preferred redshifts identified by Decarli et al. (2014), under the assumption that those line candidates are real. As an example, in the case of ID.03, multiple line detections have determined a secure redshift. Since this redshift does not lie within our surveyed volume, our nondetection is consistent with the redshift identification by Decarli et al. (2014). On the other hand, candidate ID.19 was confirmed to lie at $z = 2.0474$ based on optical grism spectroscopy (Decarli et al. 2014). Therefore, the candidate lies in our survey volume, and our line ratio limit ($r_{31} > 0.7$) is significant. This suggests moderately elevated CO excitation compared to the average ratio found for a sample of main-sequence galaxies at $z = 1.5$ ($r_{31} = 0.42$; Daddi et al. 2015) and even compared to the average ratio ($r_{31} = 0.52 \pm 0.09$) found for a sample of SMGs (Bothwell et al. 2013).

For the mid-*J* line candidates ID.15 and ID.18, our constraints on the line ratios are higher than unity. This suggests that an alternative lower-redshift mid-*J* line assignment of CO(2–1) in the PdBI data may be more likely (since it would imply a redshift outside our survey volume) if the line candidate were confirmed to correspond to real emission.

We also stack the extracted spectra to obtain more sensitive limits. In particular, we select random subsets of candidates for stacking, to take into account a possible misidentification of the correct *J* value for some CO lines. To search for lines in the stacked spectra, we match-filter using the same set of spectral templates as the main line search (Table 5). We find no line signal in the stacks above 3σ significance. Assuming an average line FWHM of 300 km s^{-1}, we therefore obtain a sensitive 3σ upper limit of 0.014 ± 0.002 Jy km s^{-1} to the line

Table 5
Template Sizes for MF3D Line Search Technique

	Spatial FWHM (arcsec)	Frequency FWHM (4 MHz channels)
COSMOS	0, 1, 2, 3, 4	4, 8, 12, 16, 20
GOODS-N	0, 1, 2, 3, 4	4, 8, 12, 16, 20

Note. Gaussian template sizes utilized in our MF3D. A spatial size of 0 stands for a single-pixel spatial extent and implements the single-pixel search that is optimal for unresolved sources. These sizes represent a uniform sampling of the parameter space that we conservatively expect to represent CO sources. The 4 MHz corresponds to $\sim$35 km s^{-1} midband.

fluxes, where the quoted uncertainties on the limit depend on the number of stacked spectra. From the same sample, we also separately stack the nine candidates for which the lines were identified as likely CO(3–2) emission by Decarli et al. (2014) and whose redshift would place their CO(1–0) line in our data cube. We obtain a 3σ limit of $\sim$0.019 Jy km s^{-1}, which implies a constraining limit of $r_{31} > 2.0$ when using the mean CO(3–2) flux in the limit (0.34 Jy km s^{-1}, using the same weights as in our stack). We estimate how many of these stacked spectra need to be removed in order for the line luminosity ratio to become smaller than unity. We find that at least six of them may not correspond to real emission, subject to the stated assumptions. In summary, we find some tentative evidence suggesting that mid-*J* blind CO searches may preferentially select galaxies with relatively high CO excitation. An alternative interpretation may be that some of the candidate mid-*J* CO emitters considered here may be spurious and do not correspond to real CO emission. In order to more strongly differentiate between these possibilities, more sensitive 3 mm observations need to be carried out.

5.2. Identification and Stacking of Galaxies with Optical Redshifts

In order to obtain additional constraints on the CO luminosity of galaxies that remain individually undetected in the volume covered by our survey, we utilize the available optical/NIR spectroscopic redshift information for galaxies in our well-studied target fields for stacking. We extract single-pixel spectra of the sources described in the following, and we stack them with a weighted average. As weights, we used the inverse of the variance of the local noise following Decarli et al. (2016). We present additional less-constraining stacks of galaxies in Appendix G, where we consider galaxies with grism redshifts and at higher redshifts for which CO(2–1) may lie within our data.

5.2.1. Spectroscopic Redshifts in the COSMOS Field

Only seven galaxies have known ground-based optical spectroscopic redshifts that place them within our COSMOS data cube, all of which were obtained as part of the zCOSMOS-deep survey (S. Lilly et al. 2018, in preparation). These galaxies have relatively low stellar masses ($\lesssim 10^{10}\,M_\odot$). Therefore, we do not expect to detect their CO emission individually. We also do not detect their averaged CO line emission down to a deep 3σ limit of $<$0.008 Jy km s^{-1} (assuming a line FWHM of 300 km s^{-1}) after stacking spectra extracted at their positions (stacked spectrum shown in Figure 8). This limit implies $L'_{\rm CO} < (1.7\text{–}2.7) \times 10^9$ K km s^{-1} pc^2 for different redshifts within our surveyed range. In order to determine the implications of this limit, we perform SED fitting of the same galaxies with MAGPHYS (da Cunha et al. 2008, 2015) to estimate their stellar masses, finding that they are compatible with the tabulated values in COSMOS2015 whenever the photo-z is similar to the spectroscopic redshift (only 3/7 cases). These stellar masses are in the range 10^9–$10^{10}\,M_\odot$. Assuming that these galaxies lie on the main sequence, we use the fitting functions from Speagle et al. (2014) to determine star formation rates (SFR) $\sim$ 2–25 $M_\odot$ yr^{-1}. These values are consistent with the lack of detections by the 3 GHz survey by Smolčić et al. (2017), which implies SFR $<$ 40–70 $M_\odot$ yr^{-1} based on the radio–FIR correlation[29] estimated by Delhaize et al. (2017). The SFRs estimated by MAGPHYS for these galaxies (with great uncertainty due to the lack of FIR detections) span the 6–150 $M_\odot$ yr^{-1} range, with a mean of 55 $M_\odot$ yr^{-1}. Assuming the star formation law found for main-sequence galaxies at high redshift[30] (Daddi et al. 2010b; Genzel et al. 2010), we can use the star formation rate estimates to infer an expected $L'_{\rm CO} \sim 8 \times 10^9$ K km s^{-1} pc^2, which is higher than our measured limit and therefore not fully consistent. The adopted chain of scaling relations, including SED fitting and the star formation law, have large scatter and therefore introduce large uncertainties in the $L'_{\rm CO}$ estimate. The apparent tension with our upper limit would disappear if the average SFR were $\sim$10–15 $M_\odot$ yr^{-1}. The nondetection of the stacked CO(1–0) emission therefore provides a valuable constraint on the CO luminosity of faint, modestly massive $z = 2$–3 galaxies, as our best estimates for their SFR appear to be in tension with the expectated $L'_{\rm CO}$ based on the star formation law by Daddi et al. (2010b) and Genzel et al. (2010).

5.2.2. Spectroscopic Redshifts in the GOODS-N Field

The 3D-*HST* catalog (Skelton et al. 2014; Momcheva et al. 2016) provides 67 galaxies in the region in GOODS-N covered in our survey area with ground-based optical spectroscopic redshift whose CO(1–0) line is covered by our data. We also include 13 more galaxies with more recent spectroscopic redshifts from the catalog by Liu et al. (2018) in our analysis. One of the galaxies in this combined sample corresponds to GN19 and is individually detected. Therefore, we exclude it from further investigation here.

One more galaxy with a spectroscopic redshift of $z_{\rm spec} = 2.320$ is perfectly matched in position and redshift with an S/N = 4.6 CO line candidate at coordinates J2000 12:36:49.10+62:18:14.0 and $z_{10} = 2.3192 \pm 0.0003$, which we name COLDz.GN.S1 (Figure 7). This CO line candidate is significantly extended (FWHM of $\sim 9\farcs0 \pm 0\farcs5 \sim 74 \pm 4$ kpc) and appears to be associated with two potentially interacting galaxies with a projected separation of $\sim$20 kpc. The galaxy to the south is associated with the spectroscopic redshift measurement, while the galaxy to the north has a compatible photometric redshift estimate. The total aperture CO flux of GN.S1 is (0.22 ± 0.11) Jy km s^{-1}, corresponding to $L'_{\rm CO} = (5 \pm 3) \times 10^{10}$ K km s^{-1} pc^2. We find a molecular gas mass of $(2.0 \pm 1.1) \times 10^{11}\,M_\odot$ when assuming $\alpha_{\rm CO} = 3.6\,M_\odot$ (K km s^{-1} pc^2)$^{-1}$. The stellar mass of the southern

[29] Recent work by Molnár et al. (2018) suggests that radio–FIR correlation for disk-dominated galaxies may not show redshift evolution and would imply a less-constraining limit of SFR $<$ 100–300 $M_\odot$ yr^{-1}.

[30] These star formation law estimates were mostly based on CO(3–2) observations; therefore, variations in the r_{31} line ratio may contribute additional uncertainty.

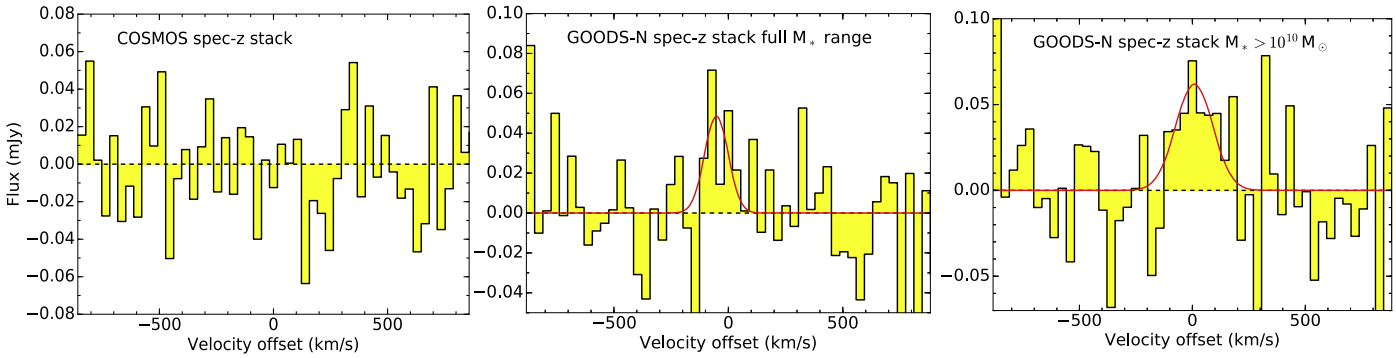

Figure 8. Spectral stacks of sets of galaxies with spectroscopic redshifts. Stack of galaxies are shown with spectroscopic redshifts in COSMOS (left, seven galaxies) and GOODS-N (right). The GOODS-N spec-z stack including the full stellar mass range (78 galaxies) displays a tentative 3.4σ detection, while the stack of massive galaxies ($M_* > 10^{10}\,M_\odot$; 34 galaxies) shows a 3.5σ detection. The best-fitting Gaussian line profiles are shown in red. The spectral resolution is the same as in Figure 5.

(brightest) component is $\sim 1.2 \times 10^{11}\,M_\odot$, and that of the northern component is $\sim 3 \times 10^9\,M_\odot$, suggesting a high gas mass fraction of $\sim$1.7. This gas fraction may be elevated due to the galaxy interaction, although the star formation rate reported by Liu et al. (2018) of $\sim 160\,M_\odot\ \mathrm{yr}^{-1}$ is approximately 2–3× lower than what may be expected from the total CO luminosity based on the star formation law (Daddi et al. 2010b; Genzel et al. 2010). Therefore, we may be witnessing a gas-rich early phase of the merger, which may precede a starburst. This source is also tentatively detected in the S2CLS map with a significance of $\sim 3.3\sigma$ and a 850 μm flux of 3.2 ± 1.0 mJy (Geach et al. 2017), which is compatible with the moderate star formation rate estimate.

Allowing for an offset of $<2''$ and $<500\,\mathrm{km\,s}^{-1}$ results in four more potential candidate associations, but they are not likely to be real due to apparent offsets in the emission and because they are not associated with the most massive or most star-forming galaxies in the sample. In the set of galaxies with spectroscopic redshifts covered by our data, nine galaxies have stellar mass estimates of $> 5 \times 10^{10}\,M_\odot$, corresponding to a gas fraction $M_{\rm gas}/M_* \sim 1$ at our approximate 3σ sensitivity limit of $L'_{\rm CO} \sim 1.5 \times 10^{10}\,\mathrm{K\,km\,s}^{-1}\,\mathrm{pc}^2$. These galaxies may therefore be expected to be individually detectable. Excluding GN19 and GN.S1, the remaining seven galaxies remain undetected, implying $M_{\rm gas}/M_* < 1$. Previous samples of main-sequence galaxies at $z = 2$–3 have shown typical molecular gas mass fractions of order $M_{\rm gas}/M_* \sim 1-1.5$ in this stellar mass range (Genzel et al. 2015; Scoville et al. 2017), i.e., higher than those limits. We note that adopting the same CO conversion factor as utilized by Genzel et al. (2015), including the correction for a stellar mass–dependent metallicity, would result in approximately 50% higher molecular gas mass estimates. Overall, the observed limits may be consistent with previous observations of the molecular gas fractions in main-sequence galaxies, although they appear to be at the low end of the expected scatter of the relation.

Stacking all 78 spectra, i.e., excluding GN19 and COLDz. GN.S1, yields a tentative ($\sim 3.4\sigma$) CO line detection in the deep stack ($6 \pm 3 \times 10^{-3}\,\mathrm{Jy\,km\,s}^{-1}$; Figure 8). The noise in this stacked spectrum is $\sim 23\,\mu\mathrm{Jy\ beam}^{-1}$ in $35\,\mathrm{km\,s}^{-1}$ wide channels. The stacked galaxies in this sample have a wide range of stellar masses and are therefore expected to show a range of CO luminosities. While constraining the average CO luminosity for this sample, we note that such an average does not represent common properties of the stacked galaxies. The measured flux in the stacked spectrum corresponds to an average CO luminosity for this galaxy sample of $L'_{\rm CO} = (1.5 \pm 0.8) \times 10^9\,\mathrm{K\,km\,s}^{-1}\,\mathrm{pc}^2$ at an average $z \sim 2.4$. These 78 galaxies have a mean stellar mass of $2.1 \times 10^{10}\,M_\odot$ and an average star formation rate of $\sim 45\,M_\odot\ \mathrm{yr}^{-1}$ (with quartiles of 3.6, 22, and $66\,M_\odot\ \mathrm{yr}^{-1}$, respectively), according to Liu et al. (2018) when available and Skelton et al. (2014) otherwise. We also stack the subset of 34 spectra corresponding to the most massive galaxies with $M_* > 10^{10}\,M_\odot$, expecting them to contribute the strongest CO signal. We detect emission in this sub-stack with a significance of 3.5σ, corresponding to a line flux of $1.2 \pm 0.7 \times 10^{-2}\,\mathrm{Jy\,km\,s}^{-1}$ and a line FWHM of $200 \pm 80\,\mathrm{km\,s}^{-1}$ (Figure 8). The line flux in the stacked spectrum corresponds to $L'_{\rm CO} = (3.2 \pm 1.8) \times 10^9\,\mathrm{K\,km\,s}^{-1}\,\mathrm{pc}^2$ at an average $z \sim 2.4$, which corresponds to a molecular gas mass of $M_{\rm gas} = (1.2 \pm 0.6) \times 10^{10}\,M_\odot$, according to our earlier choice of $\alpha_{\rm CO}$. This gas mass should be compared to the mean stellar mass $\sim 3.6 \times 10^{10}\,M_\odot$ of this subsample (with a median of $\sim 2.8 \times 10^{10}\,M_\odot$) and the average star formation rate of $\sim 66\,M_\odot\ \mathrm{yr}^{-1}$ (median of $\sim 52\,M_\odot\ \mathrm{yr}^{-1}$; Liu et al. 2018). To further investigate these star formation rate estimates, we also stack SCUBA-2 S2CLS 850 μm images and derive a 3σ upper limit of <0.7 mJy at the positions of the sample galaxies (Geach et al. 2017). Utilizing the average FIR SED from the ALESS sample at approximately the same redshift as these galaxies, the submillimeter flux upper limit implies a star formation rate constraint of $\lesssim$60–$100\,M_\odot\ \mathrm{yr}^{-1}$ (da Cunha et al. 2015). A modified blackbody model with a dust temperature of $T_d = 35$ K also implies comparable limits. We find that the star formation rate of this set of galaxies is compatible with random scatter around the star-forming main sequence reported by Speagle et al. (2014). The average expected CO luminosity based on the star formation law (Daddi et al. 2010b; Genzel et al. 2010) may be higher than our measurement by about a factor of 3 ($L'_{\rm CO} \sim 9.6 \times 10^9\,\mathrm{K\,km\,s}^{-1}\,\mathrm{pc}^2$).

5.2.3. Implications

Figure 9 summarizes our constraints on the molecular gas mass fraction in the analyzed samples of galaxies. In order to convert CO luminosity to molecular gas mass, we consider two different assumptions for $\alpha_{\rm CO}$. First, we assume a constant value of $\alpha_{\rm CO} = 3.6\,M_\odot\ (\mathrm{K\,km\,s}^{-1}\,\mathrm{pc}^2)^{-1}$ adopted by some previous studies (e.g., Daddi et al. 2010b; Decarli et al. 2014; Walter et al. 2016). We then also consider a metallicity-dependent conversion factor, evaluated by assuming a redshift

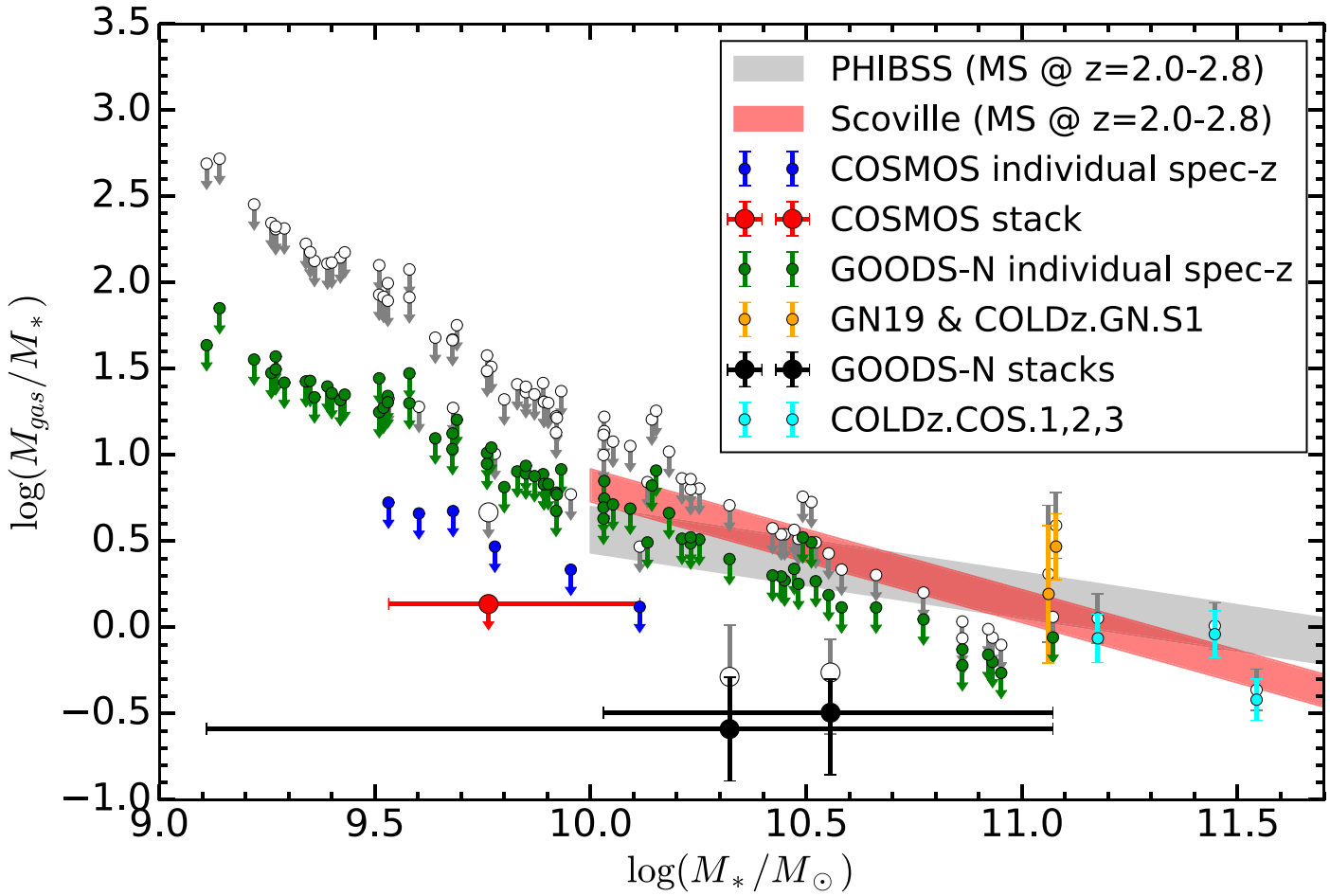

Figure 9. Molecular gas mass fraction constraints for galaxies with known spectroscopic redshifts for which the CO(1–0) emission can be constrained by the COLDz data. The points in color assume $\alpha_{CO} = 3.6\,M_{\odot}$ (K km s^{-1} pc^2)$^{-1}$, while the gray points correspond to adopting the metallicity-dependent α_{CO} from Genzel et al. (2015). All upper limits correspond to 3σ limits and assume a line FWHM of 300 km s^{-1}. The sets of individual galaxies, the origin of their stellar masses, and the constraints from stacking are described in Section 6. The gray and red shaded regions correspond to the reported average expected gas mass fraction in the range $z \sim 2.0$–2.8 for main-sequence galaxies according to Genzel et al. (2015) and Scoville et al. (2017), respectively. These regions do not show the expected scatter but rather represent the evolution within the covered redshift range. The shown scaling relations were measured over the stellar mass range $M_* \gtrsim 10^{10} M_{\odot}$.

and stellar mass–metallicity relation (Genzel et al. 2015). Previous studies investigating optically and FIR-selected galaxy samples have estimated the relationship between gas mass fraction, stellar mass, redshift, and SFR offset from the main sequence (e.g., Genzel et al. 2015; Scoville et al. 2017). While the PHIBSS project estimated molecular gas masses by measuring the CO(3–2) line emission (Tacconi et al. 2013; Genzel et al. 2015), Scoville et al. (2016, 2017) used the flux on the Rayleigh–Jeans tail of the dust continuum emission to estimate the total gas masses. We here assume that the samples of galaxies plotted, although not complete to any degree due to their preselection for having a spectroscopic redshift, may be somewhat representative of the star-forming main sequence (Figure 9). Their star formation rates are consistent with scatter around the main sequence and appear to include as many galaxies above and below the main sequence estimated by Speagle et al. (2014). Although our CO detections (both blind and with previous spectroscopic redshifts) are indicative of gas fractions compatible with or above (GN19) expectations for main-sequence galaxies, the individual CO nondetections and the stacked signal appear to be systematically lower than the predicted averages, suggesting lower gas mass fractions than might be expected (Figure 9). We can quantify the apparent deficit in stacked signal relative to expectations for the $M_* \gtrsim 10^{10}\,M_{\odot}$ sample by calculating expected gas masses for the individual stacked galaxies predicted as a function of their redshift, stellar masses, and star formation rates. The expected sample average molecular gas mass is $5.8 \times 10^{10}\,M_{\odot}$ adopting the best-fit relation by Genzel et al. (2015) and $7.5 \times 10^{10}\,M_{\odot}$ according to the relation by Scoville et al. (2017). The constant CO luminosity conversion factor above would imply a ratio between expected and observed stacked CO luminosity of 4.8 ± 2.4 and 6.3 ± 3.1, according to the relations by Genzel et al. (2015) and Scoville et al. (2017), respectively. Applying instead the metallicity-dependent CO conversion factor suggested by Genzel et al. (2015) to individual galaxies would somewhat reduce the tension, implying ratios of 3.0 ± 1.7 and 3.8 ± 2.1, according to the relations by Genzel et al. (2015) and Scoville et al. (2017), respectively. While the constraints for low stellar mass galaxies may be compatible with an evolving CO conversion factor due to low metallicity, this is unlikely to resolve the apparent conflict at the high-mass end and may point to lower-than-expected gas masses.

6. Total CO Line Brightness at 34 GHz

One additional key measurement that becomes possible with the COLDz survey is to determine the total CO line brightness

at 30–39 GHz in the survey volume. We follow the simple procedure outlined by Carilli et al. (2016) and, as a first conservative estimate, include only the independently confirmed line candidates for each field in order to derive secure lower bounds. We derive lower brightness temperature limits (since we only include the securely detected sources, without any completeness correction) at the average frequency of 34 GHz of $T_B \gtrsim 0.4 \pm 0.2\,\mu$K for the COSMOS field and $T_B \gtrsim 0.05 \pm 0.04\,\mu$K for the GOODS-N field. The uncertainties are dominated by Poisson relative uncertainties due to the limited number of sources considered. Sources near the knee of the CO luminosity function (Paper II) dominate the total surface brightness, as expected. Since the two measurements are sensitive to different parts of the CO luminosity function, we add the two values to obtain our best estimate for a lower limit on the average surface brightness of $T_B \sim 0.45 \pm 0.2\,\mu$K. Next, we attempt to include a longer list of candidates, down-weighted by their purities (evaluated in Appendix F), to estimate a plausible uncertainty range. In the COSMOS field, also including all moderate-S/N candidates presented in Appendix E, we obtain $T_B \sim 0.48$ and $0.57\,\mu$K without and with the completeness corrections evaluated in Appendix F, respectively. In the GOODS-N field, also including all candidates in Appendix E, we obtain $T_B \sim 0.18$ and $0.3\,\mu$K without and with the completeness corrections, respectively. Because the complete candidate list in GOODS-N overlaps in flux ranges with the candidates in COSMOS, it is not clear that the best estimate for this case may simply be derived by adding the two contributions; the plausible range of values from our data should therefore be considered to be the full range $T_B \sim 0.2$–$0.6\,\mu$K, with a likely lower limit of $T_B \sim 0.45 \pm 0.2\,\mu$K. These measurements are consistent with that of $T_B \sim 0.94 \pm 0.09\,\mu$K at 99 GHz by Carilli et al. (2016) within the expectation that the total (all CO lines) average surface brightness may slightly increase between 34 and 99 GHz due to adding more CO transitions together (e.g., Righi et al. 2008). Our measurement of the average surface brightness is in agreement with theoretical predictions (e.g., Righi et al. 2008; Pullen et al. 2013) that suggest a range of $T_B = 0.3$–$1\,\mu$K. Our constraints on the total CO brightness at 34 GHz suggest that the CO signal will be an important contribution to CMB spectral distortion at these frequencies, which is relevant for upcoming experimental efforts. In particular, as shown in Figure 2 by Carilli et al. (2016), our constraints at 34 GHz are already higher than the PIXIE sensitivity limit (Kogut et al. 2011, 2014), and, while lower than the low-redshift Compton distortion component, it is higher than the relativistic correction to the low-redshift signal, the primordial Silk damping distortion, and the imprint of primordial hydrogen and helium recombination radiation contributions. A measurement of these important cosmological probes will therefore necessarily require a subtraction scheme that will remove the CO line signal (also see Carilli et al. 2016).

