University of Groningen

The Grism Lens-amplified Survey from Space (GLASS). X. Sub-kiloparsec Resolution Gas-phase Metallicity Maps at Cosmic Noon behind the Hubble Frontier Fields Cluster MACS1149.6+2223

Wang, X.; Jones, T.~A.; Treu, T.; Morishita, T.; Abramson, L.~E.; Brammer, G.~B.; Huang, K.-H.; Malkan, M.~A.; Schmidt, K.~B.; Fontana, A.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Link to publication in University of Groningen/UMCG research database

CrossMark

The Grism Lens-Amplified Survey from Space (GLASS). X. Sub-kiloparsec Resolution Gas-phase Metallicity Maps at Cosmic Noon behind the Hubble Frontier Fields Cluster MACS1149.6+2223

Xin Wang[1], Tucker A. Jones[2,3,18], Tommaso Treu[1], Takahiro Morishita[1,4,5], Louis E. Abramson[1], Gabriel B. Brammer[6],
Kuang-Han Huang[3], Matthew A. Malkan[1], Kasper B. Schmidt[7], Adriano Fontana[8], Claudio Grillo[9,10], Alaina L. Henry[6,11],
Wouter Karman[12], Patrick L. Kelly[13], Charlotte A. Mason[1], Amata Mercurio[14], Piero Rosati[15], Keren Sharon[16],
Michele Trenti[17], and Benedetta Vulcani[17]

[1] Department of Physics and Astronomy, University of California, Los Angeles, CA 90095-1547, USA; xwang@astro.ucla.edu
[2] Institute of Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
[3] University of California Davis, 1 Shields Avenue, Davis, CA 95616, USA
[4] Astronomical Institute, Tohoku University, Aramaki, Aoba, Sendai 980-8578, Japan
[5] Institute for International Advanced Research and Education, Tohoku University, Aramaki, Aoba, Sendai 980-8578, Japan
[6] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[7] Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany
[8] INAF—Osservatorio Astronomico di Roma Via Frascati 33—I-00040 Monte Porzio Catone, Italy
[9] Dipartimento di Fisica, Università degli Studi di Milano, via Celoria 16, I-20133 Milano, Italy
[10] Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark
[11] Astrophysics Science Division, Goddard Space Flight Center, Code 665, Greenbelt, MD 20771, USA
[12] Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands
[13] Department of Astronomy, University of California, Berkeley, CA 94720-3411, USA
[14] INAF—Osservatorio Astronomico di Capodimonte, Via Moiariello 16, I-80131 Napoli, Italy
[15] Dipartimento di Fisica e Scienze della Terra, Universitá degli Studi di Ferrara, via Saragat 1, I-44122 Ferrara, Italy
[16] Department of Astronomy, University of Michigan, 1085 S. University Avenue, Ann Arbor, MI 48109, USA
[17] School of Physics, University of Melbourne, VIC 3010, Australia

Abstract

We combine deep *Hubble Space Telescope* grism spectroscopy with a new Bayesian method to derive maps of gas-phase metallicity for 10 star-forming galaxies at high redshift ($1.2 \lesssim z \lesssim 2.3$). Exploiting lensing magnification by the foreground cluster MACS1149.6+2223, we reach sub-kiloparsec spatial resolution and push the limit of stellar mass associated with such high-z spatially resolved measurements below $10^8 M_\odot$ for the first time. Our maps exhibit diverse morphologies, indicative of various effects such as efficient radial mixing from tidal torques, rapid accretion of low-metallicity gas, and other physical processes that can affect the gas and metallicity distributions in individual galaxies. Based upon an exhaustive sample of all existing sub-kiloparsec resolution metallicity gradient measurements at high z, we find that predictions given by analytical chemical evolution models assuming a relatively extended star-formation profile in the early disk-formation phase can explain the majority of observed metallicity gradients, without involving galactic feedback or radial outflows. We observe a tentative correlation between stellar mass and metallicity gradients, consistent with the "downsizing" galaxy formation picture that more massive galaxies are more evolved into a later phase of disk growth, where they experience more coherent mass assembly at all radii and thus show shallower metallicity gradients. In addition to the spatially resolved analysis, we compile a sample of homogeneously cross-calibrated integrated metallicity measurements spanning three orders of magnitude in stellar mass at $z \sim 1.8$. We use this sample to study the mass–metallicity relation (MZR) and find that the slope of the observed MZR can rule out the momentum-driven wind model at a 3σ confidence level.

Key words: galaxies: abundances – galaxies: evolution – galaxies: formation – galaxies: high-redshift – gravitational lensing: strong

1. Introduction

Galaxies are complex ecosystems, particularly at the peak epoch of cosmic star formation, corresponding to the redshift range of $z \sim 1$–3, also known as the "cosmic noon" (see Madau & Dickinson 2014, for a recent review). During this approximately 4 Gyr, the Hubble sequence gradually breaks down and the predominant morphology of galaxies transforms from irregular systems at high redshifts to symmetric disks and bulges at low redshifts (Mortlock et al. 2013). This complexity is to a large extent induced by the interplay between the process of star formation, and the diverse aspects of baryonic cycling, e.g.,

galactic feedback, gas inflows/outflows, and major/minor mergers (Davé et al. 2011b; Martin et al. 2012). The effect of environment surrounding galaxies can further complicate the spatial distribution of star formation, as recently revealed by Vulcani et al. (2015, 2016). Measurements of gas-phase metallicity, i.e., the chemical abundances of elements heavier than hydrogen and helium in the interstellar medium (ISM), are a powerful means to shed light on this complexity[19], because the metal enrichment history is strongly tied to the mass assembly history in galaxy evolution (Davé et al. 2011a; Lu et al. 2015b). Since detailed elemental abundances are not directly measurable

---

[18] Hubble Fellow.

[19] Hereafter, throughout the paper, we refer to gas-phase metallicity as metallicity for simplicity.

at extragalactic distances, the relative oxygen abundance in ionized gaseous nebulae, i.e., $12 + \log(O/H)$, is often chosen as the observational proxy of metallicity.

For over a decade, a tight correlation between metallicity and galaxy stellar mass (M_*), i.e., the mass–metallicity relation (MZR), has been quantitatively established, from the vast database of local galaxies observed by the Sloan Digital Sky Survey (SDSS; Tremonti et al. 2004; Zahid et al. 2012; Andrews & Martini 2013). This relation has been further extended to high redshifts, using deep near-infrared (IR) spectroscopy facilitated by large ground-based and space-based telescopes (Erb et al. 2006; Maiolino et al. 2008; Zahid et al. 2011; Henry et al. 2013; Steidel et al. 2014; Sanders et al. 2015; Guo et al. 2016). The measurements of the MZR as a function of redshift can cast useful constraints on various galaxy evolution models, since the slope of the MZR is sensitive to the properties of outflows, such as the mass loading factor and the outflow speed (see, e.g., Davé et al. 2012; Lu et al. 2015a). This slope can also be explained by variations of star-formation efficiency and gas mass fraction in galaxies with different stellar masses (see, e.g., Baldry et al. 2008; Zahid et al. 2014). The normalization of the MZR can shed light upon the stellar chemical yield across cosmic time (Finlator & Davé 2008). Mannucci et al. (2010) first suggested that there exists a so-called fundamental metallicity relation (FMR) in the 3D parameter space spanned by M_*, star-formation rate (SFR), and metallicity, such that the MZR is merely a 2D projection of this more fundamental 3D manifold (see also Hunt et al. 2016). This 3D scaling relation shows a tight scatter ($\sim$0.05 dex) in metallicity and is speculated to not evolve with z. In this context, the apparent redshift evolution of the MZR normalization originates primarily from sampling the FMR in terms of galaxies with different SFR. This concept of the FMR is in accord with the gas regulator model proposed by Lilly et al. (2013), even though mergers can also play a subtle role in shaping the form of the FMR by increasing the scatter (Michel-Dansac et al. 2008). However, at high redshifts, the validity of the FMR is still under investigation (see, e.g., Wuyts et al. 2014; Sanders et al. 2015).

Spatially resolved chemical information provides a more powerful diagnostic tool about galaxy baryonic assembly than integrated metallicity measurements, especially at high redshifts. Because for non-interacting galaxies, their metallicity radial gradients are found to be highly sensitive to the properties of gas, i.e., the surface density, the existence of inflows/outflows, and the kinematic structure (Cresci et al. 2010; Jones et al. 2013; Sánchez et al. 2014; Troncoso et al. 2014). In the past few years, metallicity radial gradients, measured from spectroscopic data acquired by ground-based instruments with natural seeing ($\gtrsim 0''.6$), have been reported at high redshifts (Queyrel et al. 2012; Stott et al. 2014; Troncoso et al. 2014). In particular, a large mass-selected sample of galaxies at $0.7 \lesssim z \lesssim 2.7$ were observed with spatially resolved gas kinematics and star formation, using the K-band Multi Object Spectrograph (KMOS) on the Very Large Telescope (VLT; i.e., the KMOS^{3D} survey, Wisnioski et al. 2015). As a result, Wuyts et al. (2016) measured metallicity gradients for 180 star-forming galaxies in three redshift intervals (i.e., [0.8, 1.0], [1.3, 1.7], and [2.0, 2.6]) and found that the majority of the metallicity gradients are flat and that there were no statistically significant correlations between metallicity gradients, and stellar, kinematic, and structural properties of the galaxy.

However, seeing-limited measurements typically have insufficient resolution to resolve the inner structure of galaxies at $z \gtrsim 1$, and the inferred metallicity gradient can potentially be biased. For example, Yuan et al. (2013) showed that coarse spatial sampling ($\geqslant$1 kpc) can result in artificially flat metallicity gradients, inferring that sufficiently high spatial resolution, i.e., sub-kiloparsec scale (corresponding to an angular resolution of $<0''.2–0''.3$), is crucial in yielding precise information of how metals are distributed spatially in extragalactic H II regions. Only a few metallicity gradient measurements meet this requirement, including a sample of nine Hα-selected galaxies at $z \in$ [0.84, 2.23] (the majority at $z \sim 1.45$), observed with the adaptive optics (AO) assisted integral field unit (IFU) spectrograph *SINFONI* on board the VLT (Swinbank et al. 2012). Even higher resolution can be attained through the combination of diffraction-limited data and gravitational lensing as shown by Jones et al. (2010, 2013), Yuan et al. (2011), targeting strongly lensed galaxies with the laser guide star AO aided IFU spectrograph *OSIRIS* at the *Keck* telescope. Following this approach, Leethochawalit et al. (2015) recently analyzed a sample of 11 lensed galaxies (3 at $z \sim 1.45$ and the rest at $z \gtrsim 2$), deriving maps of both metallicity and emission line (EL) kinematics. Although a great amount of effort has been invested in enlarging the sample size of high-z metallicity gradients obtained with sub-kiloparsec scale spatial resolution, the current sample consists of only $\lesssim$30 such gradients and is still statistically insufficient to explore trends with stellar mass and redshift.

In order to enlarge the sample of sub-kiloparsec resolution measurements, we recently demonstrated that such metallicity maps can be derived using space-based data (Jones et al. 2015b). We measured a flat metallicity gradient in a multiply lensed interacting system at $z = 1.85$, using diffraction-limited *Hubble Space Telescope* (*HST*) grism data and confirmed that flat metallicity gradients can be caused by gravitational interactions in merging systems.

The main goal of the work presented here is to collect a uniformly analyzed large sample of high-z metallicity maps obtained with sub-kiloparsec spatial resolution. To meet our goal, we improve upon the methodology proposed in our pilot paper (Jones et al. 2015b), in particular, via developing a novel Bayesian method to imply metallicity from multiple EL diagnostics simultaneously. We apply our advanced analysis to ultra-deep grism data of the massive galaxy cluster MACS1149.6+2223, thus exploiting the powerful synergy of *HST* diffraction-limited spectroscopy and lensing magnification.

The outline of this paper is as follows. In Section 2, we introduce the spectroscopic observations used in this work and the selection of our metallicity gradient sample. The photometric data and galaxy cluster lens models are briefed in Section 3. We then present our entire analysis process in Section 4, and show our results in terms of global demographics and spatially resolved analysis in Sections 5 and 6, respectively. Finally, Section 7 will conclude and discuss our study. The concordance cosmological model ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70$ km s^{-1} Mpc^{-1}) and the AB magnitude system (Oke & Gunn 1983) are used throughout this paper. Without a specific number showing the wavelength, the names of forbidden lines are simplified as [O II]$\lambda\lambda$3726,

3729 := [O II], [O III] λ5008 := [O III], [N II] λ6585 := [N II], [S II] $\lambda\lambda$6718, 6732 := [S II] in this paper.[20]

2. Spectroscopic Data and Sample Selection

In Section 2.1, we summarize our *HST* grism observations, data quality, and data reduction. Our sample selection criteria are then described in Section 2.2. We also carried out some ground-based IFU observations on part of our sample, which are presented in Appendix A.

2.1. Hubble Space Telescope Grism Spectroscopy

2.1.1. The Grism Lens-Amplified Survey from Space

The Grism Lens-Amplified Survey from Space[21] (GLASS; Proposal ID 13459; P.I. Treu, Schmidt et al. 2014; Treu et al. 2015) is an *HST* cycle 21 large general observing (GO) program. GLASS observed 10 massive galaxy clusters with the Wide Field Camera 3 Infrared (WFC3/IR) grisms (G102 and G141; 10 and 4 orbits per cluster, respectively) targeted at their centers and the Advanced Camera for Survey (ACS) Optical grism (G800L) at their infall regions. The exposure on each cluster core is split into two nearly orthogonal position angles (PAs), in order to disentangle contamination from neighboring objects. Data acquisition was completed in 2015 January. Here we focus on the grism data targeted on the center of MACS1149.6+2223 taken in 2014 February (PA = 032°) and 2014 November (PA = 125°), marked by blue squares in Figure 1 (also see Table 1). GLASS provides ∼10% of the G141 exposures used in this work, and 100% of the G102 exposures. Details on GLASS data reduction can be found in Schmidt et al. (2014) and Treu et al. (2015).

Shallow images through filters F105W or F140W were taken to aid the alignment and extraction of grism spectra. The imaging exposures are combined with the exposures obtained by the Hubble Frontier Fields (HFF) initiative and other programs to produce the deep stacks, released as part of the HFF program (see Section 3 for more details).

2.1.2. The Supernova Refsdal Follow-up Program with HST G141

The majority of the G141 exposures used in this work (30 orbits) were taken as part of the follow-up *HST* GO/DDT campaign (Proposal ID 14041, P.I. Kelly, G. Brammer et al. 2017, in preparation) of the first multiply imaged supernova, SN Refsdal (Kelly et al. 2015b). Two pointings (shown by the red squares in Figure 1) were exposed between 2014 December and 2015 January with 8° apart (i.e., at PA = 111°, and 119° respectively, as shown in Table 1), in order to optimize the spectroscopy of SN Refsdal. The analysis of the spectra of SN Refsdal, which showed that it was SN 1987A-like, is described by Kelly et al. (2015c).

2.1.3. Grism Data Reduction

The combined grism data set was reduced following the procedure of the 3D-*HST* survey (Brammer et al. 2012; Momcheva et al. 2016). An updated 3D-*HST* pipeline[22] was employed to reduce the images and spectra. The ASTRODRIZZLE software from the DRIZZLEPAC package and `tweakreg` were used to align and combine individual grism exposures, subtracting the sky images provided by Brammer et al. (2012). The time-varying sky background due to helium glow (at 10830 Å) in the Earth exosphere was also accounted for according to Brammer et al. (2014). After these initial steps, the mosaics were co-added through interlacing onto a grid of 0″065 resolution, Nyquist sampling the point-spread function (PSF) full width half maximum (FWHM). The average spectral resolution is $R \sim 210$ and $R \sim 130$ for G102 and G141, respectively. The average pixel scale after interlacing is 12 Å and 22 Å, respectively, for G102 and G141.

We used the co-added H_{160}-band mosaics as the detection image, on which SEXTRACTOR was run. An object catalog was generated and a corresponding segmentation map was created for each object, determining which spatial pixels (spaxels) belong to this object. Thus for a source registered in the catalog and falling within the WFC3 grism field of view (FOV), we extracted its spectra from grism mosaics through adding up the dispersed flux for all spaxels within an area defined by the segmentation image of this source, after flat-field correction and background subtraction. The "fluxcube" model from the AXE package (Kümmel et al. 2009) was employed to generate the 2D models of stellar spectrum, based upon the spatial profile of the source, the color information from direct image mosaics (if available), and the grism calibration configurations. For sources of interest, this model serves as the continuum model, which we re-scaled and subtracted from the observed 2D spectra (see Section 4.3 for more details) in order to obtain pure EL maps. At the same time, if the object is a bright contaminating neighbor, this 2D model also functions as spectral contamination, which is subtracted before the continuum removal.

The data used in this work contain a total of 34 orbits of G141 and 10 orbits of G102, reaching a 1σ flux limit (uncorrected for lensing magnification) of $3.5(1.2) \times 10^{-18}$ erg s^{-1} cm^{-2} over the wavelength range of G102(G141), calculated from Schmidt et al. (2016). Our data provide an uninterrupted wavelength coverage in the range of 0.8–1.7 μm and make this particular field one of the deepest fields probed by *HST* spectroscopy to date.

2.2. Sample Selection

The selection of our metallicity gradient sample is based upon the master redshift catalog for[23] MACS1149.6+2223, published by the GLASS collaboration. As described by Treu et al. (2016), redshifts were determined by combining spectroscopic information from GLASS and the SN Refsdal follow-up *HST* grism programs, ground-based *MUSE* observations and Keck *DEIMOS* data.

For the purpose of this study, we compiled an exhaustive list of spatially extended sources with secure spectroscopic redshift measurements in the range of $z \in [1.2, 2.3]$, showing strong ELs (primarily [O III] and Hβ) in the 2D grism spectra. The reason this redshift range is selected is that we require high signal-to-noise ratio (S/N) detections of [O III], Hβ, and [O II] ELs in the spectral region where grism sensitivity and throughput do not drop significantly, in order to deliver reliable metallicity estimates from well-calibrated EL diagnostics.[24] The selection results in a sample of 14 galaxies.
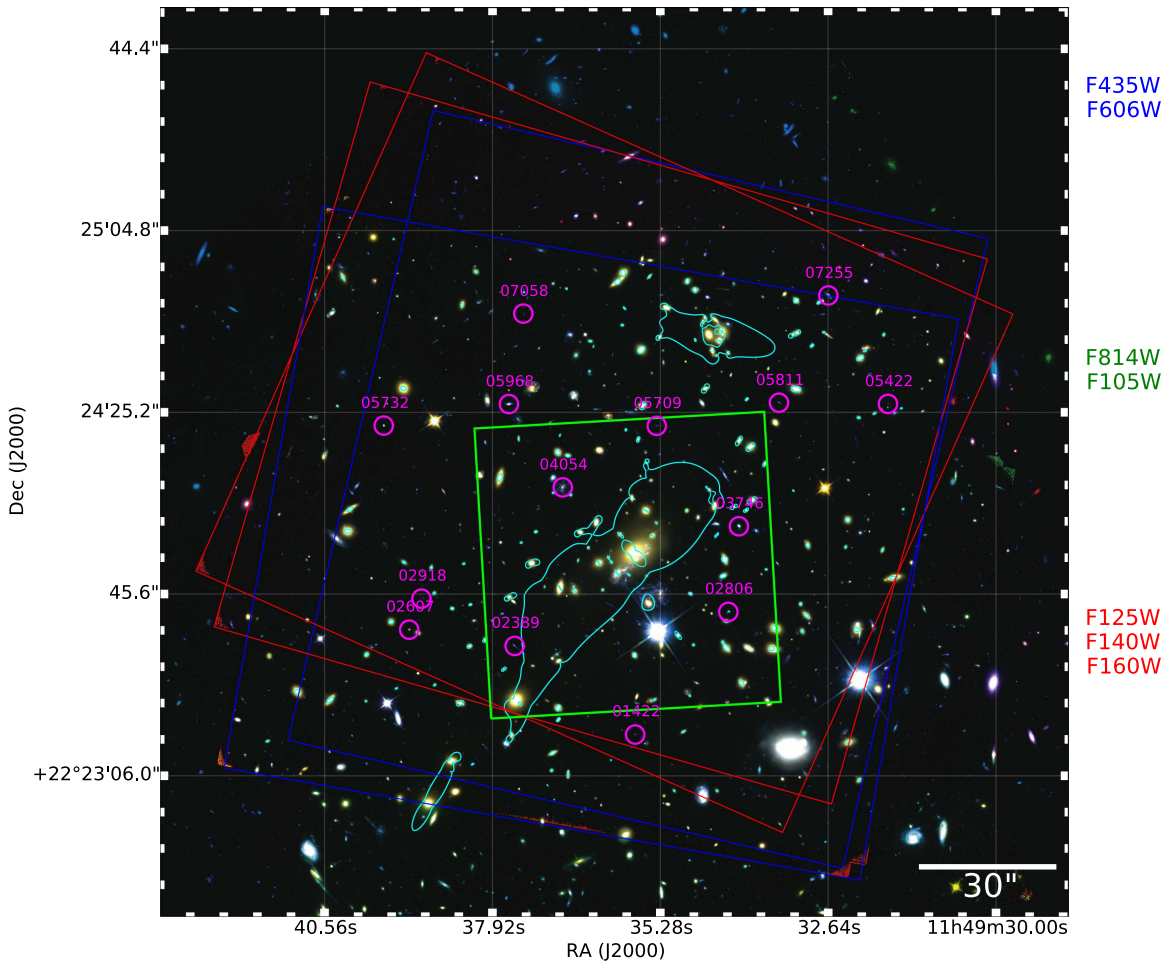
---

[20] The wavelength values are taken from http://classic.sdss.org/dr7/algorithms/linestable.html.
[21] http://glass.astro.ucla.edu
[22] http://code.google.com/p/threedhst/

[23] Available at https://archive.stsci.edu/prepds/glass/.
[24] The Hα EL is only accessible in sources at $z \leqslant 1.5$ in our sample. The GLASS spectra of an exemplary galaxy whose Hα is covered are displayed in Appendix B.

Figure 1. Color composite image of MACS1149.6+2223 from the full-depth seven-filter HFF photometry. The blue, green, and red channels comprise the *HST* broadband filters shown on the right. The blue, red, and green squares mark the footprints of GLASS, the Refsdal follow-up, and *MUSE* programs utilized in this work. The cyan contours are the critical curves at $z = 1.8$ predicted by the GLAFIC version 3 lens model. The critical curves denote the regions where lensing magnification reaches infinity. Thus the closer proximity between the object and the critical curve at the object's redshift indicates the higher magnification of the object. The magenta circles show the positions of our metallicity gradient sample. 3×3 arcsec^2 zoom-in postage stamps around these positions are shown in Figure 2.

