



Chemical Abundance Scaling Relations for Multiple Elements in $z \simeq 2\text{--}3$ Star-forming Galaxies

Allison L. Strom^{1,2,5} , Gwen C. Rudie² , Charles C. Steidel³ , and Ryan F. Trainor⁴ ¹ Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA; allison.strom@princeton.edu² Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA³ Cahill Center for Astronomy and Astrophysics, California Institute of Technology, MS 249-17, Pasadena, CA 91125, USA⁴ Department of Physics and Astronomy, Franklin & Marshall College, 637 College Avenue, Lancaster, PA 17603, USA

Received 2021 September 18; revised 2021 November 7; accepted 2021 November 9; published 2022 January 31

Abstract

The chemical abundance patterns of gas and stars in galaxies are powerful probes of galaxies’ star formation histories and the astrophysics of galaxy assembly but are challenging to measure with confidence in distant galaxies. In this paper, we report the first measurements of the correlation between stellar mass (M_*) and multiple tracers of chemical enrichment (including O, N, and Fe) in individual $z \sim 2\text{--}3$ galaxies, using a sample of 195 star-forming galaxies from the Keck Baryonic Structure Survey. The galaxies’ chemical abundances are inferred using photoionization models capable of reconciling high-redshift galaxies’ observed extreme rest-UV and rest-optical spectroscopic properties. We find that the $M_*\text{--O/H}$ relation for our sample is relatively shallow, with moderately large scatter, and is offset ~ 0.35 dex higher than the corresponding $M_*\text{--Fe/H}$ relation. The two relations have very similar slopes, indicating a high level of α -enhancement— $\text{O/Fe} \approx 2.2 \times (\text{O/Fe})_\odot$ —across two decades in M_* . The $M_*\text{--N/H}$ relation has the steepest slope and largest intrinsic scatter, which likely results from the fact that many $z \sim 2$ galaxies are observed near or past the transition from “primary” to “secondary” N production, and may reflect uncertainties in the astrophysical origin of N. Together, these results suggest that $z \sim 2$ galaxies are old enough to have seen substantial enrichment from intermediate-mass stars, but are still young enough that Type Ia supernovae have not had time to contribute significantly to their enrichment.

Unified Astronomy Thesaurus concepts: Galaxy abundances (574); High-redshift galaxies (734); Scaling relations (2031)

1. Introduction

For over four decades, scaling relations pertaining to the chemical enrichment of galaxies’ interstellar medium (ISM) have been used to study galaxy formation. Lequeux et al. (1979) investigated the correlation between galaxy luminosity and gas-phase oxygen abundance and found that more luminous galaxies hosted H II regions with higher O/H. Later, large surveys like the Sloan Digital Sky Survey (SDSS; York et al. 2000) enabled studies of large statistical samples of nearby galaxies. Tremonti et al. (2004) were among the first to complete a detailed analysis and found a strong positive correlation between stellar mass (M_*) and gas-phase O/H, which was relatively steep at lower masses and started to flatten at $M_* \gtrsim 10^{10.5} M_\odot$. The $M_*\text{--O/H}$ relation—often referred to as the “mass–metallicity relation” or simply the MZR—has continued to be studied extensively at $z \sim 0$ (see Section 5 of Maiolino & Mannucci 2019 for a review), and it has also been studied using resolved observations of nearby galaxies (e.g., González Delgado et al. 2014; Barrera-Ballesteros et al. 2017; Sánchez et al. 2019). As the size and quality of high-redshift spectroscopic surveys has increased, the MZR has been confirmed to exist up to at least $z \sim 3.5$ (e.g., Erb et al. 2006a; Maiolino et al. 2008; Steidel et al. 2014; Zahid et al. 2014; Onodera et al. 2016; Sanders et al. 2021, among many others).

Still, connecting the results from these studies and constructing a single clear picture of chemical enrichment over cosmic time has remained difficult, in large part because of the many different methods used to infer O/H. Although some studies of individual nearby objects have used “direct method” abundances, based on measurements of the electron temperature⁶ (T_e) in H II regions (e.g., Aller 1954; Pilyugin & Thuan 2005; Berg et al. 2012, 2020; Pérez-Montero 2014), the spectral lines required to determine T_e are exceedingly faint. Therefore, studies of faint, distant, and/or metal-rich galaxies have relied on more indirect tracers of chemical enrichment. The most common method is to use the ratios of strong emission lines in galaxy spectra, which can be empirically linked to gas-phase O/H using direct method abundances measured for $z \sim 0$ samples (e.g., McGaugh 1991; Pettini & Pagel 2004). Even at the same redshift as the calibration samples, however, the MZRs determined using different “strong-line” methods often do not agree and differ substantially in both shape and normalization (Kewley & Ellison 2008; Andrews & Martini 2013).

There are additional concerns about the application of strong-line methods at high redshift, where galaxies have significantly different stellar populations and conditions in their interstellar gas. The crux is that line ratio diagnostics are frequently sensitive to more than one parameter—e.g., O/H and the abundance of other elements, like nitrogen, or O/H and the ionization conditions in the gas—many of which are observed to be correlated with one another in nearby galaxies.