7. Discussion and Conclusions

In this work, we have carried out the first blind search deep field targeting CO(1–0) line emission at the peak epoch of cosmic star formation at $z = 2$–3. This allowed us to provide the least biased measurement of the molecular gas content in a representative sample of galaxies at this epoch. One of our main findings is the absence of a population of massive, gas-rich galaxies with suppressed star formation in our high S/N sample, which would have been missed by previous selection techniques. The lower-S/N, and hence lower-purity, CO line candidate sample includes several candidates without clear multiwavelength counterparts, which are therefore possible candidates for such a population of "dark," gas-rich galaxies. Nonetheless, the low purity of such candidate lines requires further independent confirmation, as the absence of a counterpart is more likely to indicate that the line feature may be spurious.

Interestingly, the CO line sources detected with confidence in this study include a mix of different galaxy populations. In particular, our subsample of independently confirmed CO emitters contains previously known starbursts like AzTEC-3 (by design), GN10, GN19, and HDF850.1 but also COLDz.COS.1, 2, and 3 and COLDz.GN.S1, which belong to the massive end of the main sequence at $z \sim 2$ (R. Pavesi et al. 2018, in preparation). This highlights the CO(1–0)-based selection, which does not preferentially select outliers in star formation such as starbursts as preferentially selected by submillimeter continuum–selected samples. Also, the total gas mass is accurately traced by these measurements, without the extinction biases that affect optical/NIR-selected samples.

Most studies of molecular gas in galaxies at high redshift to date have targeted mid-J CO lines. Although these lines have higher fluxes than the ground-state $J = 1$–0 transition and therefore are typically easier to detect, their higher critical densities and level energies imply that they do not always faithfully trace the bulk of the gas mass, but that they can be biased toward the dense and warm fraction of the gas reservoir. Therefore, in order to derive gas masses from those mid-J CO lines, an excitation correction needs to be assumed (i.e., a ratio of those lines to the CO(1–0) line brightness), which introduces a source of uncertainty. Previous blind CO searches have targeted mid-J CO lines (Decarli et al. 2014; Walter et al. 2016) and therefore relied on similar excitation correction assumptions in order to derive constraints to the total molecular gas mass. In this study, we have shown that blind CO(1–0) searches, selecting galaxies uniquely through their total gas masses, find a varied sample of galaxies belonging to a mix of different populations, which may be characterized by significant differences in CO excitation (e.g., starbursting and main-sequence galaxies; Riechers et al. 2006, 2011c, 2011a; Daddi et al. 2010b; Ivison et al. 2011; Bothwell et al. 2013; Carilli & Walter 2013). We also find significant excitation differences among the individual sources (to be described in detail by R. Pavesi et al. 2018, in preparation). Furthermore, our limits on the CO(1–0) line luminosities in the candidates previously selected by Decarli et al. (2014) indicate either that the corresponding galaxies have substantially elevated CO excitation or that a large fraction of them may not correspond to real line emission.

The so-called "wedding cake" design of the COLDz survey, targeting a shallow wide field and a deep narrower field, allows us to provide valuable independent constraints on different parts of the CO luminosity function (Paper II) that would not have been possible with a single field due to the limited accessible volume and depth. While the sensitivity of our deeper field (in COSMOS) is within a factor of two of the sensitivity that was previously achieved by ASPECS through ALMA in a comparable redshift bin (after correcting for CO excitation), the volume that we could sample in that field is six times larger. Furthermore, the volume covered in both fields

combined is >50 times as large as that covered by ASPECS-Pilot and >60 times as large as that carried out in the HDF-N with the PdBI, given the >60 times larger survey area (∼60 versus ∼1 arcmin^2).

In this study, we have also significantly further developed the methods utilized to carry out blind searches for emission lines in interferometric data sets. In particular, we have generalized the matched-filtering technique that is commonly used in the spectral dimension to identify spectral lines to the regime of interferometric data cubes where sources may be spatially extended. By taking advantage of this new source selection method, we have blindly detected significantly extended CO(1–0) line sources like COLDz.COS.3, which hosts a very large cold gas reservoir (∼30–40 kpc). Furthermore, one of our highest S/N line emitters in the GOODS-N field (GN19) and a galaxy with optical spectroscopic redshift (COLDz.GN.S1) also appear extended (∼40–70 kpc) in CO observations due to a major gas-rich merger in this galaxy (see also Ivison et al. 2011; Riechers et al. 2011b). The high incidence, two out of the eight most significant CO(1–0) sources, suggests that extended CO(1–0) sources may in fact be prevalent, in agreement with previous findings (Ivison et al. 2011; Riechers et al. 2011b). Indeed, through the blind search, we have selected other CO candidates that may be significantly extended, some of which might have been missed by previous blind line search techniques searching only for unresolved sources (Decarli et al. 2014, 2016; Walter et al. 2016).

The candidate CO lines span a large range in line FWHM, from ∼60 to ∼800 km s^{-1}, although the narrowest ones have not yet been independently confirmed. This demonstrates the need for the inclusion of a broad range in line-width templates in order not to miss a significant fraction of the signal. The spread and distribution in the line FWHM we find is comparable to those previously measured (e.g., Tacconi et al. 2013), although the occurrence of a particularly broad line (in COLDz.COS.2) in our limited highest-quality sample suggests that there may be a larger incidence of broad lines in blindly selected CO sources compared to optical/NIR selections. Nonetheless, due to the limited number of sources, this finding requires further independent study.

Although the molecular gas fraction estimates for the CO detections are comparable to expectations, the lack of CO detections for a number of massive galaxies with good-quality spectroscopic redshifts and the detection of their stacked CO signal suggests that molecular gas mass fractions for typical main-sequence galaxies may be somewhat lower than expected (Figure 9; Genzel et al. 2015; Scoville et al. 2017). A possible caveat to this interpretation may come from a higher systematic uncertainty than expected of the optical spectroscopic redshifts used in the stacking, which may lead to missing a fraction of the CO flux in the stack. While this analysis could only be carried out on samples of galaxies with previous spectroscopic redshift measurements, its conclusion is in agreement with the finding from the blind search. In particular, if the gas mass fraction and hence the CO luminosity of the known galaxies (in addition to all other galaxies in the observed cosmic volume without spectroscopic redshifts) were closer to expectations, the number of blind CO detections would have been higher. Empirical predictions based on SED fitting and scaling relations suggest an expected number of CO emitters in the range 10–20 for our field in COSMOS and 5–15 for our field in GOODS-N (da Cunha et al. 2013b). In addition, the high-mass end of the galaxy distribution in our cosmic volume provides the strongest result for lower-than-expected gas masses; also, the full sample over the shown stellar mass range provides important constraints compatible with a metallicity evolution in the CO conversion factor (Figure 9). The significant detection of CO emission from galaxies in the stack suggests that our data set is rich in additional signal that is too faint to be reliably blindly identified at the current depth but can be mined through spectroscopic observations at other wavelengths. We have therefore demonstrated the power of stacking the CO signal from galaxies with spectroscopic redshifts in order to fully take advantage of the information in CO deep-field data.

We have also developed statistical methods (presented in Appendix F) to evaluate the purity, completeness, and recovered candidate properties with higher accuracy than previous techniques. This enables us to infer the best constraints to date on the CO(1–0) luminosity function at $z \sim 2$–3 (Paper II).

With this CO deep-field study, we also further demonstrate that blind CO searches are sensitive to "optically dark," dust-obscured galaxies at very high redshift, such as GN10 and HDF850.1. In particular, the massive molecular gas reservoirs of these galaxies are among the largest in our field (Riechers et al. 2018). Our sample of new high-S/N CO(1–0) spectra for COLDz.COS1, 2, and 3 and GN.S1 provides a significant contribution to the state of current CO(1–0) measurements of main-sequence galaxies at $z > 2$[31] (see R. Pavesi et al. 2018, in preparation, for details).

Finally, the Next Generation VLA (ngVLA) is necessary to significantly improve the constraints presented here (e.g., Carilli et al. 2015; Casey et al. 2015; McKinnon et al. 2016; Selina & Murphy 2017). In particular, an equivalent survey in the 30–38 GHz range with a fivefold to tenfold sensitivity improvement for point-source detection as provided by the ngVLA will allow us to reach the depth of these observations in a small fraction of the time (∼1/50), therefore routinely reaching depths of $\log(L'_{\rm CO}/L_\odot) \sim 9.5$ in 1–2 hr of observation. The high survey speed of the ngVLA will uniquely enable the deep, wide-area surveys that are necessary to build large statistical samples, currently inaccessible to the VLA. These future surveys will constrain the luminosity function to well below the knee with percent precision for a comparable observing effort as the present survey. A significant benefit of the ngVLA will also come from the planned smaller antennas, which increase the field of view for a fixed total collecting area, therefore enhancing the survey speed. In addition, the vast bandwidth of the ngVLA will allow us to simultaneously cover CO(1–0) emission over a large fraction of the age of the universe and therefore allow us to probe CO(1–0) over the almost complete redshift range up to $z \sim 10$.

The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. D.R. and R.P. acknowledge support from the National Science Foundation under grant number AST-1614213 to Cornell University. R.P. acknowledges support through award SOSPA3-008 from the NRAO. We thank Tom Loredo for helpful discussion. I.R.S.

[31] The number of high-S/N CO(1–0) detections in non-quasar hosts or submillimeter-selected galaxies is very limited to date (Aravena et al. 2010, 2012, 2014; Riechers et al. 2010b; Bolatto et al. 2015; Spilker et al. 2016).

acknowledges support from the ERC Advanced Grant DUSTYGAL (321334), STFC (ST/P000541/1), and a Royal Society/Wolfson Merit Award. V.S. acknowledges support from the European Union's Seventh Frame-work program under grant agreement 337595 (ERC Starting Grant, CoSMass) RJI acknowledges ERC funding through Advanced Grant 321302 COSMICISM. This research has made use of the NASA/ IPAC Infrared Science Archive, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. This work is based on observations taken by the CANDELS Multi-Cycle and 3D-*HST* Treasury Program (GO 12177 and 12328) with the NASA/ESA *HST*, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. We acknowledge funding toward the 3 bit samplers used in this work from ERC Advanced Grant 321302, COSMICISM. J.A.H. acknowledges the support of the VIDI research program with project number 639.042.611, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO). M.T.S. acknowledges support from the Science and Technology Facilities Council (grant number ST/P000252/1).

Appendix A
Additional Details on the Line Search Methods

In this section, we provide additional details of our line search methods in interferometric data cubes, which were used to carry out the blind line search presented in Section 4 of the main text. We first provide a more complete description of our method of choice, MF3D[32] (which extends MF1D), and then we compare its performance to three alternative methods that we have also investigated.

A.1. Matched Filtering Interferometric Data Cubes

Since we do not expect the CO line emission in $z = 2$–3 galaxies to be resolved over more than a few beams at most, we expect our sources to be spatially well described by a family of 2D Gaussian templates. Therefore, under the prior of source shape, matched filtering is theoretically an optimal detection method.

The matched-filtering method can be thought of as concentrating all of the extended (spatially and in frequency) signal to a peak pixel that captures both the overall strength of the original signal and how closely this matches the template shape. At the same time, the smoothing of the noise in regions without signal allows us to reliably measure the noise level on the scale probed by the template size. In this way, the problem of finding emission lines with structure is effectively reduced to the problem of just examining peak heights to assess their significance.

We compute templates that are Gaussians in frequency and circular 2D Gaussians spatially (sizes given in Table 5). We then convolve the S/N cube with these templates by multiplication in Fourier space to produce multiple matched-filtered cubes, one for each template.

The main difference between the traditional application of matched filtering in astronomical images and our application to interferometric data comes from the spatially correlated nature of the noise in interferometric images. In the case of uncorrelated (i.e., white) noise, the matched filter simply corresponds to the expected source shape and size, but correlated noise introduces deviations from this matching, as described in the following.

The frequency width of a template approximately matches the line width that it selects, because the noise in different channels is uncorrelated. On the other hand, spatially, the noise has a nonzero correlation length, as determined by the synthesized beam. Therefore, the "matching" to a template is not the intuitive relation for which spatial template size matches the size that it selects. As an example, for unresolved sources, i.e., sources whose image is beam-sized, the maximal S/N is realized at the peak pixel rather than over an extended area. To calculate the relationship between template size and selected size, we therefore considered the idealized problem of circular Gaussian beams and Gaussian sources, which can be treated analytically. We calculate the correspondence between template size that maximizes the S/N and source size (see Appendix B). To carry out the calculation, we have to make the approximation of source positions being known a priori, evaluating the S/N at this position. This is not what is done in practice, since the positions are unknown. The pixel with the locally highest S/N is utilized instead. We briefly discuss the effects of this approximation on the recovered S/N in Appendix B.

The results of this calculation show that template "matching" (i.e., providing the maximum S/N) takes place approximately when

$$\sigma_A^2 = \sigma_h^2 + 2\sigma_b^2, \tag{4}$$

where σ_A is the size (radial standard deviation) of the source in the image (which is given by the sum in quadrature of the real source size and the beam), σ_h is the size of the template Gaussian, and σ_b is the beam size. For source sizes smaller than $\sqrt{2}\,\sigma_b$, the template size that maximizes the S/N is an infinitely narrow source template. Therefore, we include a single-pixel template (we call this "0-size," being the limit where the radius of the spatial Gaussian tends to 0), which implements a single-pixel search (i.e., MF1D as a subset of MF3D) and therefore selects unresolved and very slightly resolved sources. The analytical expression above only provides an indication of the matching dependence, but we do not make use of it in the following. The main simplification comes from measuring the S/N at the (in practice unknown) real position of the source rather than at the local maximum. In Appendix F, we explore the analysis of simulated artificial sources through our MF3D algorithm, and we use those simulations to numerically estimate a probabilistic connection between template sizes and injected source sizes.

The detailed steps of the blind search are summarized by the flowchart in Figure 10 and in the following. As described in Section 2, in order to correctly mosaic different fields together, we smooth every pointing to a common larger beam. This procedure reduces the S/N for point sources (see Figure 3). We therefore also run a single-pixel matched-filtering search on the Natural-mosaic, which was obtained without any smoothing. While the lack of a common beam in the Natural-mosaic would, strictly speaking, imply that the spatial structure may not be accurately calculated, this effect is negligible in the COSMOS data, where the different pointings have roughly equal resolution. Furthermore, the Natural-mosaic is sufficient

[32] We provide a Python implementation for this algorithm at https://github.com/pavesiriccardo/MF3D.

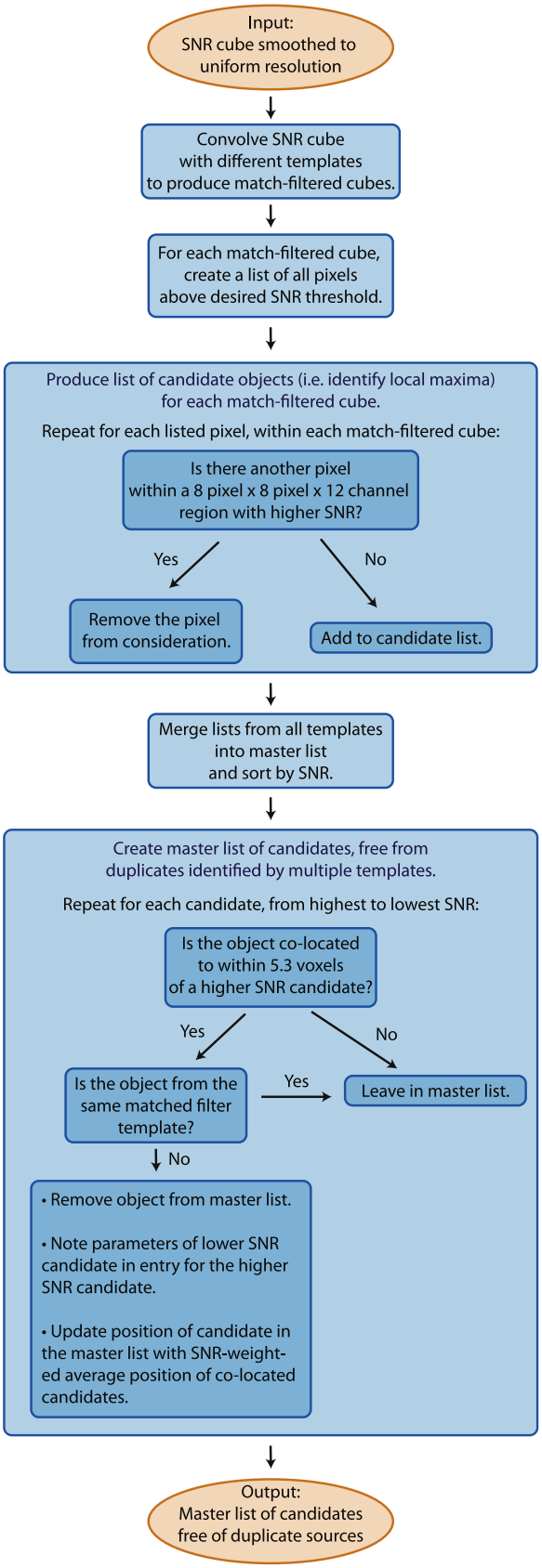

Figure 10. Flowchart describing the detailed procedure of our line search algorithm utilizing MF3D. The input is an S/N cube, and the output is a list of line features characterized by a feature S/N, which accounts for the total S/N evaluated through the template that gives the highest value.

for a search for unresolved sources, where the flux at the peak pixel represents the total flux, and is therefore correctly recorded in the Natural-mosaic. We treat the result of this Natural-mosaic matched-filtering step as an additional "spatial template," one for which less smoothing was done than even the single-pixel search in the Smoothed-mosaic. We therefore refer to it as the "−1 pixel" template. In the end, we combine the results from this search with those of the other templates, as detailed in the following.

The S/N of a detection corresponds to the ratio of the height of the peak in the matched-filtered cube to the standard deviation of empty regions. We initially normalize our templates such that the sum of the squares of the template values equals one. For independent pixel noise (which applies to the frequency channels but not to the spatial pixels), this normalization choice would imply that the noise after convolving would be the same as the noise before. This implies that the peak height corresponds to the total S/N of the candidate. For the 3D case, in particular for spatially extended templates, we need to account for the fact that the synthesized beam size results in small-scale spatial noise correlations. While the calculation for the noise in the smoothed cube is close to the measured values (see Appendix B), we decide to measure the noise in the convolved cubes directly from the standard deviation of pixel values. Since our data set is mostly free of signal, we estimate the noise in each matched-filtered cube in the COSMOS field simply by taking the standard deviation of the whole cube and normalize by dividing each pixel by this value. In the case of the GOODS-N mosaic, though, the beam size is not uniform across the mosaic, even after smoothing. In particular, the beam size in the pointings that had higher native resolution ends up being larger after smoothing than in the other pointings due to the particular nature of the *uv* coverage (see Appendix C for details). The main consequence of this slightly spatially varying beam affecting the Smoothed-mosaic for the GOODS-N field is that during matched filtering, the noise in the matched-filtered cubes is not uniform. This is expected, as the noise change during convolution is a function of the ratio of the pre- and postconvolution beam size. We therefore measure the noise in the matched-filtered cubes separately for two sets of pointings (GN29–GN42 and all the others) and use these sets to construct an approximate noise map for the matched-filtered mosaics in GOODS-N for each template (see Appendix C for details).

In order to blindly identify line features in our data and evaluate their significance using the matched-filtered cubes (one for each template), we need to locate the peaks and determine the template that provides the highest S/N to each candidate, thereby identifying the template that best matches the feature shape. The first stage is to identify, in each matched-filtered cube, the peaks, i.e., the local maxima above some significance threshold. In order to find the significance and position of a peak, we select all voxels (i.e., volume pixels) in the data cube above a fixed threshold and then retain those voxels that are local maxima by comparing to the values in a small surrounding box of 12 channels by 8 pixels by 8 pixels.

Next, we cross-match objects identified in the different match-filtered cubes (obtained from the different templates) in order to remove repeat identifications of the same object. We form a master list of all objects selected from all templates,

sorted by S/N. We then parse through each entry from highest to lowest S/N and form clusters characterized by their S/N-weighted average positions and the template with the highest S/N. We add a candidate object to a cluster if it resides within a 5.3 voxel radius of the cluster center (this threshold was found to be appropriate for our pixel size and channel width; more generally, the frequency and spatial separation thresholds may need to be different) and only if the template under consideration differs from the other templates in the cluster (to avoid clustering features identified in the same template, since they are most likely independent objects). By moving down the S/N-sorted list, we guarantee that clusters are built from their highest-significance members to their lowest. This method ensures that spatially extended/broad objects that are also identified at lower significance in smaller templates are included in the appropriate cluster as members. For neighboring point-like sources (with high significance in the smallest templates), this method maintains both objects as separate clusters and allows their corresponding low-significance/extended template candidate to be associated with both clusters. In this way, we avoid grouping separate objects into the same cluster, and we avoid splitting single objects into multiples. Each cluster then corresponds to a single galaxy candidate in our final catalog.

In order to choose the clustering thresholds and asses the independence of the result from their precise values, we test how well the algorithm performs in not clumping too much or too little by computing distances to closest neighbors. We test this both for the first stage of clump-finding in the matched-filtered cubes to check that the method to identify clump peaks works and for the second stage of matching features across different templates. We inspect the distribution of the neighbor distances and check that they behave as expected, without splitting clumps into different components (which would show up as many objects having a very close neighbor that would look like part of the same clump to visual inspection) or including different clumps (by changing the clustering thresholds and looking for any significant changes). We do not find significant issues in either phase of clustering. This technique was therefore used to refine our choice of clustering thresholds. A few objects are objectively difficult to distinguish as one or more parts, and so the algorithm performance is at a comparable level to what could be achieved through manual inspection. Overall, the method does very well in finding local peaks, appropriately splitting separate objects even when they are close together. The second stage of associating entries across different templates, while more challenging to evaluate, appears to be largely insensitive to the precise value of the thresholds within a few voxels' range.

A.2. Comparison to MF1D

A simpler version of our line search algorithm, which we call MF1D, corresponds to extracting a spectrum at each spatial pixel and running a spectral line search on each spectrum with 1D Gaussian templates. As emission lines at high-z are typically approximated as Gaussians, we note that assumptions of square profile templates are less optimal matches in the frequency dimension and therefore do not maximize the S/N for candidate emission lines, although we find the difference to be small.

We have investigated the MF1D approach, which is frequently used in the case of single-dish data, in order to provide a check on the results of our line search and to evaluate its performance. Walter et al. (2016) utilized a version of this method, which is effectively MF1D with square line templates. The main difference between the method utilized there and our implementation consists of our estimating the noise through the standard deviation of the full S/N cube rather than individual binned channel maps. This is not expected to cause a significant difference. This method, like MF3D, also requires some prescription for recognizing clusters of significant voxels as belonging to the same candidate when they are close together. We achieved this by building lists of clumps with running average positions and clustering up to a radius of 9 voxels. In the case of unresolved sources, where the peak pixel contains the maximum S/N, this method performs just as well as our more general method, since the set of templates used in this technique is a subset of those in MF3D ("0-size" templates). However, it will miss a large fraction of the extended sources by underestimating their true S/N. Although we do not expect a large fraction of resolved sources, a blind search should be as agnostic as possible with regard to the properties of the galaxies that may be selected. Indeed, since CO(1–0) traces the total cold dense gas mass, it is precisely the tracer that may reveal extended gas reservoirs. One of our top candidates, COLDz.COS.3, harbors a very extended gas reservoir, with S/N peaking in the 2″ template. A single-pixel search assigns this line an S/N = 8.2 rather than 9.2, which would imply a discrepancy of −10%. While this error would not significantly affect the significance of this candidate, such an error would be enough to move a moderate-significance candidate with S/N = 5.5 to 4.9 and therefore would effectively be missed by our search. Another advantage of the MF3D method over the 1D is that it allows us to capture a larger fraction of the signal for broader lines, because in that case, the peak signal may be substantially spread over several spatial pixels in different frequency channels due to noise. While this spreading of the signal over different pixels in different channels causes an ambiguity between spectral-S/N and moment-map-S/N for single-pixel methods, this ambiguity is resolved when the full 3D information is taken into account through MF3D. Therefore, we conclude that this method can be absorbed into our more general, improved MF3D framework and that it can be considered a subset of that technique.

A.3. Comparison to the `SExtractor` Method

We also considered modifications of existing source-finding software, such as `SExtractor`, which can effectively capture the spatial information of a line candidate while avoiding merging adjacent independent peaks (Bertin & Arnouts 1996). We used the spatial source detection part of SExtractor on individual channel maps with varying frequency binnings. We combined the detections across different binnings and at different frequencies and then established prescriptions to identify lines and their aperture-integrated S/N. These prescriptions made the results very dependent on the precise criteria used to evaluate the significance of a line. The principle is somewhat similar to matched filtering. It requires binning data cubes to multiple different velocity widths, and these binnings correspond to templates of different frequency width. Then the method relies on SExtractor for the spatial source extraction (recognizing clusters of high pixels as one unique object). It also requires finding the correct binning that maximizes the S/N. A challenge for this method is the choice

of an aperture size for the flux extraction in the channel maps to be used in separately evaluating signal and noise. Combining different aperture sizes, which imitates the range of spatial templates in MF3D, introduces additional difficulties with precisely evaluating aperture flux noise. This hybrid technique is suboptimal in the frequency dimension because binning is equivalent to filtering with a rectangular function, which is a worse match to the expected spectral line profile than a Gaussian shape, although the difference is small. Our tests show that this line search method can have similar outcomes in selecting lines to MF3D for our data. In particular, >85% of the top ∼100 candidates are matched between both methods. Comparing lists of candidates, we find that objects that were assigned a high S/N by the SExtractor method but were less significant in MF3D appeared to be less plausible candidates for visual inspection because of improbable line shapes.

The SExtractor method provides a valuable check for our use of extended templates in MF3D. In particular, the extended templates in MF3D allow finding those sources that would be missed by MF1D. The SExtractor method, which is sensitive to extended structure, confirms our extended candidate selection. Therefore, we conclude that our MF3D method coherently combines the results from single-pixel methods and other methods that are biased toward extended sources, like the use of SExtractor with fixed aperture sizes.