The positions of these 14 galaxies relative to the cluster are denoted by the magenta circles in Figure 1, while the postage stamp images of these galaxy are shown in Figure 2. The full sample is also listed in Table 2. In particular, one source in our sample, i.e., galaxy ID 04054, is one of the multiple images of SN Refsdal's host galaxy (see Figure 3 for its appearances from various perspectives). A steep metallicity gradient $(-0.16 \pm 0.02 \ \mathrm{dex\,kpc^{-1}})$ has been previously measured on another multiple image of this spiral galaxy (Yuan et al. 2011). Our work is the first metallicity gradient measurement on this particular image, which is least distorted/contaminated by foreground cluster members. We also take advantage of a more precise lens model of both the entire cluster and the SN Refsdal host, from the extensive lens modeling effort summarized by Treu et al. (2016).

3. Photometry and Lens Models from the HFF

The HFF initiative[25] (P.I. Lotz, Lotz et al. 2016) is an ongoing multi-cycle treasury program enabled by an *HST* director's discretionary time allocation. HFF targets the cores (prime fields) and infall regions (parallel fields) of six

Table 1
Properties of the Grism Spectroscopic Data Used in This Work

PA[a] (deg.)	Grism	Exposure Time (s)	Program	Time of Completion
032	G102	8723	GLASS	2014 Feb
	G141	4412	GLASS	
111	G141	36088	SN Refsdal follow-up	2014 Dec
119	G141	36088	SN Refsdal follow-up	2015 Jan
125	G102	8623	GLASS	2014 Nov
	G141	4412	GLASS	

Note. Here we only include the grism observations targetted on the prime field of MACS1149.6+2223.
[a] The position angle shown here corresponds to the "PA_V3" value reported in the WFC3/IR image headers. The position angle of the dispersion axis of the grism spectra is given by $\mathrm{PA_{disp}} \approx \mathrm{PA\_V3} - 45$.

massive galaxy clusters, reaching an ultra-faint intrinsic magnitude of 30–33, made possible by the synergy of diffraction-limited photometry and lensing magnification. The
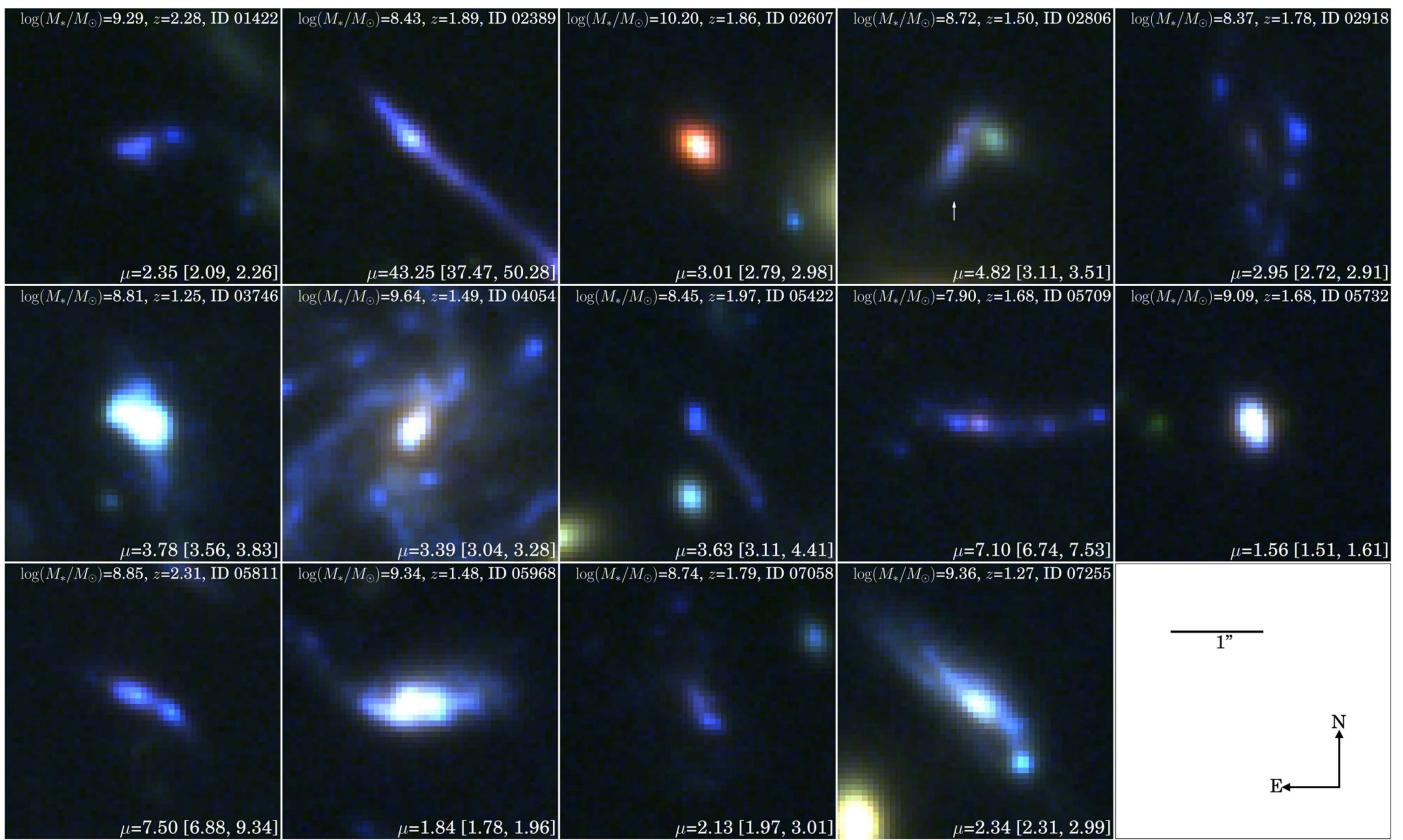
Figure 2. Zoom-in color composite stamps of our metallicity gradient sample, cut out from the full FOV image shown in Figure 1. In each stamp, we also show the spectroscopic redshift, stellar mass, and lensing magnification (best-fit value followed by the 1σ confidence interval) of the galaxy. In the panel of ID 02806, we use the white arrow to point to our galaxy of interest (ID 02806). All stamps are on the same spatial scale. The $1''$ scale bar and northeast compass are shown in the lower right corner.

large wavelength coverage, uninterrupted from B_{435} to H_{160} passbands, is crucial for photometric redshift determinations and stellar population synthesis.

Besides the deep image mosaics, the HFF collaboration also provides the community with cluster lens models.[26] Several groups of scientists were invited prior to the beginning of the campaign to provide independent models depicting the total mass distribution of the six HFF clusters, using a number of distinct techniques. As data accumulate and more multiply imaged systems are identified, the lens models are continuously improved. For our cluster of interest, MACS1149.6+2223, the most up-to-date publicly available model is the GLAFIC version 3 model, constructed using a simply parametrized method proposed by Oguri (2010). For an in-depth description of the various modeling techniques, their advantages and limitations, we refer to Treu et al. (2016) and Meneghetti et al. (2016).

In this work, we need lens models to trace the observed metallicity maps back to the source plane. Following our previous work (Jones et al. 2015b), we experiment with several publicly available models to find the ones that give the reconstruction of our target sources the highest fidelity (judging by how well the source plane morphologies of multiply imaged sources match each other). Based upon the SN Refsdal test (see also Kelly et al. 2015a), the GLAFIC version 3 model (Kawamata et al. 2015), the GRILLO model (Grillo et al. 2016), and the SHARON & JOHNSON version 3 model were the most accurate ones, so we

considered those three. As an illustration, the GLAFIC version 3 critical curves at $z = 1.8$, the median redshift of our metallicity gradient sample, are overlaid in cyan in Figure 1.

The most highly magnified source in our sample is galaxy ID 02389, one of the folded arcs that straddle the critical curve. For that particular source, we used as our default the SHARON & JOHNSON version 3 model, updated from the earlier versions presented in Johnson et al. (2014) and Sharon & Johnson (2015). The SHARON & JOHNSON version 3 model leads to a more precise reconstruction of the source plane morphology of galaxy ID 02389. We also tested that switching entirely from the GLAFIC version 3 model to the SHARON & JOHNSON version 3 model does not affect our measurements significantly, as is also pointed out by Leethochawalit et al. (2015). The lensing magnification results from the considered lens models are given in Table 2.

4. Analysis Procedures

Here we describe the stellar population synthesis analysis of our sample in Section 4.1, and source plane morphology reconstruction in Section 4.2. All of the steps necessary for extracting EL maps from 2D grism spectra combined from different PAs are detailed in Section 4.3. Our new Bayesian method to jointly infer metallicity, nebular dust extinction, and SFR from a simultaneous use of multiple strong ELs is presented in Section 4.4.

---

[26] http://archive.stsci.edu/prepds/frontier/lensmodels/

Table 2
Global Demographic Properties of the Metallicity Gradient Sample Analyzed in This Work

ID	R.A. (deg.)	Decl. (deg.)	z_{spec}	μ^a	H_{160} Magnitude (ABmag)	SED Fitting		EL Diagnostics		
						$\log(M_*{}^b/M_\odot)$	A_V^{Sc}	$12 + \log(O/H)$	SFR^b $[M_\odot \, yr^{-1}]$	A_V^{Nc}
01422	177.398643	22.387499	2.28	2.35 [2.09, 2.26]	24.46	$9.29^{+0.07}_{-0.01}$	<0.01	$8.26^{+0.11}_{-0.13}$	$15.10^{+16.76}_{-7.11}$	<1.07
02389	177.406546	22.392860	1.89	43.25 [37.47, 50.28]	23.29	$8.43^{+0.06}_{-0.07}$	<0.01	$7.88^{+0.16}_{-0.15}$	<6.37	$1.36^{+0.69}_{-0.82}$
02607[d]	177.413454	22.393843	1.86	3.01 [2.79, 2.98]	22.63	$10.25^{+0.06}_{-0.10}$	$1.70^{+0.42}_{-0.18}$	$7.70^{+0.13}_{-0.11}$	>36.51	$2.26^{+0.31}_{-0.58}$
02806	177.392541	22.394921	1.50	4.82 [3.11, 3.51]	23.35	$8.72^{+0.19}_{-0.01}$	$0.50^{+0.01}_{-0.01}$	$7.96^{+0.15}_{-0.16}$	$1.96^{+1.37}_{-0.13}$	<0.18
02918	177.412652	22.395723	1.78	2.95 [2.72, 2.91]	25.22	$8.37^{+0.15}_{-0.08}$	$0.10^{+0.34}_{-0.10}$	$8.42^{+0.11}_{-0.14}$	<70.32	$1.25^{+0.72}_{-0.80}$
03746	177.391848	22.400105	1.25	3.78 [3.56, 3.83]	22.48	$8.81^{+0.03}_{-0.01}$	<0.01	$8.11^{+0.10}_{-0.11}$	$10.95^{+1.61}_{-0.58}$	<0.17
04054	177.403393	22.402456	1.49	3.39 [3.04, 3.28]	21.27	$9.64^{+0.05}_{-0.01}$	$1.10^{+0.01}_{-0.01}$	$8.70^{+0.09}_{-0.01}$	$16.99^{+5.71}_{-2.80}$	$0.72^{+0.23}_{-0.23}$
05422	177.382077	22.407510	1.97	3.63 [3.11, 4.41]	25.05	$8.45^{+0.15}_{-0.46}$	$0.00^{+0.56}_{-0.01}$	$7.56^{+0.16}_{-0.12}$	<38.18	<1.85
05709	177.397234	22.406181	1.68	7.10 [6.74, 7.53]	25.44	$7.90^{+0.02}_{-0.03}$	<0.01	$8.21^{+0.13}_{-0.21}$	<17.97	<1.18
05732	177.415126	22.406195	1.68	1.56 [1.51, 1.61]	23.45	$9.09^{+0.01}_{-0.01}$	<0.01	$8.41^{+0.10}_{-0.12}$	<84.70	<1.31
05811	177.389220	22.407583	2.31	7.50 [6.88, 9.34]	23.79	$8.85^{+0.11}_{-0.10}$	<0.01	$8.16^{+0.14}_{-0.21}$	$4.94^{+3.13}_{-2.47}$	<0.74
05968	177.406922	22.407499	1.48	1.84 [1.78, 1.96]	22.24	$9.34^{+0.02}_{-0.03}$	$0.60^{+0.20}_{-0.01}$	$8.47^{+0.07}_{-0.08}$	$27.95^{+4.70}_{-4.41}$	$0.23^{+0.15}_{-0.13}$
07058	177.405976	22.412977	1.79	2.13 [1.97, 3.01]	24.28	$8.74^{+0.17}_{-0.15}$	$0.00^{+0.12}_{-0.01}$	$8.38^{+0.13}_{-0.19}$	<44.99	<1.85
07255	177.385990	22.414074	1.27	2.34 [2.31, 2.99]	22.47	$9.36^{+0.01}_{-0.21}$	$0.30^{+0.21}_{-0.01}$	$8.38^{+0.08}_{-0.08}$	$8.73^{+3.53}_{-3.10}$	<0.79

Notes. The error bars and upper/lower limits shown in the columns of SED fitting and EL diagnostics correspond to 1σ confidence ranges. Note that the errors on the SED fitting results do not include any systematic uncertainties associated with the Bruzual & Charlot (2003) stellar population models.
[a] Best-fit magnification values and 1σ confidence intervals. Except for galaxy ID 02389, the magnification results are from the GLAFIC version 3 model, calculated by the HFF interactive online magnification calculator available at https://archive.stsci.edu/prepds/frontier/lensmodels/webtool/magnif.html. For galaxy ID 02389, we use the SHARON & JOHNSON version 3 model instead to compute the magnification results and correct for lensing magnification.
[b] Values presented here are corrected for lensing magnification.
[c] The superscripts of "S" and "N" refer to the stellar and nebular V-band dust extinction in units of magnitude, respectively.
[d] The EL diagnostic result on this source is not trustworthy, since it is classified as an AGN candidate (see Section 5).

4.1. Spectral Energy Distribution Fitting

The full-depth seven-filter HFF photometry was fitted with the template spectra from Bruzual & Charlot (2003) using the stellar population synthesis code FAST (Kriek et al. 2009), in order to derive global estimates of stellar mass (M_*) and the dust extinction of stellar continuum (A_V^S) for each source in our sample. We take a grid of stellar population parameters that include exponentially declining star-formation histories with e-folding times ranging from 10 Myr to 10 Gyr, stellar ages ranging from 5 Myr to the age of the universe at the observed redshift, and $A_V^S = 0$–4 mag for a Calzetti et al. (2000) dust extinction law. We assume the Chabrier (2003) initial mass function (IMF) and fix the stellar metallicity to solar. Table 2 lists the best-fit M_* values, corrected for lensing magnification. Although the adopted solar metallicity is higher than the typical gas-phase abundances we infer for the sample, this has little effect on the derived stellar mass. Fixing the stellar metallicity to 1/5 solar (i.e., $Z = 0.004$ or $12 + \log(O/H) = 8.0$, corresponding to the sample median) reduces the best-fit M_* by only ~0.03 dex.

4.2. Source Plane Morphology

Measurements of metallicity gradients require knowledge of the galaxy center, major axis orientation, and inclination or axis ratio. We derive these quantities from spatially resolved maps of the stellar mass surface density following the methodology described in Jones et al. (2015b). Briefly, we smooth the HFF photometric images to a common PSF of $0\farcs2$ FWHM and fit the spectral energy distribution (SED) of each spaxel using the same procedure described in Section 4.1. The resulting M_* maps in the image plane are shown in Figures 3 and 4. We reconstruct the M_* maps in the source plane using the adopted lens models and determine the centroid, axis ratio, inclination, and major axis orientation from a 2D elliptical Gaussian fit to the M_* distribution. This allows us to measure the galacto-centric radius at each point, assuming that contours of M_* surface density correspond to constant de-projected radius (following Jones et al. 2015b).

Since the following analysis is mostly done with the robust image plane data, we reconstruct the derived radius maps in the image plane. The contours of constant source plane galacto-centric radius for each galaxy in our sample are shown in Figures 3 and 4. Note that these contours are distorted from ellipses ascribed to the shear of gravitational lensing. For nearly half of our sample where we have seeing-limited IFU data described in Appendix A, we verify that for galaxies with significant rotation support (e.g., IDs 05709 and 05968), the morphological major axis agrees at 2σ with the pseudo-slit that maximizes velocity shear according to our kinematic analysis.

4.3. Emission Line Maps from Grism Spectra

The broad grism wavelength coverage provides spatially resolved maps of multiple ELs, such as [O II], Hγ, Hβ, [O III], Hα+[N II], and [S II]. The integrated fluxes of these ELs for our sample of galaxies are presented in Appendix B. To obtain pure EL maps, we first use the AXE continuum models described in Section 2 to subtract contaminating flux from neighboring sources. The continuum model for the target object is scaled to match the locally observed levels, and then subtracted. We check that the flux residual in regions near the ELs of interest follow a normal distribution with zero mean.
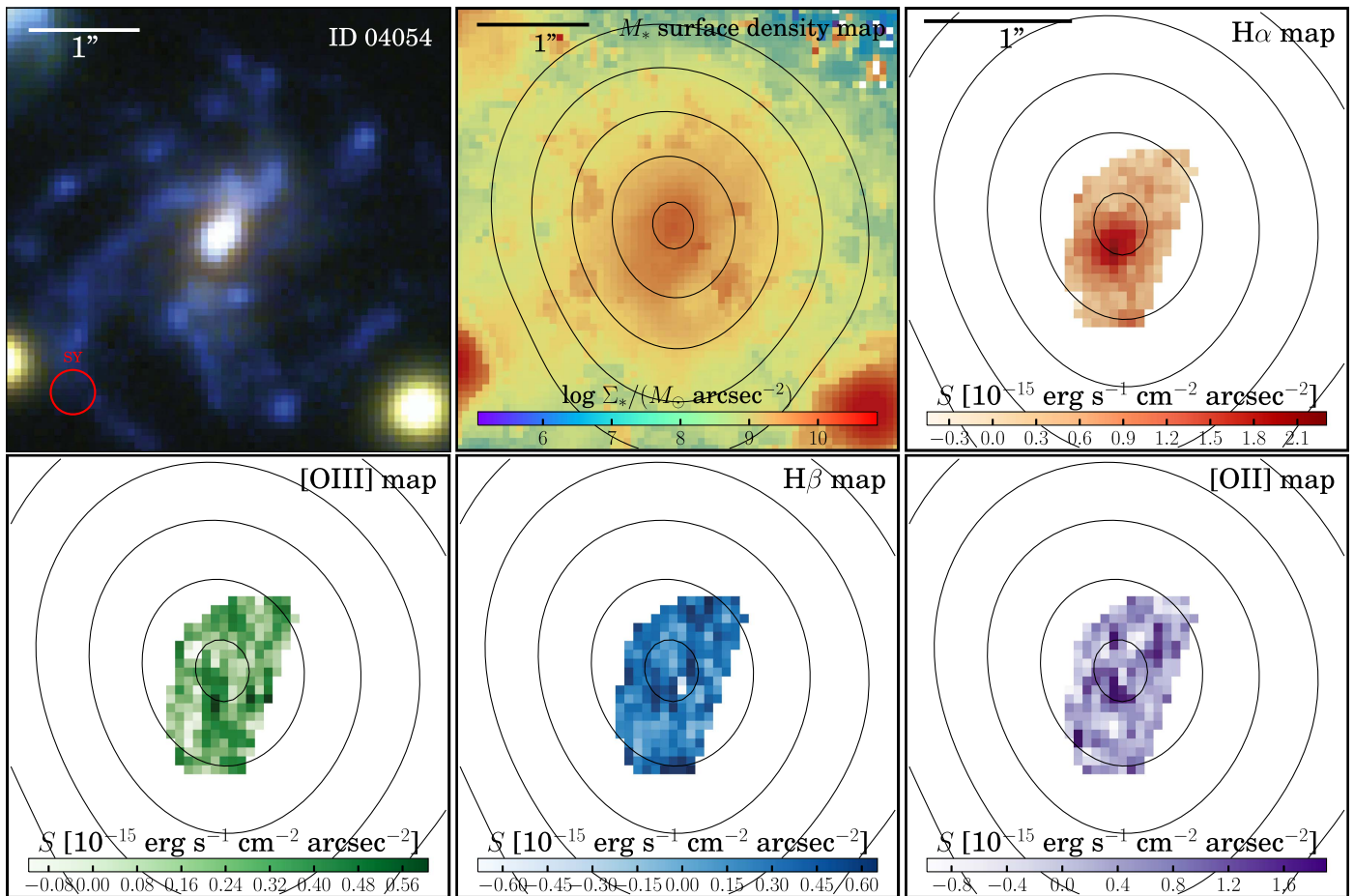
Figure 3. Multi-perspective view of the SN Refsdal host galaxy multiple image 1.3, i.e., ID 04054 in our metallicity gradient sample. Upper left: zoom-in color composite stamp cut out from Figure 1, where the estimated site of the unobserved past appearance of SN Refsdal is denoted by the red circle. Upper middle: stellar mass surface density map for this galaxy, where the source plane de-projected galactocentric radius contours are overlaid, in 2 kpc intervals, starting from 1 kpc. Upper right: Hα surface brightness map combined from the *HST* grism exposures at all PAs from the GLASS and Refsdal follow-up programs. The black contours again show the de-projected galactocentric radii in the same fashion presented in the stellar mass surface density map in the upper middle panel. Lower rows: combined surface brightness maps of [O III], Hβ, and [O II], presented in the same fashion as in the Hα emission line map in the upper right panel.

Due to the limited spectral resolution of *HST* WFC3/IR grisms, Hβ and the [O III] λλ4960,5008 doublets are partially blended in the spatially extended sources. We adopt a direct de-blending technique to separate these emission lines following the procedure used in Jones et al. (2015b). First, an isophotal contour is measured from the co-added *HST* H_{160}-band postage stamp, typically at the level of ~10% of the peak flux received from the source. We then apply this contour to the 2D grism spectra at the observed wavelengths corresponding to [O III] λλ4960,5008 and Hβ, maximizing the flux given grism redshift uncertainty (~0.01). We use the theoretical [O III] λλ4960,5008 doublet flux ratio (i.e., [O III] λ5008/[O III] λ4960 = 2.98, calculated by Storey & Zeippen 2000) to subtract the flux contribution of [O III] λ4960 corresponding to the region where [O III] λ5008 is unblended. We iterate this procedure to remove the [O III] λ4960 emission completely, resulting in pure maps of Hβ and [O III] λ5008.