⁵ Carnegie-Princeton Fellow.

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

⁶ Oxygen and other heavy elements provide the principal cooling mechanism in photoionized gas, linking temperature to enrichment.

Indeed, the implicit dependence on these underlying correlations is fundamental to the success of strong-line diagnostics. Unfortunately, we do not know a priori whether these properties are correlated with one another in the same way in high- z galaxies as in $z \sim 0$ galaxies, and so caution must be exercised when using locally calibrated line ratio diagnostics to study distant galaxies.

Ideally, we would “recalibrate” common line ratio diagnostics using a representative sample of high- z galaxies with measurements of T_e . However, despite heroic efforts from the ground (Yuan & Kewley 2009; Erb et al. 2010; Steidel et al. 2014; Sanders et al. 2016; Kojima et al. 2017; Sanders et al. 2020a), this has not yet been achieved, because measurable T_e corresponds to a relatively narrow range of O/H. Over the next several years, we can hope to revisit the task of constructing new, T_e -calibrated diagnostics that are more appropriate for measuring O/H at high redshift using observations from the James Webb Space Telescope (JWST; Gardner et al. 2006), which is scheduled to be launched later this year.

Even so, oxygen abundance is only part of the picture. In the nearby universe, it is common to study the chemical enrichment of galaxies using multiple elements. These studies frequently use observations of the stellar continuum to measure, e.g., the iron, magnesium, and carbon content of the integrated stellar population; see Section 2 of Maiolino & Mannucci (2019) for a description of the methodology, but also Gallazzi et al. (2005), which is the complementary study of stellar metallicity in the same galaxies studied by Tremonti et al. (2004). More detailed abundance patterns are also central to studies of the circumgalactic medium (CGM) at all redshifts (Tumlinson et al. 2017; Zahedy et al. 2019, 2021). More recently, very deep spectra of massive quiescent galaxies at high redshift have enabled the measurement of multiple elemental abundances (Kriek et al. 2016, 2019; Jafarizadeh et al. 2020). Using deep rest-UV spectra, it is also becoming possible to measure iron abundances in high- z star-forming galaxies (e.g., Steidel et al. 2016; Cullen et al. 2019; Topping et al. 2020; Cullen et al. 2021). These studies represent a promising direction for future high- z galaxy research, because understanding the abundance *patterns* of galaxies provides additional insight into their assembly histories—particularly the comparison between iron and α elements (including Mg and O), as the α/Fe ratio is sensitive to star formation timescales (e.g., Tinsley 1979).

The goal of this paper is to investigate the abundance patterns of a large sample of individual high- z galaxies in a way that avoids the systematic biases introduced by using locally calibrated strong-line methods. We update the method introduced by Strom et al. (2018; hereafter Strom18) and use it to self-consistently infer O/H, N/H, and Fe/H for 195 star-forming galaxies at $z \simeq 2\text{--}2.7$, the largest sample of high- z galaxies for which multiple elemental abundances have been reported. We introduce the galaxy sample and observations in Section 2. Section 3 presents the photoionization model method used for measuring the chemical abundances. In Section 4, we report the M_* -O/H relation for our $z \sim 2$ sample and compare and contrast the correlation based on our photoionization model method with the relations from more commonly used strong-line methods. Section 5 presents the corresponding M_* -N/H and M_* -Fe/H relations. Section 6 discusses the physical insights that result from measuring abundance patterns rather

than a single bulk metallicity, including constraints on galaxy star formation histories (SFHs). We summarize our findings and conclude in Section 7.

Throughout the paper, we refer to specific spectral features using their vacuum wavelengths and adopt the solar metallicity scale from Asplund et al. (2009), with

$$12 + \log(\text{O}/\text{H})_{\odot} = 8.69$$

$$12 + \log(\text{N}/\text{H})_{\odot} = 7.83$$

$$12 + \log(\text{Fe}/\text{H})_{\odot} = 7.50$$

Historically, 12 is added when reporting log abundances relative to H to ensure that values are always positive. This is not used when reporting abundance ratios such as $\log(\text{N}/\text{O})$.

2. Data

In this work, we analyze a subsample of galaxies drawn from the Keck Baryonic Structure Survey (KBSS; Rudie et al. 2012; Steidel et al. 2014), using the same criteria as Strom18 (summarized in Section 2.2) to construct the sample. KBSS is a targeted spectroscopic survey of $1.5 \lesssim z \lesssim 3.5$ galaxies in 15 separate fields, each centered on a bright quasar. Each survey field is approximately $6' \times 8'$, resulting in a total survey area of $\approx 0.24 \text{ deg}^2$. Extensive multiwavelength imaging and both optical and near-infrared (NIR) spectroscopic observations have been conducted in all fields and are described in detail in other work (e.g., Steidel et al. 2003; Reddy et al. 2012; Rudie et al. 2012; Steidel et al. 2014; Strom et al. 2017). Below, we review the aspects of the survey that are most relevant to the present analysis.