A.4. Comparison to the SPREAD Technique

We have also explored SPREAD, an algorithm developed by Decarli et al. (2014) for the PdBI blind field line search, to find emission lines in our VLA observations. This method corresponds to binning the data set in frequency and identifying channel maps with an excess of signal compared to the Gaussian noise pixel intensity distribution. This method does not take advantage of the spatial information (neither spatial extent nor position) but only of the total flux. The excess signal in a channel map does not need to come from a single source, because the SPREAD statistic is a global value that characterizes the whole channel map. This method did not perform reliably on our data set, since it relies on the small-number pixel statistics on the tails of the noise distribution, which are necessarily subject to large fluctuations. The SPREAD statistic was able to isolate the same top candidate sources as our other methods, but it loses discriminating power below an S/N of ∼8, since the SPREAD statistic does not track S/N and loses the ability to locate moderate-significance features. We conclude that MF3D captures any useful information obtained from SPREAD.

A.5. Comparison to Duchamp

We also compare our method to the sophisticated line-searching tool *Duchamp*, which was developed for SKA-precursor data cubes (Whiting 2012). *Duchamp* was extensively tested by Popping et al. (2012) and Westmeier et al. (2012) and found to provide a good blind search algorithm for both unresolved and extended emission. Because our survey is only expected to detect unresolved or slightly resolved CO emission, much of the power of *Duchamp* (e.g., "a trous" wavelet reconstruction) is not optimized for our targets of interest. The smoothing (convolution) preprocessing offered by *Duchamp* is equivalent to matched filtering with Gaussian templates in the spatial dimension and Hanning templates in the frequency dimension, although *Duchamp* only allows specifying one template size at a time and not combining results from different templates. We find that smoothing along the frequency axis is necessary in order to recover even the most significant line emitters in our cubes, as expected due to the wide line widths relative to channel widths. On the other hand, *Duchamp* does not allow us to smooth in the frequency and spatial dimensions simultaneously, thereby preventing optimal recovery of the full S/N for slightly extended sources. While the "a trous" wavelet reconstruction is designed to perform well on the extended structure of a general shape, it is not optimal for recovering only slightly extended spatial structure and hence does not yield the same S/N recovery as matched filtering in this specific case of interest. *Duchamp* offers two choices for peak identification algorithms leading to candidate identification, with pixel clustering being predominantly carried out spatially rather than spectrally. While both of these algorithms perform equally well in recovering all of our top-line candidates, the simple 3D peak identification algorithm implemented in MF3D simultaneously utilizes the full 3D information.

Based on all these considerations, we find that the best use of *Duchamp* in our data is achieved by manually adopting different frequency-width smoothing templates and combining the resulting S/N to select unresolved line candidates of different velocity widths. This procedure directly mimics our MF3D method, and therefore we do not adopt *Duchamp* for the COLDz survey data.

Appendix B
Matched Filtering Interferometric Images

We here discuss the analytical results of our investigation of MF3D in the specific case when it is adapted to interferometric images. We study idealized noise and source conditions in order to derive an approximate relationship between the template spatial size and the "matched" size of a feature that would display the highest S/N for that template. The purpose is to demonstrate the effect of correlated interferometric noise on the sizes that are selected through this technique.

If, for simplicity, we assume the synthesized beam to be a circular Gaussian with standard deviation σ_b, then the noise correlation function can be shown to be $\langle n(\boldsymbol{x})n(\boldsymbol{x}')\rangle = \sigma_0^2 e^{-|\boldsymbol{x}-\boldsymbol{x}'|^2/4\sigma_b^2}$, where σ_0 is the noise of the image. Let us define our idealized data as containing a Gaussian source of peak intensity s and convolved size σ_A in addition to additive, zero-mean Gaussian noise in the image. We also assume the spatial template to be a circular Gaussian of size σ_h, i.e., $h(\boldsymbol{x}) = \frac{1}{2\pi\sigma_h^2} e^{-x^2/2\sigma_h^2}$. The expectation value of the template-convolved image at the (assumed known) position of the source is then given by $\frac{s\sigma_A^2}{\sigma_A^2+\sigma_h^2}$. Furthermore, the standard deviation of the convolved image is given by $\frac{\sigma_0\sigma_b}{\sqrt{\sigma_b^2+\sigma_h^2}}$. Therefore, the S/N measured in the matched-filtered image is given by $\mathrm{S/N} = \frac{s}{\sigma_0}\frac{\sigma_A^2\sqrt{\sigma_b^2+\sigma_h^2}}{\sigma_b(\sigma_A^2+\sigma_h^2)}$. For a fixed source size σ_A, this S/N has a maximum at $\sigma_h^2 = \sigma_A^2 - 2\sigma_b^2$, or $\sigma_h = 0$ (i.e., the delta function limit of a Gaussian, corresponding to a single-pixel template), in case $\sigma_A^2 < 2\sigma_b^2$, i.e., if the intrinsic deconvolved source size is smaller than the beam size.

We ran simulations to compare these analytical results to the discretized case of pixels and did not find significant

differences. We also explored the effect of the realistic implementation of matched filtering, i.e., where the source position is not known a priori but the peak of the convolved image is taken instead. The main result appears to be that for σ_A^2 up to $2\sigma_b^2$; the S/N is almost flat as a function of template size. Therefore, a single-pixel template and slightly extended templates maximize the S/N with a smooth and slow transition, as the source size becomes more important relative to the beam size. For the purpose of our measurement, a precise formula for the match between template size and source size is not needed, and a probabilistic assignment based on artificial source recovery suffices. In particular, the results from our artificial sources show that in the full 3D case, the matching may be complex, and it depends on S/N as well as on line velocity width. The extra dependence on the line width can be understood as due to the use of the peak value rather than the value at the known source position because a wider-velocity line allows for a larger area over which the peak may be found (due to the combination of positive noise and real signal). To conclude, the matching of sources and templates at the basis of matched filtering can be approximately estimated from the previous calculation. For unresolved or slightly resolved sources, the S/N is a weak function of template size. As the $\sigma_A^2 \sim 2\sigma_b^2$ threshold is approached and crossed, the S/N becomes a rapidly increasing function of template size with a clear peak for extended templates. Therefore, in order to avoid missing extended sources in blind line searches in interferometric data, we recommend the inclusion of extended (hence 3D) templates, as described in this work.

Appendix C
Accounting for the Beam Inhomogeneity in our GOODS-N Mosaic

In order to mosaic pointings together, it is preferable to smooth all pointings to a common beam. This is not straightforward for the wide-area part of this survey (in the GOODS-N field) due to the large number of pointings observed over the course of several months, which caused a range of array configurations to be utilized.

For the pointings in the COSMOS field, each pointing was observed in every track. Therefore, the mosaic has uniform beam size properties. In the GOODS-N field, on the other hand, the beam differences potentially cause nonuniformity in the mosaic. This is most significant for pointings GN29 to GN42, which were mostly observed in the DnC configuration. In preparation for mosaicking, the individual pointings are all smoothed in the image plane to a common beam size with CASA `imsmooth`, but pointings for which a larger amount of smoothing was required end up with slightly larger beams than the target beam size.

The reason why smoothing DnC data seems to necessarily produce slightly different beams than D-configuration data can be appreciated from a look at the Fourier transform of a typical image from these pointings, effectively their *uv* coverage, in Figure 11. Smoothing multiplies the *uv* plane by a tapering Gaussian of the appropriate size, calculated from the starting beam size and target beam size. However, the D-configuration *uv* coverage does not look the same as the Gaussian-tapered DnC-configuration *uv* coverage; hence, the final beam is always going to be an imperfect match (unless both data sets are smoothed to a very large beam, at which point the initial shape of either *uv* coverage does not matter).

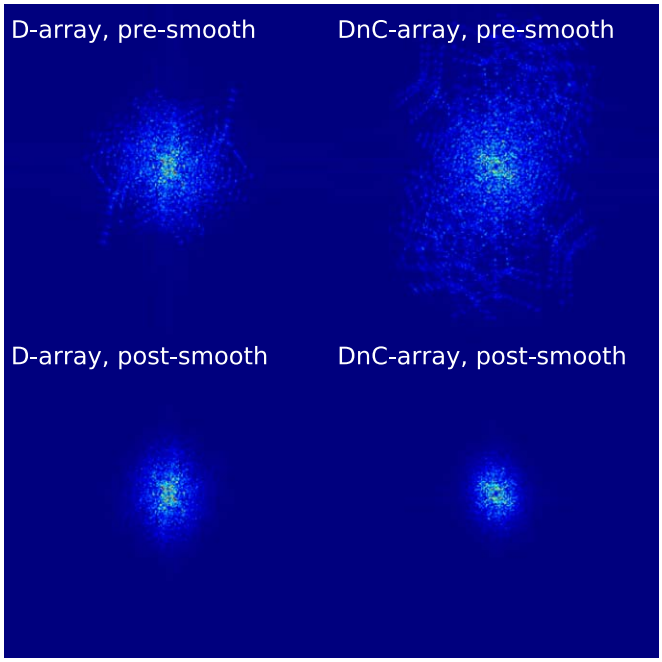

Figure 11. Absolute value of the Fourier transform of channel maps (in this example, channel 100) of individual pointings (in this example, pointing GN3, which has D-array data only, and GN34, which has mostly DnC-array data) before and after smoothing. The post-smoothing image shows that, although the smoothing is supposed to bring the two pointings to a common beam, instead it makes the resolution of the DnC-array pointings coarser; i.e., it produces a larger beam than in the pointings with D-array data only.

While the slight spatial inhomogeneity of the beam size is inconsequential in producing the S/N cube (as the noise in the mosaic can be calculated analytically and accounted for by Equation (2)), the beam size difference causes spatially varying noise in the matched-filtered cubes (see Appendix B). The main difference is between the set of 14 pointings (GN29–GN42) and the rest, so we also mosaic and match-filter them separately, in addition to working with mosaics with and without this set. Exploiting the improved uniformity within these sub-mosaics, we can measure the noise post–matched filtering. The objective is using the noise in the matched-filtered sub-mosaics to calculate the noise in the matched-filtered full GOODS-N mosaic. The S/N in the full mosaic is related to the sub-mosaics by

$$\begin{aligned} &\mathrm{S/N_{TOT}} \\ &= \frac{\sqrt{\sum_A A_i^2/\sigma_i^2}}{\sqrt{\sum_{\mathrm{TOT}} A_i^2/\sigma_i^2}}\mathrm{S/N}_A + \frac{\sqrt{\sum_B A_i^2/\sigma_i^2}}{\sqrt{\sum_{\mathrm{TOT}} A_i^2/\sigma_i^2}}\mathrm{S/N}_B \\ &\doteq f_A\,\mathrm{S/N}_A + f_B\,\mathrm{S/N}_B, \end{aligned} \tag{5}$$

where $\mathrm{TOT} = A \cup B$ represents the set of GN29–GN42 pointings and the set of the remaining pointings. Match-filtering then corresponds to convolving with template h. This can be expressed as

$$\begin{aligned} (f \cdot \mathrm{S/N}) * h &= \int f(x) S/N(x) h(y-x)\,dx \\ &\simeq f(y) \int S/N(x) h(y-x), \end{aligned} \tag{6}$$

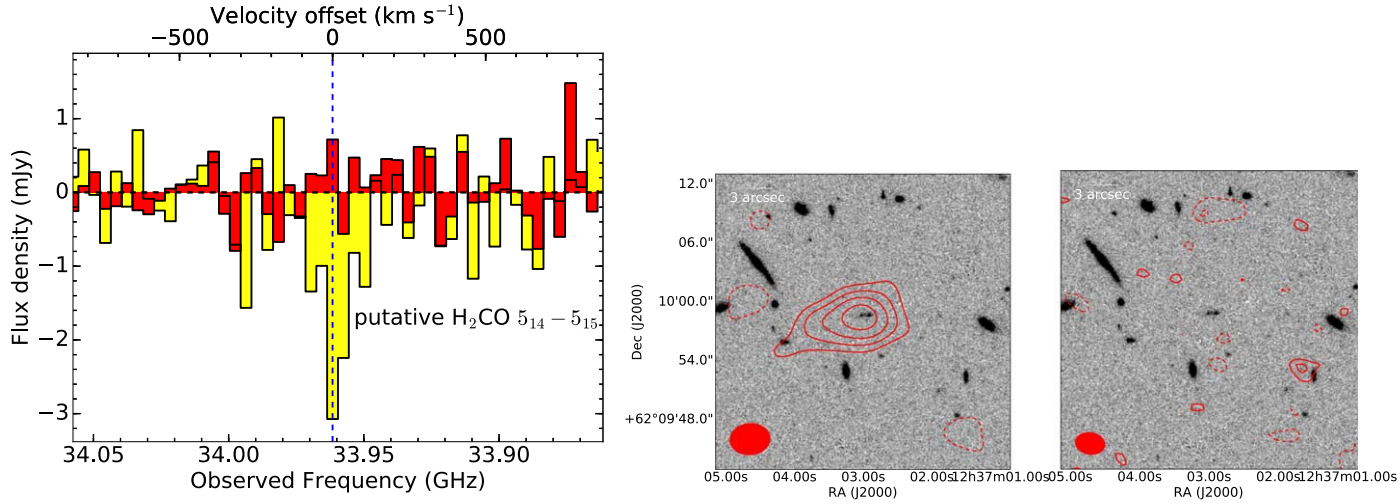

Figure 12. Top: aperture spectra of the most significant negative feature in the GOODS-N data in the original data (yellow histogram; where the feature was selected) and the newer observations (red histogram; pointing GN57). Bottom: The feature, a putative formaldehyde absorption line against the CMB (contours), appeared to be compatible with the 72.4 GHz (5_{14}–5_{15}) line of formaldehyde at the photo-$z \sim 1.13$ of the galaxies shown in the *HST H*-band image. The left image shows the line map in the original data, and the right image shows the same frequency range in the newer data where the line is not present, which suggests that it was simply due to noise. The contours are shown in steps of 1σ starting from $\pm 2\sigma$ with negative signal as solid contours to show absorption.

with $f(x) \simeq f(y) + O(f' \cdot \mathrm{FWHM}_h)$. To zeroth order, we can take f as constant over the scale of template h, which allows us to pull it out of the integral. This approximation is appropriate because the fraction functions f change slowly over the size of a template. Therefore, the noise after convolving with the template is

$$\mathrm{std}[(f_A(x)\mathrm{S/N}_A) * h] \simeq f_A \cdot \mathrm{std}[\mathrm{S/N}_A * h], \quad (7)$$

and we can calculate the noise in the matched-filtered mosaic by summing the standard deviations from the two terms in Equation (5) in quadrature. This method only requires measuring the noise in the matched-filtered sub-mosaics, in which the noise is uniform, and the fraction functions f_A and f_B, which can be calculated. Therefore, this process allows us to calculate noise maps of the matched-filtered cubes in GOODS-N, thereby accounting for the noise inhomogeneity due to spatially varying beam sizes.

Appendix D
Search for Negative Features as Potential Formaldehyde Absorption

D.1. Putative Feature

To better understand the characteristics of the noise in our survey data, we have also used our MF3D line-searching algorithm to detect negative line features. In order to constrain spurious line features due to noise, we take advantage of the symmetry around zero of interferometric noise in the absence of strong sources in the field. Although negative line features can usually be assumed to be due to noise, line absorption against the uniform CMB has been suggested to be a potential source of such negative lines. In particular, formaldehyde in dense molecular gas in galaxies has been confirmed at low-z to have the potential to produce such absorption against the CMB (Zeiger & Darling 2010; Darling & Zeiger 2012).

The most significant negative feature in the initial GOODS-N data cube (pointings GN1–GN56) had a high significance of $\sim 6.6\sigma$, and it appeared to be coincident with a pair of local interacting galaxies, GOODS J123702.92+620959.0 with a photo-z of $z = 1.13 \pm 0.05$ (Figure 12). Intriguingly, the strong absorption feature would be consistent with the 72.4 GHz (5_{14}–5_{15}) line of formaldehyde (H_2CO) at $z \sim 1.13$. The energy-level structure of formaldehyde allows collisional population anti-inversion in dense molecular clouds, making the line excitation temperature lower than the CMB temperature and producing absorption against the CMB itself. Formaldehyde silhouettes of galaxies in absorption offer both a novel probe of cosmological size and distance and a measurement of dense gas masses, density, and excitation properties (Darling & Zeiger 2012). Therefore, in order to investigate the possibility that our most significant negative feature may be a real absorption line against the CMB, we obtained additional VLA data, both at the same frequency (as an additional pointing, GN57 as part of our main survey) and at 22.6 GHz, in order to target the 4_{13}–4_{14} line of formaldehyde. We observed this additional tuning with the VLA *K* band (project ID: 15B-370; PI: Pavesi) on 2015 November 6. The observations lasted approximately 3 hr (130 minutes on source) in the D-array configuration, with a spectral setup consisting of a single tuning of the two 1 GHz 8 bit samplers (2 GHz total, dual polarization), with central frequencies of 21.58 and 22.5815 GHz for the two intermediate frequencies (IFs), respectively. The same calibrators were observed as for the main survey observations. We calibrated the data using CASA v.4.5 using the VLA pipeline and minor manual flagging, and we imaged the visibilities using natural weighting. The data cube was produced with 1 MHz channels corresponding to $\sim 13\,\mathrm{km\,s^{-1}}$, which is small compared to the expected line width. The data cube has a beam size of $4\farcs4 \times 3\farcs4$ and an rms noise of ~ 0.2 mJy beam^{-1} in 1 MHz wide channels.

We did not detect the lower-frequency line (Figure 13), and the new observations in pointing GN57 at the same frequency do not show any evidence for absorption at the same position and frequency (Figure 12). We therefore rule out the presence of an absorption line, and we conclude that the original feature was simply due to noise. By excluding the possibility that the most significant negative feature may correspond to real absorption, we strengthen our confidence in the assumption

Table 6
Formaldehyde Lines, Redshift Ranges, and Volumes Covered by the COLDz Survey

Transition	ν_0 (GHz)	$z_{\rm min}$	$z_{\rm max}$	$\langle z \rangle$	Volume (Mpc^3)	$\Delta T_{\rm Obs}$ Limit (K)	Volume Density (Mpc^{-3})
			COSMOS				
4_{13}–4_{14}	48.285	0.24	0.56	0.44	1,850	−0.03	$<6.2 \times 10^{-4}$
5_{14}–5_{15}	72.409	0.86	1.34	1.12	9,253	−0.03	$<1.2 \times 10^{-4}$
			GOODS-N				
4_{13}–4_{14}	48.285	0.27	0.61	0.47	13,690	−0.11	$<8.3 \times 10^{-5}$
5_{14}–5_{15}	72.409	0.90	1.42	1.18	62,329	−0.11	$<1.8 \times 10^{-5}$

Note. The quoted $\Delta T_{\rm Obs}$ limits correspond to 6σ over a line FWHM of 200 km s^{-1}. The volume density limit represents the 68% quantile of the probability distribution for the space abundance.

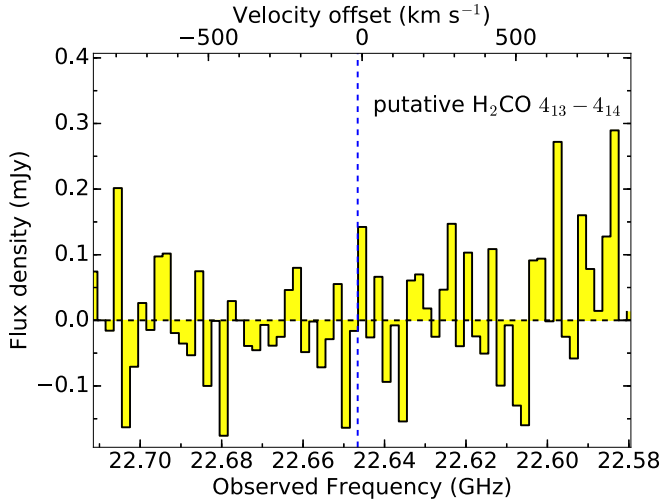

Figure 13. The VLA *K*-band spectrum at the expected redshifted frequency of the 48.3 GHz 4_{13}–4_{14} line of formaldehyde at $z \sim 1.13$. If the absorption feature we detected in our original data had been a real formaldehyde absorption line, we would expect to detect strong absorption, which is not seen.

that all (or at least most) negative line features are due to noise, which is crucial for the purity assessment of our positive line candidates.

D.2. H_2CO Deep-field Limits

Our lack of detections of significant formaldehyde absorption lines allows us to place some of the first constraints on the cosmic abundance of such absorption lines. By assuming a line FWHM of 200 km s^{-1}, we derive median 6σ limits (with no negative line candidates found above this threshold) of 0.18 and 0.55 mJy beam^{-1} for the COSMOS and GOODS-N fields, respectively, corresponding to $\Delta T_{\rm Obs}$ of −0.03 and −0.11 K at the average frequency and beam size of our survey. We note that the beam size of our observations ($\sim 3''$) is likely to be larger than the absorbing molecular regions ($\sim 0\farcs25$–$1\farcs25$ at $z \sim 1$; Darling & Zeiger 2012), implying a dilution of the expected signal strength due to the beam filling factor of ∼0.025–0.1. We use the absence of significant negative detections to infer a probability distribution for their space abundance by assuming a uniform uncorrelated distribution of sources over the cosmic volume covered by our survey and, therefore, a Poisson number count. The probability distribution for the space abundance is then an exponential distribution with a mode at zero and a mean equal to the inverse of the volume sampled (the 68th percentile upper limit to the space density is listed in Table 6). We use the model results from Darling & Zeiger (2012) to derive, from our $\Delta T_{\rm Obs}$ limit, a constraint on the line optical depth. These models imply that, at $z \sim 1$, the maximal expected temperature decrement with respect to the CMB is $\Delta T_{\rm Obs}/(1 - \exp^{-\tau}) \sim -1.2$ K for the 5_{14}–5_{15} and 4_{13}–4_{14} lines covered by our survey. This implies limits on the line optical depth of $\tau \lesssim 0.025$ and 0.1 for the COSMOS and GOODS-N fields, respectively. Although these values are comparable to the optical depths previously measured for the lower-frequency formaldehyde transitions, our results may be weaker by an order of magnitude or more due to the beam filling factor (Mangum et al. 2008, 2013; Darling & Zeiger 2012).

Appendix E
Description of the Individual Line Candidates

In this section, we briefly describe the remaining CO line candidates and potential counterpart associations (Table 7). These candidates are currently not independently confirmed and thus are only used in our statistical analysis. Because we only expect a small fraction of these candidates to correspond to real CO line emission, we advise caution in interpreting these lower-significance candidates on a per-source basis until they are independently confirmed. Quoted photometric redshift ranges are the 1σ uncertainties reported in the COSMOS2015 (Laigle et al. 2016) and CANDELS catalogs (Brammer et al. 2012; Skelton et al. 2014; Momcheva et al. 2016). The positional search radius considered is $3''$, which is dictated by the positional uncertainties (which are larger for extended sources) and the possibility of real physical offsets in the stellar emission, e.g., due to differential dust obscuration. The visual counterpart inspection was carried out utilizing *HST* (*H* band in GOODS-N and *I* band in COSMOS, where the *H* band was not available) and IRAC 3.6 μm images (band 1), which are shown in Figures 14 and 15. We have also inspected images from the other IRAC bands and find no evidence for additional counterpart matches relative to IRAC band 1.

E.1. COSMOS

COLDz.COS.4.—This is the highest-S/N candidate in COSMOS without a secure counterpart. There is a potential match at $0\farcs5 \pm 0\farcs6$ to the NW in the COSMOS2015 catalog with an uncertain photo-$z = 1.5$–2.3 matching the CO(1–0) redshift of $z = 2.30$. Two *I*-band and 3.6 μm sources are aligned with the elongated CO candidate emission (Figure 14).