Note that this direct de-blending will be compromised if the source is so extended that its [O III] λ5008 and Hβ are blended. In practice, we fine-tune the isophotal contour level to avoid this contamination. As a result, no cases in our sample have problems with our direct de-blending method, and the resulting pure [O III] λ5008 and Hβ maps are uncontaminated.

In the case of Hα and [N II], the emission lines are separated by less than the grism spectral resolution. Therefore, we cannot use the direct technique to de-blend Hα and [N II] emission, and instead treat the [N II]/Hα ratio as an unknown parameter in Section 4.4.

After obtaining pure EL maps, we rotate the maps from each PA to the same orientation and then combine the data, in order to utilize the full depth of the grism exposures. We note that EL maps must first be extracted from grism spectra at each PA before being combined, in order to account for the different rotation and spectral resolution. There are, in total, six different PA-grism combinations as given in Table 1. At certain redshifts, ELs are visible in the overlapping region between the two grisms at $\lambda \in [1.08, 1.15]\,\mu$m resulting in up to six individual maps of the same EL. We use the data taken at PA = 119° as a reference, as this minimizes errors arising from rotating the various PAs into alignment. We first apply a small vertical offset to each EL map to account for the slight tilt of the spectral trace (i.e., known offset between the dispersion direction and the *x*-axis of extracted spectra; Brammer et al. 2012). Next, we apply a small horizontal shift to correct for uncertainties in the wavelength calibration and redshift. Finally, we rotate each EL map to align it with the reference orientation at PA = 119°. The best values of these
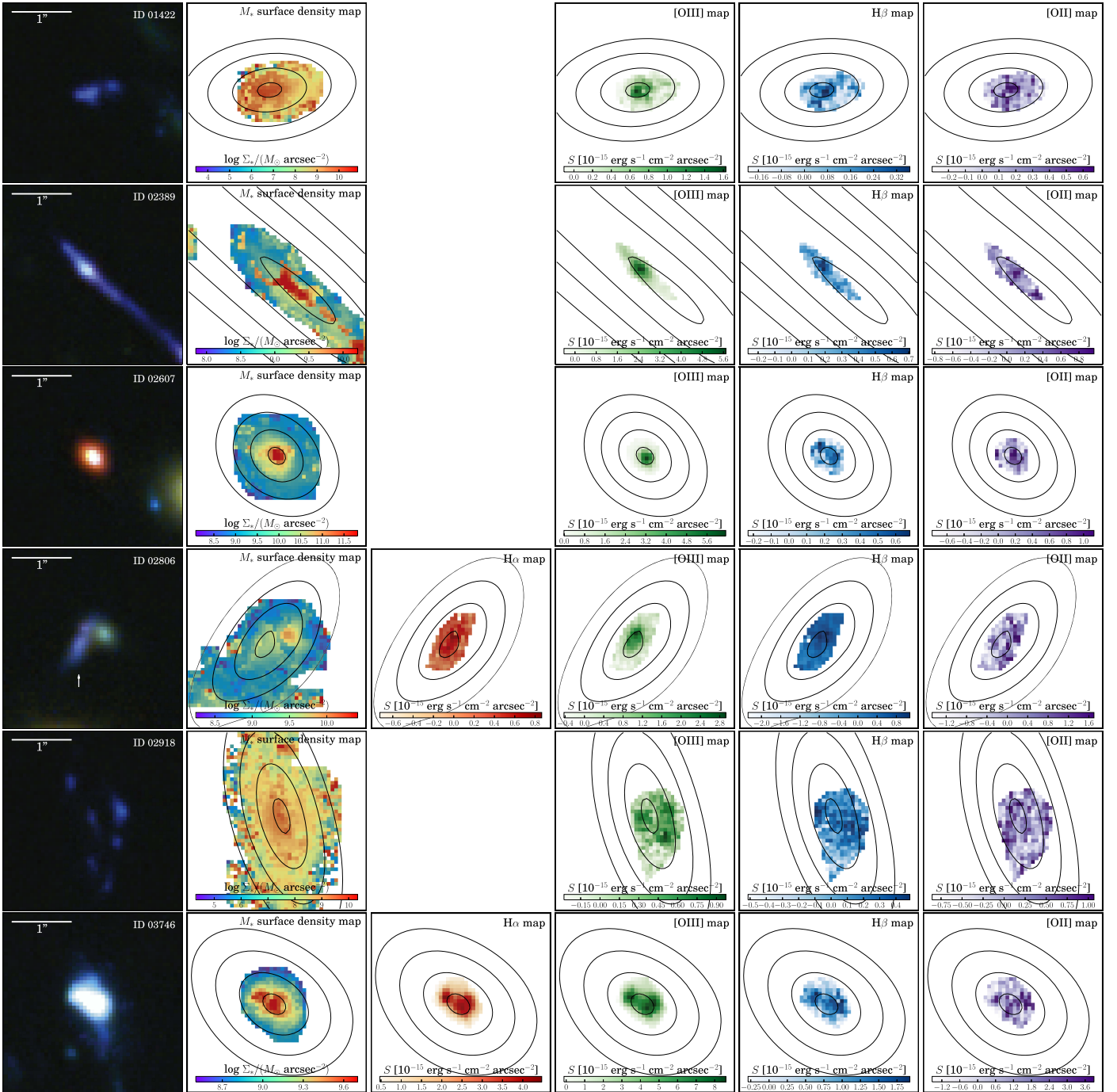
Figure 4. Multi-perspective view of our metallicity gradient sample except for the SN Refsdal host, which is shown in Figure 3. In each row, we show the zoom-in color composite stamp (the same as that in Figure 2), the stellar mass surface density map, the combined surface brightness maps of Hα (if available given object redshift), [O III], Hβ, [O II] for each object. The black contours overlaid represent the source plane de-projected galactocentric radii in a 2 kpc interval, starting from 1 kpc. The spatial extent and orientation of the stamps for each object have been fixed. The typical surface brightness 1σ uncertainty in these combined EL maps is 2×10^{-16} erg s^{-1} cm^{-2} arcsec^{-2}.

three alignment parameters are found by maximizing the normalized cross-correlation coefficient (NCCF) in order to achieve optimal alignment of multiple PAs:

$$\mathrm{NCCF} = \frac{1}{n} \Sigma_{x,y} \frac{(S_{\mathrm{ref}}(x, y) - \langle S_{\mathrm{ref}} \rangle)(S_{\mathrm{PA}}(x, y) - \langle S_{\mathrm{PA}} \rangle)}{\sigma_{S_{\mathrm{ref}}} \sigma_{S_{\mathrm{PA}}}}.$$

(1)

Here S is the surface brightness profile for a specific EL, S_{ref} corresponds to the reference alignment stamp adopted as the PA = 119° data, and n is the number of spaxels in the surface brightness profile. $\langle S \rangle$ and σ_S represent the average and standard deviation of the surface brightness, respectively.

Once the data from each PA are aligned, we vet the EL maps from each PA-grism combination and reject those that show
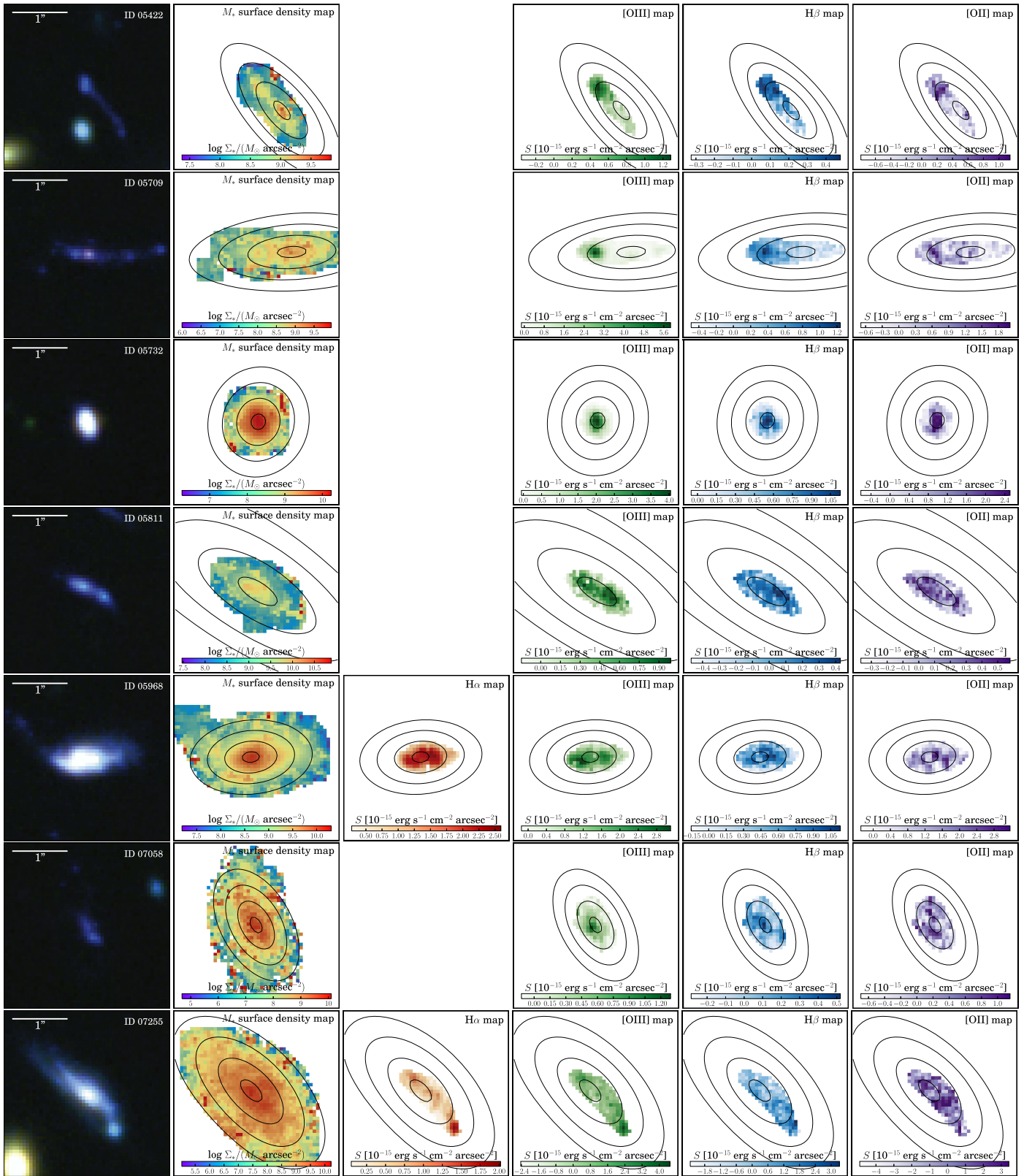
Figure 4. (Continued.)

significant contamination-subtraction residuals or grism reduction defects. We combine the remaining maps of a given EL using an inverse variance weighted average in order to properly account for the different exposure times and noise levels. The final combined maps of each EL for each galaxy are shown in Figures 3 and 4.

<div align="center">

Table 3
Sampling Parameters and Their Prior Information

</div>

Set	Parameter	Symbol (Unit)	Prior
Vanilla	gas-phase metallicity	$12 + \log(\text{O/H})$	flat: [7.0, 9.3]
	nebular dust extinction	A_V^N	flat: [0, 4]
	de-reddened Hβ flux	$f_{H\beta}$ (10^{-17} erg s^{-1} cm^{-2})	Jeffrey's: [0, 100]
Extended	[N II] to Hα flux ratio	[N II]/Hα	Jeffrey's: [0, 0.5]
Derived	star-formation rate	SFR ($M_\odot$ yr^{-1})	...

Note. Most constraints presented in this work are obtained under the "Vanilla" parameter set with the extended parameter fixed as [N II]/Hα = 0.05.

4.4. Line-flux-based Bayesian Inference of Metallicity

Generally speaking, two methods exist in deriving gas-phase oxygen abundance in star-forming galaxies at high redshifts. "Direct" determinations based on electron temperature measurements, which require high S/N detections of auroral lines (e.g., [O III] $\lambda 4364$), are still extremely challenging beyond the local universe (see Sanders et al. 2016, for a rare example). We therefore rely on the calibrations of strong collisionally excited EL flux ratios to estimate metallicities.

One of the most popular strong EL diagnostics is the flux ratio of [N II]/Hα calibrated by Pettini & Pagel (2004). However, a large offset (0.2–0.4 dex) between the loci of high-z and present-day star-forming galaxies in the Baldwin–Phillips–Terlevich (BPT, Baldwin et al. 1981) diagram has recently been discovered, indicating that extending the locally calibrated [N II]/Hα to high-z can be problematic (Sanders et al. 2015; Shapley et al. 2015). The interpretation of this offset is still a topic of much debate. It has been interpreted as the existence of a fundamental relation between the nitrogen–oxygen abundance ratio and M_* (Masters et al. 2016), a systematic dependence of strong EL properties with respect to Balmer line luminosity (Cowie et al. 2016), a combination of harder ionizing radiation and higher ionization states of the ISM gas at high redshifts. (Steidel et al. 2014, 2016), or an enhancement of nitrogen abundance in hot H II regions (Pilyugin et al. 2010). Fortunately, the alpha-element BPT diagrams show no offset with z (Sanders et al. 2015; Shapley et al. 2015), and therefore the diagnostics based upon those ELs are more reliable.

In this work, we adopt the calibrations of Maiolino et al. (2008, M08), including the flux ratios of [O III]/Hβ, [O II]/Hβ, and [N II]/Hα. Given the potential systematics related to nitrogen ELs, we do not use the [N II]/Hα calibration of M08. M08 fit the relation between these EL ratios and gas-phase oxygen abundance based upon "direct" measurements (from [O III] $\lambda 4364$) for $12 + \log(\text{O/H}) < 8.3$, and photoionization model results for higher metallicities. This provides a continuous framework valid over the wide range of $12 + \log(\text{O/H}) \in (7.1, 9.2)$. We note that although there are alternative calibrations of metallicity available in the literature and the applicability of these recipes is currently a hotly debated topic (see, e.g., Dopita et al. 2013; Blanc et al. 2015). Here we are primarily interested in relative metallicity measurements. Even though the absolute measurements of metallicity may change if we used a different calibration, the gradients and morphological features in the maps should be more robust to changes in the calibration. We will restrict our comparisons with other samples to only include studies assuming the M08 calibrations.

Unlike previous work in which calibrated relations are applied to various EL flux ratios (see, e.g., Pérez-Montero 2014; Bianco et al. 2015), we design a Bayesian statistical approach that directly uses the individual EL fluxes as input, such that the information from one EL is only used once. Our approach presents several advantages over those based on line flux ratios. First, it properly accounts for flux uncertainties for lines that are weak or undetected. Second, multiple lines can be taken into consideration without the risk of double counting any information.

Our approach simultaneously infers gas-phase metallicity ($12 + \log(\text{O/H})$), nebular dust extinction (A_V^N), and expected de-reddened Hβ flux in units of 10^{-17} erg s^{-1} cm^{-2}. Details of the prior assumptions applied to these parameters are explained in Table 3. For values of these three parameters, we can compute the expected line fluxes, and compare with measured values to compute the likelihood and then the posterior probability. The χ^2 statistic in our inference procedure is calculated as

$$\chi^2 = \sum_i \frac{(f_{\text{EL}_i} - R_i \cdot f_{H\beta})^2}{(\sigma_{\text{EL}_i})^2 + f_{H\beta}^2 \cdot (\sigma_{R_i})^2}. \quad (2)$$

Here EL_i denotes the set of emission lines [O II], Hγ, Hβ, [O III], Hα, and [S II]. f_{EL_i} and σ_{EL_i} represent the observed de-reddened EL_i flux and its uncertainty given a value of A_V^N. We adopt the Cardelli et al. (1989) galactic extinction curve with $R_V = 3.1$. R_i is the dust-corrected flux ratio between EL_i and Hβ, with σ_{R_i} being the associated intrinsic scatter. Note that the content of R_i varies depending on the corresponding EL_i in the following ways.

1. For $\text{EL}_i \in \{H\alpha, H\gamma\}$, R_i corresponds to the Balmer decrement. We use the intrinsic Balmer ratios of Hα/Hβ = 2.86 and Hγ/Hβ = 0.47, appropriate for case B recombination and fiducial H II region conditions (see, e.g., Hummer & Storey 1987). We set $\sigma_{R_i} = 0$, assuming that the intrinsic Balmer ratios are fixed with no scatter. These ratios are proxies of the nebular dust extinction A_V^N.
2. For $\text{EL}_i \in \{[\text{O II}], [\text{O III}]\}$, R_i and σ_{R_i} are taken from the M08 calibrations. These ratios are diagnostics of the oxygen abundance $12 + \log(\text{O/H})$.
3. For $\text{EL}_i = [\text{S II}]$, we compute the values of R_i and σ_{R_i} using the Balmer ratio of Hα/Hβ and our new calibration of the ratio [S II]/Hα, derived following Jones et al. (2015a). Jones et al. (2015a) used the same data as the low-metallicity calibrations of M08. Although M08 do

Table 4
Coefficients for the EL Flux Ratio Diagnostics Used in Our Bayesian Inference

R	c_0	c_1	c_2	c_3	c_4	c_5
Hα/Hβ	0.4564	...	...	...	...	...
Hγ/Hβ	−0.3279	...	...	...	...	...
[O III]/Hβ	0.1549	−1.5031	−0.9790	−0.0297	...	...
[O II]/Hβ	0.5603	0.0450	−1.8017	−1.8434	−0.6549	...
[S II]/Hα	−0.5457	0.4573	−0.8227	−0.0284	0.5940	0.3426

Note. The value of the EL flux ratio is calculated by the polynomial functional form, i.e., $\log R = \sum_{i=0}^{5} c_i \cdot x^i$, where $x = 12 + \log(\text{O/H}) - 8.69$. The ratio of [S II]/H$\alpha$ is calibrated by the work of Jones et al. (2015a), and the Balmer line ratios correspond to the Balmer decrement, whereas the ratios between oxygen lines and Hβ are from M08.

not provide any calibrations for [S II], we expect our calibration to be self-consistent and reliable for $12 + \log(\text{O/H}) \lesssim 8.4$. [S II]/H$\alpha$ is primarily useful as a diagnostic to differentiate between the upper and lower branches of the [O III]/Hβ calibration. This is valuable because [N II] is not directly measured and [O II] is typically of low signal-to-noise in the lower redshift galaxies for which [S II] is available, due to decreasing grism throughput at $\lambda < 0.9\ \mu m$.

4. For $\text{EL}_i = \text{H}\beta$, $R_i = 1$ with $\sigma_{R_i} = 0$. This is just comparing the observed and expected Hβ flux, corrected for dust extinction.

These EL flux ratios can be computed given a value of $12 + \log(\text{O/H})$, using a universal polynomial functional form, i.e., $\log R = \sum_{i=0}^{5} c_i \cdot x^i$, where $x = 12 + \log(\text{O/H}) - 8.69$. We summarize the coefficients (c_i) for each ratio diagnostic used in our statistical inference in Table 4.

In addition, the [N II]/Hα ratio is included as an additional parameter for galaxies at $z \lesssim 1.5$ where these lines are observed. However, we do not attempt to determine [N II]/Hα using the M08 calibrations, as locally calibrated diagnostics tend to underestimate the [N II]/Hα ratio in high-z galaxies whereas oxygen and other α-element line ratios remain constant with redshift (Steidel et al. 2014; Jones et al. 2015a; Shapley et al. 2015). Instead, we either leave this parameter free ("extended" priors), or fixed to [N II]/H$\alpha = 0.05$ ("vanilla" priors). The vanilla value is typical of galaxies with similar z and oxygen EL ratios as our sample. Fixing the value of [N II]/Hα provides faster convergence and does not significantly affect the inferred metallicity. We therefore report constraints for the "vanilla" parameter set. For galaxy ID 04054, the SN Refsdal host galaxy, we fix its [N II]/Hα to be 0.11, as measured by Yuan et al. (2011).

We use the Markov Chain Monte Carlo sampler EMCEE (Foreman-Mackey et al. 2013) to explore the parameter space, setting the number of walkers to 100 with 5000 iterations for each walker, and a burn-in of 2000. The autocorrelation times are computed for each parameter and are used to make sure chains have converged. An example constraint result for the SN Refsdal host galaxy's global EL fluxes is shown in Figure 5. While this paper is primarily concerned with gas-phase metallicities, our method simultaneously provides constraints on A_V^N and SFR, which can be compared to the results of SED fitting. We then calculate the lensing-corrected de-reddened Hα luminosity from $f_{\text{H}\beta}$ (assuming the fiducial Balmer ratio) and the magnification estimate of the source. Then the instantaneous SFR can be derived, via the
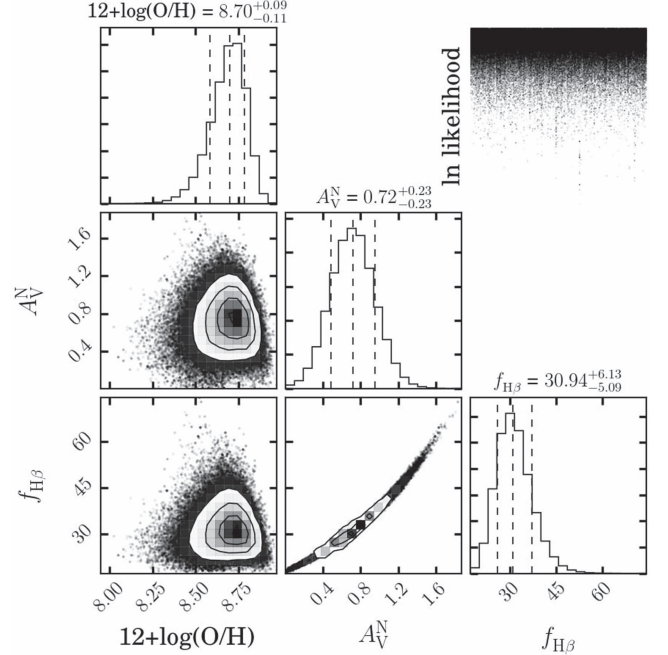
Figure 5. Marginalized 1D and 2D constraints on ($12 + \log(\text{O/H})$, A_V^N, and $f_{\text{H}\beta}$), which are drawn from the integrated line fluxes of the Refsdal host galaxy (ID 04054), after all grism exposures are combined. The values on top of each column are the medians with 1σ uncertainties for each parameter. The plot in the upper right corner illustrates a good convergence of the sampling.

commonly used calibration of Kennicutt (1998) and Moustakas et al. (2010), i.e.,

$$\text{SFR} = 4.6 \times 10^{-42} \frac{L(\text{H}\alpha)}{\text{erg s}^{-1}} \quad [M_\odot\ \text{yr}^{-1}], \tag{3}$$

which is valid for a Chabrier (2003) IMF. This SFR estimate does not rely on any assumptions about star-formation history, and thus is a more direct measure of ongoing star formation than values given by SED fitting.