2.1. Photometry and Parent Sample Selection

The majority of KBSS galaxies are selected on the basis of their rest-UV colors, using observed optical imaging in the U_n , G , and R bands and the color selection criteria introduced by Adelberger et al. (2004) and Steidel et al. (2004) to identify Lyman Break Galaxy analogs at $z \simeq 2\text{--}2.7$ (sometimes referred to as “BX” and “BM” galaxies). This color selection largely translates to a star formation rate-selected sample and is successful at identifying both young galaxies and older galaxies with significant current star formation or with rising SFHs. Relying on the shape and brightness of galaxies’ rest-UV spectral energy distributions (SEDs) can introduce a bias against very massive galaxies and galaxies with heavily reddened ($E(B - V)_{\text{cont}} > 0.3$) UV spectra. To mitigate these biases, we incorporate knowledge of the shape of galaxies’ rest-optical spectra by relaxing the rest-UV color selection and pairing it with a cut in $R - K_s$ color, which probes the 4000 Å and Balmer breaks across the redshift range of interest. Strom et al. (2017) provide a more thorough description of these extended selection criteria, but in brief: these “RK” galaxies occupy a region of rest-UV color space where $z \sim 2$ galaxies selected using other NIR methods, such as distant red galaxies (DRGs; Franx et al. 2003) and BzK galaxies (Daddi et al. 2004), are found (see also the discussion by Reddy et al. (2005)).

2.2. Near-infrared Spectroscopy

Beginning in 2012, NIR spectroscopic observations of KBSS galaxies have been conducted in J -, H -, and K -band

using the Multi-Object Spectrometer for InfraRed Exploration (MOSFIRE; McLean et al. 2012) installed at the Cassegrain focus of the Keck I telescope on Maunakea. We presented the first results from the MOSFIRE component of KBSS in Steidel et al. (2014; hereafter Steidel14), with a comprehensive analysis of an expanded sample later being reported in Strom et al. (2017). We refer the reader to those two papers for complete details regarding the NIR spectroscopic data acquisition, reduction, and analysis.

The spectra were acquired using $0''.7$ slits, with typical seeing of $0''.5$ – $0''.8$. The data were reduced using the publicly available data reduction pipeline,⁷ and subsequently corrected to vacuum heliocentric velocity. If multiple spectra of the same object were acquired, they were shifted to match in the spatial direction and combined using inverse variance weighting. Slit losses were determined to first-order by comparing observations of a relatively bright star placed on a mask slit with its photometric magnitude in the same band; individual object slit losses were determined by comparing observations of the same object on multiple masks. The typical slit loss corrections are factors of ~ 2 , consistent with the estimates based on comparisons of detected continua with broadband photometry and SED models.

The reduced 2D spectrograms were analyzed using the custom IDL package *mospec*⁸, developed by A. Strom specifically for interacting with MOSFIRE spectra of emission line galaxies (Strom et al. 2017). Galaxy spectra were extracted using boxcar extraction apertures, then fit using the best-fit SED model as the stellar continuum and a Gaussian emission line model with a single redshift and line width per band. The ratios of the [O III] $\lambda\lambda 4959, 5008$ and [N II] $\lambda\lambda 6549, 6585$ doublets are fixed at 3:1, where the longer-wavelength lines are $3\times$ the strength of their shorter-wavelength counterparts. This greatly improves emission line measurements when one of the two features is impacted by an OH sky line.

To date, ~ 1500 KBSS galaxies have been observed with MOSFIRE, ~ 800 of which fall in the redshift range $1.9 \leq z \leq 2.7$, where many of the key rest-optical emission features are accessible from the ground at NIR wavelengths. As in Strom18, galaxies are selected for analysis when there are observations of the spectral regions near $H\alpha$, $H\beta$, [O III] $\lambda 5008$, and [N II] $\lambda 6585$. Measurements of or limits on [O II] $\lambda\lambda 3727, 3729$, [O III] $\lambda 4363$, [S II] $\lambda\lambda 6718, 6732$, and [Ne III] $\lambda 3869$ are incorporated when present. Objects are included in the sample regardless of the signal-to-noise (S/N) ratio of any single line measurement, but we require the ratio of the band-to-band slit corrections to be less than a factor of 2, as well as an $S/N > 5$ measurement of the Balmer decrement $H\alpha/H\beta$, which is used to account for reddening due to dust.⁹ Galaxies with evidence of significant contamination from active galactic nuclei (in either their rest-UV or rest-optical spectra) or whose $H\alpha$ or $H\beta$ is compromised by a nearby OH line are excluded. In total, these criteria result in a sample of 196 galaxies with $\langle z \rangle = 2.3$. Of these, 195 galaxies have good abundance measurements, as described below in Section 3.¹⁰