Table 7
Catalog of the Line Candidates Identified in Our Analysis that Have Not Been Independently Confirmed to Date

ID	R.A. (J2000.0)	Decl. (J2000.0)	Frequency (GHz)	Redshift	Flux (Jy km s^{-1})	FWHM (km s^{-1})	S/N	Opt/NIR c.part?	Comments
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
COSMOS									
COLDz.COS.4	10:00:22.34	+02:34:14.0	34.887 ± 0.007	2.3041 ± 0.0007	0.12 ± 0.04	600 ± 150	5.71	Possible	
COLDz.COS.5	10:00:17.63	+02:34:36.0	34.814 ± 0.005	2.3110 ± 0.0005	0.08 ± 0.03	360 ± 100	5.62	N	
COLDz.COS.6	10:00:23.27	+02:34:22.0	31.989 ± 0.003	2.6034 ± 0.0003	0.037 ± 0.013	250 ± 70	5.59	Possible	
COLDz.COS.7	10:00:21.60	+02:33:56.0	35.823 ± 0.002	2.2178 ± 0.0002	0.12 ± 0.04	140 ± 40	5.56	Possible?	Extended
COLDz.COS.8	10:00:25.07	+02:35:56.0	35.291 ± 0.004	2.2663 ± 0.0004	0.24 ± 0.09	250 ± 70	5.56	N	Extended
COLDz.COS.9	10:00:22.44	+02:36:16.0	36.593 ± 0.003	2.1501 ± 0.0003	0.13 ± 0.04	220 ± 50	5.57	N	Extended
COLDz.COS.10	10:00:23.44	+02:36:29.0	34.132 ± 0.001	2.3772 ± 0.0001	0.11 ± 0.03	92 ± 18	5.56	Possible	Extended
COLDz.COS.11	10:00:20.43	+02:34:56.00	37.81 ± 0.002	2.0487 ± 0.0002	0.05 ± 0.02	120 ± 40	5.49	N	Blended $z = 0.3$, slightly extended
COLDz.COS.12	10:00:17.53	+02:35:11.00	35.354 ± 0.002	2.2605 ± 0.0002	0.025 ± 0.009	160 ± 40	5.43	N	
COLDz.COS.13	10:00:14.26	+02:35:02.50	32.423 ± 0.004	2.5552 ± 0.0004	0.09 ± 0.03	300 ± 90	5.43	N	
COLDz.COS.14	10:00:21.73	+02:35:57.00	35.005 ± 0.002	2.293 ± 0.0001	0.018 ± 0.008	90 ± 30	5.42	N	Blended $z = 0.9$
COLDz.COS.15	10:00:20.20	+02:35:31.50	34.681 ± 0.006	2.3237 ± 0.0005	0.033 ± 0.015	340 ± 110	5.41	Possible	Matched photo-z
COLDz.COS.16	10:00:25.50	+02:35:35.00	36.934 ± 0.004	2.121 ± 0.0004	0.07 ± 0.03	260 ± 80	5.34	Possible	Matched photo-z
COLDz.COS.17	10:00:19.70	+02:35:01.50	33.917 ± 0.007	2.3986 ± 0.0007	0.06 ± 0.02	570 ± 150	5.33	N	
COLDz.COS.18	10:00:24.60	+02:34:38.00	31.296 ± 0.002	2.6832 ± 0.0002	0.037 ± 0.012	180 ± 40	5.32	Possible	Matches photo-z, slightly extended
COLDz.COS.19	10:00:21.97	+02:34:54.50	37.083 ± 0.002	2.1085 ± 0.0002	0.046 ± 0.017	130 ± 30	5.29	N	M star nearby
COLDz.COS.20	10:00:17.03	+02:34:59.50	31.437 ± 0.002	2.6667 ± 0.0003	0.021 ± 0.008	170 ± 50	5.28	N	
COLDz.COS.21	10:00:23.47	+02:34:58.50	36.349 ± 0.004	2.1712 ± 0.0004	0.02 ± 0.02	110 ± 80	5.28	N	Slightly extended
COLDz.COS.22	10:00:22.90	+02:34:10.00	35.301 ± 0.005	2.2653 ± 0.0004	0.12 ± 0.04	350 ± 90	5.27	Possible	Very uncertain photo-z, extended
COLDz.COS.23	10:00:20.57	+02:34:01.00	38.504 ± 0.001	1.9937 ± 0.0001	0.14 ± 0.04	60 ± 10	5.26	Possible	Matches photo-z, slightly extended
COLDz.COS.24	10:00:14.99	+02:35:41.00	35.362 ± 0.002	2.2597 ± 0.0002	0.07 ± 0.02	130 ± 30	5.25	Possible	Close separation and photo-z
COLDz.COS.25	10:00:21.07	+02:34:30.50	33.579 ± 0.005	2.4328 ± 0.0006	0.049 ± 0.018	410 ± 110	5.25	Possible	Close separation and photo-z, extended
GOODS-N									
COLDz.GN.1	12:36:59.79	+62:11:09.50	37.485 ± 0.003	2.0751 ± 0.0002	0.405 ± 0.137	200 ± 50	6.38	Possible	Slightly extended
COLDz.GN.2	12:36:27.94	+62:14:09.78	32.518 ± 0.002	2.5448 ± 0.0002	0.109 ± 0.033	210 ± 50	6.14	N	
COLDz.GN.4	12:36:54.77	+62:17:28.00	35.937 ± 0.002	2.2076 ± 0.0002	0.382 ± 0.12	180 ± 40	6.08	N	Extended
COLDz.GN.5	12:37:00.00	+62:15:21.00	37.229 ± 0.005	2.0962 ± 0.0005	0.283 ± 0.073	520 ± 100	6.06	N	Photo-z at lower z
COLDz.GN.6	12:37:01.50	+62:12:35.50	32.878 ± 0.003	2.5059 ± 0.0003	0.116 ± 0.038	260 ± 60	6.0	Possible	
COLDz.GN.7	12:36:28.89	+62:13:00.80	30.615 ± 0.002	2.7652 ± 0.0003	0.189 ± 0.047	290 ± 50	5.97	N	
COLDz.GN.8	12:36:37.31	+62:15:03.39	37.737 ± 0.004	2.0546 ± 0.0003	0.823 ± 0.272	270 ± 70	5.89	Possible	Close photo-z, very extended
COLDz.GN.9	12:36:56.35	+62:18:19.50	36.348 ± 0.001	2.1713 ± 0.0001	0.077 ± 0.022	100 ± 20	5.88	Possible	
COLDz.GN.10	12:36:33.87	+62:15:29.36	32.841 ± 0.002	2.51 ± 0.0002	0.143 ± 0.05	120 ± 30	5.83	Possible	Slightly extended
COLDz.GN.11	12:36:53.20	+62:14:34.49	35.073 ± 0.001	2.2866 ± 0.0001	0.323 ± 0.101	90 ± 20	5.81	N	Extended
COLDz.GN.12	12:37:03.50	+62:12:52.00	37.93 ± 0.007	2.039 ± 0.0006	0.317 ± 0.114	490 ± 130	5.81	Possible	photo-$z = 4$
COLDz.GN.13	12:36:43.18	+62:14:22.44	36.743 ± 0.005	2.1372 ± 0.0005	0.199 ± 0.071	390 ± 100	5.8	N	
COLDz.GN.14	12:36:59.07	+62:14:48.00	34.133 ± 0.006	2.3771 ± 0.0006	0.605 ± 0.192	490 ± 120	5.65	Possible	Very close photo-z, extended
COLDz.GN.15	12:36:41.67	+62:15:47.93	33.34 ± 0.009	2.4574 ± 0.0009	0.178 ± 0.069	640 ± 190	5.64	Possible	Blended photo-$z \sim 3$
COLDz.GN.16	12:36:49.42	+62:12:17.98	32.807 ± 0.002	2.5136 ± 0.0002	0.062 ± 0.022	140 ± 40	5.63	Possible	Local foreground?
COLDz.GN.17	12:37:01.22	+62:13:04.50	33.2 ± 0.002	2.472 ± 0.0002	0.084 ± 0.036	140 ± 50	5.63	N	
COLDz.GN.18	12:36:51.12	+62:15:54.99	36.938 ± 0.003	2.1206 ± 0.0003	0.152 ± 0.043	270 ± 60	5.62	Possible	Matched photo-z
COLDz.GN.19	12:37:08.45	+62:14:23.48	36.017 ± 0.002	2.2005 ± 0.0002	0.207 ± 0.09	120 ± 40	5.61	N	Local foreground, extended
COLDz.GN.20	12:37:04.88	+62:17:44.49	30.727 ± 0.003	2.7514 ± 0.0003	0.205 ± 0.062	280 ± 60	5.6	Possible	Close photo-z
COLDz.GN.21	12:36:54.93	+62:11:29.50	35.265 ± 0.004	2.2687 ± 0.0004	0.194 ± 0.057	390 ± 90	5.6	N	
COLDz.GN.22	12:37:00.14	+62:11:58.50	33.441 ± 0.003	2.447 ± 0.0003	0.075 ± 0.037	140 ± 50	5.59	N	

Table 7
(Continued)

ID	R.A. (J2000.0)	Decl. (J2000.0)	Frequency (GHz)	Redshift	Flux (Jy km s^{-1})	FWHM (km s^{-1})	S/N	Opt/NIR c.part?	Comments
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
COLDz.GN.23	12:36:56.71	+62:13:19.50	37.925 ± 0.002	2.0394 ± 0.0002	0.156 ± 0.051	160 ± 40	5.59	N	
COLDz.GN.24	12:36:46.85	+62:12:18.97	31.165 ± 0.008	2.6988 ± 0.0009	0.222 ± 0.095	560 ± 180	5.56	N	Slightly extended
COLDz.GN.25	12:37:03.35	+62:08:59.00	35.598 ± 0.002	2.2381 ± 0.0002	0.119 ± 0.043	130 ± 30	5.55	Possible	Close photo-z
COLDz.GN.26	12:37:00.29	+62:16:31.50	36.733 ± 0.006	2.1381 ± 0.0005	0.169 ± 0.056	490 ± 120	5.54	Possible	Very faint counterpart
COLDz.GN.27	12:36:45.09	+62:18:00.46	37.373 ± 0.001	2.0843 ± 0.0001	0.096 ± 0.025	90 ± 20	5.54	Possible	Close photo-z
COLDz.GN.28	12:36:47.77	+62:12:57.47	32.289 ± 0.003	2.57 ± 0.0003	0.113 ± 0.04	210 ± 60	5.51	N	Different spec-z, $z = 2.932$, slightly extended
COLDz.GN.29	12:37:06.38	+62:16:34.49	32.861 ± 0.002	2.5078 ± 0.0002	0.089 ± 0.027	190 ± 40	5.51	Possible	
COLDz.GN.30	12:36:50.43	+62:10:29.48	31.297 ± 0.002	2.6832 ± 0.0002	0.236 ± 0.088	150 ± 40	5.51	N	Slightly extended

Note. We advise caution in interpreting these lower-significance candidates on a per-source basis until they are independently confirmed. Columns are: (1) line ID; (2)–(3) R.A. and decl. (J2000); (4) central frequency and uncertainty based on Gaussian fitting; (5) CO(1–0) redshift and uncertainty, unless otherwise noted; (6) velocity-integrated line flux and uncertainty; (7) line FWHM as derived from a Gaussian fit; (8) S/N measured by MF3D; (9) presence of a spatially coincident optical/NIR counterpart; and (10) comments.

COLDz.COS.5.—No counterpart is found in the COSMOS2015 catalog or the images at the position of this candidate.

COLDz.COS.6.—The images show an IRAC 3.6 μm source $3'' \pm 0.''2$ to the SE of the candidate that has a photo-$z = 2.9$–3 and might therefore be associated with our candidate, although the offset appears significant (the CO(1–0) redshift is $z = 2.60$). At the CO position, there is an *I*-band source with photo-$z = 0.44$–0.48, although its 3.6 μm image is contaminated by the brighter, higher-z galaxy. It is unclear if the CO candidate may be related.

COLDz.COS.7.—The CO(1–0) redshift of this candidate is $z = 2.22$. It is at the position of an *HST I*-band source that is not in the COSMOS2015 catalog and is not visible in the 3.6 μm image.

COLDz.COS.8.—Candidate is at the position of a faint *HST I*-band and IRAC 3.6 μm source that is listed in the COSMOS2015 catalog $2.''3 \pm 0.''4$ to the N ($z_{\rm phot} = 0.2$–0.7).

COLDz.COS.9.—Candidate shows spatially extended CO emission centered on an *HST I*-band source with a photo-$z = 0.1$–0.8, which is therefore unlikely to be associated with the candidate.

COLDz.COS.10.—Candidate is spatially extended and cospatial with multiple faint *HST I*-band galaxies. The COSMOS2015 catalog only reports a faint galaxy $1.''3 \pm 0.''7$ to the SW with a very uncertain photo-z of 0.8–4.3 that may be associated with our candidate.

COLDz.COS.11.—Candidate is near the position of a low-z galaxy ($z_{\rm phot} = 0.32$–0.35). There is a brighter IRAC 3.6 μm source $1.''6 \pm 0.''4$ to the NW, $z_{\rm phot} = 1.5$–1.6, that may be related to the CO candidate with a CO redshift of 2.0.

COLDz.COS.12.—No counterpart is found in the COSMOS2015 catalog or the images at the position of this candidate.

COLDz.COS.13.—No counterpart is found in the COSMOS2015 catalog or the images at the position of this candidate.

COLDz.COS.14.—Candidate is affected by foreground contamination that prevents any counterpart assessment. In particular, there is a bright photo-$z = 0.9$ galaxy at $1.''6 \pm 0.''3$ to the NW.

COLDz.COS.15.—Candidate has a potential counterpart match. The COSMOS2015 catalog lists a galaxy $1.''8 \pm 0.''4$ to the NW with a photo-$z = 1.8$–2.8 that is very faint in the *I*-band and IRAC 3.6 μm images. Assuming CO(1–0) would place this candidate at $z = 2.32$.

COLDz.COS.16.—Candidate has a potential counterpart match, but it appears confused with a bright galaxy $1.''6$ to the SE with a photo-$z = 1.0$–1.2. The potential counterpart has photo-$z = 1.3$–2.6 and is located about $2'' \pm 0.''5$ to the NE.

COLDz.COS.17.—No counterpart is found in the COSMOS2015 catalog or the images.

COLDz.COS.18.—Candidate has a potential counterpart match $1.''6 \pm 0.''4$ to the N with a photo-$z = 2.4$–2.5; assuming CO(1–0) would place it at $z = 2.68$. This candidate is contaminated by a local bright galaxy to the NE.

COLDz.COS.19.—Candidate does not appear to have a counterpart. An M star is located $0.''8 \pm 0.''7$ to the NE and partly prevents counterpart identification.

COLDz.COS.20.—Candidate does not have a counterpart. The COSMOS2015 catalog lists two galaxies at separations of $2.''2 \pm 0.''3$ and $2.''3 \pm 0.''3$. The first galaxy has a photo-$z = 0.6$–0.9, and the second one is at photo-$z = 1.9$–2.5. This latter galaxy may be associated with our candidate, which has a CO(1–0) redshift of $z_{10} = 2.67$.

COLDz.COS.21.—No counterpart is found in the COSMOS2015 catalog or the images at the position of this spatially extended candidate, but the IRAC 3.6 μm images are contaminated by bright nearby stars and galaxies.

COLDz.COS.22.—A faint galaxy is visible in the *HST I*-band image $1.''5 \pm 0.''7$ to the NW that may be associated with our line candidate. The catalog lists a very uncertain photo-$z = 1.5$–5.5, which is compatible with the CO(1–0) redshift of $z_{10} = 2.27$.

COLDz.COS.23.—Candidate has a potential counterpart association. This is a galaxy $1.''5 \pm 0.''5$ to the SW that is compatible with the position of at least part of the slightly spatially extended line emission. The photometric redshift for this galaxy is photo-$z = 1.7$–2.8, which is compatible with the CO(1–0) redshift of $z_{10} = 1.99$.

COLDz.COS.24.—Candidate has a potential counterpart association. The potential counterpart is only $0.''8 \pm 0.''9$ to the N and has a photometric redshift of $z_{\rm phot} = 1.9$–2.0, which is close to the CO(1–0) redshift of $z_{10} = 2.26$. A second galaxy is seen $1.''6 \pm 0.9$ to the E that has a photometric redshift of photo-$z = 0.89$–0.92 and contaminates the emission in the IRAC 3.6 μm images.

COLDz.COS.25.—This spatially extended candidate has a potential counterpart. This potential counterpart is $1.''4 \pm 0.''7$ to the SE and has a photometric redshift of photo-$z = 2.6$–3.0, which is close to the CO(1–0) redshift of $z_{10} = 2.43$. The IRAC 3.6 μm images are contaminated by a nearby star, which makes it difficult to identify faint sources reliably.

E.2. GOODS-N

COLDz.GN.1.—This spatially extended line candidate has potential counterpart matches. There are multiple galaxies that are compatible with the line emission position blended in the IRAC 3.6 μm image but visible in the *HST H* band with photometric redshifts in the CANDELS catalog (Skelton et al. 2014). The closest catalog match has a separation of only $0.''4 \pm 0.''6$ to the NE and an uncertain photo-$z = 0.8$–2.2. The catalog lists three more galaxies within $3''$ (separations of $1.''3 \pm 0.''6$, $2'' \pm 0.''6$, and $2.''6 \pm 0.''6$) with photo-zs of 1.5–1.7, 0.9–2.5, and 1.1–1.8, respectively. The CO(1–0) redshift of our candidate is $z_{10} = 2.08$, which makes it compatible with at least two of these potential counterparts.

COLDz.GN.2.—This is the highest-S/N candidate in GOODS-N without a clear counterpart. The CANDELS catalog lists a faint source $2.''6 \pm 0.''3$ to the SE with uncertain photo-$z = 1.0$–2.0 (Skelton et al. 2014). The CO(1–0) redshift of our candidate ($z = 2.54$) makes it a possible but unlikely counterpart.

COLDz.GN.4.—Candidate is unlikely to have a counterpart. There are no galaxies in the CANDELS catalog within $3''$, and no galaxies are visible in the *HST H*-band or IRAC 3.6 μm images.

COLDz.GN.5.—There are two galaxies in the images within $2''$ of the line candidate with separations of $1.''5 \pm 0.''3$ and $1.''7 \pm 0.''3$. They are unlikely to be counterparts because they have photometric redshifts of $z_{\rm phot} = 1.1$–1.3 and 0.4–0.5, respectively, while the CO(1–0) redshift of our candidate is $z_{10} = 2.1$.

COLDz.GN.6.—Candidate has a potential match $2\farcs8 \pm 0\farcs4$ to the SE, in the direction where the CO emission is slightly spatially extended. The catalog lists a photo-z of 2.4–2.5, which is compatible with the CO(1–0) redshift of $z_{10} = 2.51$, suggesting a possible counterpart match.

COLDz.GN.7.—Candidate has an unlikely but possible match $2\farcs9 \pm 0\farcs3$ to the SE that appears to be at a significant offset. The galaxy has a photo-$z = 1.8$–2.0, which is not compatible with the CO(1–0) redshift of $z_{10} = 2.76$; therefore, we do not consider this to be a match.

COLDz.GN.8.—This spatially extended CO candidate has a possible match $3\farcs0 \pm 0\farcs8$ to the SE with an uncertain photo-z of 1.4–2.4, which is compatible with the CO(1–0) redshift of $z_{10} = 2.05$. This is a potential match because the line emission appears to be very spatially extended and may be compatible with coming from a dust-obscured part of the optical galaxy.

COLDz.GN.9.—Candidate is unlikely to have a counterpart. It appears near a spec-$z = 0.516$ galaxy that is $2\farcs8 \pm 0\farcs3$ to the NW. The catalog also lists a faint photo-$z = 1.9$–2.1 galaxy $2\farcs4 \pm 0\farcs3$ to the SW (which appears to be significantly offset from the CO line emission) that could be consistent with the CO(1–0) redshift of $z_{10} = 2.17$.

COLDz.GN.10.—This slightly spatially extended candidate is unlikely to have a counterpart. The closest catalog association is $1\farcs5 \pm 0\farcs5$ to the NE and has a photo-z of 4.4–5. The CO(2–1) redshift for our candidate would be $z_{21} = 6.0$ and is therefore an unlikely match. The CANDELS catalog lists two more galaxies just below 3″ to the NE with photo-zs of 1.7–2.0 and 1.9–2.4 that may be compatible with the CO(1–0) redshift of $z_{10} = 2.51$.

COLDz.GN.11.—Candidate appears spatially extended and elongated. No objects are seen in the *HST H*-band and IRAC 3.6 μm images. The CANDELS catalog lists a galaxy $1\farcs8 \pm 0\farcs7$ to the NW that has a photo-z of 0.6–1.6. This is inconsistent with the CO(1–0) redshift of $z_{10} = 2.29$, so a match is unlikely.

COLDz.GN.12.—Candidate has no likely match. The catalog lists a faint galaxy $1\farcs7 \pm 0\farcs4$ to the NE with photo-$z = 4.1$–4.4 that may potentially be associated. The counterpart status is difficult to evaluate due to blending with the bright local (spec-$z = 0.784$) galaxy at a separation of just $2\farcs4 \pm 0\farcs4$.

COLDz.GN.13.—Candidate is spatially extended and elongated and unlikely to have a counterpart association. The CANDELS catalog lists two potential matches within 3″ with separations of $1\farcs7 \pm 0\farcs5$ and $3'' \pm 0\farcs5$. The photometric redshifts listed by the catalog are $z_{\rm phot} = 0.2$–0.4 and 0.6–1.4, respectively, which makes them unlikely counterparts given the CO(1–0) redshift of our candidate ($z_{10} = 2.14$).

COLDz.GN.14.—This is a spatially extended CO candidate, and it has a possible counterpart that is faint but visible in the IRAC 3.6 μm image. We identify this counterpart with the catalog listing of a photo-$z = 2.4$–2.7 galaxy that is displaced by $2\farcs8 \pm 0\farcs7$ to the SW. This counterpart is compatible with the CO(1–0) redshift of $z_{10} = 2.38$. The offset may not be significant because the IR-detected galaxy appears to be compatible with the position of this spatially extended candidate.

COLDz.GN.15.—Candidate may have a counterpart. The 3.6 μm image is partly blended with a spec-$z = 0.453$ galaxy $2\farcs3 \pm 0\farcs6$ to the W, which makes the identification difficult. The catalog lists two possible counterparts with photo-$z = 1.6$–1.8 and 3.1–3.9 offset, respectively, $0\farcs9 \pm 0\farcs6$ and $1\farcs7 \pm 0\farcs6$ to the NW and NE, that are not compatible with the CO(1–0) redshift of $z_{10} = 2.46$.

COLDz.GN.16.—Candidate may have a counterpart. The image is partly blended with a spec-$z = 0.961$ galaxy $2\farcs3 \pm 0\farcs6$ to the NE, which makes identification difficult. The CANDELS catalog lists three more galaxies within 2″ and 3″ from our candidate with photo-$z = 0.9$–2.0, 2.2–2.4, and 1.4–2.3, all of which may be compatible with the CO(1–0) redshift of $z_{10} = 2.51$.

COLDz.GN.17.—Candidate appears spatially extended and elongated. The catalog lists a galaxy $1\farcs5 \pm 0\farcs5$ to the SE that is visible in the IRAC 3.6 μm images. This galaxy has a photo-z of 1.2–1.9, which is only somewhat inconsistent with the CO(1–0) redshift of $z_{10} = 2.47$. Therefore, a counterpart association cannot be ruled out.

COLDz.GN.18.—Candidate appears to be closely associated with other lower-significance candidates that are visible in the line maps. It has a potential match, a faint galaxy with photo-$z = 1.3$–2.2 only $0\farcs8 \pm 0\farcs5$ to the SE that is compatible with the CO(1–0) redshift of $z_{10} = 2.12$. The catalog also lists three galaxies $0\farcs4 \pm 0\farcs5$, $1\farcs8 \pm 0\farcs5$, and $2\farcs7 \pm 0\farcs5$ to the SE with photo-zs of 0.3–0.9, 2.5–2.8, and 4.3–5.2, respectively, that may be associated in case of incorrect photometric redshifts.

COLDz.GN.19.—This spatially extended candidate is blended with a local foreground galaxy at $z = 0.564$. No continuum emission is detected in our data at this position; therefore, we exclude the possibility that the line candidate may be spurious and due to noise superposed to continuum emission. The presence of the bright foreground contaminates the *HST H*-band and IRAC 3.6 μm images, making it difficult to evaluate the counterpart status. Lensing of a faint $z = 2.20$ galaxy by the foreground galaxy is a possibility.

COLDz.GN.20.—Candidate may have a counterpart. It appears close to a foreground galaxy that partly contaminates the IRAC 3.6 μm image. The *HST H*-band image shows a potential match that the catalog identifies as a galaxy $1\farcs16 \pm 0.3$ to the NW with a photo-z of 1.6–2.5. The association is not ruled out by our CO(1–0) redshift of $z_{10} = 2.75$.

COLDz.GN.21.—Candidate is unlikely to have a counterpart. The images show a galaxy $1\farcs0 \pm 0\farcs6$ to the NE that has a grism-z of 0.86–0.94 from Momcheva et al. (2016). This makes it incompatible with the CO(1–0) redshift of $z_{10} = 2.27$.

COLDz.GN.22.—No counterpart is found in the images or the CANDELS catalog at the position of this candidate.

COLDz.GN.23.—No counterpart is found in the images or the CANDELS catalog at the position of this candidate.

COLDz.GN.24.—This spatially extended candidate may have a counterpart that is visible in the IRAC 3.6 μm image but not in the *HST H*-band image. The separation is $2\farcs9 \pm 0\farcs5$ to the S, but this may not be significant due to the extent of the emission. The photo-z is uncertain and ranges from 0.9 to 2.3; therefore, an association is not strongly ruled out by our CO(1–0) redshift of $z_{10} = 2.7$.

COLDz.GN.25.—Candidate has a potential counterpart $1\farcs7 \pm 0\farcs7$ to the NE. The galaxy is well visible in the *H*-band and IRAC 3.6 μm images and has a photo-z of 1.9–2.2, which is compatible with the CO(1–0) redshift of $z_{10} = 2.24$.

COLDz.GN.26.—This spatially extended CO candidate has potential matches that appear very faint in the IRAC 3.6 μm image. The catalog lists two galaxies at $1\farcs8 \pm 0\farcs5$ and $1\farcs9 \pm 0\farcs5$ to the NE with uncertain photo-zs of 0.6–4.2 and 2.1–3.7, respectively, which makes them compatible with the

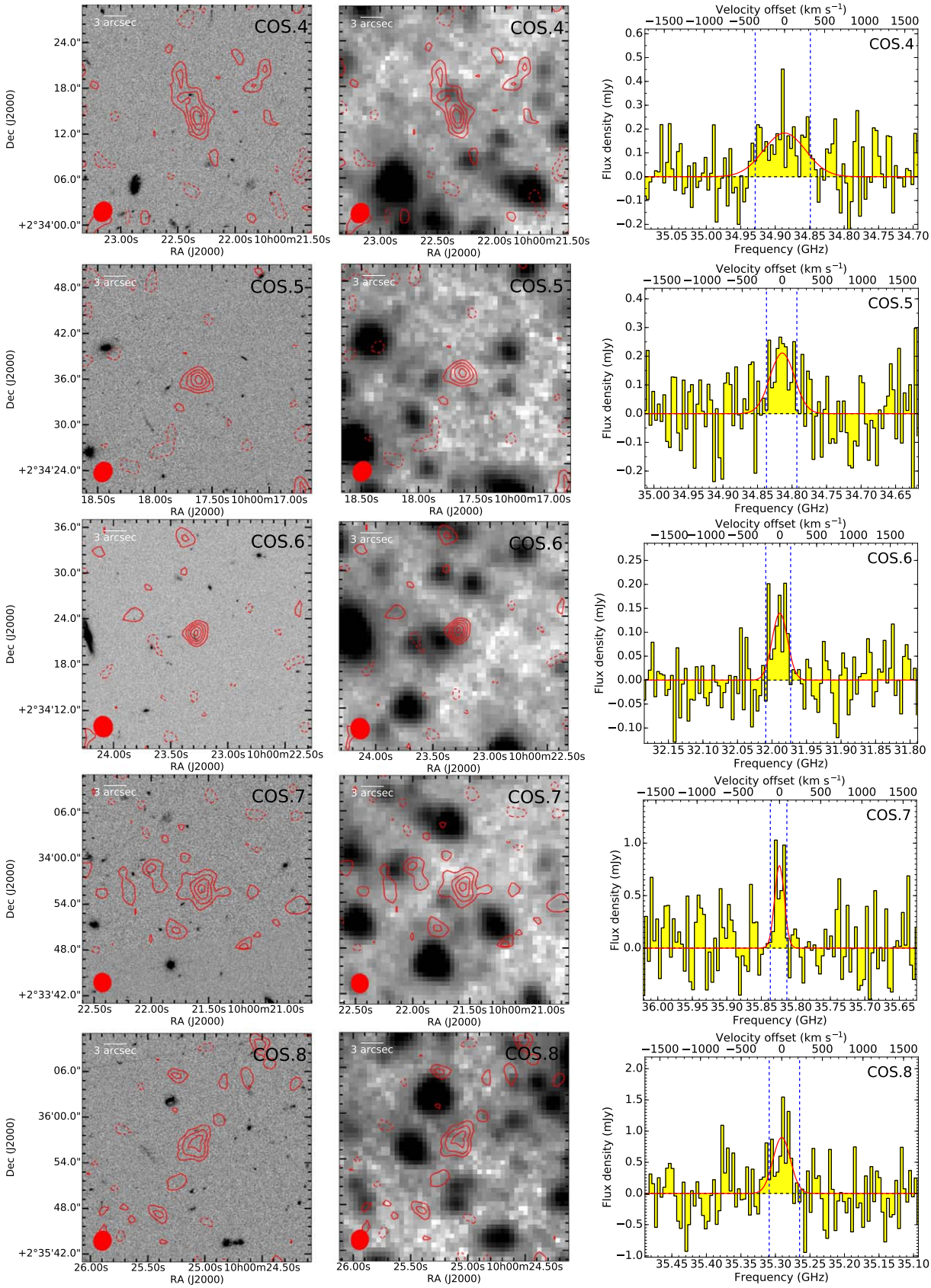

Figure 14. Additional candidate integrated line map overlaid (contours) on *HST* *I*-band (left) and IRAC channel 1 images (middle) from SPLASH (gray scale; Steinhardt et al. 2014). The *HST* images were obtained from the online IRSA/IPAC database. Contours are shown in steps of 1σ starting at $\pm 2\sigma$. Right: line candidate aperture spectra (histograms) and Gaussian fits (red curves) to the line features. The observed frequency resolution is the same as in Figure 5. The velocity range used for the overlays is shown by the dashed blue lines.