We apply this inference procedure to both galaxy-integrated line fluxes (to be discussed in Section 5) as well as individual spaxels in the EL maps (which will lead to metallicity gradient measurements described in Section 6).

5. Global Demographic Properties

Based upon the Bayesian analysis, we obtain the global measurements of gas-phase metallicity, nebular dust extinction, and SFR, for all sources in our sample, as shown in Table 2. First, we check for the presence of any active galactic nucleus (AGN) contamination in our sample. Because not all galaxies
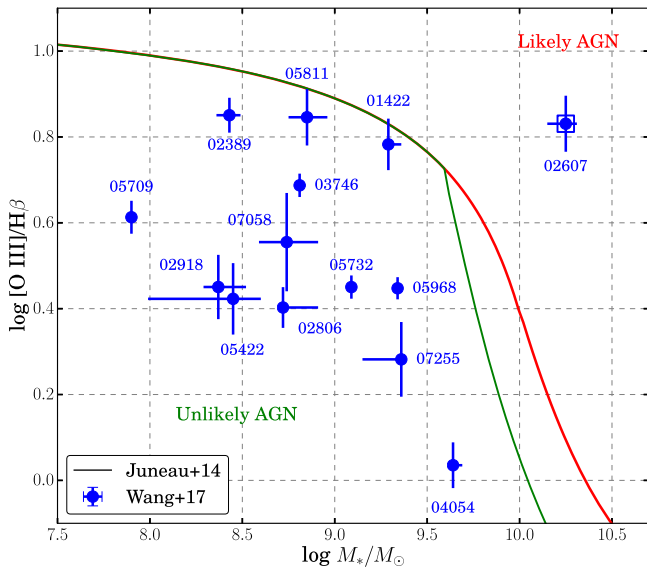
Figure 6. Mass-excitation diagnostic diagram for our metallicity gradient sample. The curves are the demarcation lines updated by Juneau et al. (2014). The regions below the green curve have a low probability ($<10\%$) of being classified as AGNs, whereas galaxies located above the red curve are secure AGN candidates. In our sample, the only AGN candidate, i.e., ID 02607, is marked by a square.

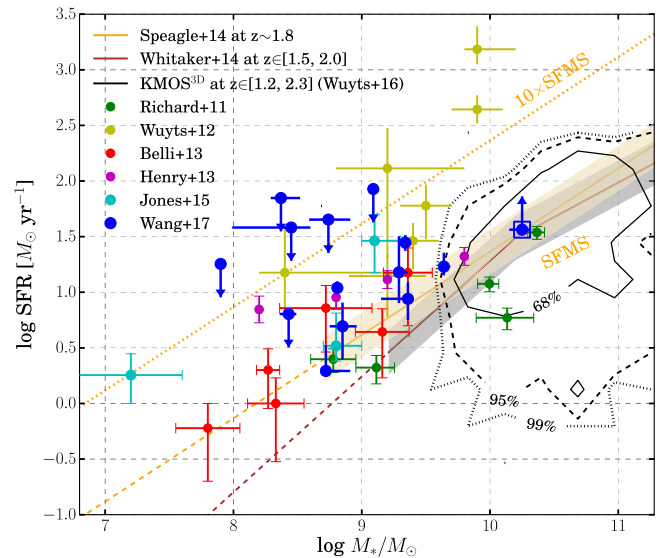**Figure 7.** Star-formation rate as a function of stellar mass for emission-line galaxies at $z \in [1.2, 2.3]$. Our metallicity gradient sample is marked by blue points; the literature sample is color-coded by reference according to the legend. The star-forming main sequence compiled by Speagle et al. (2014) and Whitaker et al. (2014) is represented by brown and orange lines, respectively, where the dotted part is extrapolated at masses below the mass completeness limit. The shaded regions denote the typical scatter of the star-forming main sequence, i.e., 0.2–0.3 dex. We also show the source density contours for the KMOS^{3D} survey at the same redshift range, where 68%, 95%, and 99% of all KMOS^{3D} galaxies at $z \in [1.2, 2.3]$ can be found within the black solid, dashed, and dotted contours respectively. It is found that our sample is highly complementary to the KMOS^{3D} sample in terms of stellar mass.

in our sample reside in the redshift range where all BPT lines are available, we resort to the mass-excitation diagram showing the [O III]/Hβ flux ratio as a function of M_*, first proposed by Juneau et al. (2011) to differentiate star-forming galaxies and AGNs in local SDSS observations. Juneau et al. (2014) further revised the demarcation scheme by applying line luminosity evolution and selection effects to a more complete SDSS galaxy sample, and tested that this scheme agrees well with the bivariate distributions found in a number of high-z galaxy samples. According to the Juneau et al. (2014) classification scheme, all but one galaxy (ID 02607) have a negligible probability of being AGNs (see Figure 6). Recently, Coil et al. (2015) found that a $+0.75$ dex M_* shift in the Juneau et al. (2014) scheme is required to avoid serious AGN contamination in the *MOSDEF* galaxy sample at $z \sim 2.3$. We verified that galaxy 02607 can still be safely classified as an AGN candidate, even considering the Coil et al. (2015) shifted classification scheme. Because the M08 calibrations originate from H II region observations and theoretical models, not designed for AGNs, our method will not predict a reliable metallicity constraint for the AGN candidate ID 02607. We thereby exclude this source in the rest of our population analysis. For the rest of our sample, comprising 13 star-forming galaxies, the constraint on gas-phase metallicity is robust.

In Figure 7, we plot the measured SFR as a function of M_* for our sample. In comparison, the loci of mass-selected galaxies observed by the KMOS^{3D} survey and the "star-formation main sequence" (SFMS, Whitaker et al. 2014; Speagle et al. 2014) at similar redshifts are also shown. We can see that selecting lensed galaxies based upon EL strength can probe deep into the low-mass regime at high redshifts, which is currently inaccessible to mass complete surveys. As expected, our emission line selected targets probe the upper envelope of the SFMS, so that a detailed comparison is non-trivial and should take into account the selection functions of each sample. The advantage of the emission line selection technique is that it is the most cost-effective way to obtain gas metallicities for

stellar masses down to $10^7 M_*$ at $z \sim 2$. Using the same technique, we discovered an analog of local group dwarf spheroidals (e.g., Fornax, Coleman & de Jong 2008) at $z = 1.85$ experiencing active star formation (with 1 dex offset from the SFMS), in our previous work (Jones et al. 2015b).

We also collect all published gas-phase metallicity measurements in emission-line galaxies in the redshift range of our sample $1.2 \lesssim z \lesssim 2.3$, obtained exclusively from the strong EL calibrations prescribed by M08. The reason we only select measurements based upon the M08 calibrations is that adopting different strong EL calibrations can give rise to different absolute metallicity scales offset by up to ~ 0.7 dex at the high-mass end, according to the comparative study conducted by Kewley & Ellison (2008). In order to minimize the systematic uncertainty associated with EL calibrations, we thus focus on this homogeneous M08 sample alone.

As shown in Figure 8, the M08-based sample includes five galaxies from Richard et al. (2011), seven from Wuyts et al. (2012), six from Belli et al. (2013), four stack measurements by Henry et al. (2013), three interacting galaxies at $z = 1.85$ analyzed by Jones et al. (2015b), and 13 star-forming galaxies presented in this work. In total, this sample consists of 38 independent measurements at $1.2 \lesssim z \lesssim 2.3$, with a median redshift of $z_{\mathrm{median}} = 1.84$. The mutual agreement between our sample and that from the literature provides an independent confirmation that our new Bayesian method leads to constraints on gas-phase metallicity consistent with previous studies also adopting the same calibrations. With the combined sample, we are able to derive the MZR at this redshift, spanning three orders of magnitude in stellar mass.
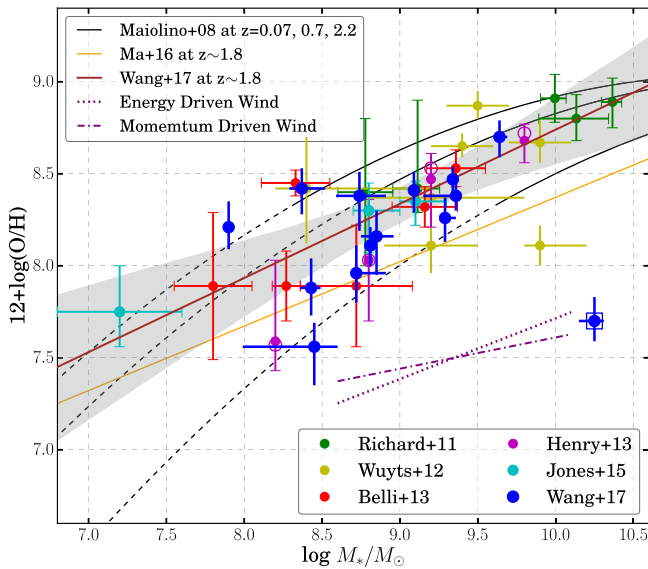
Figure 8. Relation between stellar mass and gas-phase oxygen abundance, from observations and simulations at $z \sim 1.8$. The color-coding for all the points is the same as in Figure 7. For the sake of consistency, all metallicity measurements are derived assuming the Maiolino et al. (2008) strong EL calibrations. The black curves represent the second order polynomials fitted by Maiolino et al. (2008) to data sets at different redshifts, where dashed parts correspond to extensions beyond mass completeness limits of those data sets. The thick brown line denotes our best-fit linear relation based upon all data points except the AGN candidate (i.e., ID 02607) marked by a square. The shaded band marks the 1σ confidence region. For the Henry et al. (2013) stack data points, we also show the measurements without dust correction as open circles, which are not included in the fit. The thin orange line is the prediction from the cosmological zoom-in simulations conducted by Ma et al. (2016a) at the same redshift range. The purple dotted and dashed–dotted lines illustrate the slopes given by the energy and momemtum driven models, respectively.

We fit the following functional form to this sample of 38 galaxies in order to derive the MZR, i.e.,

$$12 + \log(\text{O/H}) = \alpha \log M_*/M_\odot + \beta + N(\sigma), \qquad (4)$$

where α, β, and σ represent the slope, the intercept and the intrinsic scatter, respectively. The Python package LINMIX[27] was employed to perform a Bayesian linear regression following the method proposed by Kelly (2007). The median values and 1σ uncertainties for these three parameters drawn from the marginalized posteriors are

$$\alpha = 0.40^{+0.07}_{-0.07}, \quad \beta = 4.71^{+0.62}_{-0.64}, \quad \sigma = 0.03^{+0.02}_{-0.01}. \qquad (5)$$

The resulting MZR together with its 1σ confidence region is shown in Figure 8.

Recently, Guo et al. (2016) presented a comprehensive study of the MZR and its scatter at $z = 0.5$–0.7 from data acquired by large spectroscopic surveys in the *CANDELS* fields (Grogin et al. 2011; Koekemoer et al. 2011). Guo et al. (2016) also summarized a variety of theoretical predictions of the MZR slope (i.e., α) given by diverse approaches. The slopes predicted by the energy-driven wind ($\alpha \sim 0.33$) and the momentum-driven wind ($\alpha \sim 0.17$) equilibrium models proposed by Finlator & Davé (2008) and Davé et al. (2012) are shown schematically in Figure 8. We see that our inferred slope is only marginally compatible with the prediction given by the energy-driven wind model at 1σ confidence level, whereas strongly disfavors the momentum-driven wind model. Considering the majority of the

galaxies in our sample have relatively lower masses, our result confirms the finding by Henry et al. (2013) that momentum-driven winds cannot be the primary workhorse that shapes the MZR in the low-mass regime (below $10^{9.5} M_\odot$) at $z \gtrsim 1$.

Our constrained value of α is consistent at a 1σ confidence level with that given by Ma et al. (2016a), from the FIRE (Feedback in Realistic Environment) cosmological zoom-in simulations (Hopkins et al. 2014). Evaluated at $z = z_{\text{median}}$, their MZR reads $12 + \log(\text{O/H}) = 0.35 \log M_*/M_\odot + 4.87$. The FIRE simulations reaches a spatial resolution as high as 1–10 pc, which is one to two orders of magnitude better than that in conventional large-volume cosmological hydrodynamic simulations (e.g., Davé et al. 2013). The high spatial resolution enables a more realistic and self-consistent treatment of multiphase ISM, star formation, galactic winds, and stellar feedback, than those in semi-analytic models (see Somerville & Davé 2015, for a recent review). As a result, these zoom-in simulations are capable of reproducing many observed relations, e.g., the galaxy stellar mass functions, the evolution of cosmic SFR density and specific SFR (sSFR). The consistency between the slopes of our MZR and that by Ma et al. (2016a) indicates that small-scale physical processes are important in producing the cumulative effects of galactic feedback. However, we note that there is a 0.16 dex offset in the values of the MZR intercept (β). One possible cause is that the absolute metallicity scale in the M08 calibrations is offset. As shown by Kewley & Ellison (2008), the intercepts given by different metallicity calibrations can vary up to 0.7 dex. Other potential causes lie in the assumptions adopted by the Ma et al. (2016a) simulations, e.g., the stellar yield is too small, the gas fractions in their simulated galaxies are too high, etc. Nevertheless, a discrepancy of 0.1–0.2 dex is not unusual given the uncertainties in stellar yield and metallicity calibrations.

Aiming at testing the validity of the FMR using the M08-based sample at $z \sim 1.84$ compiled in this work, we calculate the predicted values for metallicity from the measurements of M_* and SFR, according to Equation (2) in Mannucci et al. (2011), who extended the FMR to masses down to $10^{8.3} M_\odot$. Figure 9 shows the difference between the FMR-predicted values and the actual measurements of metallicity. We find that the entire M08-based sample is consistent with the FMR given the measurement uncertainties and intrinsic scatter. The median offset for the entire sample is 0.07 dex, with a median uncertainty of 0.14 dex. The median offset and uncertainty for our galaxy sample analyzed in this work are −0.07 dex and 0.12 dex, respectively. Among other individual data sets, the only one that shows a marginally significant deviation from the FMR is that by Wuyts et al. (2012), which has a median offset of 0.22 dex and a median uncertainty of 0.11 dex. We speculate that it is due to the fact that the Wuyts et al. (2012) data set is the only one in our compiled M08-based sample that relies solely upon the [N II]/Hα flux ratio, which can be a biased tracer of metallicity at high redshifts (Sanders et al. 2015; Shapley et al. 2015). This result reiterates the necessity of combining multiple EL flux ratio diagnostics simultaneously in order to suppress the systematics associated with local calibrations in the accurate measurement of high-z metallicity.

6. Spatially Resolved Analysis

In this section, we present and discuss in depth the new and unique results obtained from the spatially resolved analysis of our sample, beyond the reach of the conventional integrated measurements in Section 5. The high spatial resolution maps of

---

[27] Available at https://github.com/jmeyers314/linmix.
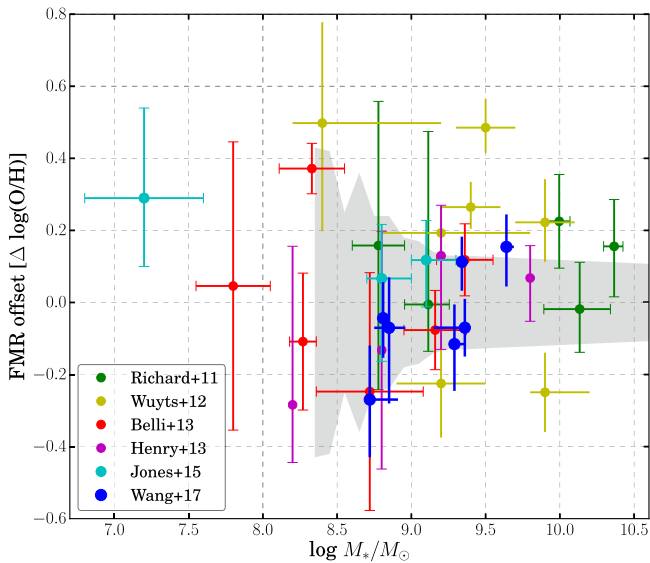
Figure 9. Offset from the fundamental metallicity relation first proposed by Mannucci et al. (2010) for the $z \sim 1.8$ metallicity measurements derived assuming the Maiolino et al. (2008) calibrations. The color-coding for the points is the same as in Figure 7. The shaded region shows the intrinsic scatter of the FMR extended to low-mass regime given by Mannucci et al. (2011).

gas-phase metallicity from *HST* grism spectroscopy are described in Section 6.1. We make notes on individual galaxies in Section 6.2. The maps of EL kinematics on part of our sample are presented in Appendix A.

6.1. Gas-phase Metallicity Maps at Sub-kiloparsec Resolution

To estimate the spatial distribution of the constrained parameters, we apply the analysis described in Section 5 to each individual spaxel in the EL maps. In Figure 10, we show the maps of best-fit metallicity and the corresponding conservative uncertainty (the larger side of asymmetric error bars given by the Bayesian inference) measured for galaxies in our metallicity gradient sample. We bin spaxels 2 by 2 to regain the native WFC3/IR pixel scale. In the maps, we only include spaxels with at least one EL (primarily $H\alpha$ or [O III]) detected with $\geqslant 5\sigma$ significance. The metallicity gradients measured from these individual spaxels are shown in the right column of Figure 10. We also measure metallicity gradients using another independent method, i.e., via radial binning. The range of each radial annulus is determined by requiring its S/N $\gtrsim 15$. In order to avoid biasing toward low metallicities, we put the S/N threshold on $H\alpha$ whenever available (i.e., $z \lesssim 1.5$) instead of on [O III], since the latter correlates tightly with $12 + \log(O/H)$, in the lower branch of [O III]/$H\beta$, where the majority of our sample resides. Based upon our S/N threshold criterion, three sources in our sample, IDs 02918, 05422, and 07058, do not have enough spaxels to constitute a trustworthy metallicity map, and therefore were removed from the spatially resolved analysis. So in total, we present 10 star-forming galaxies with sub-kiloparsec resolution metallicity maps. In order to measure radial gradients, we perform Bayesian linear regressions in the same manner as described in Section 5. We take into account the conservative uncertainties for individual metallicity constraints in each spaxel or annulus. The resulted metallicity gradients are given in Table 5. We note that the metallicity maps are not always symmetric around the center of stellar mass. Therefore, while gradients are a useful statistic to describe the overall behavior and to compare with numerical

simulations, they do not contain all the available information about the apparent diversity of morphologies. Thus, it is helpful to keep the 2D maps in mind while interpreting the gradients.

As a result, seven galaxies have gradients consistent with being flat at 2σ, among which three show almost uniformly distributed metals (IDs 02806, 03746, and 05968), two have marginally positive gradients (IDs 01422 and 05732), and the other two display mildly negative gradients (IDs 02389 and 04054). Apart from these, the other three galaxies in our sample have very steep negative gradients: IDs 07255, 05709, and 05811. In particular, after the lensing magnification correction, the stellar mass of galaxy ID 05709 is estimated to be $\sim 10^{7.9} \, M_\odot$. This is the first time a sub-kiloparsec scale spatially resolved analysis has been done on such low-mass systems at high redshifts.

In order to verify that our results are not contaminated by ionizing radiation from AGNs, we examine spatially resolved BPT diagrams. This approach is only possible for sources at $z \lesssim 1.5$, where $H\alpha$ is detectable. Furthermore, due to the low spectral resolution of *HST* grisms, we slightly modify the BPT diagram to plot a flux ratio of [O III]/$H\beta$ as a function of [S II]/$H\alpha$ + [N II], using the "vanilla" value for [N II]/$H\alpha$. As Figure 11 shows, we do not identify any strong correlation between source plane galactocentric radius and deviation from the H II region loci, suggesting that there are no AGNs hidden in the center of these galaxies. Moreover, the integrated EL fluxes of all three galaxies lie in close proximity to the extreme starburst model prescribed by Kewley et al. (2006), which confirms our working assumption that AGN contributions are negligible for our sample, except in one case (i.e., ID 02607, see Figure 6).

We also test our metallicity estimates using the pure empirical EL calibrations based on a large sample of "direct" metallicity determinations in SDSS local H II regions recently published by Curti et al. (2016), replacing the M08 hybrid EL calibrations. It is found out that our results on integrated metallicities and metallicity gradients of the galaxies in our sample remain unaffected.

The radial range where metallicity gradients are measured is critical since H II regions at far outer/inner regions of galaxies are found to show significantly elevated/truncated oxygen abundances (see, e.g., Sánchez et al. 2014). A reasonable radial range for metallicity gradient measurements is believed to be between the disk effective radius (Freeman 1970; Sánchez et al. 2014) and the optical radius (Mollá & Díaz 2005; Mollá et al. 2016c), i.e., roughly twice the size of the disk effective radius. We verified that all of our metallicity gradients are derived in this radial range, which enhances the significance of the comparative studies (of our measurements at least) presented in Sections 6.1.1 and 6.1.2.