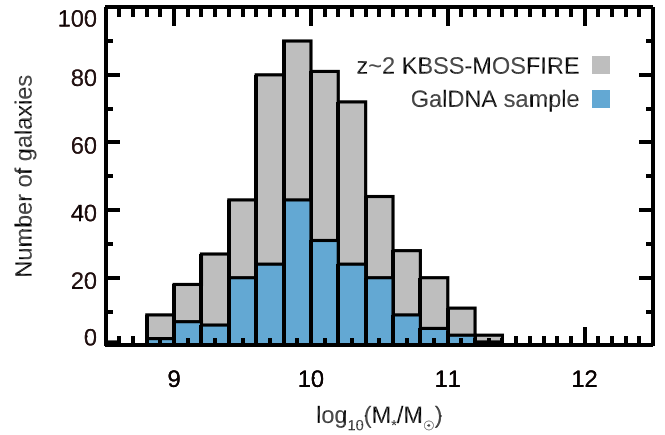


Figure 1. The distribution of stellar mass M_* for the 195 galaxies in our final sample, shown in blue. For comparison, the gray histogram shows the distribution for the complete parent sample of KBSS galaxies with $1.9 \leq z \leq 2.7$. The paper sample is consistent with the full KBSS sample and has a median $M_* = 10^{10.0} M_\odot$.

2.3. Stellar Masses

Stellar mass (M_*) estimates for the KBSS galaxies are inferred from reddened stellar population synthesis models fit to broad- and intermediate-band photometry spanning the rest-UV to the rest-IR. The rest-optical (observed-NIR) magnitudes are corrected for line emission using the measurements from the MOSFIRE spectra described above. The general SED fitting methodology is explained by Reddy et al. (2012), with a description of its application to the current KBSS sample being found in Strom et al. (2017). Theios et al. (2019) examined the impact of using other stellar population models (i.e., those with lower Z_* , as expected for stellar populations at high redshift) on the parameters inferred from the SED fitting. Fortunately, rank ordering in M_* is generally preserved regardless of the specific choice of model. As a result, we use the stellar masses from Strom et al. (2017), based on the Bruzual & Charlot (2003) stellar population synthesis models with a Chabrier (2003) initial mass function, to facilitate a more straightforward comparison with other analyses that use similar SED fitting techniques and models. The M_* distributions for the full KBSS sample (the gray histogram) and for the 195 galaxies with well-measured abundances (the blue histogram) are shown in Figure 1. These two samples are consistent with being drawn from the same parent population, based on a two-sample Kolmogorov–Smirnov test. The typical statistical uncertainty on $\log(M_*/M_\odot)$ is ≈ 0.16 dex (Shapley et al. 2005; Erb et al. 2006b).

3. Chemical Abundances with GaDNA

There are a number of ways to determine the level of chemical enrichment or metallicity of distant galaxies. Broadly speaking, all methods rely on measurements of emission lines from the photoionized gas in galaxies’ star-forming regions and fall into two categories: (1) empirical diagnostics based on “direct” T_e -based measurements of a calibration sample, where both bright nebular lines and faint auroral lines can be detected; or (2) photoionization model methods, where observed emission lines are compared to predictions made by models assuming an input ionizing source and set of physical conditions in the gas.

⁷ <https://www2.keck.hawaii.edu/inst/mosfire/drp.html>

⁸ <https://github.com/allisonstrom/mospec>

⁹ The S/N calculation for the Balmer decrement accounts for uncertainties in the relative flux calibration between NIR bands.

¹⁰ It was not possible to identify a photoionization model solution for Q1623-BX449, although the Markov Chain Monte Carlo chains did converge.

Empirical methods implicitly depend on the correlation of various quantities, including, e.g., the ionization state of the gas and the relative abundances of elements like oxygen and nitrogen. The existence of these underlying correlations causes galaxies and H II regions to form tight sequences in line-ratio space—for example, the Baldwin–Phillips–Terlevich or “BPT” diagram comparing $[\text{O III}]/\text{H}\beta$ and $[\text{N II}]/\text{H}\alpha$ (Baldwin et al. 1981; Veilleux & Osterbrock 1987)—and allow many different line ratios to be used to infer O/H. Even if a line ratio is only weakly correlated with O/H, so long as the quantity that it is most strongly correlated with also correlates with O/H, the diagnostic can still be used.

In contrast, photoionization model methods allow the chosen parameters to vary alongside gas-phase metallicity and thus be explicitly determined for individual objects. Examples include IZI (Blanc et al. 2015), which returns ionization parameter U and O/H; BOND (Vale Asari et al. 2016), which measures both O/H and N/O; and NebulaBayes (Thomas et al. 2018), which measures ISM pressure (related to gas density) in addition to U and O/H. Because we are interested in determining multiple chemical abundances for the galaxies in our sample, we use the photoionization model method introduced by Strom18, which measures O/H, N/H, and Fe/H, in addition to U . The general method, including recent updates and a revised parameter estimation technique, is described here.