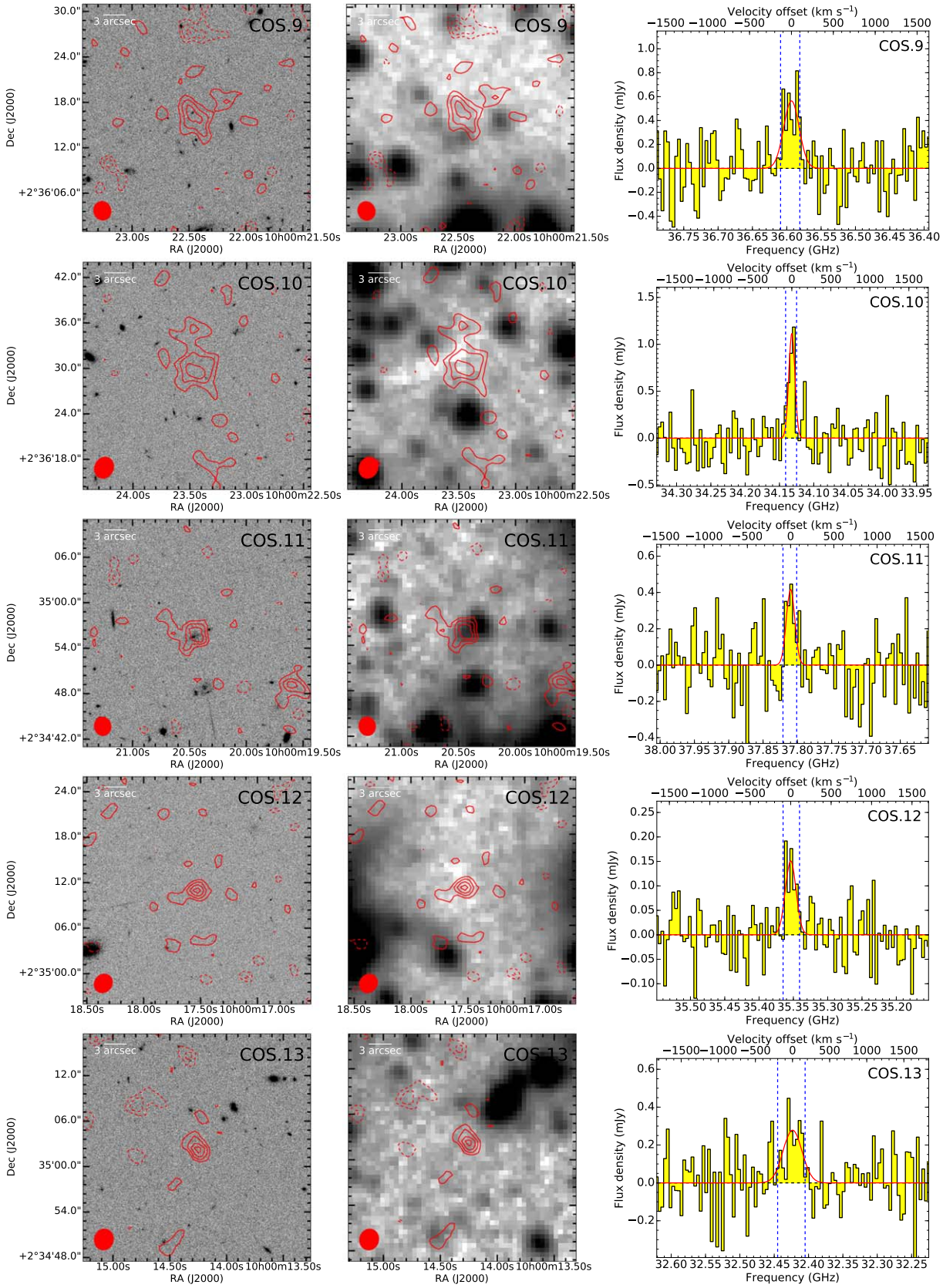

Figure 14. (Continued.)

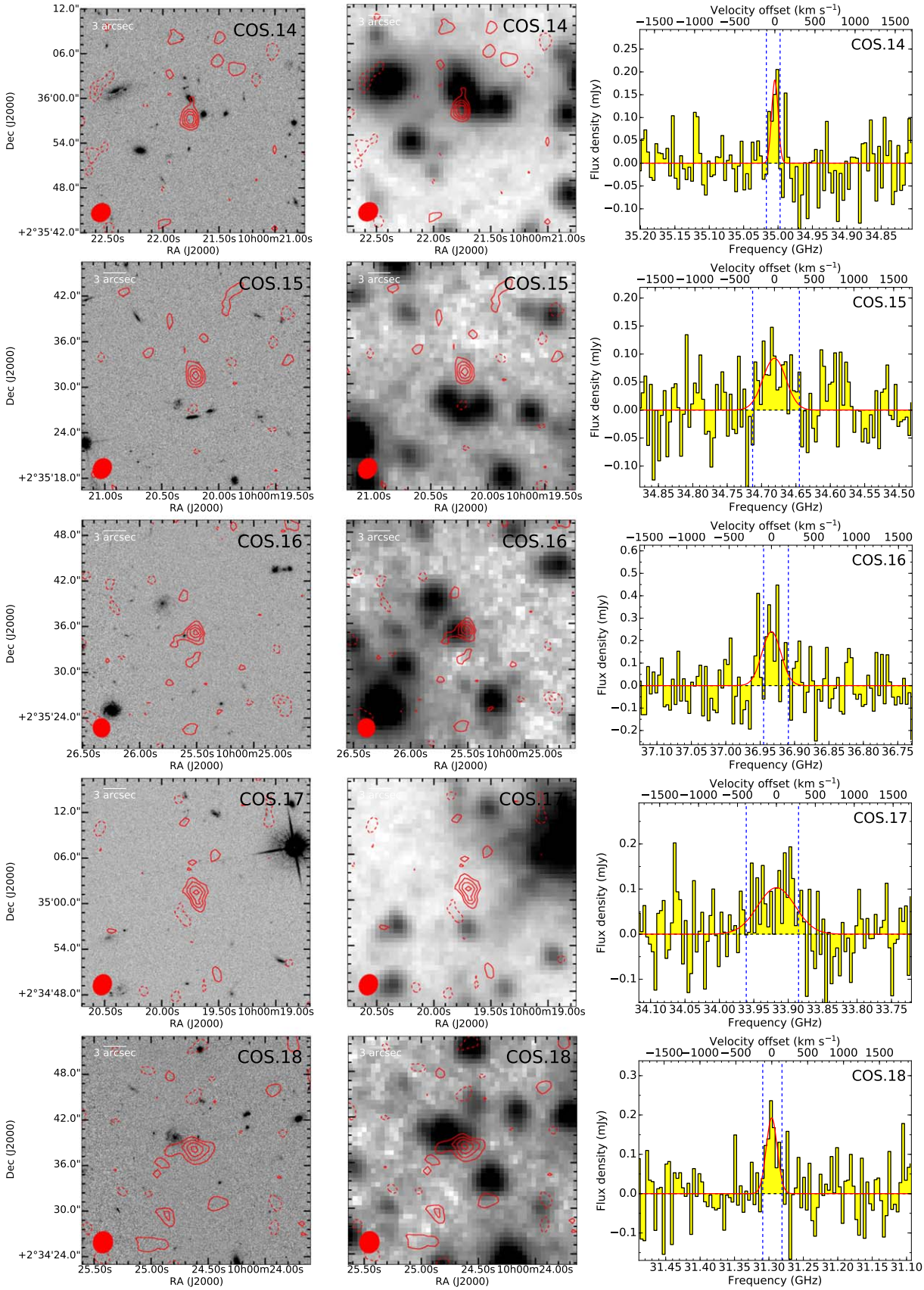

Figure 14. (Continued.)

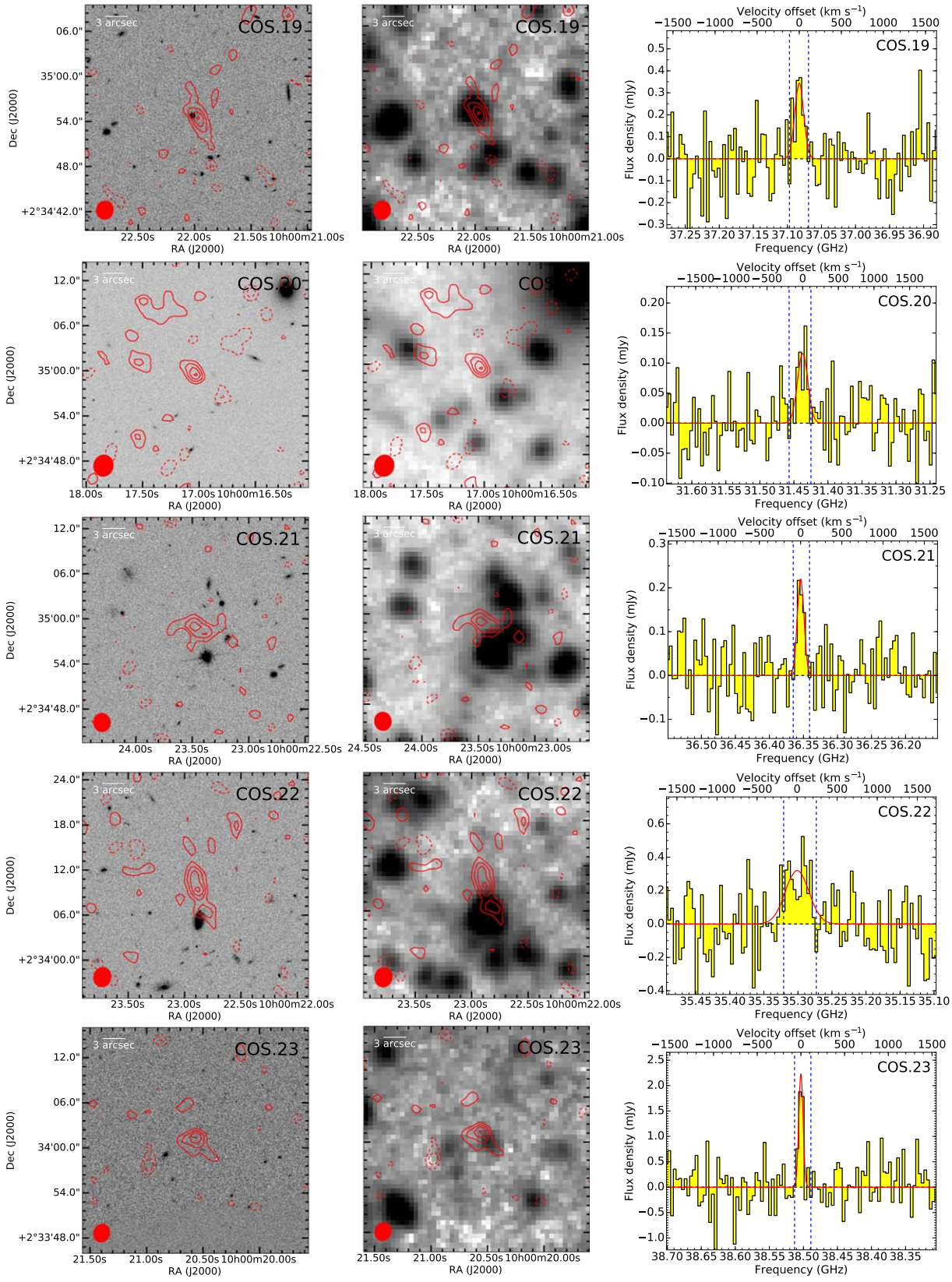

Figure 14. (Continued.)

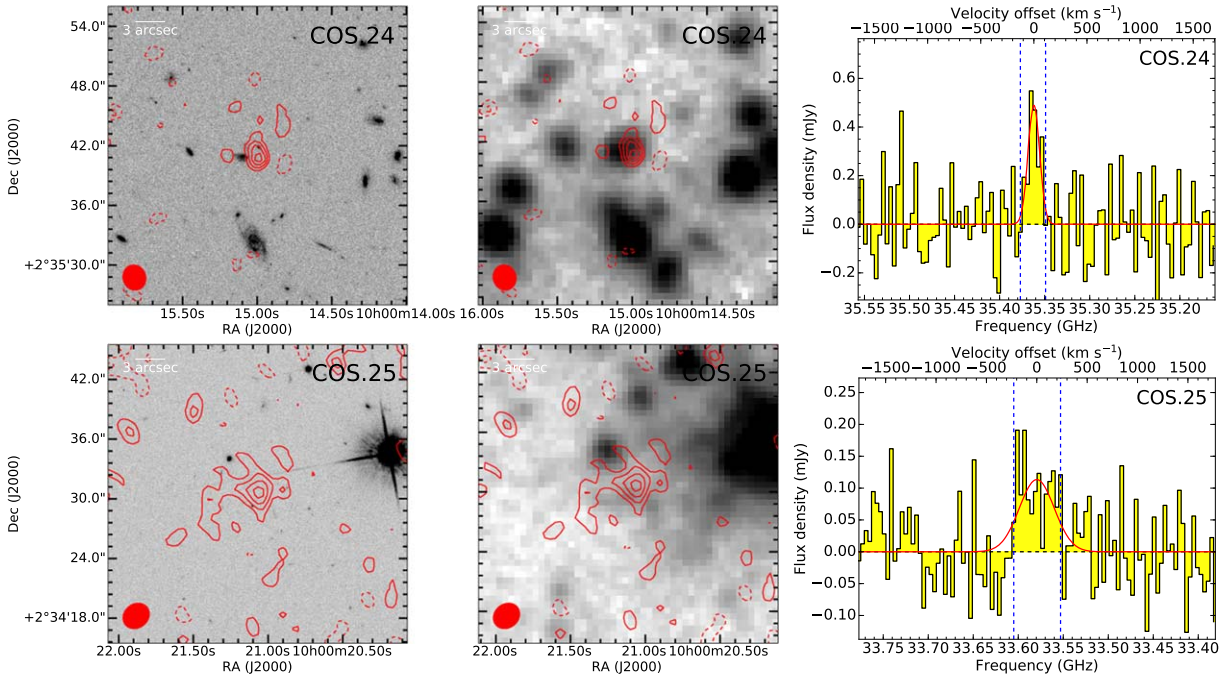

Figure 14. (Continued.)

CO(1–0) redshift of $z_{10} = 2.14$. The spatial offset may not be significant because of the spatial extent of the emission, and these are therefore potential counterpart matches.

COLDz.GN.27.—Candidate has three potential counterparts in the catalog with close photo-zs. The first is $2\farcs3 \pm 0\farcs4$ to the SE with a photo-z of 1.5–2.3. The second is $2\farcs7 \pm 0\farcs4$ to the SE with a photo-z of 1.5–2. Both of these are compatible with the CO(1–0) redshift of $z_{10} = 2.08$. The third potential counterpart is located $2\farcs9 \pm 0\farcs4$ to the NE and has a photo-z in the range 5.2–5.7, which is consistent with the CO(2–1) redshift of $z_{21} = 5.16$.

COLDz.GN.28.—The *HST* *H*-band and IRAC 3.6 μm images show a potential match $1\farcs6 \pm 0\farcs3$ to the S, but this galaxy was reported to have a spec-$z = 2.932$ (Skelton et al. 2014). Assuming CO(1–0) would imply $z_{10} = 2.57$, which implies either a lack of counterpart or an incorrect spectroscopic redshift.

COLDz.GN.29.—Candidate has a possible counterpart. Both the *HST* *H*-band and the IRAC 3.6 μm images show multiple sources within $2''$. The catalog lists two faint galaxies with photo-$z = 0.8$–3.5 and 1.1–1.7 just $1\farcs4 \pm 0\farcs4$ and $1\farcs5 \pm 0\farcs4$ to the SE and NE, respectively. The first of these is compatible with the CO(1–0) redshift of $z_{10} = 2.51$.

COLDz.GN.30.—This spatially extended candidate does not appear to have counterparts. The IRAC 3.6 μm image is blended with a bright foreground galaxy, and the only catalog association (offset by $2\farcs7 \pm 0\farcs9$ to the SW) has a photo-$z = 0.7$–0.9, which is incompatible with the CO(1–0) redshift of $z_{10} = 2.68$.

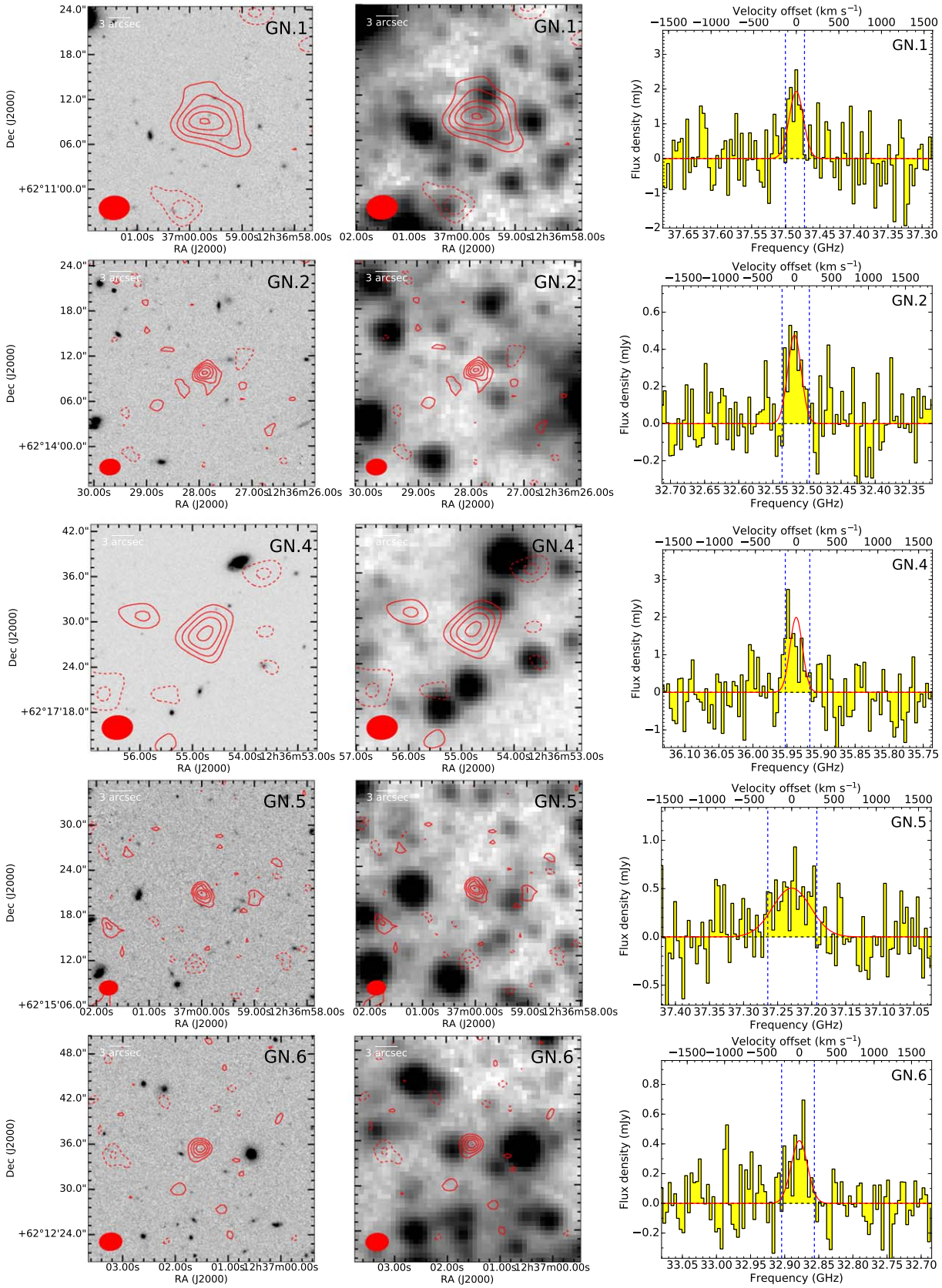

Figure 15. Additional candidate integrated line map overlaid (contours) on *HST H*-band (left) and IRAC channel 1 images (middle; gray scale). The *HST* and *Spitzer* images were obtained from the CANDELS database. The CO line data were taken from the Natural-mosaic or Smoothed-mosaic when the line emission was unresolved/resolved, respectively. Contours are shown in steps of 1σ starting at $\pm 2\sigma$. Right: line candidate single-pixel/aperture spectra (histograms; for unresolved/resolved emission) and Gaussian fits to the line features (red curves). The observed frequency resolution is the same as in Figure 5. The velocity range used for the overlays is shown by the dashed blue lines.

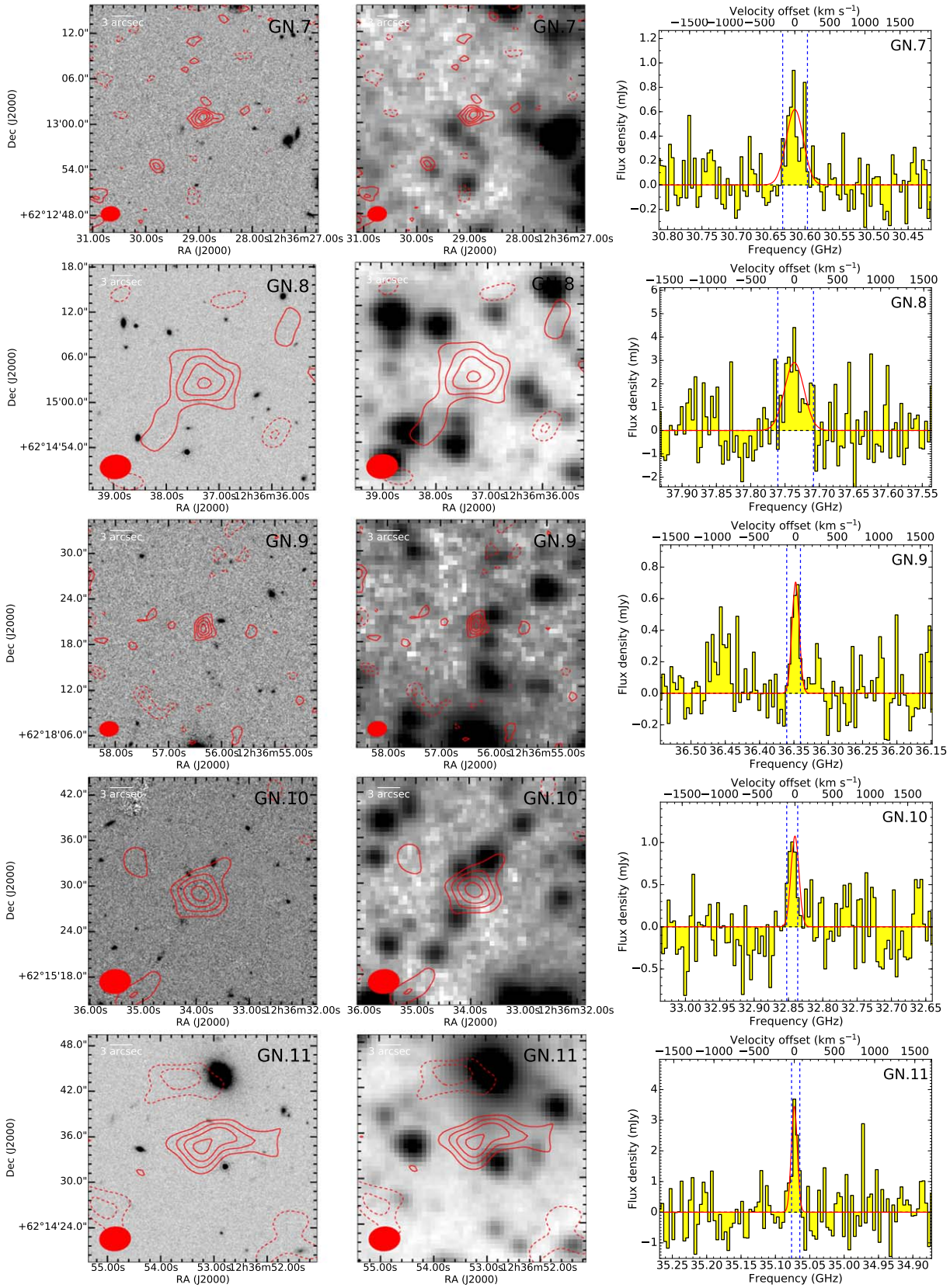

Figure 15. (Continued.)

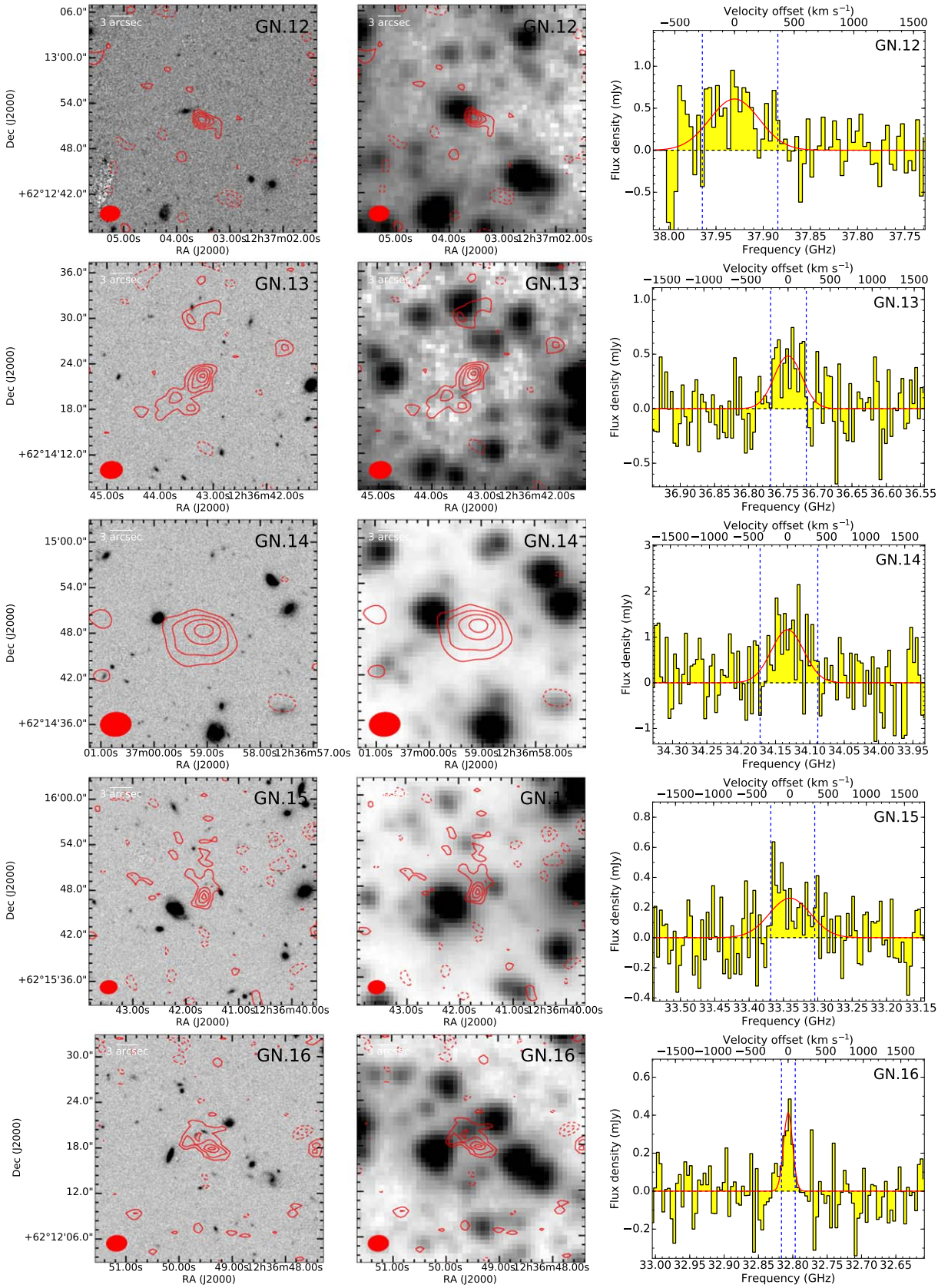

Figure 15. (Continued.)