6.1.1. Cosmic Evolution of Metallicity Gradients

We plot the evolution of the metallicity gradient across cosmic time in Figure 12. Here we include all existing high-z gradient measurements obtained with sufficiently high spatial sampling, i.e., finer than kpc resolution in the source plane. Together with the 10 new metallicity gradient measurements from the individual spaxel method presented here, in Figure 12, we show one highly magnified galaxy of an interacting triple at $z = 1.85$ analyzed in our previous work (Jones et al. 2015b). We also include 11 measurements by Leethochawalit et al. (2015; 5 isolated and 6 merging), 4 by Jones et al. (2013; 3 isolated and 1 merging),
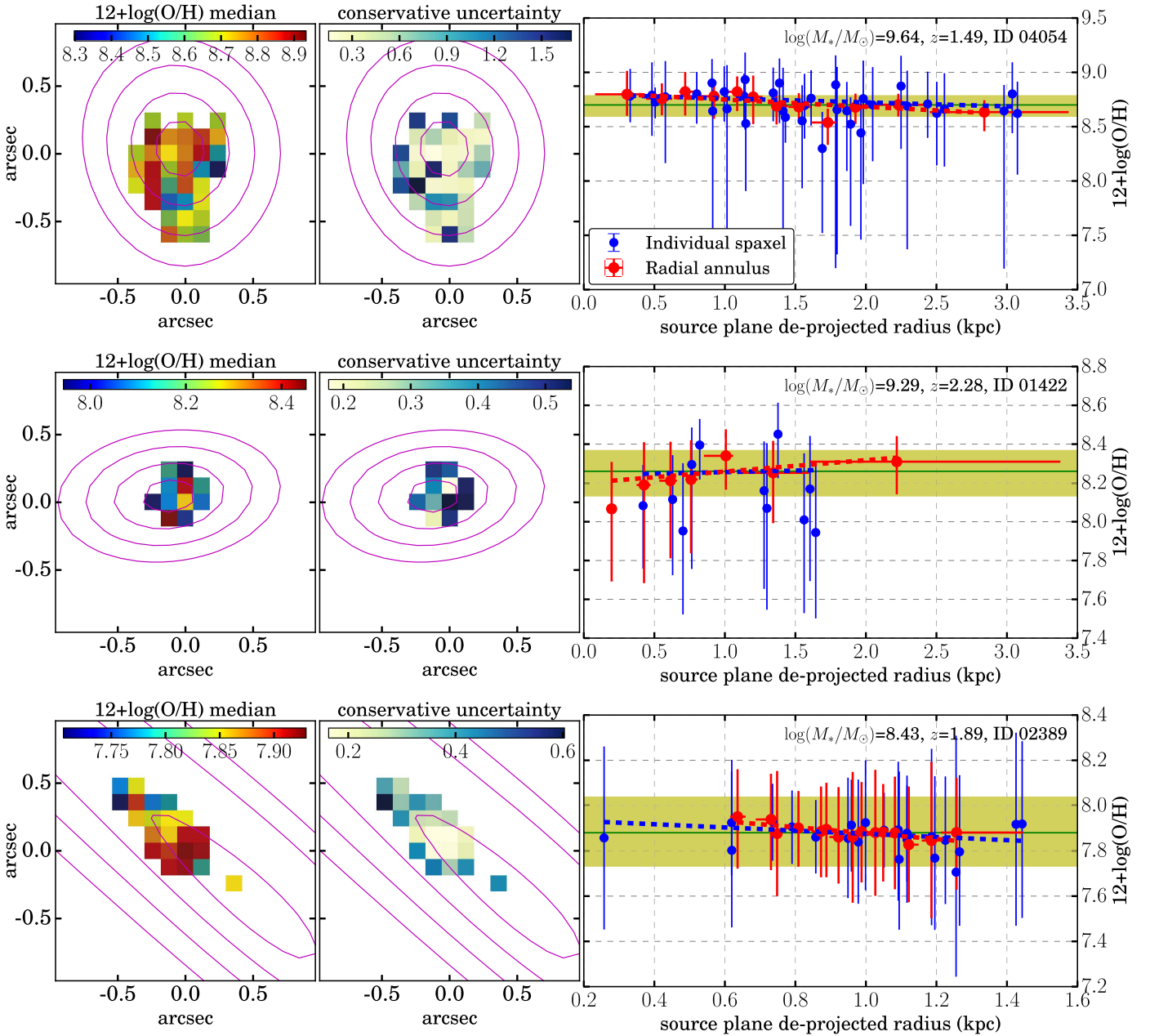
Figure 10. Maps of metallicity constraints (median value in the left column and conservative uncertainty—the larger side of asymmetric 1σ error bars—in the center column) and plots of radial metallicity gradients in the right column. Note that unlike what we show in the combined EL maps (Figures 3 and 4), here the metallicity maps are rebinned to a scale of $0\rlap{.}''13$, corresponding to the native resolution of WFC3/IR. In the maps, only spaxels with the signal-to-noise ratio of Hα or [O III] larger than five are shown. The same source plane de-projected galactocentric radii that are denoted by black contours in Figures 3 and 4 are represented by magenta contours, with the only difference being that contours are in 1 kpc intervals now. In the panels in the right column, blue and red points correspond to metallicity measurements for individual spaxels and radial annuli, respectively. The radial gradients derived based upon these measurements are shown by the dashed lines in corresponding colors. The yellow band and green horizontal line mark the global constraint on $12 + \log(\mathrm{O/H})$, from integrated line fluxes.

1 sub-L_* post-merger at $z \sim 1$ by Frye et al. (2012), and 7 data points from Swinbank et al. (2012; 5 isolated and 2 interacting). The scatter of these high-resolution gradient measurements is ~0.12 dex and ~0.22 dex at $z \sim 1.5$ and $z \sim 2.3$ respectively. The spread of recent gradient measurements from the seeing-limited KMOS[3D] survey is also overlaid in Figure 12. Note that the PSF FWHM given by the median seeing of their observations is $0\rlap{.}''6$, which results in a 5 kpc resolution at $z \sim 2$, twice the size of the half-light radius of an L_* galaxy at this redshift. We thus only focus on interpreting the sub-kiloparsec resolution gradient measurements, in order to avoid possible biases toward null values associated with poor spatial sampling (see, e.g., Yuan

et al. 2013). Although corrected for beam smearing, the KMOS[3D] gradient measurements still display small scatter than that of the high-resolution results. When possible, we also divide each data set in terms of sources being isolated or kinematically disturbed (i.e., undergoing merging). In part of our sample, signatures of post-merger tidal remnants can be clearly identified, which strongly indicates that gravitational interaction plays a key role in shaping the spatial distribution of metals in these galaxies (IDs 02806, 03746, and 05968).

In order to compare theoretical predictions with observations, we incorporate into Figure 12 the predictions for the metallicity gradient cosmic evolution from canonical chemical
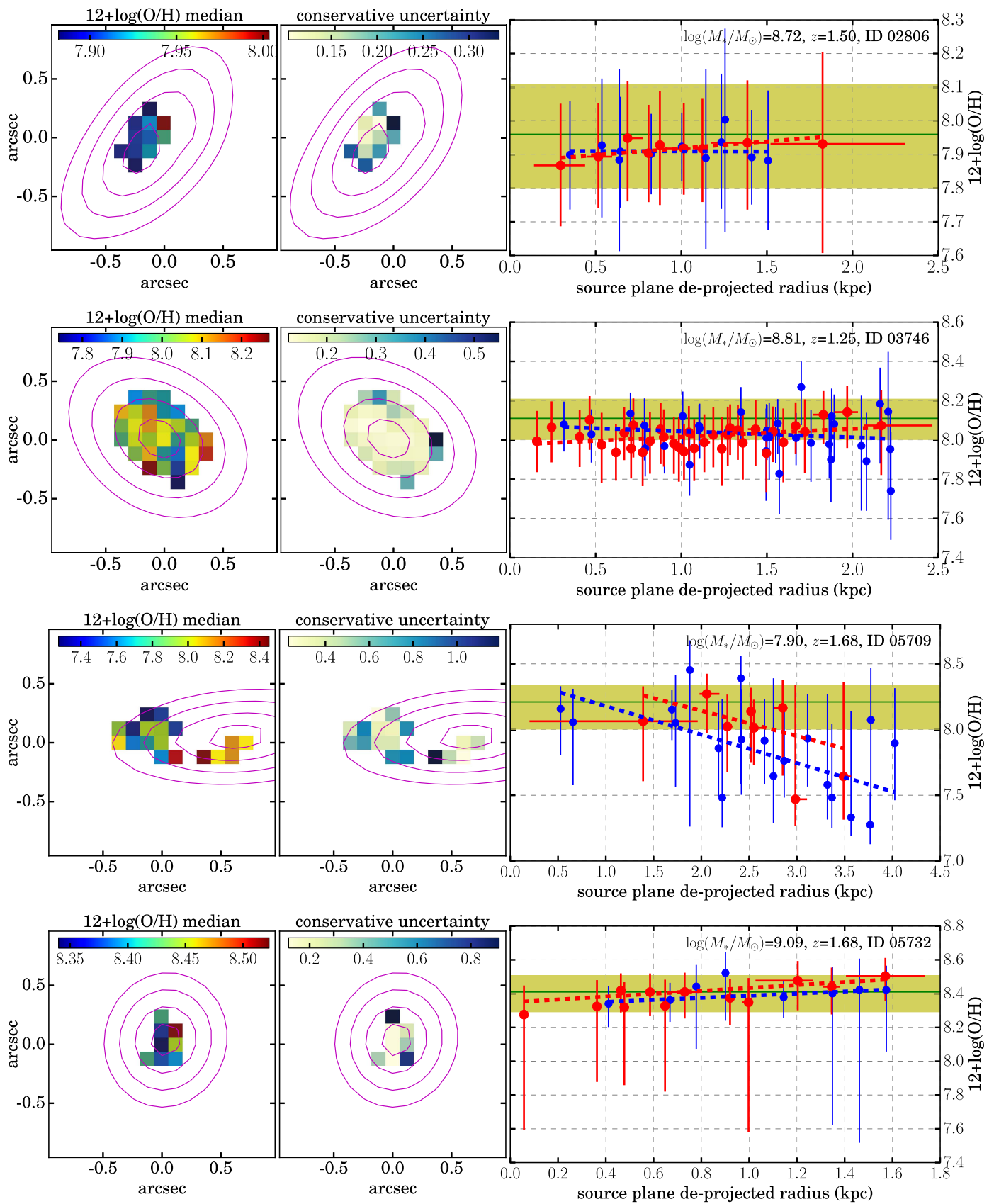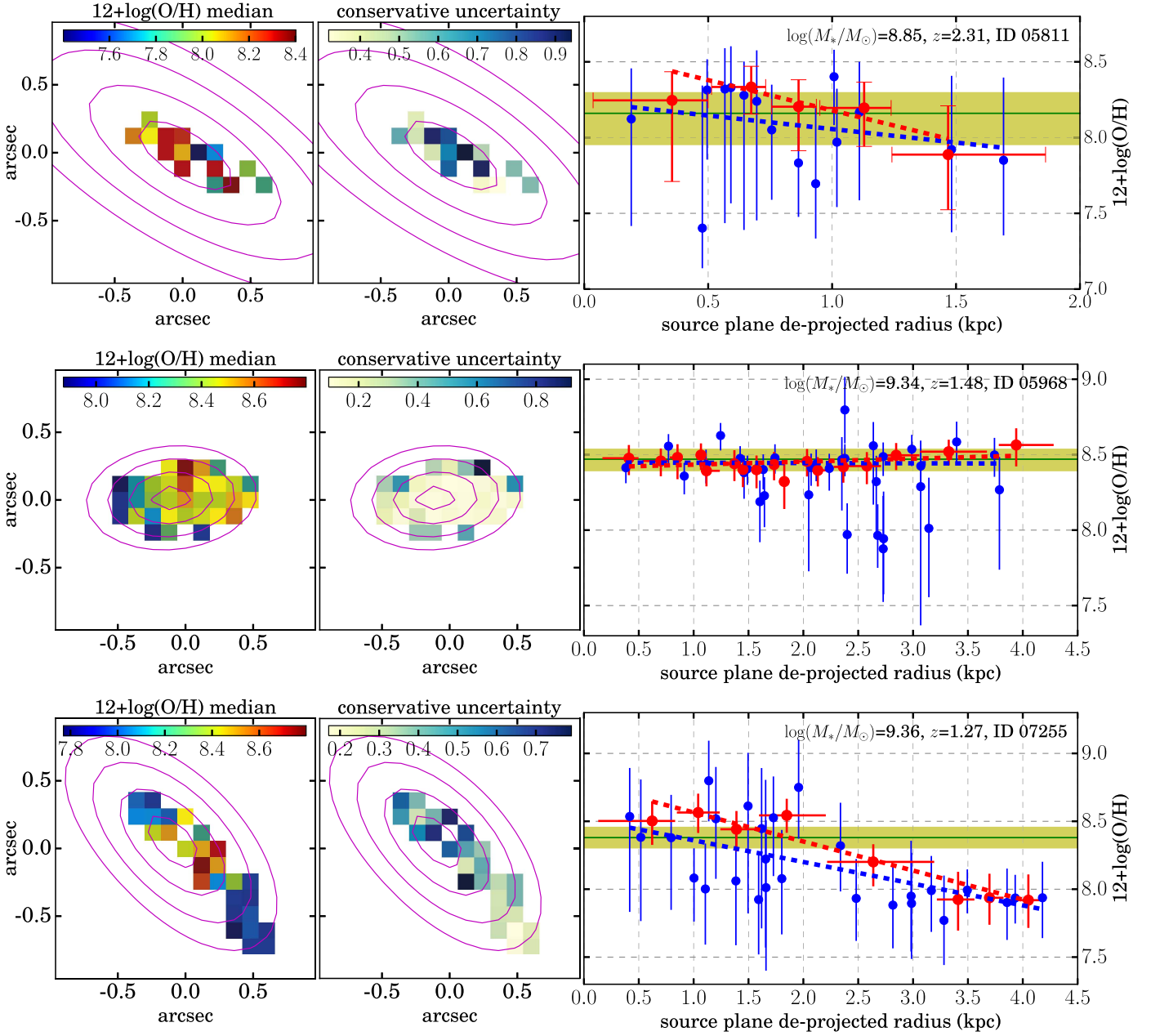
Figure 10. (Continued.)

Figure 10. (Continued.)

Table 5
Gas-phase Metallicity Gradients Measured by Two Different Methods

ID	Metallicity Gradient (dex kpc^{-1})	
	Individual Spaxel	Radial Annulus
01422	0.02 ± 0.08	0.06 ± 0.04
02389	-0.07 ± 0.04	-0.13 ± 0.03
02806	-0.01 ± 0.02	0.04 ± 0.02
03746	-0.03 ± 0.03	0.04 ± 0.02
04054	-0.04 ± 0.02	-0.07 ± 0.02
05709	-0.22 ± 0.05	-0.19 ± 0.06
05732	0.06 ± 0.05	0.08 ± 0.02
05811	-0.18 ± 0.08	-0.40 ± 0.07
05968	-0.01 ± 0.02	0.02 ± 0.01
07255	-0.16 ± 0.03	-0.21 ± 0.03

evolution models (CEMs) based upon the "inside-out" disk growth paradigm given by Chiappini et al. (2001, C01) and Mollá et al. (2016a, 2016b, M16). C01 constructed a two-phase accretion model for galaxy formation. The first infall period corresponds to the formation of the dark matter halo and the thick disk, with an exponentially declining infall rate at a fixed e-folding timescale ($\sim$1 Gyr). In this phase, the infall of pristine gas is rapid and isotropic, giving birth to a relatively extended profile of star formation. The thin disk is formed in the second phase, where gas is preferentially accreted to the periphery of the disk, with the e-folding timescale proportional to galacto-centric radius (vis-á-vis the "inside-out" growth). The foundation of M16 is Mollá & Díaz (2005), who proposed their prescription by treating the ISM as a multiphase mixture of hot diffuse gas and cold condensing molecular clouds, and
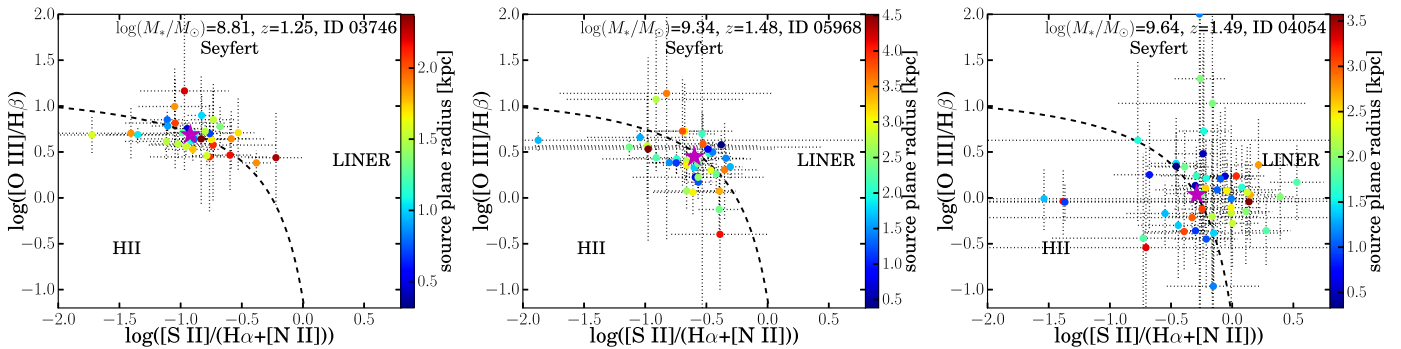
Figure 11. Spatially resolved BPT diagrams for three representative sources in our metallicity gradient sample. The points in each panel correspond to spaxels in the combined EL maps for each source, color coded by the source plane de-projected galactocentric radius. Note that similar to what we show in Figure 10, all spaxels represented by colored points here are rebinned to recover the native WFC3/IR pixel scale (0″.13). The magenta star denotes the position where the entire galaxy would lie, calculated from integrated EL fluxes, with the length of the error bar comparable to the size of the symbol. The dashed curve is adapted from the extreme starburst scenario given by Kewley et al. (2006), assuming the "vanilla" value for [N II]/Hα. Given the 1σ uncertainties, all the spaxels are broadly consistent with being H II regions.

calibrating against different star-formation efficiencies. Popular nucleosynthesis prescriptions for SNe Ia/II and AGB stars are employed by both to construct the yields of dominant isotopes (e.g., H, He, C, N, O, Fe). Improving over what is done in the Mollá & Díaz (2005) multiphase CEMs, M16 adopted newly calibrated gas infall rates from Mollá et al. (2016c) and a variety of different prescriptions for H I-to-H_2 conversion, with the best being the ASC technique (Y. Ascasíbar et al. 2017, in preparation). We thus show this model solution in the black dashed–dotted curve. As shown in Figure 12, the majority of the high-z measurements are generally compatible with both model predictions, with the M16 model showing better consistency, largely ascribed to its more accurate assumptions of ISM structure and more realistic treatment of star formation. Note that these CEMs do not take galactic feedback or radial gas flows into account. So without the consideration of strong physical mechanisms that stir and mix up the enriched gas with relatively unenriched ambient ISM, conventional analytical models are still able to match the metallicity gradients observed at high redshifts.

In addition to the predictions from these analytical models, we also compare our measurements with results from numerical simulations. First, we show the fiducial representative (i.e., the realization g15784) of the McMaster Unbiased Galaxy Simulations (MUGS, Stinson et al. 2010) using the gravitational N-body and SPH code Gasoline. This simulation employed the "conventional" feedback scheme, i.e., depositing $\sim$10%–40% kinetic energy from SN explosions into heating ambient ISM. Furthermore, it adopted a high star-formation threshold (gas particle density $>1\,cm^{-3}$), which makes star-forming activities highly concentrated in the galaxy center at early stages of disk formation. As a result, a steep gradient (i.e., $\gtrsim -0.2$ dex kpc^{-1}) is persisting at $z > 1$ and then significantly flattens with time. Although greatly offset from most of the measurements (and other theoretical predictions) as seen in Figure 12, this particular evolutionary track is actually quite consistent with some of the observed steep gradients (galaxies 07255, 05709, and 05811 in our analysis). In comparison, the same galaxy (g15784) is re-simulated with an enhanced feedback prescription of sub-grid physics as part of the Making Galaxies In a Cosmological Context (MaGICC, Stinson et al. 2013) program. A factor of 2.5 difference in the energy output rate from SN feedback diverges the evolutionary track of metallicity gradients given by the MUGS and MaGICC simulations, making the evolution of the MaGICC

gradient less dramatic and more similar to the M16 model solution. This demonstrates that amplifying the feedback strength can significantly flatten the metallicity gradients in early star-forming galaxies (Gibson et al. 2013). This is actually adopted by some recent studies (Leethochawalit et al. 2015; Wuyts et al. 2016) to explain their metallicity gradient measurements, which are also shown in Figure 12. However, as we have already mentioned, this is not the only explanation possible.

Lastly, in Figure 12, we also show another set of numerical results, from an ensemble of 19 galaxies in different environments (9 in the isolated field and 10 in loose groups), simulated by the RAMSES Disk Environment Study (RaDES, Few et al. 2012) using the adaptive mesh refinement code RAMSES. Unlike what is shown for the MUGS and MaGICC simulations (i.e., just one single realization), the full range of the whole suite of RaDES simulations is marked in Figure 12, which provides a statistical comparison sample, though the scatter of the RaDES simulations cannot capture the scatter seen in actual measurements. Generally speaking, the evolutionary trends of the abundance gradients of RaDES galaxies lie somewhat in-between the two tracks of g15784 in MUGS and MaGICC simulations, but are still quite consistent with the majority of the high-z gradient measurements. Because no mass loading is assumed in the RaDES simulations, outflows do not play any role in this observed consistency. Moreover, since a lower star-formation threshold is used ($>0.1\,cm^{-3}$), which is similar to what the analytical CEMs assumed, star formation in RaDES galaxies occurs more uniformly in the early disk-formation phase. This strongly suggests that having more extended star formation can also result in galaxies possessing shallower gradients, confirming the tight link between star-formation profile and metallicity gradient (Pilkington et al. 2012).