3.1. Photoionization Model Grid

We use the same photoionization model predictions for the strong nebular lines in galaxies’ rest-optical spectra as Strom18, which are generated using Cloudy (v13.02; Ferland et al. 2013), with stellar population synthesis models from the Binary Population and Spectral Synthesis¹¹ code (BPASSv2; Eldridge & Stanway 2016; Stanway et al. 2016) as the input ionizing radiation field. The BPASS models were chosen because of their relative success in reconciling the rest-UV and rest-optical observations of the same $z \sim 2$ galaxies, specifically because the implementation of binary evolution physics in BPASS produces harder ionizing radiation fields at fixed Z_* than, e.g., including stellar rotation or adopting a more top-heavy stellar initial mass function (Steidel et al. 2016). This “boost” of higher energy photons is needed to match the collisional-to-recombination line ratios such as R23.¹² In high- z galaxies, these line ratios tend to be larger than can be explained using stellar models with softer ionizing radiation fields, even at moderate gas-phase O/H, corresponding to peak emission from O ions.

We adopt a plane parallel geometry and assume constant SFHs with an age of 100 Myr and a constant gas density of $n_{\text{H}} = 300 \text{ cm}^{-3}$. This density was originally chosen to be consistent with the electron density n_e ¹³ determined using the $[\text{O II}]$ ratio measured from a stack of $z \sim 2$ –3 KBSS galaxies (Strom et al. 2017), but it is also representative of the typical electron density for the individual galaxies in the current sample, which ranges from $n_e \simeq 150$ –400 cm^{-3} . Using models with $n_{\text{H}} = 100 \text{ cm}^{-3}$ or $n_{\text{H}} = 1000 \text{ cm}^{-3}$ for galaxies with lower or higher densities than this, respectively, does not significantly change their derived abundances or the overall distribution of abundances in the sample. Dust grains are

implemented using the “Orion” mixture, with a dust-to-gas ratio that scales linearly with the metallicity of the gas, Z_{neb} . Model grids are calculated for multiple stellar population synthesis models with increasing stellar enrichment (Z_*), allowing $\log(Z_{\text{neb}}/Z_{\odot})$ and $\log(U)$, defined as $\log(n_{\gamma}/n_{\text{H}})$, to vary in 0.1 dex steps. Initially, we assume a solar abundance pattern in the gas, but interpolate between the fixed grid points and scale the strength of the nitrogen lines a posteriori to accommodate nonsolar N/O. These choices allow us to predict emission line intensities (relative to $\text{H}\beta$) for the following parameter space:

$$\begin{aligned} Z_* &= [0.001, 0.014] \text{ or } Z_*/Z_{\odot} \approx [0.07, 1.00], \\ Z_{\text{neb}}/Z_{\odot} &= [0.1, 2.0], \\ \log(U) &= [-3.5, -1.5], \\ \log(\text{N/O}) &\geq -2.0. \end{aligned}$$

The lower limit on $\log(\text{N/O})$ is ~ 0.5 dex lower than the primary plateau observed for local H II regions (e.g., van Zee et al. 1998; Izotov & Thuan 1999).

These base model parameters can then be converted to abundances:

$$\begin{aligned} [\text{Fe/H}] &= \log(Z_*/Z_{\odot}) \\ 12 + \log(\text{O/H}) &= \log(Z_{\text{neb}}/Z_{\odot}) + 8.69 \\ 12 + \log(\text{N/H}) &= \log(Z_{\text{neb}}/Z_{\odot}) + 8.69 + \log(\text{N/O}), \end{aligned}$$

where the bracket notation for Fe refers to the enrichment relative to solar, $[\text{Fe/H}] = \log(\text{Fe/H}) - \log(\text{Fe/H})_{\odot}$. These translations are motivated by several considerations. First, the effect of changing Z_* on the emission line predictions results from differences in the shape of the ionizing radiation field for a given stellar model family, and the shape of the ionizing spectrum of massive stars is most impacted by the amount of Fe present in their atmospheres. Because this method relies on Fe enrichment in massive stars, we avoid concerns about depletion onto dust that impact the abundance of Fe in the gas. In contrast to the important role Fe plays in determining the properties of massive stars and the amount of gas heating, it is O that provides the majority of the gas cooling via collisionally-excited lines, due to its relatively high abundance with respect to other elements and many low-lying transitions. Thus, we can assume that changes in Z_* correspond to changes in Fe/H, and changes in Z_{neb} correspond to changes in O/H.

We use a Markov Chain Monte Carlo approach to sample the posterior within multivariate parameter space (c.f. Section 3.2 of Strom18), adopting the additional prior that $0.0 \leq \log(Z_{\text{neb}}/Z_*) \leq 0.73$. These limits roughly correspond to a range of α -enhancement (i.e., O/Fe) between the solar value and Fe-poor core-collapse supernovae (CCSNe) yields (Nomoto et al. 2006).