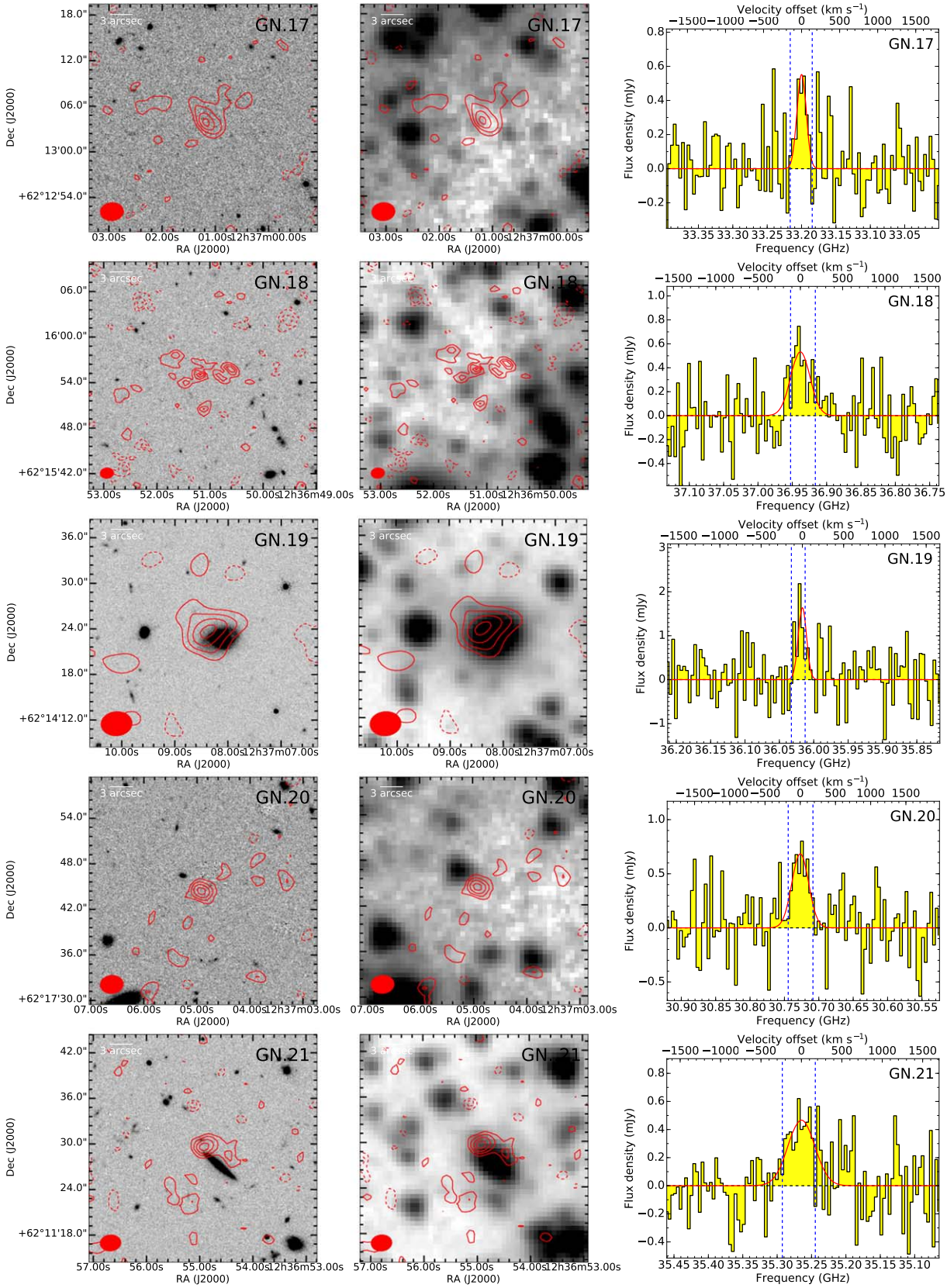

Figure 15. (Continued.)

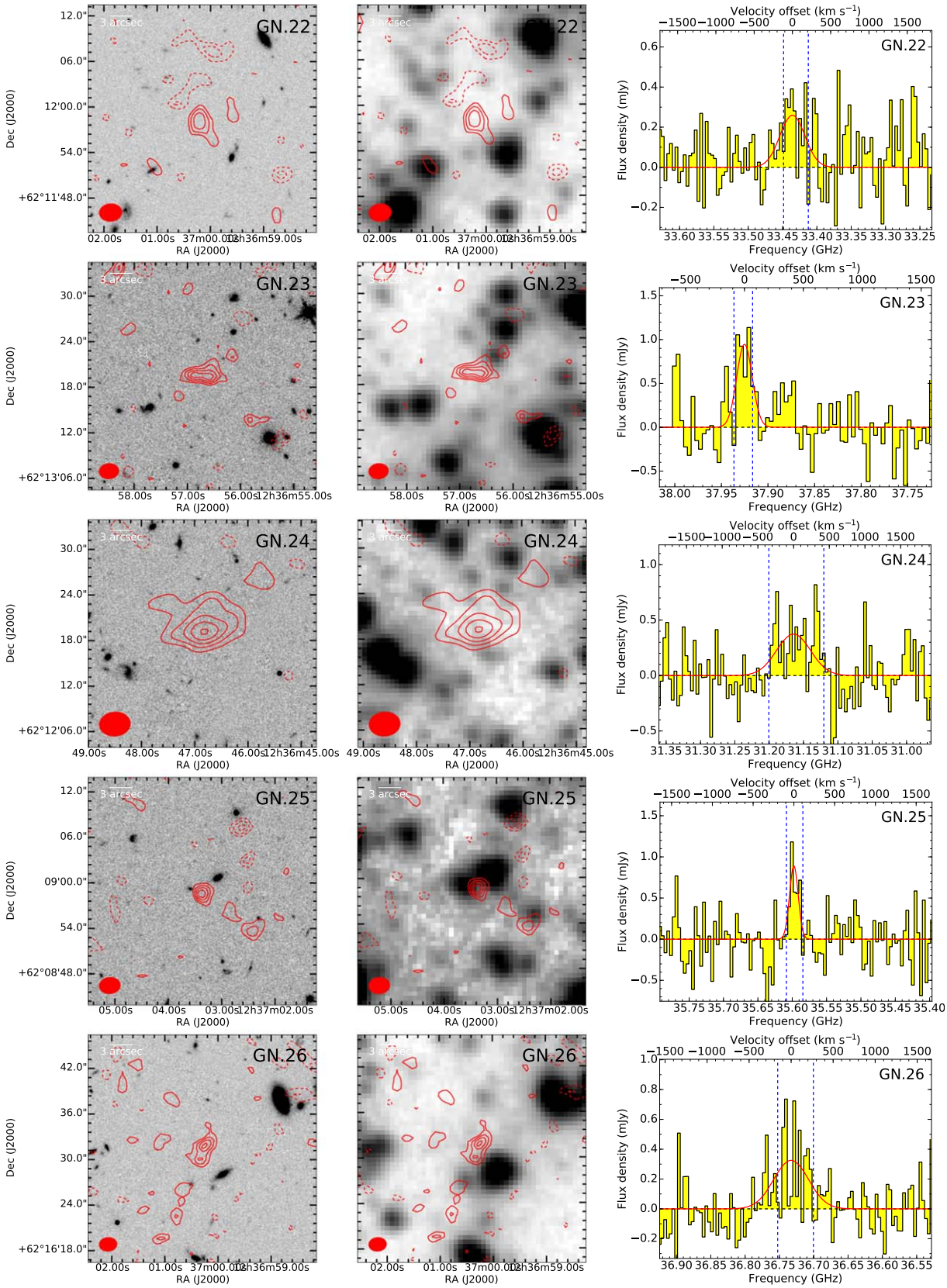

Figure 15. (Continued.)

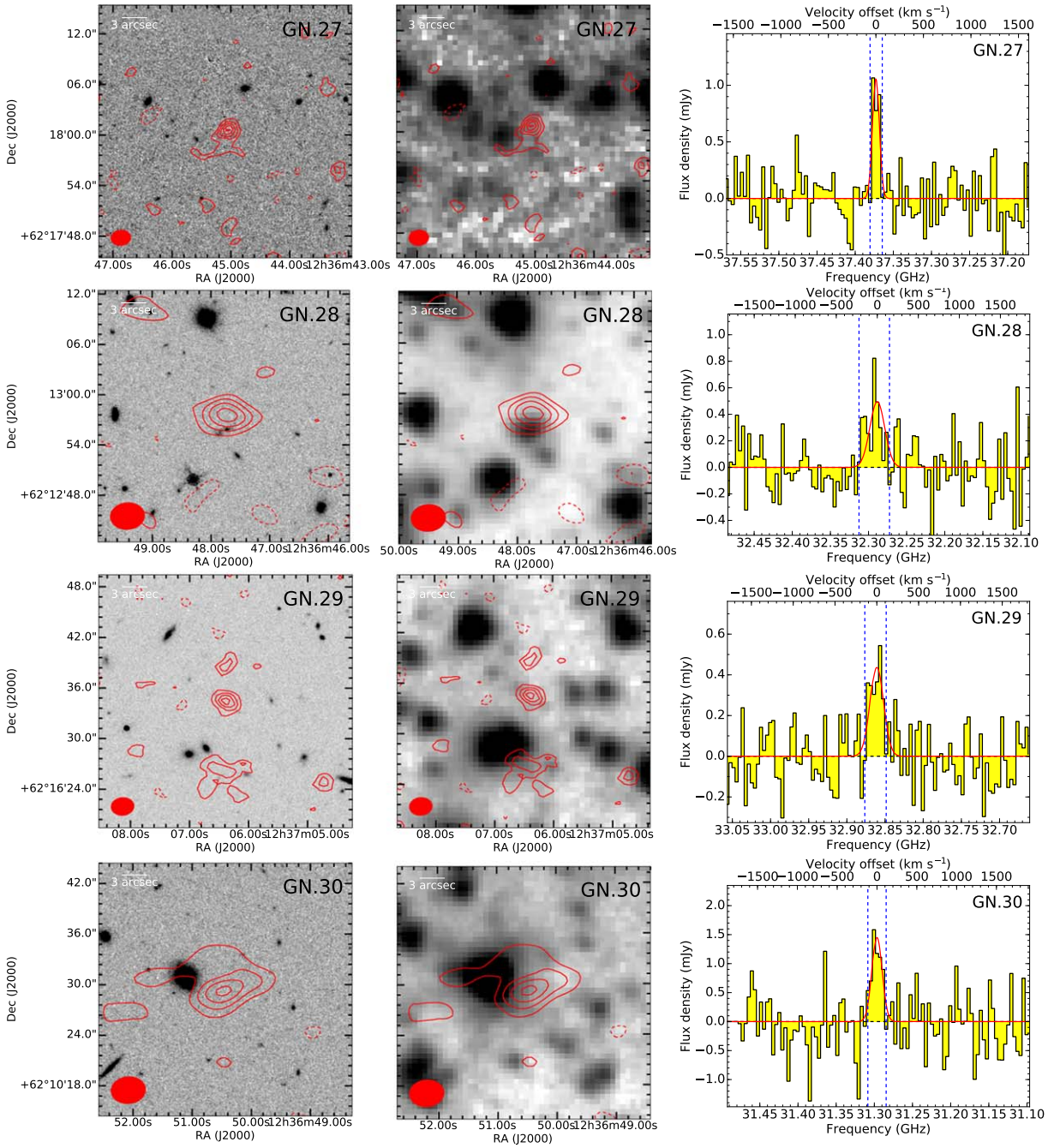

Figure 15. (Continued.)

Appendix F
Statistical Properties of the Candidate CO Emitter Sample

In order to extract as much statistical information as possible from our CO candidate list, we have to evaluate (1) the probability of each line candidate being real, (2) the line luminosity probability function, and, (3) for each luminosity bin, the completeness of our line search, i.e., the probability that a galaxy would in fact be detected by our line search, as a function of the line emission luminosity, spatial size, and velocity width.

In the following subsections, we will describe the methods we have developed to evaluate each of these separate components, which enter the luminosity function calculation (Paper II).

F.1. Reliability Analysis

The purpose of a reliability (also called purity or fidelity) analysis is to consistently assign probability estimates to each line candidate to represent a real line source. In this section, we attempt to provide a general solution to the problem of evaluating purities in the case of blind interferometric line searches that builds the foundation for our analysis.

The most accurate way to tackle this problem is to utilize the symmetry around zero of the noise distribution provided by

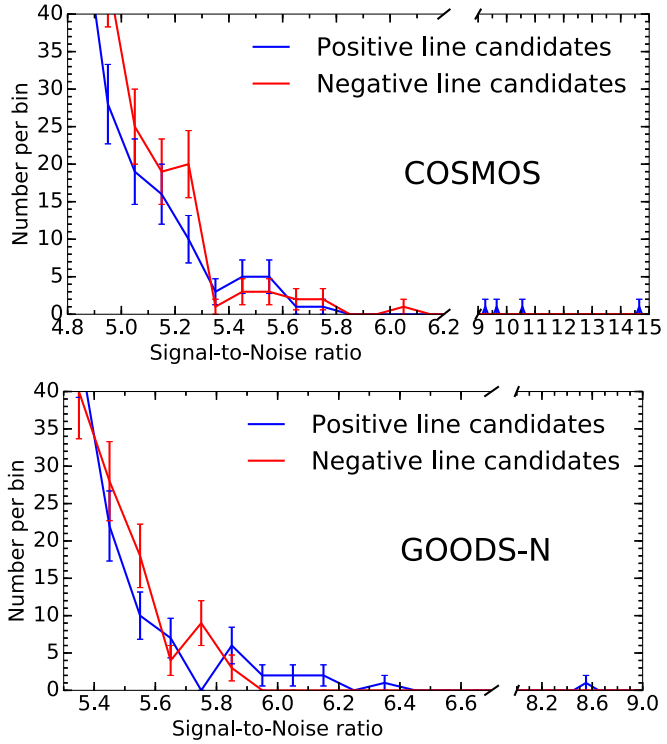

Figure 16. The S/N distributions of our line search candidates. The blue line shows the histogram of positive line features, and the red shows negative line features. Poisson errors per bin are shown. The statistical evaluation of the "excess" of positive over negative line features at a given S/N provides a measure of how many candidates may be expected to be real sources.

interferometric data. This is subject to the caveat of imperfect calibration and sidelobes of bright sources, which, however, should be negligible in our case because the continuum sources in our field are not very bright ($<0.3\sigma$ and $<2\sigma$ per 4 MHz channel for the brightest source in COSMOS and GOODS-N, respectively). An alternative approach would be to try and reproduce many instances of the noise distribution by a well-defined, simplified noise model and to evaluate the rate of false-positive detections as a function of S/N. However, this procedure may be strongly dependent on precisely capturing the statistical correlation properties of the noise (e.g., González-López et al. 2017). We therefore run an equivalent blind line search for negative line features in our data in order to estimate the contamination due to noise. We show the comparison of the distributions of S/N for positive and negative lines in Figure 16 that were used in the following to estimate the reliability for each positive line candidate. In the following, we apply Bayesian techniques to obtain estimates of the purities that are subject to well-controlled assumptions.

The basic idea is to estimate the significance of the excess of positive over negative features at a given S/N. Any excess can be considered an indication that a fraction of the positive features may correspond to real line signal. Some previous studies have taken a "cumulative" approach to this problem and used the ratio of the number of positive and negative features with S/N greater than the S/N of the line under consideration, utilizing this ratio to estimate purities (e.g., Walter et al. 2016). This may cause a substantial bias for purities that refer to individual candidates. In particular, the presence of high-S/N real candidates would raise the purity of moderate-S/N positive features. We therefore choose a "differential" approach, but we also choose not to use bins in S/N. This choice is motivated by the small number of candidates in the bins of interest, which would make the results highly dependent on the precise binning of the S/N axis. We therefore model the occurrence rate of lines as an inhomogeneous Poisson process along the S/N axis with a parameterized mean occurrence rate per unit S/N interval (see Section 14.5 of Gregory 2010 for an introduction). We can then use the machinery of Bayesian inference to study the posterior probability distribution for the rate of real sources and noise spikes and therefore infer purities for each line candidate.

In order to derive our final likelihood function, we first consider a case where we group line candidates in bins of S/N. While the result of this calculation already has wide applicability and offers certain benefits (e.g., by avoiding any parametric assumptions for the source and noise distributions), binning introduces an unnecessary dependence on bin choice and does not allow us to capture the intrinsic continuity of the source and noise rates as a function of S/N. Therefore, we will follow the standard procedure and take the limit in which the bins are small, such that each bin contains at most one detection, thereby eliminating the bias introduced by binning (e.g., Gregory & Loredo 1992). In each S/N bin, the task at hand is to determine the probability distribution for the fraction of line detections that are real sources rather than noise.

As a starting point, we infer a model for the noise distribution by fitting a Poisson process to the distribution of negative line features. Complex modeling for the noise feature occurrence rate is not necessary for estimating purities because in the moderate S/N regime of interest, the uncertainty will be dominated by shot noise due to the small number of candidate features. We therefore assume the Poisson rate (i.e., the expected number of negative lines per bin) to be well described by the tail of a Gaussian as a function of S/N centered at zero. We fit for the normalization and width of this Gaussian and thereby obtain a probabilistic description of the noise. The adopted two-parameter Gaussian tail model provides an excellent fit to the distribution of negative features. We stress that this method does not rely on the assumption of a Gaussian noise distribution but rather represents a convenient fitting function that takes advantage of the smoothness of the underlying noise distribution as a function of S/N. This method avoids using discontinuous bins or cumulative functions and allows us to exploit the symmetry between positive and negative noise features to generalize the noise realization provided by the negative features, as well as to estimate the probability of any positive line candidate to also be due to noise.

In the following, we derive purities using S/N bins. We then consider the continuum limit, as explained above. The quantity of interest is the probability of having $N_{s,i}$ real sources in the ith S/N bin, given that we observed $N_{o,i}$ lines, $p(N_{s,i}|N_{o,i}, \mu_{b,i})$. Here $\mu_{b,i}$ is the mean number of noise lines expected in the ith bin. By explicitly introducing the dependence on the real source rate (for the Poisson process), $\mu_{s,i}$, we can calculate this

probability as follows:

$$p(N_{s,i}|N_{o,i},\,\mu_{b,i}) = \int d\mu_{s,i}\, p(N_{s,i}|N_{o,i},\,\mu_{b,i},\,\mu_{s,i}) p(\mu_{s,i}|N_{o,i},\,\mu_{b,i}). \quad (8)$$

The first term, i.e., the probability of $N_{s,i}$ real sources once we assume a source rate, is the same as the product probability for $N_{s,i}$ sources given a source rate $\mu_{s,i}$ times $N_{o,i} - N_{s,i}$ noise features, given a noise rate of $\mu_{b,i}$:

$$p(N_{s,i}|N_{o,i},\,\mu_{b,i},\,\mu_{s,i}) = \frac{\mathrm{Pois}(N_{s,i},\,\mu_{s,i}) \cdot \mathrm{Pois}(N_{o,i} - N_{s,i},\,\mu_{b,i})}{\sum_{k=0}^{N_{o,i}} [\mathrm{Pois}(k,\,\mu_{s,i})\mathrm{Pois}(N_{o,i} - k,\,\mu_{b,i})]}. \quad (9)$$

Here $\mathrm{Pois}(N,\,\mu)$ stands for the Poisson probability for N events given a mean μ, and the denominator in the previous expression is a normalization factor. The second term in Equation (8) is the probability for the source rate given the observed number and noise rate, and it is therefore given by

$$p(\mu_{s,i}|N_{o,i},\,\mu_{b,i}) \propto p(N_{o,i}|\,\mu_{s,i},\,\mu_{b,i}) p(\mu_{s,i}) = \mathrm{Pois}(N_{o,i},\,\mu_{s,i} + \mu_{b,i})\, p(\mu_{s,i}) \quad (10)$$

by a straightforward application of the Bayes theorem.

We then follow the standard prescription for inhomogeneous Poisson processes, considering it as the case where the equally distributed bins are so small that each bin either contains a single line or not. In this section, we use the term rate of the Poisson process to indicate the number of line feature occurrences per unit S/N interval. In the limit of small bins containing at most one line detection, the probability for a Poisson rate μ (which can be assumed to take the form of a parametric function of S/N) given the list of detection S/Ns is calculated by the standard formula for the likelihood of an inhomogeneous Poisson process:

$$\log p(\{\mathrm{S/N}_i\}|\,\mu) = \sum_i \log \mu(\mathrm{S/N}_i) - \int_a^b d(\mathrm{S/N}')\, \mu(\mathrm{S/N}'). \quad (11)$$

Here $\{\mathrm{S/N}_i\}$ refers to the list of line detection S/Ns, the a and b integration limits reflect the range of S/N that is considered for fitting, and μ is our parametric model function for the rate of lines as a function of S/N.

In the next steps, we use the occurrence rate of background noise lines measured from the negatives by maximizing the likelihood for the noise model. A more complex approach would include the full probability distributions for the noise model parameters in the purity evaluation. We have tested this approach and confirmed that it does not affect our purity results. In particular, using MCMC samples from the probability distribution for the noise model parameters, we have evaluated the purity of one of our moderate-S/N candidates. We found that the median purity coincides with the purity evaluated with our simpler method and that the relative scatter in the purity introduced by this uncertainty on the noise model is $\lesssim$10%. This is much smaller than our conservative estimate of the systematic uncertainty, which we adopt in the following. Therefore, we maximize the probability for the complete set of negative lines (range of the integral S/N $\in [4, \infty)$) while assuming a Gaussian tail model for the rate function $\mu_b = N \exp(-\frac{S/N^2}{2\sigma_b^2})$ in order to determine the parameters N and σ_b. To determine the purity/reliability of each object, we calculate the probability that its "small bin" contains one real source and zero noise lines. Equation (9) therefore gives

$$p(N_s = 1|N_o = 1,\,\mu_b,\,\mu_s) = \frac{\mu_s}{\mu_s + \mu_b}, \quad (12)$$

and hence Equation (8) becomes

$$\mathrm{purity_k} = p(\mathrm{S/N_k\ is\ real}|\{\mathrm{S/N_i}\},\,\mu_b) = \int d\mu_s(\mathrm{S/N_k})\, \frac{\mu_s(\mathrm{S/N_k})}{\mu_s(\mathrm{S/N_k}) + \mu_b(\mathrm{S/N_k})}\, p(\mu_s|\{\mathrm{S/N_i}\},\,\mu_b). \quad (13)$$

The last term is important and represents the probability distribution for the source rate parameters (replacing Equation (10)). It can be written as the product of the probability in Equation (11) (for a rate equal to $\mu_b + \mu_s$) multiplied by priors on the source rate parameters (i.e., the last term above).

In order to compute these purities, we implement a posterior probability function for the source rate μ_s computed by Equation (11) as a function of the model parameters. We sample it using an MCMC technique, making use of the python package `emcee` (Foreman-Mackey et al. 2013). The integral in Equation (13), which corresponds to the purity of the kth detection, is equivalent to averaging the ratio

$$\frac{\mu_s(\mathrm{S/N_k})}{\mu_s(\mathrm{S/N_k}) + \mu_b(\mathrm{S/N_k})} \quad (14)$$

over these MCMC samples of source rate parameters. It may be seen as a weighted average of this ratio weighted by the posterior probability for μ_s.

The simple parameterization adopted for μ_s(S/N) is $\mu_{s0}\left(\frac{\mathrm{S/N}}{6}\right)^{-\alpha}$. Thus, we normalize the occurrence rate of real sources at S/N = 6 and allow for a shallow power-law increase of the rate toward lower S/N values, as we expect that there may be more real faint sources than bright sources. We impose uniform unconstraining priors on μ_{s0} and α. This parameterization is intended only to accurately describe the source rate over a small range of S/N, because the line candidates of dominant interest for the purity estimation are those with $5 < \mathrm{S/N} < 6.5$.