6.1.2. Mass Dependence of Metallicity Gradients

In Figure 13, we plot a metallicity gradient as a function of stellar mass for a subset of all gradient measurements shown in Figure 12, where reliable M_* estimates are available. Notably, in the low-mass regime of $M_* \lesssim 10^9\,M_\odot$, all the current sub-kiloparsec resolution metallicity gradient measurements at $z \gtrsim 1$ are obtained from our work, and our measurements even extend to below $10^8\,M_\odot$. This illustrates the power of combining the space-based grism spectroscopy with gravitational lensing as a means to recover reliable metal abundance distributions in early star-
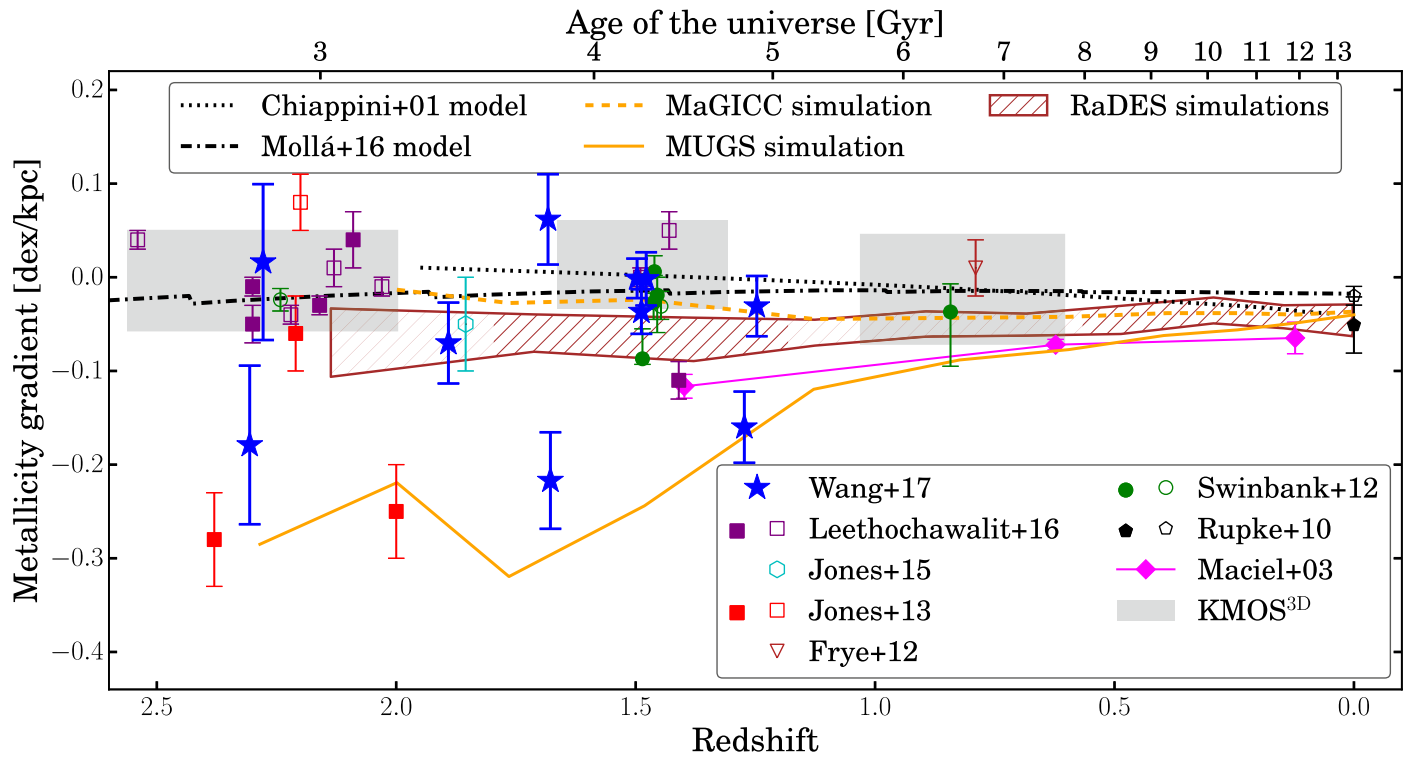
Figure 12. Evolution of metallicity gradients with redshift. The blue stars represent the metallicity gradients measured in this work, and the cyan hexagon is from our previous work (Jones et al. 2015b). We include all the high-z metallicity gradient measurements with sub-kiloparsec resolution in the current literature: lensed galaxies analyzed by Frye et al. (2012), Jones et al. (2013), and Leethochawalit et al. (2015), non-lensed galaxies observed with AO (Swinbank et al. 2012). The spread of the KMOS[3D] measurements obtained with seeing-limited conditions (Wuyts et al. 2016) is marked as gray shaded regions. We also show the averages of local metallicity gradients (Rupke et al. 2010), and the trend of the Milky Way gradient evolution based upon planetary nebula estimates (Maciel et al. 2003). Except for the blue stars, hollow and filled symbols correspond to interacting/merging and isolated systems, respectively. In addition, the predictions from different analytical chemical evolution models and numerical simulations are also shown as comparisons to the observational results (see Section 6.1 for more details).
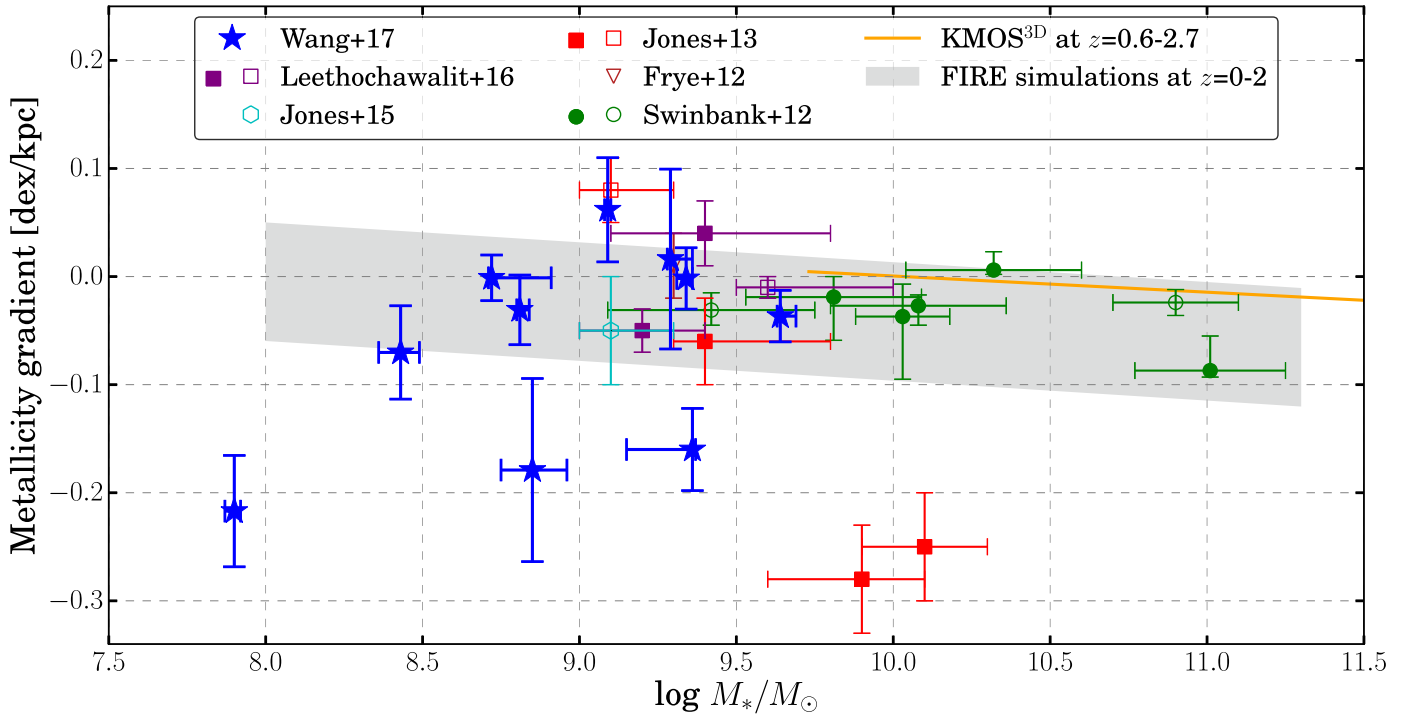
Figure 13. Observed metallicity gradients as a function of M_*. The color-coding of symbols is the same as in Figure 12, except that the blue stars, hollow, and filled symbols correspond to interacting/merging and isolated systems, respectively. Most of the Leethochawalit et al. (2015) gradient measurements do not have reliable M_* estimates due to the lack of near-IR photometry. The orange line shows a simple linear fit to the metallicity gradient results at $z = 0.6$–2.7 from the KMOS[3D] observations, only covering the stellar mass range of $\log(M_*/M_\odot) \in [9.7, 11.5]$ (Wuyts et al. 2016). The 2σ interval of the FIRE simulation results presented by Ma et al. (2016b) is shown as the shaded region. The only work to date that presents metallicity gradient measurements in the $M_* \lesssim 10^9 M_\odot$ regime is this work, which even extends to below $10^8 M_\odot$.

forming disks in the low-mass regime. From these measurements, we can tentatively observe an intriguing correlation between M_* and metallicity gradient: more massive galaxies tend to have shallower (less negative) gradients. This is consistent with the galaxy formation "downsizing" picture (see, e.g., Brinchmann et al. 2004; Fontanot et al. 2009) such that more massive galaxies are more evolved and their metal distributions are flatter since they are in a later phase of disk mass assembly where star formation occurs in a more coherent mode throughout the disk. This correlation between mass and gradient slope has also been seen in a variety of theoretical studies including the RaDES simulations (Few et al. 2012) and the multiphase CEMs by (Mollá & Díaz 2005).

Several groups of authors have reported a characteristic oxygen abundance gradient for nearby disk galaxies, such that the normalized gradient in units of dex per scale radius is approximately invariant with galaxy global properties (M_*, morphology, etc.; e.g., Sánchez et al. 2014; Bresolin & Kennicutt 2015; Ho et al. 2015). Consequently, larger and more massive galaxies have shallower gradients in physical units (dex kpc^{-1}; e.g., Ho et al. 2015), as we tentatively see in Figure 13. We verify that 8/10 of our gradient measurements, if expressed in terms of disk effective radii, are consistent with the common gradient slope reported by Sánchez et al. (2014). This common normalized abundance gradient may reflect a more fundamental relation between local surface mass density and metallicity (e.g., Bresolin & Kennicutt 2015). Both of these quantities decrease more rapidly with radius in smaller galaxies, in line with our "downsizing" picture of metallicity gradients. Downsizing is further reinforced by the evolution in the size–mass relation recovered by van der Wel et al. (2014).

As a comparison, we include the linear fit to the low spatial resolution metallicity gradient measurements from the KMOS^{3D} survey in the similar z range (see Figure 6 in Wuyts et al. 2016). Since the mass completeness limit for the KMOS^{3D} data set is $\sim 10^{9.7} M_\odot$, our analysis is highly complementary to the KMOS^{3D} result in terms of M_*, albeit our metallicity gradients are measured at much higher spatial resolution. However, the KMOS^{3D} fitting relation deviates systematically from the "downsizing" mass dependency of metallicity gradients and is also in conflict with some sub-kiloparsec resolution gradient measurements obtained by Jones et al. (2013) at an over 3σ confidence level. This can be attributed to possible biases associated with the low spatial resolution of the KMOS^{3D} observations and the small mass range of their sample.

We also show the 2σ spread of the galaxy metallicity gradients from the FIRE cosmological zoom-in simulations (Ma et al. 2016b). This suite of simulations has been used to study the MZR as we discussed in Section 5 (Ma et al. 2016a). The simulations show a diversity of gradient behaviors at four redshift epochs ($z = 0, 0.8, 1.4, 2.0$). At $z = 2$, the scatter of their simulation results can reach 0.2 dex when radial gradients are measured in the central 2 kpc region, quite consistent with the scatter (~ 0.22 dex) seen in high-resolution gradient measurements at $z \sim 2.3$. However, their simulations still cannot reproduce some of the steep gradient measurements, especially some of our results in the low-mass regime. According to Ma et al. (2016b), galactic feedback from intense starburst episodes on a 0.1–1 Gyr timescale can effectively flatten metallicity gradients. Hence for the galaxies measured with steep metallicity gradients, galactic feedback must be of

ID	log ($M_*{}^a/M_\odot$)	Instrument	Emission Feature	$\sigma_{obs}{}^b$ (km s^{-1})	$\sigma_{int}{}^c$ (km s^{-1})
02389	$8.43^{+0.06}_{-0.07}$	*MUSE*	C III]	63 ± 4	19^{+11}_{-19}
02806	$8.72^{+0.19}_{-0.01}$	*MUSE*	[O II]	42 ± 3	24 ± 5
03746	$8.81^{+0.03}_{-0.01}$	*MUSE*	[O II]	74 ± 9	63^{+10}_{-11}
04054	$9.64^{+0.05}_{-0.01}$	*MUSE*	[O II]	68 ± 6	59 ± 7
05709	$7.90^{+0.02}_{-0.03}$	KMOS	[O III]	41 ± 7	17^{+13}_{-17}
05968	$9.34^{+0.02}_{-0.03}$	KMOS	[O III]	85 ± 6	76 ± 7

Notes.
[a] Values taken from Table 2.
[b] Not corrected for instrument resolution.
[c] Corrected intrinsic velocity dispersion.

limited influence, at least in affecting metal distribution in H II regions on such a short timescale.

In short, we observe a tentative mass dependence of metallicity gradients that low-mass galaxies have steeper gradients compared with high-mass galaxies at high redshifts. A larger set of accurately measured and uniformly analyzed metallicity gradients well sampled in M_* (in the regime of $M_* \lesssim 10^9 M_\odot$ in particular) is needed in order to fully explore the validity of the metallicity gradient "downsizing" picture and whether galactic feedback plays a crucial role in altering this picture.

6.2. Notes on Individual Galaxies

In this section, we highlight some noteworthy sources in our metallicity gradient sample. For reference, all the measurements of global properties, metallicity gradients, and kinematic decompositions in these sources are presented in Tables 2, 5, and 6, respectively.

02389: This is the most highly magnified source in our sample, with $\mu = 43.25^{+7.03}_{-5.78}$, given by the SHARON & JOHNSON version 3 model. It is one of two connected double images that straddle the lensing critical curve at a source redshift of $z = 1.89$, forming a highly sheared arc more than 5 arcsec. This is confirmed by the velocity shear of ~ 50 km s^{-1} detected in both components of the fold arc. After lensing correction, the stellar mass of this source is $\log(M_*/M_\odot) = 8.43^{+0.06}_{-0.07}$, consistent with the very low dynamical mass indicated by both the velocity shear and integrated velocity dispersion. The relatively high $V/\sigma \gtrsim 2$ of this source shows that it is a disk galaxy. The metallicity gradient is mildly negative, i.e., -0.07 ± 0.04, consistent with the prediction by analytical CEMs for disk galaxies at similar redshifts.

02806: This is a star-forming galaxy at $z = 1.50$ with a very flat metallicity gradient of -0.01 ± 0.02. We removed the contamination from a marginally overlapping cluster member (with greenish color in the color composite stamp shown in Figure 4). Note that such a process would not be possible with seeing-limited data quality. The *MUSE* observation does not spatially resolve this source well, and its [O II] emission line is detected with low S/N near the red edge of *MUSE* spectral coverage. A tentative velocity shear is seen at low significance. The faint fuzzy elongation to the southeast of this galaxy suggests a likely recent merging event, which accounts for its gradient being flat. (See sources 03746 and 05968 for more outstanding cases where post-merger features can be identified.) The global metallicity of this source ($12 + \log(O/H) = 7.91^{+0.16}_{-0.16}$) is slightly lower than the average given by the MZR at similar M_*, in

agreement with the study of Michel-Dansac et al. (2008), suggesting that the infall of low-metallicity gas during mergers reduces the average metallicity.

03746: This is an intense starburst galaxy with SFR = $10.95^{+1.61}_{-0.58}\,M_\odot\,\mathrm{yr}^{-1}$ at $z = 1.25$, as revealed by the BPT diagram in Figure 11. It also displays a flat metallicity gradient of -0.03 ± 0.03. Moreover, the EL maps of this source show highly clumpy star-forming regions, which do not have direct correspondence in the broadband photometry. The *MUSE* observation gives a velocity shear of $\sim 20\,\mathrm{km\,s}^{-1}$, and an integrated dispersion of $\sim 60\,\mathrm{km\,s}^{-1}$, which suggests that this system is dispersion dominated ($V/\sigma < 1$). There is an extended feature southward to the source center, which is probably a tidal remnant, consistent with this source being in a recent post-merger phase. Besides, given the irregularities seen in the central region of the M_* map, this system is still long before dynamical relaxation, and we likely capture this merger system almost right after the coalescence. Therefore, we conclude that the flat metallicity gradient recovered in this system is from tidal interactions. See galaxy 05968 below for an even more outstanding post-merger case.

04054: This is one of the multiple images of the SN Refsdal host galaxy, a highly magnified almost face-on spiral at $z = 1.49$. It is labeled as image 1.3 by Treu et al. (2016).[28] Our global demographic analysis yields the following measurements: $\log(M_*/M_\odot) = 9.64^{+0.05}_{-0.01}$, SFR $= 16.99^{+5.71}_{-2.80}\,M_\odot\,\mathrm{yr}^{-1}$, and $12 + \log(\mathrm{O/H}) = 8.70^{+0.09}_{-0.11}$. Compared with the previous analysis of this galaxy (Livermore et al. 2015; Rawle et al. 2016, not necessarily the same image though), our work benefits from a more complete wavelength coverage of imaging data (from B_{435} to H_{160} bands) at a greater depth. In particular, we used the most updated lens model, optimized specifically for this spiral galaxy (Treu et al. 2016), giving a reliable estimate of the magnification. Rawle et al. (2016) measured SFR $\sim 5\,M_\odot\,\mathrm{yr}^{-1}$ on image 1.1 of the SN Refsdal host from *Herschel* IR photometry. However, their value is derived assuming a magnification factor of $\mu = 23.0$, given by an older lens model for MACS1149.6+2223 (Smith et al. 2009). If the magnification value of $\mu = 8.14$ given by the latest GLAFIC model is adopted instead, their measurement yields SFR ~ 14, comparable to our SFR measurement. As a result, our measurements are fairly consistent with the SFMS, the MZR and the FMR at similar redshifts. Given the high M_* and $12 + \log(\mathrm{O/H})$, this galaxy is significantly evolved by $z \sim 1.5$ and with the still relatively high SFR, it will likely turn into a galaxy more massive than our Milky Way at $z = 0$. According to the kinematic map derived from the *MUSE* spectroscopy of the [O II] EL doublet, we obtain the velocity shear to be $\sim 110\,\mathrm{km\,s}^{-1}$ (see also Figure 8 in Grillo et al. 2016) and the integrated dispersion $\sim 60\,\mathrm{km\,s}^{-1}$, in good agreement with Livermore et al. (2015) who found $2v_{2.2} = 118 \pm 6\,\mathrm{km\,s}^{-1}$, and local (spatially resolved) velocity dispersion of $50 \pm 10\,\mathrm{km\,s}^{-1}$, based upon AO data.

The measured metallicity gradient for the SN Refsdal host based upon our analysis of image 1.3 is -0.04 ± 0.02 from the individual spaxel method, and -0.07 ± 0.02 from the radial annulus method. This measurement is very close to the usual abundance gradient of local spirals, including our Milky Way,

measured from either H II regions or planetary nebulae (Smartt & Rolleston 1997; van Zee et al. 1998; Henry et al. 2004; Esteban et al. 2015). It is worth noting that our measurement is not as steep as the previous result reported by Yuan et al. (2011), i.e., -0.16 ± 0.02, measured from the AO assisted *OSIRIS* IFU observation of image 1.1 of the SN Refsdal host, assuming the Pettini & Pagel (2004) [N II]/Hα empirical calibration. After the discovery of SN Refsdal, Kelly et al. (2015c) obtained a 1 hr *MOSFIRE* spectra on the nuclear region and the SN explosion site in image 1.1, and measured $12 + \log(\mathrm{O/H}) = 8.6 \pm 0.1$ and $12 + \log(\mathrm{O/H}) = 8.3 \pm 0.1$, respectively, using the same [N II]/Hα calibration. The spatial offset between these two measurements in terms of galactocentric radius is $8.2 \pm 0.5\,\mathrm{kpc}$, given by the GLAFIC model. As a result, the recent *MOSFIRE* observations indicate a gradient being -0.04 ± 0.02, highly consistent with our measurements.

Moreover, we obtain a robust constraint on the nebular dust extinction, i.e., $A_V^N = 0.72^{+0.23}_{-0.23}$. Compared with the stellar dust extinction retrieved from SED fitting, i.e., $A_V^S = 1.10^{+0.01}_{-0.01}$, we see that in this galaxy, the dust reddening of the ionized gas in H II regions is less severe than that of the stellar continuum (see another source, i.e., ID 05968, from which the same conclusion is drawn). This is different from what has been assumed previously: the color excess of nebular ELs is larger than that of the stellar continuum, i.e., $E_S(B - V) = (0.44 \pm 0.03) E_N(B - V)$ (Calzetti et al. 2000). Recently, Reddy et al. (2015) conducted a systematic study of dust reddening using the early observations from the *MOSDEF* survey and found that only 7% of their sample is consistent with $E_N(B - V) \gtrsim E_S(B - V)/0.44$. They also discovered that as sSFR (derived from Balmer EL luminosities) increases, $E_N(B - V)$ diverges more significantly from $E_S(B - V)$, with a scatter of ~ 0.3–0.4 dex at sSFR $\sim 3 \times 10^{-9}$ yr^{-1}, which can account for the difference we see in the total dust attenuation (A_V) of the line and continuum emissions of this galaxy.

05709: This galaxy has the lowest stellar mass in our metallicity gradient sample ($\log(M_*/M_\odot) = 7.90^{+0.02}_{-0.03}$) after lensing correction.[29] This is for the first time any diffraction-limited metallicity gradient analysis and seeing-limited gas kinematics measurement have been achieved in star-forming galaxies with such small M_* at the cosmic noon. This galaxy shows a steep metallicity gradient, i.e., -0.22 ± 0.05. The velocity shear is measured to be $\sim 25\,\mathrm{km\,s}^{-1}$, with integrated dispersion being $\sim 20\,\mathrm{km\,s}^{-1}$. The inferred low dynamical mass is consistent with the result from SED fitting mentioned above. In light of the kinematic result $V/\sigma \gtrsim 1$, this galaxy is probably a clumpy thick disk. More interestingly, observed from a multi-perspective point of view (see Figure 4), this galaxy shows a clumpy star-forming region at the periphery of its disk (having a distance of 3–4 kpc to the mass center), which has a tremendous amount of line emission, yet very little stellar mass, and has low gas-phase metallicity. Also see galaxy ID 07255 for a more dramatic case.

05968: With a tidal tail more appreciably shown to the northeast, and almost perpendicular to the galactic plane, this galaxy (at $z = 1.48$) is safely classified as a post-merger. However, note that this identification would not be possible and this source would have surely been classified as a thick disk due to its symmetric morphology and kinematic properties,

---

[28] It is located at the minimum of its lensing time delay surface, any intrinsic variabilities in the source are observed first in this image. Unfortunately, in the case of SN Refsdal, its image is estimated to appear ~ 12 years ago, long before the deployment of WFC3.