The photoionization model grids and IDL routines used to implement this method, which we collectively refer to as GaldNA (referencing the analogy of abundance patterns as “galaxy DNA”), are publicly available.¹⁴ We caution, however, that some of the built-in assumptions, including that all of the emission from ionized gas observed in a galaxy’s spectrum is produced in H II regions, may be inappropriate for some samples. This is especially true for galaxies at lower redshifts

¹¹ <https://bpass.auckland.ac.nz/>

¹² $R23 = \log\left(\frac{[\text{O III}] + [\text{O III}]}{\text{H}\beta}\right)$.

¹³ For ionized gas, $n_{\text{H}} \approx n_e$.

¹⁴ <https://github.com/allisonstrom/galdna>

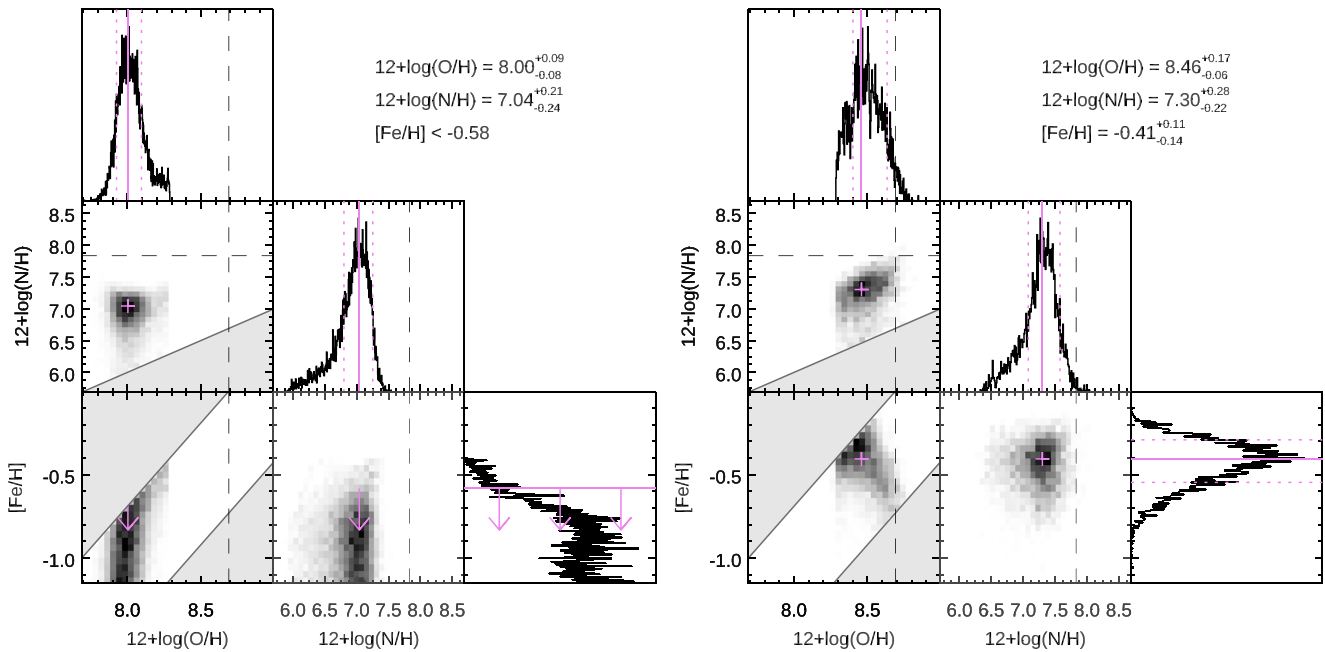


Figure 2. The 1D and 2D posterior PDFs for $12+\log(\text{O}/\text{H})$, $12+\log(\text{N}/\text{H})$, and $[\text{Fe}/\text{H}]$ for Q0105-BX147 ($M_* = 10^{9.5} M_\odot$), where the low-O/H and high-O/H peaks have been separated into the left and right panels, respectively. These portions of the posterior are considered separately when estimating the abundance patterns in each case (denoted by the purple symbols and solid lines, with the dashed lines indicating the 68% HDIs). Disallowed regions in N/O and O/Fe are illustrated by the solid gray shading (c.f. Section 3.1), and the solar values of O/H and N/H are identified by the dashed black lines. To decide between potential solutions, we utilize the corresponding ionization parameter estimates to place the galaxy close to the $z \sim 2$ locus in mass-ionization space (Figure 4). In this case, the low-O/H solution is preferred.

and those with lower star formation rate (SFR) surface densities, where contributions from, e.g., diffuse ionized gas and shocks are likely to be nonnegligible (Sanders et al. 2017).

3.2. Parameter Estimation

In Strom18, we reported the maximum a posteriori (MAP) values for the four univariate posteriors and required the associated 68% highest density intervals (HDIs) to be separated from the boundaries of the allowed parameter space for all parameters except N/O. Here, we revisit the manner in which the upper limits on both N/H and Fe/H are determined, as well as how we treat galaxies with multiple peaks in their posterior probability distribution function (PDF). These changes allow us to better capture the abundance patterns in the largest possible sample.