By applying this procedure, we face a choice of the S/N range to be fitted. In the COSMOS field, we start by including all the line candidates with S/N > 5. This results in purities of 100% for the top candidates (with secure counterparts) and <7% for the next objects down the list. This is caused by the large gap between S/N = 5.7 and 9 where no candidates were found and that favors a low source rate for the source distribution. Our simplified Poisson model, with slowly varying source rates as a function of S/N, may only be assumed to be an accurate description of the data over a limited range in S/N. We therefore also attempt to exclude the brightest sources and the large S/N gap without detections in the model fitting. Therefore, to obtain an upper limit on the purities, we exclude the brightest candidates and only fit the range $5 < \mathrm{S/N} < 5.8$. This yields an upper limit on the purities of up to ~10%–20% of the top few remaining objects

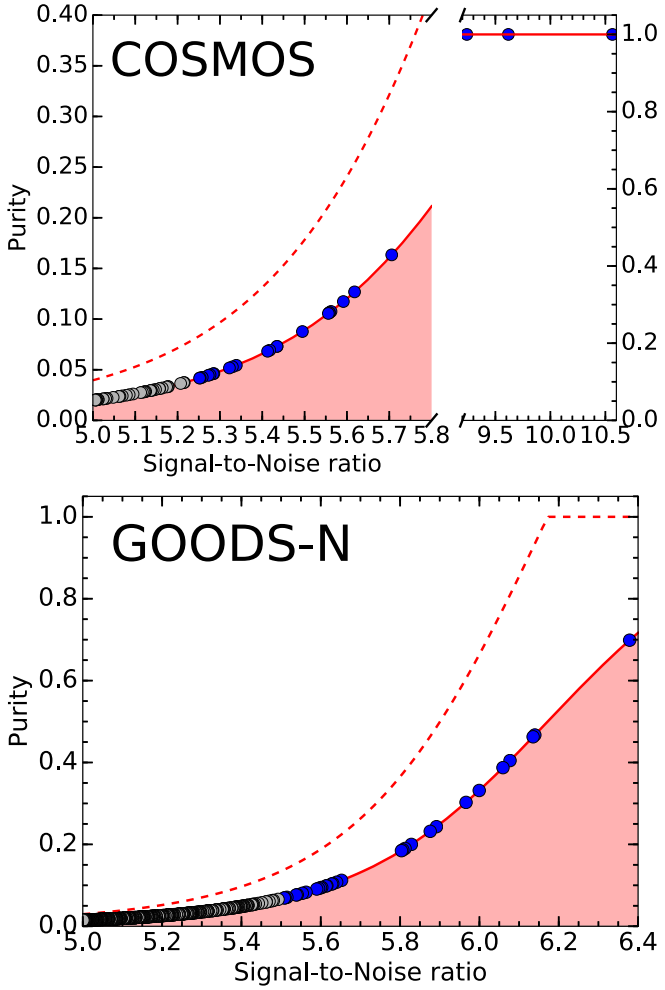

Figure 17. Estimated purity as a function of candidate S/N. The blue points indicate the higher-significance candidates reported in this work. The purity estimation utilizes an extended candidate list down to an S/N of 5. The red shaded area indicates the S/N range that we consider for calculating purities. We also show the more conservative prescription adopted in applying purity corrections, i.e., assuming a uniform distribution for the purity, treating the calculated value as an upper limit. Our alternative prescription instead uses the calculated purities as having a Gaussian uncertainty of 100% (the upper 1σ limit is indicated by the dashed line). We do not show the highest-S/N line detections corresponding to AzTEC-3 and GN10 because they correspond to CO(2–1) line emission, and they were not included in the purity estimation because of their high S/N (see the text for details).

to be real (Figure 17). In the GOODS-N field, there is no gap in the S/N distribution of the line candidates (the highest-S/N source is GN10 at $z > 5$). Therefore, we include all candidates in the range $5 < \mathrm{S/N} < 6.4$.

The procedure we have described would attribute a purity of 70% for the candidate COLDz.GN.1, 50% to GN19 (which we manually correct to be 100% because we know it to be a real line), and in the 30%–50% range for the other S/N ∼ 6 candidates, subsequently decreasing to about 7% at S/N = 5.5 (Figure 17).

When utilizing these purities to assemble the CO luminosity function, we consider two possible alternative strategies that allow us to estimate the effects of the systematic uncertainties introduced by our purity computation. In the first approach, we treat these purities as having 100% uncertainty; i.e., we will draw purities (for the Monte Carlo sampling used to estimate the allowed range of the luminosity function) as independent random numbers normally distributed around the estimated values, with standard deviation equal to the purity estimate themselves and truncating at zero. The alternative approach is to implement these purities as upper limits and to draw purities from a uniform distribution between zero and the calculated values. The latter provides a more conservative purity estimation. Therefore, the luminosity function constraints are somewhat lower in this method, although they are compatible between the two methods. This conservative approach attempts to implement the additional information coming from the lack of clear multiwavelength counterparts to our moderate-S/N candidates. We will present the detailed results of both approaches in Paper II.

The S/N thresholds adopted in Section 4 and Table 3 correspond to approximate purities of ∼4% and ∼7% for COSMOS and GOODS-N, respectively. We emphasize that previously employed definitions of purity have differed significantly. In particular, we attempt to assess the fidelity defined by Walter et al. (2016) for our candidate selection. The comparison is not straightforward because the definition of fidelity used in that work relies on the details of their line search algorithm, but an approximate implementation of their method indicates an equivalent fidelity of approximately 80%–90% for COSMOS and 50%–60% for GOODS-N in their method.

F.2. Estimating Noise Tail Extent from Data Cube Sizes

Due to the short-scale noise correlation intrinsic to interferometric noise (over the synthesized beam length scale), the calculation of the highest expected S/N due to noise (both positive and negative) is not straightforward, as the counting of "independent elements" is nontrivial. Vio & Andreani (2016) and Vio et al. (2017) independently discussed a similar analysis of this case. We have reached the same conclusions, although we take a slightly different approach, as we describe below. A detailed analysis of extreme value statistics in the case of smooth Gaussian random fields (which is a good approximation for interferometric noise) was developed by Bardeen et al. (1986) and Bond & Efstathiou (1987), among others, and expanded upon by Colombi et al. (2011). Here we only summarize the main results as relevant to our data and discuss the implications. The objective is a description of the probability distribution function for the highest S/N in a data cube that is uniquely due to noise and how this varies as a function of cube "size." If we consider the original data cube, then noise is not correlated across different channels, and a noise realization is equivalent to a 2D case with spatial correlation only and an area equivalent to the total area across the full cube (i.e., the sum of the areas over the independent channels). In this case, the approximate cumulative distribution function for the highest S/N (ν) to be expected from such a noise realization is given by

$$P(\nu_{\max} < \nu) \simeq \exp\left(-\frac{1}{4\sqrt{2\pi}} N_n \nu e^{-\nu^2/2}\right), \quad (15)$$

where N_n is the "naive" counting given by the total area divided by the "beam area" (defined by a radius equal to the beam standard deviation). The second case regards the case where correlation of the noise across channels has been introduced (for example, by convolution with a spectral template in order to match-filter) and is also relevant to line searches in the form of the noise properties of matched-filtered cubes. In this case, the correlation takes place in 3D, and a slightly different

approximate formula describes the cumulative distribution function for the highest S/N (ν) to be expected,

$$P(\nu_{\max} < \nu) \simeq \exp\left(-\frac{1}{6\pi\sqrt{2}} N_n \nu^2 e^{-\nu^2/2}\right), \quad (16)$$

where N_n is the "naive" counting given by the total cube volume divided by the "effective beam volume" (an ellipsoid with a radius equal to the beam standard deviation in the spatial dimension and the standard deviation of the template used in the spectral dimension). We have verified that the highest-significance noise peaks measured as negative features in our data are compatible with these probabilistic predictions. We note that these distribution functions are quite broad and only predict the highest S/N expected due to noise to be approximately in the S/N = 5.5–6 range for our deeper mosaic and S/N = 5.7–6.4 for our wider mosaic. This is a manifestation of the strong intrinsic stochasticity of the noise tails. We also note that the "effective number of independent elements" implied by these estimates is ∼10 and ∼20 times higher than the naive counting in the 2D and 3D cases, respectively, and that these ratios are themselves increasing functions of data-cube size. The naive counting of independent elements would therefore lead to a significant underestimation of the extent of the noise tails. This conclusion is compatible with the results of Vio & Andreani (2016) and Vio et al. (2017). However, we here report equations that explicitly describe the distribution of the maximum S/N to be expected from noise rather than implicitly through the probability distribution function of local maxima. The analysis above is only approximately equivalent to the analysis presented in Vio et al. (2017), because they express the distribution function of interest as a function of N_p, i.e., the number of local maxima in the noise realization, which is itself a random variable with its own probability distribution.

F.3. Artificial Source Analysis

In order to estimate the completeness and biases introduced by our line search and flux extraction methods, we perform an extensive probabilistic analysis of artificially injected sources into our maps. The main goal of this analysis is to establish a probabilistic connection between recovered candidate properties and intrinsic properties such as spatial size, velocity width, and line flux. This will provide some control over the uncertainties that affect the analysis of the CO luminosity function (Paper II). We also develop a method to correct the luminosity function by the completeness of our line search, which avoids a purely "per-source" completeness estimation as far as possible (due to the bias of "per-source" corrections) while avoiding assumptions that would significantly affect the result.

Since the large majority of the data cube contains very little signal, we use the data themselves as our model for the noise and inject artificial sources of varying size, velocity width, and flux at random positions in the data cube (Table 8). We inject sources in each cube (500 in COSMOS and 2500 in GOODS-N), estimating that this will not cause crowding of the field—therefore not causing overlaps between different sources—during the line search and effectively simulating the recovery of each injected source individually. We then analyze each injected cube following the same steps of MF3D that we applied to the real data, and in the end, we search for the injected sources to determine the recovered S/N and the line parameters that would have been measured. We define the "flux-factor" as the ratio of the measured line flux to the injected flux. Therefore, the distribution of flux-factors captures both flux corrections and uncertainties on our flux estimations. The purpose of the flux-factor analysis is not just to correct for potential biases in our flux extraction procedure but also to estimate the uncertainty of the flux recovery. We subsequently utilize these flux probability distributions to inform our luminosity function estimates (Paper II). We ignore dependencies on frequency or position of the injected source flux-factors (determined as a function of local S/N) and completeness, thereby obtaining average values that correctly sample the data for an approximately uniform distribution of real sources in our cube.

Table 8
Artificial Source Injected Sizes

	COSMOS	GOODS-N
Intrinsic spatial size (arcsec)	∼0.5, ∼3.0, ∼4.7	1.0, 2.5, 4.5
Convolved spatial size (arcsec)	2.6, 4.0, 5.3	2–3, 3–4, 4.8–5.4
Frequency width (MHz)	23.2, 46.8, 70	23.2, 46.8, 70
Velocity width (km s^{-1})	∼200, ∼400, ∼600	∼200, ∼400, ∼600

Note. Gaussian sizes are utilized for the injected artificial sources. All sizes refer to the Gaussian FWHM. The convolved sizes are the injected sizes in the Natural-mosaic. These are fixed in COSMOS, while in GOODS-N, because of the larger beam differences across the mosaic, we injected sources of fixed intrinsic sizes and convolved them to the local beam size appropriate for each mosaic position.

Both the completeness and the flux-factors are dependent on the source size and line FWHM. Since we only inject sources of three spatial sizes and three frequency widths, we develop a probabilistic framework to relate each line candidate to the different injected sizes (Table 8). For each detected candidate, based on the template size and velocity width where their S/N peaks, we determine a probability distribution of belonging to each category of "injected" spatial size and line width, therefore matching in a continuous and probabilistic way the measured sizes to a discrete grid of intrinsic properties, as explained in detail in the following.

F.3.1. Flux-factors Based on Artificial Sources

The artificial source analysis allows us to estimate how well our measured fluxes correspond to the injected flux for candidates of different S/N, spatial size, and velocity width. The objective of the flux-factor analysis is to characterize the uncertainty and bias of our flux estimates in order to correctly estimate the uncertainty of our luminosity function measurement.

We confirm that aperture fluxes have a slight bias toward higher fluxes because positive noise adjacent to a candidate source tends to enlarge the fitted sizes and therefore contribute spurious flux to the candidate (Condon 1997). We thus need to estimate the magnitude of this bias at the S/N range of interest (∼5–6) to correct the measured fluxes accordingly. A correction factor relies on an estimate of how likely a measured extended source is to be due to noise rather than real extended structure. In order to determine this bias, we need to assume an expected approximate size distribution for our sources to be combined with information from the artificial sources regarding how the flux is affected by the interplay of real and measured sizes.

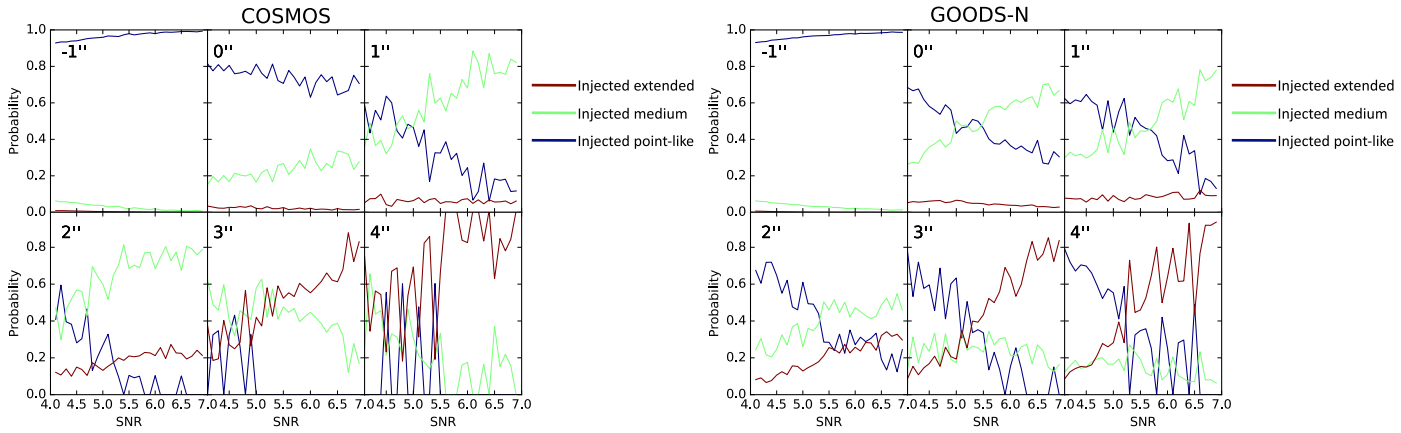

Figure 18. Probability of injected spatial size (proxy for "real" size) as a function of S/N for different measured sizes (injected sizes are color-coded: smallest in blue, intermediate in green, significantly extended in red; sizes of the artificial sources listed in Table 8). Measured sizes are indicated by the spatial size of the peak template (see Table 5). Left: results for the COSMOS field. Right: results for the GOODS-N field. A measured (peak template) size of $-1''$ corresponds to a source that achieves its peak S/N in the Natural-mosaic, i.e., before any smoothing, at the native resolution, while the $0''$ template is to a point-source template applied to the Smoothed-mosaic.

We use the artificial sources to determine the probabilities of spatial extension (probability of being like spatial bin 0, 1, or 2 of the injected sources; Table 8) given the measured size as traced by the size of the spatial template that gives the highest S/N. This probability is used in the following to relate the measured properties of each line candidate to the flux-factor and completeness, which are computed for the bins of injected properties. We use the Bayes theorem to relate the probability of a given real size conditional to a measured size: p(real-injected size|measured size) to the probability distribution that we can measure from the artificial sources, which is the probability of measuring a given size conditional to a certain injected size p(measured size|injected size), by employing a prior on the expected real size distribution.[33] We must employ a prior for the probabilities of the real sizes that captures our expectation that most sources would be unresolved, while allowing for a fraction of resolved sources coming from extended gas reservoirs and merging and/or blended objects. We adopt 88%, 10%, and 2% for the size bins in Table 8, respectively. We stress that although these relative fractions are uncertain, their precise choice does not significantly affect any of the results, because their effect is simply to modulate our assignment between line candidates and injected sources. The main effect of this assignment is in the estimation of completeness corrections, where the uncertainty introduced by these priors is small compared to the systematic uncertainty introduced by choosing a weighting based on the detected size distribution.

We estimate p(measured size|injected size) by measuring the fraction of injected sources of a given size recovered at different match-filter spatial template sizes. In this way, we compute the final posterior probability for a given measured size to originate from an unknown "injected" size by combining this with the prior (Figure 18):

$$p(\text{real/injected size}|\text{measured size, S/N}) \propto p(\text{real size}) \times p(\text{measured size}|\text{injected size, S/N}). \quad (17)$$

We follow the same procedure for relating the measured velocity width (from the peak template) to the injected line widths by assuming a flat prior for the line width over the three injected bins (Table 8). These are required to compute the completeness of line candidates by relating their measured properties to the injected sources. The derived probability distribution functions are shown in Figure 19.

In order to calculate flux-factors from the artificial sources, we employ an analogous technique. We calculate probability distributions of the flux-factors, given source spatial template size and S/N, by weighing the distribution of flux-factors found for given measured sizes by the probability that the given measured size originates from the different possible injected sizes. Specifically, the correction ratio depends on both the injected and the measured size, as a larger ratio is needed to correct for a compact source that appears extended:[34]

$$p(\text{flux ratio}|\text{measured size, S/N}) = \sum_{\text{injected}} p(\text{flux ratio}|\text{measured, injected, S/N}) \times p(\text{injected}|\text{measured, S/N}). \quad (18)$$

These flux-factor distributions (Figure 20) can be approximated by lognormal distributions peaking near a factor of one but with a tail to larger ratios to correct for the bias toward larger spatial size, which is introduced by including positive noise as part of the candidate source. We stress that this is just a convenient parameterization of the measured distributions. We calculate the probability distribution of flux ratios for each measured spatial size in bins of S/N. Since the distributions are noisy due to the limited number of artificial sources and do not strongly depend on S/N in the narrow 5–6 range of interest for our candidates, we consider the mean distribution over the $5 < \text{S/N} < 6$ range (Figure 20). The fitted lognormal curves to these mean distributions, which provide a good interpolation to the noisy distribution estimates, will be utilized in the construction of the luminosity function. We can understand the general trend seen in these shapes as follows. The smallest template selects point sources. Therefore, the distribution peaks near 1. Slightly extended sources have a larger mean flux correction, reflecting the finding that they are most likely noise-smeared point sources, and so their flux needs to be reduced.

[33] Note that the artificial sources can only be used to estimate distributions conditional to a given injected size, because the relative frequency of injected sizes that was utilized (uniform) is not representative of the expected distribution of real sizes.

[34] The only distribution that can be directly estimated through counting artificial sources is conditional to injected size. We cannot marginalize over those sizes without specifying a distribution of expected source sizes first.

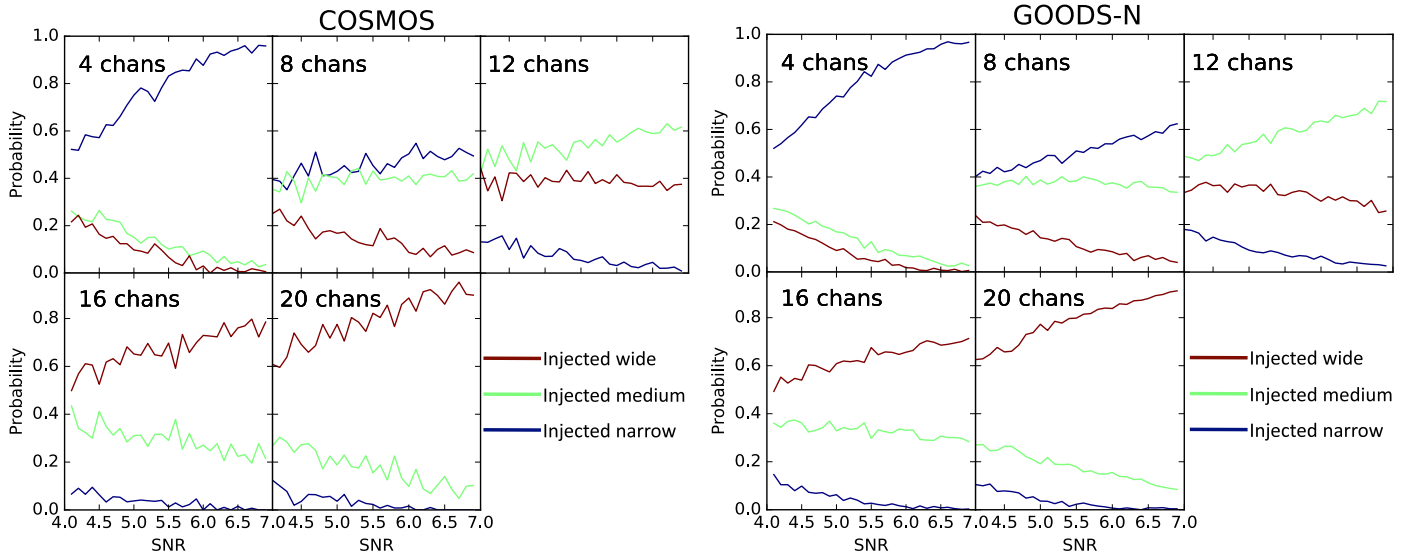

Figure 19. Probability of injected velocity width (proxy for "real" width) as a function of S/N for different measured widths (injected sizes are color-coded: narrowest in blue, intermediate in green, wide in red; FWHMs of the artificial sources are listed in Table 8) in the COSMOS (left) and GOODS-N (right) fields. The measured FWHMs are indicated by the frequency size of the peak template (see Table 5).

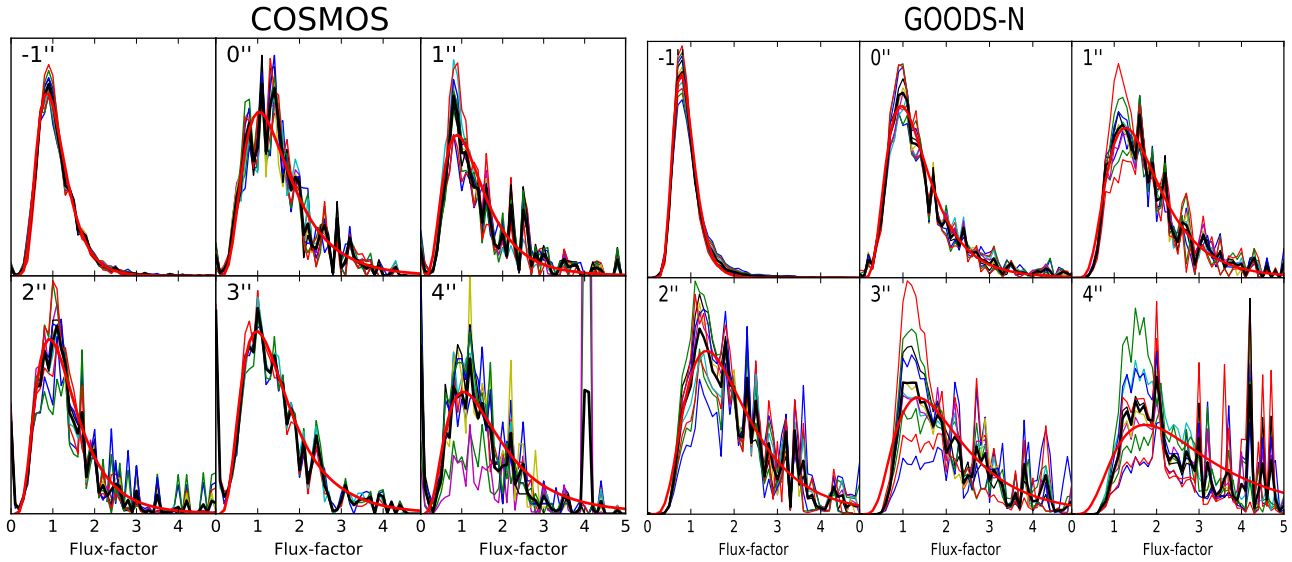

Figure 20. Probability distribution functions of the "flux-factor" (i.e., the ratio of measured-to-injected line flux) for the range $5 < S/N < 6$ (in steps of 0.1σ) conditional to the measured spatial size, as indicated by the spatial size of the peak template. We also show (black lines) the mean distribution taken over the full S/N range to obtain a more representative estimate. We fit this probability distribution by a lognormal distribution and show the best-fit model in red. On the left, we show the results for the COSMOS field, and on the right is the GOODS-N field. A measured (peak template) size of $-1''$ corresponds to a source that achieves its peak S/N in the Natural-mosaic, i.e., before any smoothing, at the native resolution, while the $0''$ template is to a point-source template applied to the Smoothed-mosaic.

Slightly larger (intermediate-size) sources then require less correction because it becomes more likely that they are somewhat extended in reality. Even larger sources show a long tail of larger flux-factors because it is extremely unlikely that the real source is very extended. Therefore, their fluxes need to be significantly corrected (or rather, there is significant uncertainty as to their real flux, and we need to account for this in constructing the luminosity function).

F.3.2. Completeness

In order to estimate the completeness of our detection process, we utilize the artificial sources to measure the fraction of the injected sources that are detected. The objective of the completeness correction is to account for the fraction of the mosaic volume where a given line candidate would be detectable and for the fraction of objects of a given intrinsic line luminosity that would be missed by a fixed S/N threshold.

We assume that the fraction of detected lines (a proxy for the probability of detection) only depends on the integrated line flux, spatial size, and velocity width. By injecting artificial sources that uniformly sample random positions within the edges of the mosaic, we derive completeness corrections that account for the effects of the spatial and frequency variation of the sensitivity, as previously adopted by Walter et al. (2016).

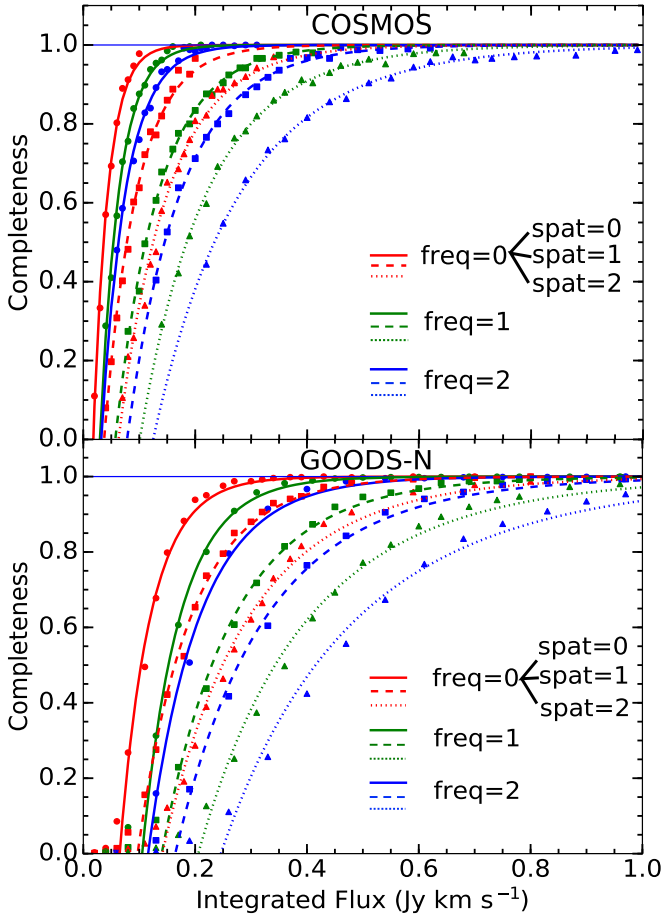

Figure 21. Completeness corrections calculated as a function of the injected flux of the artificial sources. Top: completeness in the COSMOS field. Bottom: completeness in the shallower GOODS-N field. Colors distinguish the three different velocity widths of the artificial sources, and the line (marker) style distinguishes different spatial sizes. Markers represent the measured completeness in bins of line flux, and lines represent the best-fitting interpolating functions.