[29] The magnification value we used in correcting for lensing is $7.10^{+0.43}_{-0.36}$, given by the GLAFIC model. The magnification value given by the SHARON & JOHNSON model is similar, i.e., $8.43^{+0.73}_{-0.60}$.

without the ultra-deep HFF imaging data. Compared to the M_* and EL maps of ID 03746, the M_* map of this galaxy is more smooth and EL maps are less clumpy, which indicates that merging took place longer before than that in source ID 03746. Still, the star formation is high, i.e., SFR $= 27.95^{+4.70}_{-4.41}\,M_\odot\,\mathrm{yr}^{-1}$. The global metallicity ($12 + \log(\mathrm{O/H}) = 8.47^{+0.07}_{-0.08}$) and stellar mass ($\log(M_*/M_\odot) = 9.34^{+0.02}_{-0.03}$) are quite comparable to the MZR at similar redshifts. The metallicity gradient, as expected for systems that have experienced recent mergers (see, e.g., Kewley et al. 2010), is constrained to be robustly flat (-0.01 ± 0.02). The EL kinematic analysis yields a velocity shear of $\sim 120\,\mathrm{km\,s}^{-1}$, and an integrated dispersion of $\sim 75\,\mathrm{km\,s}^{-1}$. The velocity field appears typical of intermediate-mass disk galaxies at this redshift, which infers that this post-merger system has been significantly relaxed and regained at least some of the rotation support. This is also one of the few cases where robust constraints on nebular dust extinction can be derived, i.e., $A_V^N = 0.23^{+0.15}_{-0.13}$, which is less than the dust extinction of stellar continuum obtained from SED fitting $A_V^S = 0.60^{+0.20}_{-0.01}$, as already seen in the case of ID 04054.

07255: This galaxy is highly consistent with the SFMS at $z = 1.27$ formulated by Speagle et al. (2014), with $\log(M_*/M_\odot) = 9.36^{+0.01}_{-0.21}$ and SFR $= 8.73^{+3.53}_{-3.10}\,M_\odot\,\mathrm{yr}^{-1}$. By $z = 0$, it will turn into a sub-Milky-Way-sized galaxy according to Behroozi et al. (2013). Its slightly lower than average metallicity of $12 + \log(\mathrm{O/H}) = 8.38^{+0.08}_{-0.08}$ can be a result of low-metallicity gas inflow, following the popular interpretation of the FMR at high z. This hypothesis would also explain its slightly enhanced SFR. In the metallicity map, the inflow can be identified with a bright star-forming clump in the lower right corner, which consists of very little M_*, but is dominating in the EL maps, similar to the case of galaxy 05709 discussed above. This indicates that in this system, the metal enrichment timescale is much larger than the star-forming timescale, which is in turn much larger than the gas infall timescale. Unfortunately, at the moment, we do not have EL kinematic observation on this system.

The metallicity gradient for this galaxy is constrained to be very steep, i.e., -0.16 ± 0.03 using the individual spaxel method, and -0.21 ± 0.03 from the radial annulus method. We also dissected this galaxy to separate the star-forming clump from the bulk part and re-do the analysis. It is found that the H II regions closely related to the clump (with at least 2.5 kpc distance to the galaxy mass center) have significantly lower metallicity and provide more than one-third of the entire star formation, i.e., $12 + \log(\mathrm{O/H}) = 8.04^{+0.14}_{-0.18}$, $A_V^N < 0.78$, SFR $= 3.05^{+1.21}_{-1.10}$, compared to the quantities measured in the bulk of the galaxy (within 2 kpc to the mass center), i.e., $12 + \log(\mathrm{O/H}) = 8.51^{+0.08}_{-0.10}$, $A_V^N = 0.63^{+0.59}_{-0.41}$, and SFR $= 5.78^{+3.41}_{-2.47}$. In both regions, EL Hα is detected at a significance of $\gtrsim 30\sigma$. This again demonstrates that measurements of global quantities can be to some extent misleading and also demonstrates the importance of detailed spatially resolved information. Cases like this galaxy should be exceedingly interesting, since their steep metallicity gradients at such high z are not predicted by current mainstream numerical simulations (equipped with strong galactic feedback) or conventional analytical CEMs. Through studying systems like this, we can also learn how gas is accreted into the inner disk, and more importantly whether the accretion happens alongside star formation or long before the gas clouds can collapse.

7. Summary and Discussion

7.1. Summary

In this paper, we took advantage of the deep *HST* grism spectroscopic data, acquired by the GLASS and SN Refsdal follow-up programs, and obtained gas-phase metallicity maps in high-z star-forming galaxies. The *HST* grisms provide an uninterrupted window for observing the cosmic evolution of metallicity gradients through z from 2.3 to 1.2 (corresponding to the age of the universe from roughly 2.5 to 5.5 Gyr), including the redshift range of $1.7 < z < 2.0$, which is unattainable from ground-based observations due to low atmospheric transmission. Another advantage of the *HST* grisms is that multiple ELs are available within a single setup and all the potential issues related to combining data with different setups and atmospheric conditions are eliminated. Our main conclusions can be summarized as follows.

1. We presented 10 maps of gas-phase metallicity (as shown in Figure 10), measured at sub-kiloparsec resolution in star-forming galaxies in the redshift range of $1.2 \lesssim z \lesssim 2.3$, in the prime field of MACS1149.6+2223. This field is so far one of the deepest fields exposed by *HST* spectroscopy, at a depth of 34(10) orbits of G141(G102), reaching a 1σ flux limit of $1.2(3.5) \times 10^{-18}\,\mathrm{erg\,s}^{-1}\,\mathrm{cm}^{-2}$, without lensing correction.

2. We developed a novel Bayesian statistical approach to jointly constrain gas-phase metallicity, nebular dust extinction, and Hα-based dust-corrected SFR. In determining these quantities, our method does not rely on any assumptions about star-formation history, nor the conversion of dust reddening from the stellar phase to the nebular phase. Unlike the majority of previous work on deriving metallicity from strong EL calibrations, our method does not compare directly the observed EL flux ratios with the calibrated ones. Instead, we work directly with EL fluxes, which effectively avoids using redundant information and circumvents possible biases in the low S/N regime.

3. The metallicity maps we obtained show a large diversity of morphologies. Some maps display disk-like shapes and have metallicity gradients consistent with those predicted by analytical CEMs (i.e., IDs 04054, 01422, 02389, 05732), whereas other disk-shaped galaxies have exceedingly steep gradients (i.e., IDs 07255, 05709, 05811) and bright star-forming clumps at the periphery of their disks. These large-offset star-forming clumps, containing very little stellar mass, are estimated to have lower metallicity than the corresponding global values, which can be interpreted as evidence for the infall of low-metallicity gas enhancing star formation. We also recovered three systems with nearly uniform spatial distribution of metals (i.e., IDs 02806, 03746, 05968). In these systems, we can identify possible signatures of post-merger tidal remnants, which suggests that the observed flat gradients are likely caused by gravitational interactions.

4. Collecting all existing sub-kiloparsec resolution metallicity gradient measurements at high redshifts, we study the cosmic evolution and mass dependence of metallicity gradients.

   (i) We found that predictions given by analytical CEMs assuming a relatively extended star-formation profile

in the early phase of disk growth can reproduce the majority of observed metallicity gradients at high redshifts, without involving strong galactic feedback or radial outflows. This confirms the tight link between star formation and metal enrichment.

(ii) We tentatively observed a correlation between stellar mass and metallicity gradient, which is in accord with the "downsizing" galaxy formation scenario that more massive galaxies are more evolved into a later phase of disk growth, where they experience more coherent mass assembly and metal enrichment than lower mass galaxies.

A larger sample of uniformly analyzed metallicity maps well sampled in the full M_* range (especially in the low-mass regime, i.e., $M_* \lesssim 10^9\,M_\odot$) at high redshifts is needed to further investigate the cosmic evolution and mass dependence of metallicity gradients.

5. We also compiled a sample of 38 global metallicity measurements all derived from the Maiolino et al. (2008) EL calibrations in the current literature (including 13 galaxies presented in this work). This sample, at $z \sim 1.8$, spans three orders of magnitude in M_*. We measured the MZR and tested the FMR using this sample.

(i) The slope of the MZR constrained by this sample rules out the momentum-driven wind model by Davé et al. (2012) at a $3\,\sigma$ confidence level. Our MZR is consistent with the theoretical prediction given by recent cosmological zoom-in simulations, suggesting that high spatial resolution simulations are favorable in reproducing the metal enrichment history in star-forming galaxies.

(ii) Given the intrinsic scatter and measurement uncertainties, we do not see any significant offset from the FMR except the subsample of Wuyts et al. (2012). This discrepancy is likely due to the fact that the Wuyts et al. (2012) data set is the only subsample that relies exclusively on [N II]/Hα to estimate metallicity. We therefore advocate to avoid nitrogen lines when estimating gas-phase metallicity in high-z H II regions.

7.2. Interpretation of Flat/Steep Metallicity Gradients at High Redshifts

We now turn to the interpretation of our observed metallicity gradients and maps. Apart from the flattening of metallicity gradients through natural evolution of galaxies that brings the abundances of any elements in the ISM asymptotically to the average stellar yields at the end of the evolution everywhere in the galaxy, there are four physical mechanisms on relatively short timescales that could significantly flatten a star-forming galaxy's pre-existing negative metallicity gradient (not exclusively in high-z scenarios).

1. Efficient radial mixing by tidal effects from gravitational interactions, as seen in galaxy IDs 02806, 03746, and 05968 in our sample and the triply imaged arc 4 in our previous work (Jones et al. 2015b).
2. Rapid recycling of metal-enriched material by enhanced galactic feedback, as seen in the results of the MaGICC numerical simulations, for example.

3. Turbulence mixing driven by thermal instability as elucidated by Yang & Krumholz (2012).
4. Funnelling of cold streams of low-metallicity gas directly into the galaxy center, as suggested by Dekel et al. (2009).[30]

Besides pre-existing negative gradients being flattened by these processes, galaxies can also inherit a broadly flat metallicity gradient from a relatively extended star-formation profile in the early disk-formation phase and coherent mass assembly for the disk growth. Because there is strong evidence that the mass assembly for Milky Way progenitors takes place uniformly at all radii, maintaining almost the same M_* profile from $z \sim 2.5$ to $z \sim 1$ (van Dokkum et al. 2013; Morishita et al. 2015). However, the uncertainties associated with current measurements are still insufficient to distinguish numerical simulations with enhanced feedback prescriptions from conventional analytical CEMs assuming more uniform star formation (see Figure 12).

One perhaps obvious but nevertheless important caveat common to most galaxy evolution work is that we only have access to cross-sectional data, i.e., observations of different objects at one specific epoch. Therefore, any interpretations of plots like Figure 12, which try to tie these data with longitudinal theories (describing the evolution of one specific object), should be made with caution.

Perhaps an even more intriguing interpretation of Figure 12, which is robust against the aforementioned caveat, is that there exist some really steep metallicity gradients at high z, such as our galaxy IDs 07255, 05709, and 05811. Based upon current measurements, the occurrence rate of high-z metallicity gradients more negative than $-0.1\,\mathrm{dex\,kpc^{-1}}$ is $5/34$. We can derive the following constraints on the physical processes that can contribute to these steep metallicity gradients observed beyond the local universe:

1. galactic feedback (including supernova explosions and stellar winds) must be of limited influence (via being confined for instance),
2. star-formation efficiency is low, or at least lower than their local analogs (in simulations, this can be achieved by having a higher star-formation threshold or less concentrated molecular gas, as shown by the MUGS runs),
3. star formation at early times is highly centrally concentrated, and the mass assembly of the inner regions is much faster than that of the outer regions, and
4. alternatively, steep gradients are the result of low-metallicity gas falling onto the outskirts of their disks, implying that the dynamic timescale is much shorter than the star-forming timescale, which is much shorter than the metal enrichment timescale, as seen in our galaxy 07255.

We need larger samples with robust measurements in order to more accurately quantify the occurrence rate of these steep metallicity gradients in high-z star-forming disks. The existence of a distinct population of galaxies with steep metallicity gradients, if confirmed by future observations constraining their abundance, is a powerful test of galaxy formation models, because this is not predicted by most current theoretical models. Hence, these galaxies with steep metallicity gradients should be the prime

---

[30] This mechanism is invoked to explain the inverted gradients (lower abundances in the inner disk regions) seen in some of the AMAZE/LSD observations (Cresci et al. 2010). However, we note that in their observations the near-IR integral field spectrometer *SINFONI* was used in seeing-limited mode.
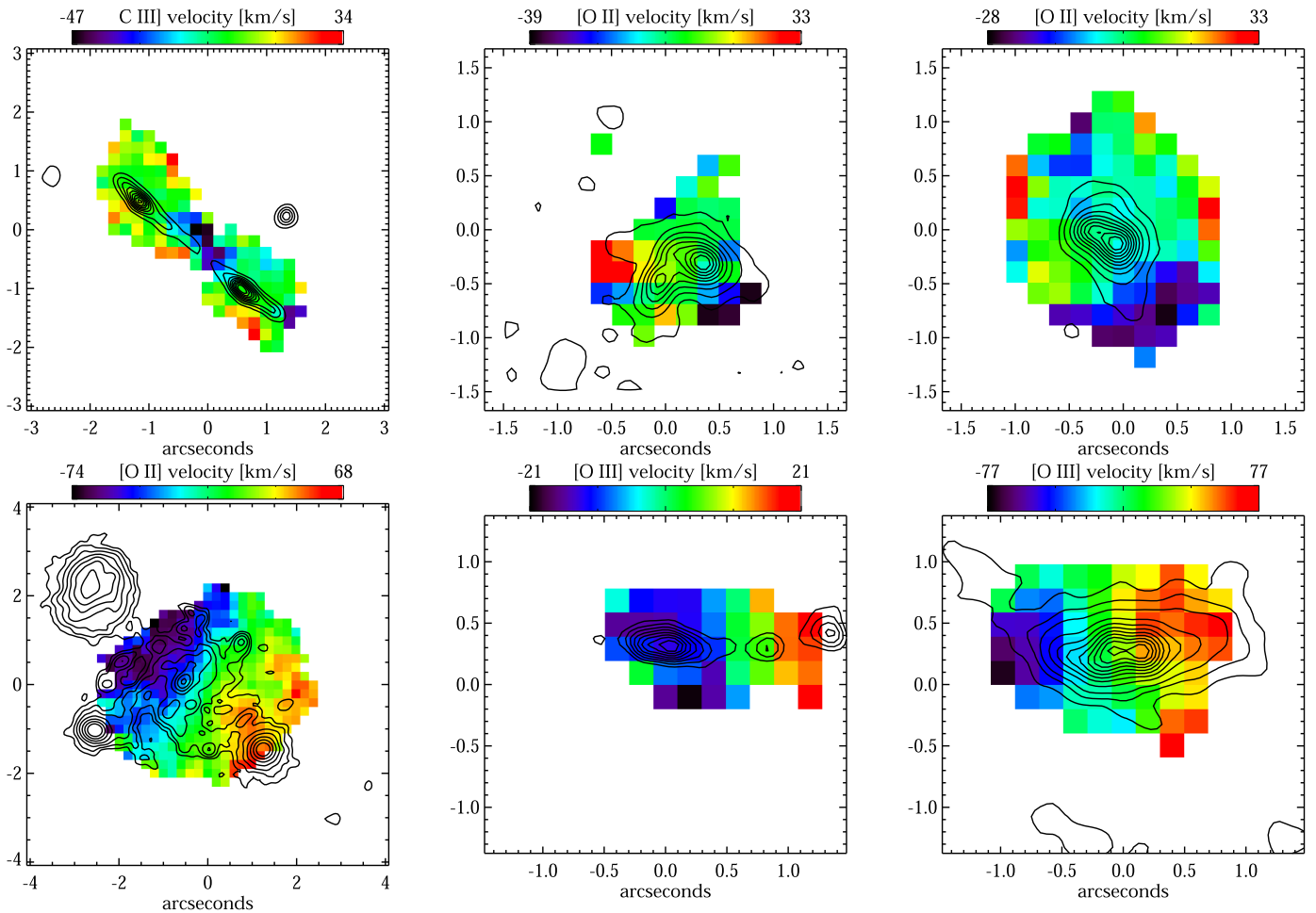
Figure 14. Velocity fields derived from integral field spectra. Top row: IDs 02389, 02806, 03746 (left to right) observed with *MUSE*. Bottom row: ID 04054 observed with *MUSE* (see also Figure 8 in Grillo et al. 2016) and IDs 05709, 05968 (left to right) observed with KMOS. In all cases, the color scale shows EL velocity centroids in each spaxel, with contours showing *HST*/ACS I_{814} band continuum. The kinematics map for galaxy 05709 (shown in the lower middle panel) is the first one to date presented for star-forming galaxies with $M_* \lesssim 10^8 M_\odot$ at $z \gtrsim 1$.

targets for new numerical predictions: future *JWST* observations will constrain their properties in great detail, allowing discriminating tests of current theories.

Furthermore, most current numerical simulations are aimed at reproducing present-day Milky Way analogs, surrounded by a dark matter halo with virial mass $\sim 10^{12} M_\odot$. It would be useful to have large samples of simulated sub-L_* galaxies (M_* below $5 \times 10^9 M_\odot$ at $z \sim 2$) in order to investigate how metals are recycled in these less massive systems and provide a more suitable comparison to our metallicity gradient measurements at high z. With improvements on both the observational and theoretical sides, we will be able to answer why in these high-z star-forming disk galaxies, the metallicity gradients can be established so early and why they are so steep compared to those found in local analogs.

Appendix A
Emission Line Kinematics from Ground-based IFU Observations

When available, we use ground-based spectroscopic IFU data to investigate the 2D kinematics of the sources for which we
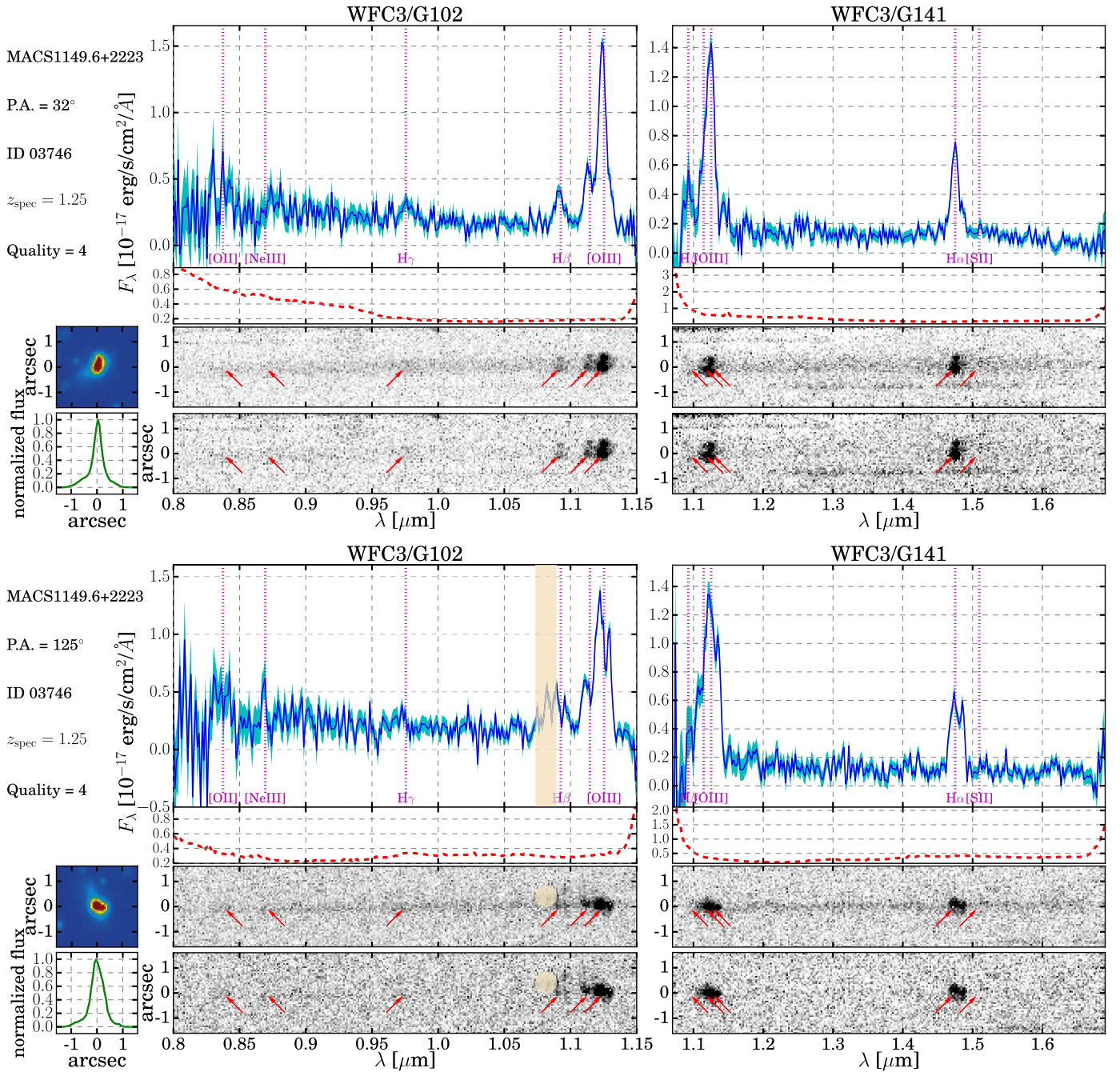
Figure 15. *HST* grism spectra for one exemplary source in our sample, ID 03746, seen at the two GLASS PAs displayed in the two sub-figures respectively. The Hα EL of this galaxy is covered by G141 because of its redshift $z = 1.25$. In each sub-figure, the two small square panels on the left refer to the 2D postage stamp (top) and the 1D collapsed image (bottom). The 2D postage stamp is created from the HFF H_{160}-band co-adds. The eight panels in the middle and right rows show, from top to bottom, the contamination subtracted 1D spectra (where flux is represented by the blue solid line and the noise level by the cyan shaded band), the 1D contamination model is represented by the red dashed line, the contamination subtracted 2D spectra and the 2D spectra having source continuum further removed. In the 1D and 2D spectra, the locations of emission lines are highlighted by vertical dotted magenta lines and red arrows, respectively, and the wheat colored patches cover contamination model defects. Some ancillary information about the source can be found in the upper left corner in each sub-figure.