3.2.1. Bimodal Posteriors

The connection between specific emission lines and the model parameters in GalDNA is discussed at length in Strom18, but there are a few key results that inform our physical intuition about how differences in galaxy spectra correspond to differences in model parameters and, thus, differences in chemical abundances. Foremost is the fact that ratios of collisionally-excited lines to recombination lines (e.g., $[\text{O III}]/\text{H}\beta$ and $[\text{S II}]\lambda\lambda 6718, 6732/\text{H}\alpha$) effectively trace the amount of photoionization heating per ionization of an H atom. Thus, all else being equal, these ratios will increase with decreasing Z_* (i.e., Fe/H), because Fe-poor stars produce more energetic ionizing photons on average than Fe-rich stars. At the same time, many of these ratios are also double-valued: they increase as gas enrichment (Z_{neb} or, equivalently, O/H) increases, until gas cooling mediated by the heavy elements

in the gas causes the gas temperature to decrease, at which point these line ratios also decrease.¹⁵ This behavior is well-known, and is one of the challenges inherent in using line diagnostics like R23 to infer O/H, because a single value of R23 can correspond to two different values of O/H. Further, R23 remains at $\sim 0.8\text{--}1.0$ (which is where most high- z galaxies fall) for a large range in O/H near the “turnover” where gas cooling begins to outpace increasing gas enrichment. The combination of these effects mean that maximal values of line ratios like $[\text{O III}]\lambda 5008/\text{H}\beta$, $[\text{O II}]/\text{H}\beta$, and $[\text{S II}]/\text{H}\beta$ are only achieved at low Fe/H and moderate O/H. Lower values of these line ratios will usually correspond to either lower or higher O/H, which manifest as bimodal posteriors in GalDNA. Figure 2 presents the posterior PDFs for one galaxy (Q0105-BX147) where this occurs, with the total posterior being divided into low-O/H (left panel) and high-O/H (right panel) “solutions.”

Figure 3 demonstrates why properly accounting for galaxies with bimodal posteriors is important for characterizing the correlation between O/H and other quantities, such as M_* or strong-line ratios. Galaxies with a single local maximum in their O/H posterior (the gray histogram) range from $12 + \log(\text{O}/\text{H}) \approx 8$ (roughly 20% solar) to just above solar O/H. There is overlap between the upper edge of this distribution and the higher-O/H maxima for galaxies with bimodal posteriors (the open purple histogram). In contrast, lower-O/H maxima only occur in bimodal posteriors, and are not seen for galaxies with single-peaked posteriors.

To determine which combination of parameters—and thus which O/H—is preferred for individual galaxies, we rely on

¹⁵ A notable exception is $[\text{N II}]\lambda 6585/\text{H}\alpha$, which continues to increase with increasing O/H, due to the rapid increase in N/O at higher O/H.

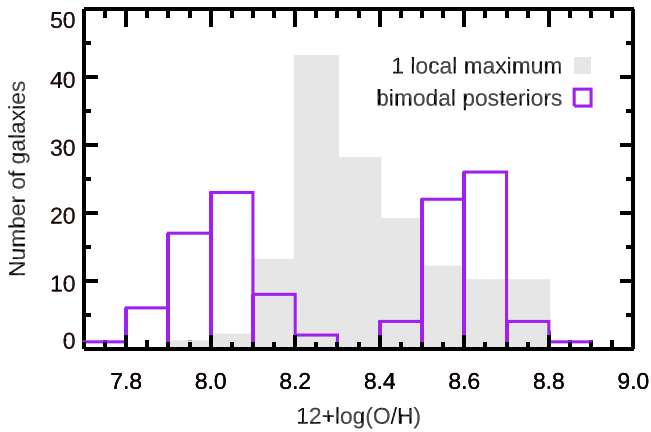


Figure 3. The distribution of GalDNA-inferred O/H for the $z \sim 2$ galaxies in our sample. The light gray histogram represents the set of 138 objects having either a single local maximum in the posterior PDF or only one valid solution, if the posterior PDF is double-peaked. The open purple histogram shows the distribution of model-inferred O/H for both solutions in the 57 remaining galaxies with bimodal posteriors. While the distribution for galaxies with single-peaked posteriors extends to relatively high O/H, the lowest abundances are only probed by galaxies with more than one potential model solution.

the anticorrelation between U and M_* . This relation is one of the strongest and most significant correlations between galaxy properties, and it has been observed in galaxy samples at multiple redshifts using both ionization-sensitive line ratios (Sanders et al. 2015, 2020b; Jeong et al. 2020) and the inferred ionization parameter (e.g., Kaasinen et al. 2018). The tight locus of galaxies in the mass-excitation diagram (e.g., Juneau et al. 2011, 2014), which compares $[\text{O III}]/\text{H}\beta$ and M_* , is another manifestation of this relationship.