While completeness is a property of the overall number counts in a luminosity bin, we partially adopt the spirit of the $1/V_{\rm max}$ method of calculating a "per-source" completeness only insofar as this depends on the line velocity width and especially spatial size of the detection. While this may potentially introduce a bias[35] (see, e.g., Hogg et al. 2010), it saves us from additional assumptions about the size distribution. We therefore point out the important caveat that our luminosity function estimate does not correct for missed objects due to poor sampling of the size distribution. The effect can be seen by noticing, for example, that the completeness for extended objects at low flux values drops very quickly. This is reflected in the fact that all of our low-flux objects are point sources (therefore, the completeness correction in the lowest line luminosity bins misses the potential contribution from undetected extended sources).

The completeness is measured from the artificial sources as a function of "injected" properties, i.e., injected integrated flux, and spatial size and frequency width (in three bins each; see Table 8) as the ratio of injected sources recovered with S/N above a threshold value of 5σ to the total number of injected sources (the precise choice of a threshold does not change the result appreciably). These measured completeness values are shown in Figure 21, together with the interpolating functions that we use in deriving the luminosity function: the two-parameter (I_0 and I_1) family of functions $1 - e^{-I/I_0}/(I + I_1)$, where I is the integrated line flux. While this chosen family of interpolating functions has no specific significance, we found it to provide an appropriate description of the measured completeness.

The optimal way to correct for the completeness of a luminosity function bin would require calculating the mean completeness within the bin over the full "internal" parameter space (in our case, these are spatial and velocity line sizes, frequency, and precise luminosity within the bin) weighted by the model expectation for the distribution of sources within this space. We adopt an intermediate approach between this "mean completeness" approach and a purely "per-candidate" approach. We assume that the distribution of line luminosity within a bin and the frequency distribution are uniform and average over this subset of the internal space by randomly sampling it. On the other hand, we do not assume an intrinsic spatial size and velocity width distribution in order to avoid biasing our result. We therefore adopt the sizes and line widths of the candidates to calculate the appropriate completeness, thereby letting the data determine the size and velocity width distributions.

In summary, for each detected line candidate, the size properties (i.e., the spatial size and frequency width) of the "real" underlying source are estimated probabilistically based on the measured size and width (from the peak template). Then, probability-weighted completenesses are the factors that enter the evaluation of the luminosity function. The dependences of these completenesses on the line flux are mean values evaluated for the full luminosity function bin, rather than depending on the candidate line flux measurement. This hybrid approach helps us mitigate the bias that would derive from a purely "per-source" correction (e.g., Hogg et al. 2010).

F.4. Implementation of the Statistical Corrections

In order to assemble the luminosity function, we use a variation on the method used by Decarli et al. (2014, 2016). We weigh the contribution of each line candidate to the luminosity bin by its purity and inversely by its completeness and use the total cosmic comoving volume covered within the edges of the mosaic. The completeness correction converts this volume to an effective $V_{\rm max}$ for each galaxy, also accounting for the spatial variation in sensitivity. In order to estimate the systematic uncertainty introduced by our assumptions, we evaluate the luminosity function with many random realizations of flux- factor, purity assignment, and luminosity bin widths and boundaries. One of the differences between our approach and the approach employed by Decarli et al. (2014, 2016) consists of including a larger number of candidates. We have also calculated the luminosity function using the same method employed by Decarli et al. (2016), and the result is consistent with our more extensive method. The advantage of our approach consists of relying less heavily on the properties of the few moderate-S/N individual candidates that happen to be located near

[35] Regions of parameter space that have very low completeness tend to be poorly sampled and hence cannot properly be accounted for in a completeness calculation that is weighted by the detected candidates.

the top of the S/N list but that still have a limited probability of being real. Although a fraction of the moderate-S/N candidates are expected to be real, it is not clear that those near the top (of the set of uncertain candidates) of the S/N list have a significantly greater likelihood of being real given the limited range in S/N considered (∼5–6). By utilizing a larger sample of candidates in deriving constraints on the luminosity function down-weighted by appropriate purities, we do not introduce additional bias but rather better explore the implications of the systematic uncertainties. In particular, the statistical justification for a "per-source" purity and completeness correction (which we cannot fully avoid) only holds for large enough samples. By better sampling the "internal" space of possible candidate sizes and line widths and adopting average per-bin completenesses (i.e., not using the uncertain measured fluxes), we aim to achieve a more accurate completeness correction and evaluation of the systematic uncertainties introduced by these factors.

In detail, for each $L'_{\rm CO}$ bin, we calculate a completeness factor appropriate for each of the nine bins in the spatial size–frequency width grid by averaging over 1000 random realizations of values of $L'_{\rm CO}$ in each bin, using random redshifts to calculate the corresponding integrated flux and hence the appropriate completeness correction for each. We average over this frequency distribution and precise $L'_{\rm CO}$ within each bin. We therefore enforce a uniform prior and maintain the dependence on the spatial-frequency size information separate. We subsequently apply them for each line candidate as weighted by the candidate probability distribution for its spatial-frequency size assignment. In this way, we use the measured relative occurrence of different spatial and velocity sizes as weights for the appropriate completeness for each luminosity function bin.

In order to explore the range of luminosity function values allowed by our systematic uncertainty, we use 10,000 Monte Carlo realizations of the luminosity function calculation for each bin width and shift, where we vary the purity assignment independently for each candidate and the flux-factor to be applied. We therefore "move candidates around" among adjacent luminosity bins, simulating the effect of the uncertainty in their intrinsic fluxes. We separately implement the purity in one of the two ways we described in Appendix F.1, either as normally or uniformly distributed. We also implement the flux correction as taking a random value drawn from the appropriate lognormal distribution, which was derived for each spatial size in the previous section. We also add a normal uncertainty of 20% to the measured flux to reproduce the uncertainty in our flux calibration.

Finally, in order to describe the range of values for the luminosity function (in log comoving volume density space) spanned by our 10,000 Monte Carlo realizations, we calculate the median value for each bin and a measure of the scatter around the median. We evaluate the scatter conservatively by quoting luminosity function ranges that include 90% of the probability. We also evaluate the statistical Poisson uncertainty as appropriate for each bin as a relative uncertainty of $1/\sqrt{N}$, where N corresponds to the number of candidates in an $L'_{\rm CO}$ bin.

Appendix G
Additional Stacking Results of Galaxies with Optical Redshifts

In addition to the stacking of individual sets of galaxies described in the main text, we have attempted to stack three more sets of galaxies based on optical redshift information. We do not detect significant signal in these stacked spectra.

G.1. COSMOS Protocluster at z = 5.3

We search for CO(2–1) from nine member galaxies of the AzTEC-3 protocluster, as identified through the Lyman Break (LBG) technique and color selection by Capak et al. (2011). While two of them show a hint of a CO emission line signal (∼2.5σ), all others (including LBG-1, the one with the strongest [C II] emission; Riechers et al. 2014 and 2018 Capak et al. 2015) are consistent with noise. The positions of the LBGs with tentative CO detections are J2000 10:00:21.96+02:36:08.5 and 10:00:20.13+02:35:53.9. Assuming a line FWHM of 250 km s^{-1} (i.e., the width of the [C II] line in LBG-1), we derive CO line fluxes of approximately 0.03 ± 0.012 Jy km s^{-1}, corresponding to $L'_{\rm CO21} = 7 \pm 3 \times 10^9$ K km s^{-1} pc^2. We do not claim any detections due to the low significance and only quote these as approximate limits for reference. We also stack all spectra at the positions of LBGs (limited to the seven galaxies that are not contaminated by emission from the bright CO(2–1) line in AzTEC-3 at the resolution of our survey) and do not detect a significant signal (Figure 22). We therefore place a 3σ limit of <0.012 Jy km s^{-1} on their average CO(2–1) emission, corresponding to $L'_{\rm CO21} < 3 \times 10^9$ K km s^{-1} pc^2. The average stellar mass for the stacked LBGs, as reported by the COSMOS2015 catalog, is $4 \times 10^9\,M_\odot$. If we assume an $\alpha_{\rm CO} = 3.6\,M_\odot$ (K km s^{-1} pc^2)$^{-1}$, we obtain a gas mass upper limit of $<1.1 \times 10^{10}\,M_\odot$ and therefore a gas fraction of $M_{\rm gas}/M_* < 3$, which is not strongly constraining. None of the LBGs are detected in 3 GHz radio continuum emission (Smolčić et al. 2017). This would place a limit of $<165\,M_\odot$ yr^{-1} on their star formation rate if we adopted the redshift evolution of the radio–FIR correlation measured by Delhaize et al. (2017). This may be converted to an expected limit of $L'_{\rm CO} < 2.2 \times 10^{10}$ K km s^{-1} pc^2 by assuming the star formation law (Daddi et al. 2010b; Genzel et al. 2010). This limit is higher than what we derive from our CO nondetection, implying that our deep observations provide strong constraints on the CO luminosity of $z > 5$ LBGs.

G.2. Grism Redshifts in the GOODS-N Field

The 3D-*HST* catalog (Momcheva et al. 2016) contains 694 galaxies in GOODS-N with grism redshifts for which the CO(1–0) line is covered by our data. Nevertheless, the majority of these grism spectra do not significantly improve the redshift determination over the photometric redshift and are therefore not usable for stacking. We search for matches to our line search candidates (down to 4σ) within a radius of 2″ and ∼500 km s^{-1}. We find 16 potential matches and assess the contamination by chance association by also matching our blind detection catalog to 694 random positions within the signal data cube. We find that the distribution due to random associations is well described by a Gaussian with a mean 8 associations and a standard deviation of 3.5 associations. Therefore, the majority of our line associations are likely to be random, but some may be expected to be real. The measured

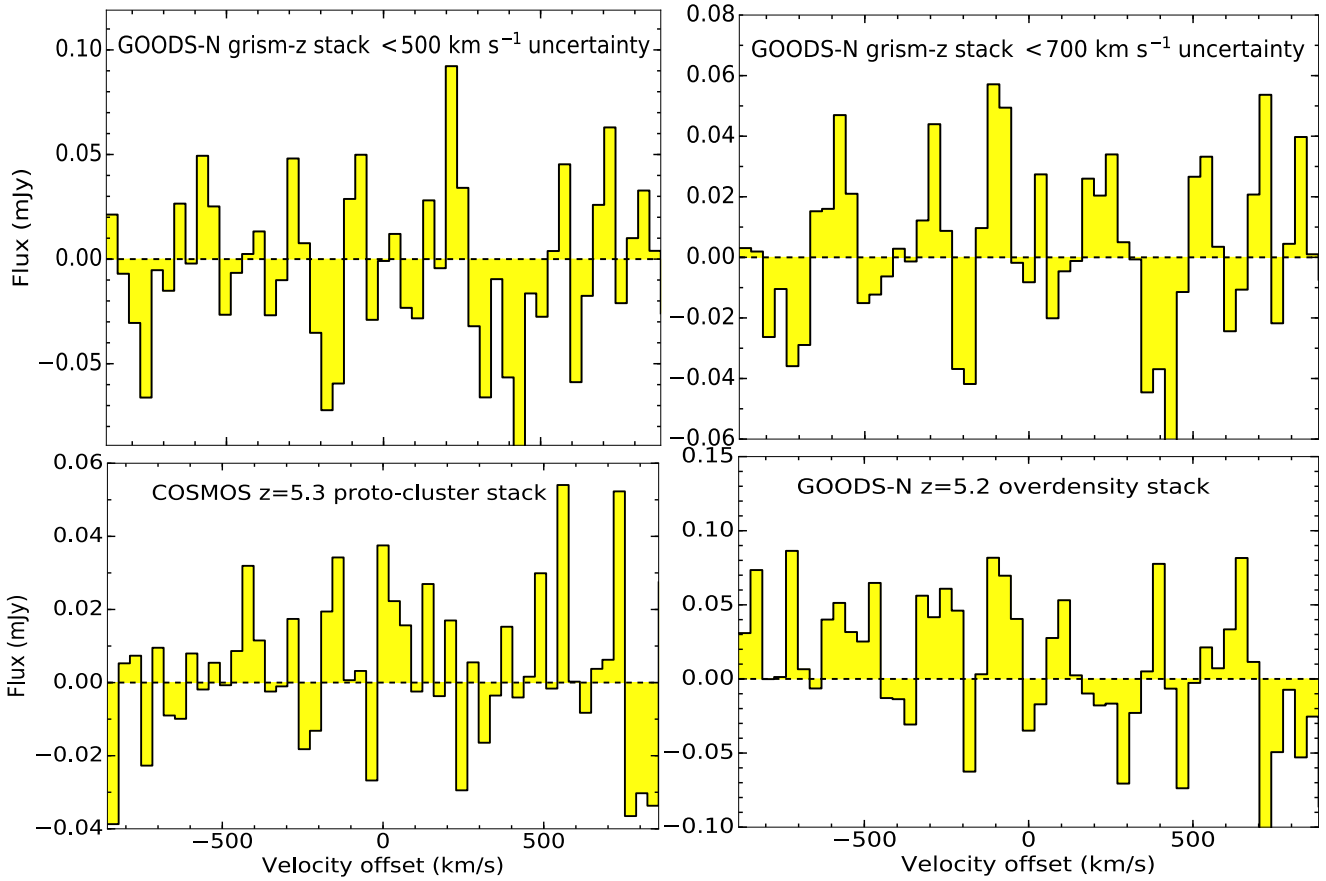

Figure 22. Additional spectral stacks of sets of galaxies with spectroscopic or grism redshifts. Top: CO(1–0) stacks of subsets of galaxies with the highest-quality grism spectra with redshift uncertainty <500 (left) and <700 km s^{-1} (right). Bottom: CO(2–1) stacks for samples of potential $z > 5$ galaxies belonging to previously identified overdensities in our fields, i.e., the AzTEC-3 protocluster at $z = 5.3$ (left) and the overdensity around HDF850.1 (right; Capak et al. 2011; Walter et al. 2012). The spectral resolution is the same as in Figure 5.

line fluxes of these candidate counterparts would imply gas masses that are sometimes larger than the stellar masses. This is possible, but we consider it more likely that those cases may be random noise associations.

We also stack the spectra extracted at the positions of the galaxies with high-quality grism redshifts (Figure 22). Thirty-seven of them have redshift uncertainties less than 500 km s^{-1} (based on the 95th percentile of the redshift probability distribution reported by Momcheva et al. 2016), and 35 more have redshift uncertainties less than 700 km s^{-1}. At this level of uncertainty, it would be likely that a fraction of the line signal contributes to the stacked spectrum. Both of these stacks show no detection. Assuming a line FWHM of 300 km s^{-1}, this implies 3σ upper limits of <0.011 and <0.008 Jy km s^{-1}, corresponding to $\lesssim$3 and 2×10^9 K km s^{-1} pc^2, respectively.

G.3. HDF850.1 $z = 5.2$ Galaxy Overdensity in GOODS-N

We also search for potential CO(2–1) emission from galaxies in the $z \sim 5.2$ overdensity around the SMG HDF850.1 (Walter et al. 2012), taking advantage of the abundance of available spectroscopic redshifts in this region. Twenty-four of the 105 galaxies with spectroscopic redshifts presented in that work fall within our data, but none of them are individually detected. We stack the spectra to obtain an average spectrum (Figure 22). No significant emission is found after stacking. Assuming a line FWHM of 300 km s^{-1} implies a 3σ upper limit of <0.015 Jy km s^{-1}, and $L'_{\rm CO21} < 3.5 \times 10^9$ K km s^{-1} pc^2 at $z \sim 5.2$. We match these 105 galaxies to galaxies within 1″ in the photometric catalog by Skelton et al. (2014), finding 83 matches, but we do not adopt their stellar mass estimates because the redshifts of these galaxies were often greatly underestimated.

ORCID iDs

Riccardo Pavesi https://orcid.org/0000-0002-2263-646X
Chelsea E. Sharon https://orcid.org/0000-0002-6250-5608
Jacqueline A. Hodge https://orcid.org/0000-0001-6586-8845
Roberto Decarli https://orcid.org/0000-0002-2662-8803
Fabian Walter https://orcid.org/0000-0003-4793-7880
Chris L. Carilli https://orcid.org/0000-0001-6647-3861
Emanuele Daddi https://orcid.org/0000-0002-3331-9590
Ian Smail https://orcid.org/0000-0003-3037-257X
Rob J. Ivison https://orcid.org/0000-0001-5118-1313
Mark Sargent https://orcid.org/0000-0003-1033-9684
Elisabete da Cunha https://orcid.org/0000-0001-9759-4797
Manuel Aravena https://orcid.org/0000-0002-6290-3198
Jeremy Darling https://orcid.org/0000-0003-2511-2060
Vernesa Smolčić https://orcid.org/0000-0002-3893-8614
Nicholas Z. Scoville https://orcid.org/0000-0002-0438-3323

References

Aravena, M., Carilli, C., Daddi, E., et al. 2010, ApJ, 718, 177
Aravena, M., Carilli, C. L., Salvato, M., et al. 2012, MNRAS, 426, 258
Aravena, M., Hodge, J. A., Wagg, J., et al. 2014, MNRAS, 442, 558
Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, ApJ, 769, 80
Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15
Berta, S., Lutz, D., Genzel, R., Förster-Schreiber, N. M., & Tacconi, L. J. 2016, A&A, 587, A73
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Blain, A. W., Smail, I., Ivison, R. J., Kneib, J.-P., & Frayer, D. T. 2002, PhR, 369, 111
Bolatto, A. D., Warren, S. R., Leroy, A. K., et al. 2015, ApJ, 809, 175
Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207
Bond, J. R., & Efstathiou, G. 1987, MNRAS, 226, 655
Bothwell, M. S., Smail, I., Chapman, S. C., et al. 2013, MNRAS, 429, 3047
Bourne, N., Dunne, L., Bendo, G. J., et al. 2013, MNRAS, 436, 479
Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, ApJS, 200, 13
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Capak, P. L., Carilli, C., Jones, G., et al. 2015, Natur, 522, 455
Capak, P. L., Riechers, D., Scoville, N. Z., et al. 2011, Natur, 470, 233
Carilli, C. L., Chluba, J., Decarli, R., et al. 2016, ApJ, 833, 73
Carilli, C. L., McKinnon, M., Ott, J., et al. 2015, arXiv:1510.06438
Carilli, C. L., & Walter, F. 2013, ARA&A, 51, 105
Casey, C. M., Hodge, J. A., Lacy, M., et al. 2015, arXiv:1510.06411
Casey, C. M., Narayanan, D., & Cooray, A. 2014, PhR, 541, 45
Chen, C.-C., Hodge, J. A., Smail, I., et al. 2017, ApJ, 846, 108
Colombi, S., Davis, O., Devriendt, J., Prunet, S., & Silk, J. 2011, MNRAS, 414, 2436
Condon, J. J. 1997, PASP, 109, 166
Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
da Cunha, E., Groves, B., Walter, F., et al. 2013a, ApJ, 766, 13
da Cunha, E., Walter, F., Decarli, R., et al. 2013b, ApJ, 765, 9
da Cunha, E., Walter, F., Smail, I. R., et al. 2015, ApJ, 806, 110
Daddi, E., Bournaud, F., Walter, F., et al. 2010a, ApJ, 713, 686
Daddi, E., Cimatti, A., Renzini, A., et al. 2004, ApJ, 617, 746
Daddi, E., Dannerbauer, H., Elbaz, D., et al. 2008, ApJL, 673, L21
Daddi, E., Dannerbauer, H., Krips, M., et al. 2009, ApJL, 695, L176
Daddi, E., Dannerbauer, H., Liu, D., et al. 2015, A&A, 577, A46
Daddi, E., Elbaz, D., Walter, F., et al. 2010b, ApJL, 714, L118
Dannerbauer, H., Walter, F., & Morrison, G. 2008, ApJL, 673, L127
Darling, J., & Zeiger, B. 2012, ApJL, 749, L33
Decarli, R., Walter, F., Aravena, M., et al. 2016, ApJ, 833, 69
Decarli, R., Walter, F., Carilli, C., et al. 2014, ApJ, 782, 78
Delhaize, J., Smolčić, V., Delvecchio, I., et al. 2017, A&A, 602, A4
Eales, S., Smith, M. W. L., Auld, R., et al. 2012, ApJ, 761, 168
Faisst, A. L., Capak, P. L., Yan, L., et al. 2017, ApJ, 847, 21
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, MNRAS, 465, 1789
Genzel, R., Tacconi, L. J., Gracia-Carpio, J., et al. 2010, MNRAS, 407, 2091
Genzel, R., Tacconi, L. J., Lutz, D., et al. 2015, ApJ, 800, 20
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJL, 600, L93
González-López, J., Bauer, F. E., Aravena, M., et al. 2017, A&A, 608, A138
Gregory, P. 2010, Bayesian Logical Data Analysis for the Physical Sciences (Cambridge: Cambridge Univ. Press)
Gregory, P. C., & Loredo, T. J. 1992, ApJ, 398, 146
Greve, T. R., Bertoldi, F., Smail, I., et al. 2005, MNRAS, 359, 1165
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Groves, B. A., Schinnerer, E., Leroy, A., et al. 2015, ApJ, 799, 96
Guidetti, D., Bondi, M., Prandoni, I., et al. 2017, MNRAS, 471, 210
Hodge, J. A., Swinbank, A. M., Simpson, J. M., et al. 2016, ApJ, 833, 103
Hogg, D. W., Myers, A. D., & Bovy, J. 2010, ApJ, 725, 2166
Ivison, R. J., Papadopoulos, P. P., Smail, I., et al. 2011, MNRAS, 412, 1913
Keating, G. K., Bower, G. C., Marrone, D. P., et al. 2015, ApJ, 814, 140
Keating, G. K., Marrone, D. P., Bower, G. C., et al. 2016, ApJ, 830, 34
Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kogut, A., Chuss, D. T., Dotson, J., et al. 2014, Proc. SPIE, 9143, 91431E
Kogut, A., Fixsen, D. J., Chuss, D. T., et al. 2011, JCAP, 7, 025
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
Lentati, L., Wagg, J., Carilli, C. L., et al. 2015, ApJ, 800, 67
Liu, D., Daddi, E., Dickinson, M., et al. 2018, ApJ, 853, 172
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, ApJ, 760, 6
Mangum, J. G., Darling, J., Henkel, C., & Menten, K. M. 2013, ApJ, 766, 108
Mangum, J. G., Darling, J., Menten, K. M., & Henkel, C. 2008, ApJ, 673, 832
McKinnon, M., Carilli, C., & Beasley, T. 2016, Proc. SPIE, 9906, 990627
McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in ASP Conf. Ser. 376, Astronomical Data Analysis Software and Systems XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco, CA: ASP), 127
Miettinen, O., Novak, M., Smolčić, V., et al. 2017, A&A, 602, A54
Molnár, D. C., Sargent, M. T., Delhaize, J., et al. 2018, MNRAS, 475, 827
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Morrison, G. E., Owen, F. N., Dickinson, M., Ivison, R. J., & Ibar, E. 2010, ApJS, 188, 178
Narayanan, D., & Krumholz, M. R. 2014, MNRAS, 442, 1411
Pope, A., Scott, D., Dickinson, M., et al. 2006, MNRAS, 370, 1185
Popping, A., Jurek, R., Westmeier, T., et al. 2012, PASA, 29, 318
Popping, G., Somerville, R. S., & Trager, S. C. 2014, MNRAS, 442, 2398
Popping, G., van Kampen, E., Decarli, R., et al. 2016, MNRAS, 461, 93
Pullen, A. R., Chang, T.-C., Doré, O., & Lidz, A. 2013, ApJ, 768, 15
Riechers, D. A., Capak, P. L., Carilli, C. L., et al. 2010a, ApJL, 720, L131
Riechers, D. A., Carilli, C. L., Capak, P. L., et al. 2014, ApJ, 796, 84
Riechers, D. A., Carilli, C. L., Maddalena, R. J., et al. 2011a, ApJL, 739, L32
Riechers, D. A., Carilli, L. C., Walter, F., et al. 2011b, ApJL, 733, L11
Riechers, D. A., Carilli, C. L., Walter, F., & Momjian, E. 2010b, ApJL, 724, L153
Riechers, D. A., Hodge, J., Walter, F., Carilli, C. L., & Bertoldi, F. 2011c, ApJL, 739, L31
Riechers, D. A., Pavesi, R., Sharon, C. E., et al. 2018, ApJ, submitted (arXiv:1808.04371)
Riechers, D. A., Walter, F., Carilli, C. L., et al. 2006, ApJ, 650, 604
Righi, M., Hernández-Monteagudo, C., & Sunyaev, R. A. 2008, A&A, 489, 489
Rudnick, G., Hodge, J., Walter, F., et al. 2017, ApJ, 849, 27
Sandstrom, K. M., Leroy, A. K., Walter, F., et al. 2013, ApJ, 777, 5
Schinnerer, E., Groves, B., Sargent, M. T., et al. 2016, ApJ, 833, 112
Scoville, N., Lee, N., Vanden Bout, P., et al. 2017, ApJ, 837, 150
Scoville, N., Sheth, K., Aussel, H., et al. 2016, ApJ, 820, 83
Selina, R., & Murphy, E. 2017, ngVLA Reference Design Development and Performance Estimates, http://library.nrao.edu/public/memos/ngvla/NGVLA_17.pdf
Sharon, C. E., Riechers, D. A., Hodge, J., et al. 2016, ApJ, 827, 18
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2015, ApJ, 807, 128
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Smolčić, V., Novak, M., Bondi, M., et al. 2017, A&A, 602, A1
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Spilker, J. S., Bezanson, R., Marrone, D. P., et al. 2016, ApJ, 832, 19
Steidel, C. C., Shapley, A. E., Pettini, M., et al. 2004, ApJ, 604, 534
Steinhardt, C. L., Speagle, J. S., Capak, P., et al. 2014, ApJL, 791, L25
Tacconi, L. J., Genzel, R., Neri, R., et al. 2010, Natur, 463, 781
Tacconi, L. J., Genzel, R., Saintonge, A., et al. 2018, ApJ, 853, 179
Tacconi, L. J., Genzel, R., Smail, I., et al. 2008, ApJ, 680, 246
Tacconi, L. J., Neri, R., Chapman, S. C., et al. 2006, ApJ, 640, 228
Tacconi, L. J., Neri, R., Genzel, R., et al. 2013, ApJ, 768, 74
Vio, R., & Andreani, P. 2016, A&A, 589, A20
Vio, R., Vergès, C., & Andreani, P. 2017, A&A, 604, A115
Walter, F., Decarli, R., Aravena, M., et al. 2016, ApJ, 833, 67
Walter, F., Decarli, R., Carilli, C., et al. 2012, Natur, 486, 233
Walter, F., Decarli, R., Sargent, M., et al. 2014, ApJ, 782, 79
Westmeier, T., Popping, A., & Serra, P. 2012, PASA, 29, 276
Whiting, M. T. 2012, MNRAS, 421, 3242
Zeiger, B., & Darling, J. 2010, ApJ, 709, 386