obtain metallicity maps from the *HST* grism data. A first set of ground-based data were obtained in 2015 with the instrument *MUSE* on the VLT, taken as part of the Director's Discretionary Time program 294.A-5032 (P.I. Grillo) aimed at assisting with the modeling of SN Refsdal. The *MUSE* pointing is denoted by the green square in Figure 1. The observations, data reduction, first results, and applications to the host galaxy of Refsdal and mass modeling of the cluster are described by Grillo et al. (2016),

Karman et al. (2016). Five objects in our metallicity gradient sample fall within the *MUSE* pointing, with one on the edge of the FOV. Hence, four maps showing complete kinematic structures can be extracted for sources IDs 02389, 02806, 03746, and 04054. A second set of ground-based observations were obtained with the instrument KMOS also on the VLT, as part of our GLASS follow-up KMOS Large Program 196.A-0778 (P.I. Fontana). Two objects (IDs 05709 and 05968) in our sample were
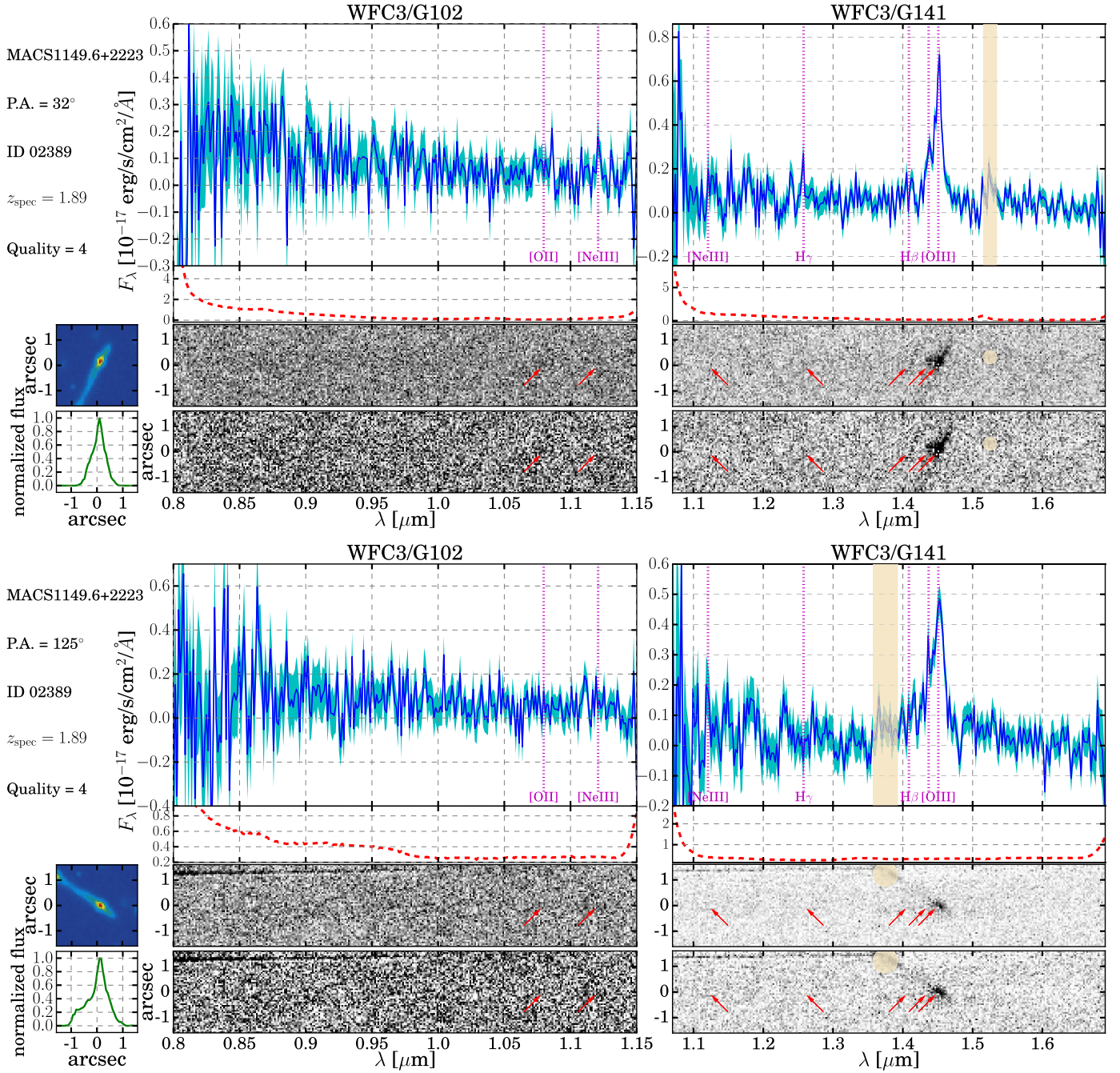
Figure 16. Same as Figure 15, except that source ID 02389 is shown. The Hα EL of this galaxy has shifted out of the grism coverage because of its redshift $z = 1.89$.

targeted with the deployable IFUs, using the YJ grating. The observations, data reduction, and first results from this program are described by Mason et al. (2016). Thanks to lensing magnification, we push for the first time the mass limit of EL kinematic analyses to below $10^8 M_\odot$ (see Table 6).

Kinematic information is valuable for interpreting metallicity gradients and their evolution. We derive gas-phase kinematics for the aforementioned galaxies, by fitting the strongest available emission feature. This is typically C III] $\lambda\lambda1907,1909$, [O II], or [O III] $\lambda\lambda4960,5008$. We fit each doublet as a sum of two Gaussian components with equal velocity dispersion (σ) and redshift, plus a constant continuum level. The fits are weighted by the inverse-variance spectrum estimated from sky regions in the data cubes. Best-fit values for the spatially integrated spectra

are given in Table 6. The intrinsic velocity dispersion $\sigma_{\rm int}$ is calculated by subtracting the instrument resolution (determined from sky ELs) in quadrature from the observed best-fit velocity dispersion $\sigma_{\rm obs}$. For C III] and [O II], we determine the best-fit doublet ratio of the integrated spectra, while the [O III] doublet ratio is fixed to its intrinsic atomic value. For individual spaxels, we fix the doublet ratios to their integrated values in order to avoid spurious fits. We also spatially smooth each data cube with a 3-pixel ($0''\!\!.6$) kernel to improve S/N in the diffuse extended regions. We attempt to fit every spaxel in the smoothed data cube surrounding each object. Bad fits are rejected on the basis of low S/N and unrealistic values (typically requiring $20\,{\rm km\,s^{-1}} < \sigma_{\rm obs} < 200\,{\rm km\,s^{-1}}$, and velocities within $200\,{\rm km\,s^{-1}}$ of the integrated systemic value). Figure 14 shows

Table 7
Measured Emission Line Fluxes

ID	z_{spec}	$f_{[S\ II]}$	$f_{H\alpha}$	$f_{[O\ III]}$	$f_{H\beta}$	$f_{H\gamma}$	$f_{[O\ II]}$
01422	2.28	⋯	⋯	18.08 ± 0.60	2.98 ± 0.61	3.60 ± 0.86	8.70 ± 1.18
02389	1.89	⋯	⋯	53.24 ± 0.66	7.51 ± 0.70	3.65 ± 1.25	4.65 ± 2.03
02607	1.86	⋯	⋯	37.86 ± 0.65	5.59 ± 0.83	<1.68	4.23 ± 1.94
02806	1.50	7.00 ± 2.39	12.75 ± 0.78	30.21 ± 1.17	11.95 ± 1.22	9.29 ± 2.69	5.17 ± 4.08
02918	1.78	⋯	⋯	22.96 ± 1.26	8.14 ± 1.33	2.18 ± 1.70	14.71 ± 3.79
03746	1.25	11.69 ± 0.77	97.29 ± 0.81	170.40 ± 1.39	35.00 ± 2.18	14.93 ± 3.29	49.40 ± 8.37
04054	1.49	29.38 ± 3.14	57.26 ± 1.06	17.96 ± 1.45	16.56 ± 1.53	<5.16	34.08 ± 5.10
05422	1.97	⋯	⋯	10.79 ± 0.71	4.08 ± 0.73	3.91 ± 0.88	1.82 ± 1.63
05709	1.68	⋯	⋯	54.03 ± 1.06	13.17 ± 1.13	8.59 ± 1.60	18.69 ± 3.54
05732	1.68	⋯	⋯	35.12 ± 0.71	12.45 ± 0.73	6.57 ± 1.07	22.92 ± 2.88
05811	2.31	⋯	⋯	21.09 ± 0.61	3.01 ± 0.64	7.71 ± 0.93	7.31 ± 1.25
05968	1.48	17.19 ± 1.36	68.13 ± 0.81	59.08 ± 1.09	21.08 ± 1.19	21.36 ± 3.10	55.99 ± 3.95
07058	1.79	⋯	⋯	12.87 ± 0.96	3.59 ± 0.91	3.96 ± 1.25	8.73 ± 2.77
07255[a]	1.27	11.70 ± 1.21	37.19 ± 0.92	41.81 ± 2.59	21.85 ± 4.16	⋯	11.84 ± 9.28

Note. The fluxes are observed values in units of 10^{-17} erg s^{-1} cm^{-2}, not accounting for dust extinction nor lensing magnification. The flux error bars denote 1σ measurement uncertainties. The 2σ upper limit is given for undetected lines. Only sources at $z \leqslant 1.5$ have Hα and [S II] line flux measurements due to the grism spectral coverage.
[a] The missing of Hγ flux for this source is due to the grism defect.

velocity fields of each source from all spaxels with acceptable EL fits. We discuss the kinematics of individual galaxies in the context of their metallicity gradients and other properties in Section 6.2. For purposes of this discussion, we estimate velocity shear V_{shear} as the total change in velocity across the kinematic major axis of each source shown in Figure 14. Estimates of V/σ in the text refer to V_{shear}. This quantity serves as a useful proxy for the gas orbital velocity: $V_{shear} = 2V_{max}\sin i$, i.e., twice the observed maximum orbital velocity (not corrected for inclination).

Appendix B
Emission Line Flux Measurements from *HST* Grisms

After a careful selection, we compiled a list of sources most promising for the spatially resolved spectroscopic analysis in the redshift range of $z \in [1.2, 2.3]$. For instance, we show the GLASS spectra of two of these sources (one at $z \leqslant 1.5$ with Hα covered and the other at $z \geqslant 1.5$ without Hα) in Figures 15 and 16. The SN Refsdal follow-up exposures, albeit only covering G141, have better S/Ns compared with the spectra shown here by a factor of $\sqrt{15/2}$ at each PA. Combining all the available grism exposures, we measured the EL fluxes on all our sources and presented the measurements in Table 7.

References

Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140
Baldry, I. K., Glazebrook, K., & Driver, S. P. 2008, MNRAS, 388, 945
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57
Belli, S., Jones, T. A., Ellis, R. S., & Richard, J. 2013, ApJ, 772, 141
Bianco, F. B., Modjaz, M., Oh, S. M., et al. 2015, 54 arXiv:1505.06213
Blanc, G. A., Kewley, L. J., Vogt, F. P. A., & Dopita, M. A. 2015, ApJ, 798, 99
Brammer, G. B., Pirzkal, N., McCullough, P., & MacKenty, J. 2014, Instrument Science Report, WFC3 2014-03 (Baltimore, MD: StScI)
Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, ApJS, 200, 13
Bresolin, F., & Kennicutt, R. C. J. 2015, MNRAS, 454, 3664
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, AJ, 345, 245
Chabrier, G. 2003, PASP, 115, 763

Chiappini, C., Matteucci, F., & Romano, D. 2001, ApJ, 554, 1044
Coil, A. L., Aird, J., Reddy, N. A., et al. 2015, ApJ, 801, 35
Coleman, M. G., & de Jong, J. T. A. 2008, ApJ, 685, 933
Cowie, L. L., Barger, A. J., & Songaila, A. 2016, ApJ, 817, 57
Cresci, G., Mannucci, F., Maiolino, R., et al. 2010, Natur, 467, 811
Curti, M., Cresci, G., Mannucci, F., et al. 2016, arXiv:1610.06939
Davé, R., Finlator, K., & Oppenheimer, B. D. 2011a, MNRAS, 416, 1354
Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, MNRAS, 421, 98
Davé, R., Katz, N., Oppenheimer, B. D., Kollmeier, J. A., & Weinberg, D. H. 2013, MNRAS, 434, 2645
Davé, R., Oppenheimer, B. D., & Finlator, K. 2011b, MNRAS, 415, 11
Dekel, A., Birnboim, Y., Engel, G., et al. 2009, Natur, 457, 451
Dopita, M. A., Sutherland, R. S., Nicholls, D. C., Kewley, L. J., & Vogt, F. P. A. 2013, ApJS, 208, 10
Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006, ApJ, 644, 813
Esteban, C., García-Rojas, J., & Pérez-Mesa, V. 2015, MNRAS, 452, 1553
Few, C. G., Gibson, B. K., Courty, S., et al. 2012, A&A, 547, A63
Finlator, K., & Davé, R. 2008, MNRAS, 385, 2181
Fontanot, F., de Lucia, G., Monaco, P., Somerville, R. S., & Santini, P. 2009, MNRAS, 397, 1776
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Freeman, K. C. 1970, ApJ, 160, 811
Frye, B. L., Hurley, M., Bowen, D. V., et al. 2012, ApJ, 754, 17
Gibson, B. K., Pilkington, K., Brook, C. B., Stinson, G. S., & Bailin, J. 2013, A&A, 554, A47
Grillo, C., Karman, W., Suyu, S. H., et al. 2016, ApJ, 822, 78
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Guo, Y., Koo, D. C., Lu, Y., et al. 2016, arXiv:1603.04863
Henry, A. L., Scarlata, C., Domínguez, A., et al. 2013, ApJL, 776, L27
Henry, R. B. C., Kwitter, K. B., & Balick, B. 2004, AJ, 127, 2284
Ho, I. T., Kudritzki, R.-P., Kewley, L. J., et al. 2015, MNRAS, 448, 2030
Hopkins, P. F., Kere, D., Onorbe, J., et al. 2014, MNRAS, 445, 581
Hummer, D. G., & Storey, P. J. 1987, MNRAS, 224, 801
Hunt, L., Dayal, P., Magrini, L., & Ferrara, A. 2016, arXiv:1608.05417
Johnson, T. L., Sharon, K., Bayliss, M. B., et al. 2014, ApJ, 797, 48
Jones, T. A., Ellis, R. S., Jullo, E., & Richard, J. 2010, ApJL, 725, L176
Jones, T. A., Ellis, R. S., Richard, J., & Jullo, E. 2013, ApJ, 765, 48
Jones, T. A., Martin, C., & Cooper, M. C. 2015a, ApJ, 813, 126
Jones, T. A., Wang, X., Schmidt, K. B., et al. 2015b, AJ, 149, 107
Juneau, S., Bournaud, F., Charlot, S., et al. 2014, ApJ, 788, 88
Juneau, S., Dickinson, M., Alexander, D. M., & Salim, S. 2011, ApJ, 736, 104
Karman, W., Grillo, C., Balestra, I., et al. 2016, A&A, 585, A27
Kawamata, R., Oguri, M., Ishigaki, M., Shimasaku, K., & Ouchi, M. 2015, arXiv:1510.06400
Kelly, B. C. 2007, ApJ, 665, 1489
Kelly, P. L., Brammer, G. B., Selsing, J., et al. 2015c, arXiv:1512.09093
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015a, arXiv:1512.04654
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015b, Sci, 347, 1123

Kennicutt, R. C. J. 1998, ARA&A, 36, 189
Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183
Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. M. 2006, MNRAS, 372, 961
Kewley, L. J., Rupke, D. S. N., Zahid, H. J., Geller, M. J., & Barton, E. J. 2010, ApJL, 721, L48
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kriek, M. T., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Kümmel, M., Walsh, J. R., Pirzkal, N., Kuntschner, H., & Pasquali, A. 2009, PASP, 121, 59
Leethochawalit, N., Jones, T. A., Ellis, R. S., et al. 2015, arXiv:1509.01279
Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y.-j. 2013, ApJ, 772, 119
Livermore, R. C., Jones, T. A., Richard, J., et al. 2015, MNRAS, 450, 1812
Lotz, J. M., Koekemoer, A. M., Coe, D., et al. 2016, arXiv:1605.06567
Lu, Y., Blanc, G. A., & Benson, A. 2015a, ApJ, 808, 129
Lu, Z., Mo, H. J., & Lu, Y. 2015b, MNRAS, 450, 606
Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016a, MNRAS, 456, 2140
Ma, X., Hopkins, P. F., Feldmann, R., et al. 2016b, arXiv:1610.03498
Maciel, W. J., Costa, R. D. D., & Uchida, M. M. M. 2003, A&A, 397, 667
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Maiolino, R., Nagao, T., Grazian, A., et al. 2008, A&A, 488, 463
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
Mannucci, F., Salvaterra, R., & Campisi, M. A. 2011, MNRAS, 414, 1263
Martin, C. L., Shapley, A. E., Coil, A. L., et al. 2012, ApJ, 760, 127
Mason, C. A., Treu, T., Fontana, A., et al. 2016, arXiv:1610.03075
Masters, D., Faisst, A., & Capak, P. 2016, arXiv:1605.04314
Meneghetti, M., Natarajan, P., Coe, D., et al. 2016, arXiv:1606.04548
Michel-Dansac, L., Lambas, D. G., Alonso, M. S., & Tissera, P. 2008, MNRAS, 386, L82
Mollá, M., Cavichia, O., Costa, R. D. D., et al. 2016a, arXiv:1612.03064
Mollá, M., & Díaz, A. I. 2005, MNRAS, 358, 521
Mollá, M., Díaz, A. I., Ascasíbar, Y., et al. 2016b, arXiv:1612.03348
Mollá, M., Díaz, A. I., Gibson, B. K., Cavichia, O., & López-Sánchez, Á.-R. 2016c, MNRAS, 462, 1329
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Morishita, T., Ichikawa, T., Noguchi, M., et al. 2015, ApJ, 805, 34
Mortlock, A., Conselice, C. J., Hartley, W. G., et al. 2013, MNRAS, 433, 1185
Moustakas, J., Kennicutt, R. C. J., Tremonti, C. A., et al. 2010, ApJS, 190, 233
Oguri, M. 2010, PASJ, 62, 1017
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Pérez-Montero, E. 2014, MNRAS, 441, 2663
Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59
Pilkington, K., Few, C. G., Gibson, B. K., et al. 2012, A&A, 540, A56
Pilyugin, L. S., Vilchez, J. M., & Thuan, T. X. 2010, ApJ, 720, 1738

Queyrel, J., Contini, T., Kissler-Patig, M., et al. 2012, A&A, 539, A93
Rawle, T. D., Altieri, B., Egami, E., et al. 2016, MNRAS, 459, 1626
Reddy, N. A., Kriek, M. T., Shapley, A. E., et al. 2015, ApJ, 806, 259
Richard, J., Jones, T. A., Ellis, R. S., et al. 2011, MNRAS, 413, 643
Rupke, D. S. N., Kewley, L. J., & Chien, L. H. 2010, ApJ, 723, 1255
Sánchez, S. F., Rosales-Ortega, F. F., Iglesias-Páramo, J., et al. 2014, A&A, 563, A49
Sanders, R. L., Shapley, A. E., Kriek, M. T., et al. 2015, ApJ, 799, 138
Sanders, R. L., Shapley, A. E., Kriek, M. T., et al. 2016, arXiv:1606.04107
Schmidt, K. B., Treu, T., Bradac, M., et al. 2016, ApJ, 818, 38
Schmidt, K. B., Treu, T., Brammer, G. B., et al. 2014, ApJL, 782, L36
Shapley, A. E., Reddy, N. A., Kriek, M. T., et al. 2015, ApJ, 801, 88
Sharon, K., & Johnson, T. L. 2015, ApJL, 800, L26
Smartt, S. J., & Rolleston, W. R. J. 1997, ApJL, 481, L47
Smith, G. P., Ebeling, H., Limousin, M., et al. 2009, ApJL, 707, L163
Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, ApJ, 795, 165
Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, arXiv:1605.07186
Stinson, G. S., Bailin, J., Couchman, H., et al. 2010, MNRAS, 408, 812
Stinson, G. S., Brook, C. B., Macciò, A. V., et al. 2013, MNRAS, 428, 129
Storey, P. J., & Zeippen, C. J. 2000, MNRAS, 312, 813
Stott, J. P., Sobral, D., Swinbank, A. M., et al. 2014, MNRAS, 443, 2695
Swinbank, A. M., Sobral, D., Smail, I., et al. 2012, MNRAS, 426, 935
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Treu, T., Brammer, G. B., Diego, J. M., et al. 2016, ApJ, 817, 60
Treu, T., Schmidt, K. B., Brammer, G. B., et al. 2015, ApJ, 812, 114
Troncoso, P., Maiolino, R., Sommariva, V., et al. 2014, A&A, 563, A58
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28
van Dokkum, P. G., Leja, J., Nelson, E., et al. 2013, ApJL, 771, L35
van Zee, L., Salzer, J. J., & Haynes, M. P. 1998, ApJL, 497, L1
Vulcani, B., Treu, T., Nipoti, C., et al. 2016, arXiv:1610.04615
Vulcani, B., Treu, T., Schmidt, K. B., et al. 2015, ApJ, 814, 161
Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104
Wisnioski, E., Förster Schreiber, N. M., Wuyts, S., et al. 2015, ApJ, 799, 209
Wuyts, E., Kurk, J., Förster Schreiber, N. M., et al. 2014, ApJL, 789, L40
Wuyts, E., Rigby, J. R., Sharon, K., & Gladders, M. D. 2012, ApJ, 755, 73
Wuyts, E., Wisnioski, E., Fossati, M., et al. 2016, arXiv:1603.01139
Yang, C.-C., & Krumholz, M. R. 2012, ApJ, 758, 48
Yuan, T., Kewley, L. J., & Rich, J. A. 2013, ApJ, 767, 106
Yuan, T., Kewley, L. J., Swinbank, A. M., Richard, J., & Livermore, R. C. 2011, ApJL, 732, L14
Zahid, H. J., Bresolin, F., Kewley, L. J., Coil, A. L., & Davé, R. 2012, ApJ, 750, 120
Zahid, H. J., Dima, G. I., Kudritzki, R.-P., et al. 2014, ApJ, 791, 130
Zahid, H. J., Kewley, L. J., & Bresolin, F. 2011, ApJ, 730, 137