Using the ionization parameter is also practical because it avoids directly imposing a prior on the distribution of abundances, and because U is generally the most precisely determined model parameter when using GalDNA. Further, the distributions of M_* and U for galaxies with a single maximum in their posterior and for galaxies with bimodal posteriors (including both possible combinations of parameters) are statistically consistent with one another, unlike the distributions of O/H seen in Figure 3. In other words, not only can both sets of galaxies be expected to follow the same locus in the M_* – U plane, but the two U values for galaxies with bimodal posteriors are distinct enough to meaningfully discriminate between the possible combinations of parameters. To accomplish this, we use the single-peaked galaxies to define the best-fit M_* – U relation, shown by the solid black line in Figure 4. For galaxies with bimodal posteriors, we then choose the combination of parameters with U closest to this locus. The preferred “solutions” determined in this manner are represented by the dark blue squares (low-O/H maxima) and red triangles (high-O/H maxima).

The resulting distributions of U for the low-O/H and high-O/H solutions are significantly different from one another, with the low-O/H galaxies having generally higher U than the high-O/H galaxies. For the galaxies that are outliers with respect to these U clusters (Q0821-MD5, Q1700-MD103, and Q1623-BX293), we choose the combination of parameters that minimizes this separation. There are 57 galaxies with bimodal posteriors; we select the low-O/H solution for 31 of these and the high-O/H solution for 26. The portion of the multivariate

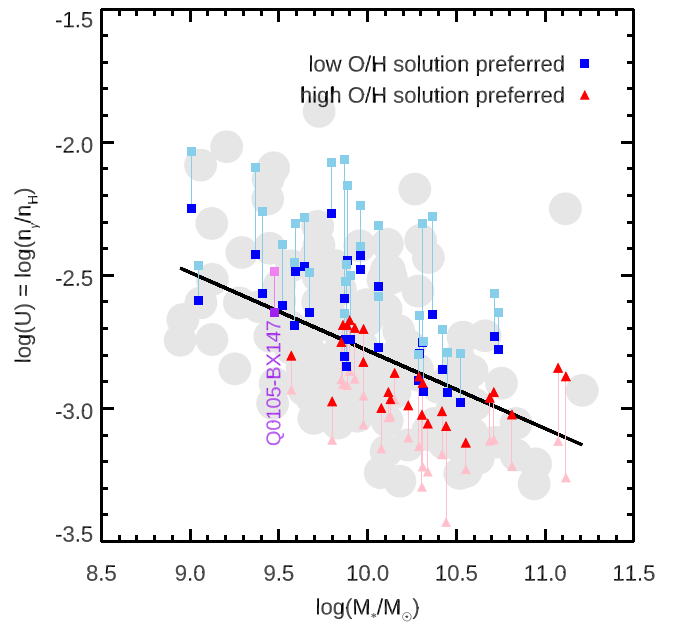


Figure 4. Model-inferred $\log(U)$ and M_* for galaxies with a single peak in the posterior PDF (the large gray points) and for galaxies with multimodal posteriors (the colored points). The distributions of both quantities are consistent between the two samples. The best-fit linear relation for the locus of the gray points is shown by the solid black line. For galaxies with multiple possible abundance patterns, the darker symbols indicate the solutions that are most consistent with this locus, and the lighter symbols show the locations of the second, dispreferred solutions. In general, a lower U corresponds to the low-O/H peak in the posterior, and a higher U corresponds to the high-O/H peak in the posterior. The parameter values for Q0105-BX147, which were also shown in Figure 2, are marked by the purple squares; lower U is preferred and corresponds to the lower O/H peak in the posterior (the left panel in Figure 2).

posterior corresponding to the preferred peak is analyzed on its own in order to determine the MAP values for O/H, N/H, and Fe/H for the individual galaxies.

We also investigated other criteria for deciding between unique combinations of parameters for galaxies with bimodal posteriors. These strategies, including assessing the proximity to the N/H–O/H locus of local H II regions or deciding based on posterior density alone, do not change the qualitative results presented below. We adopt and endorse the method based on the M_* – U relation because it is easily applied in all cases, and uses additional knowledge about the individual galaxies themselves to make a more informed choice.

3.2.2. Nitrogen and Iron Limits

Occasionally, the $[\text{N II}]\lambda\lambda 6549, 6585$ doublet is poorly detected or undetected in the spectra of the galaxies in our sample, even when other emission lines (including nearby $\text{H}\alpha$) are well-detected. This can occur as a result of low nitrogen abundance and/or high levels of ionization. For our analysis, we do not require the significant detection of any single emission line (just an $\text{S/N} \geq 5$ measurement of the Balmer decrement), but these low S/N measurements of $[\text{N II}]$ correspond to posterior PDFs that about the lower limit of $\log(\text{N/O}) \geq -2$. Because the imposed prior is on N/O and not N/H, the lowest allowed N/H is a function of O/H. This will tend to produce a well-defined local maximum in N/H space, even when N/O is an upper limit. To account for this, we test whether both the N/H and N/O posterior PDFs are bounded

(i.e., detached from the limits of the sampled parameter space). If neither the N/H nor the N/O posteriors are bounded, or if the N/O posterior alone is not bounded, the nitrogen abundance is assumed to be an upper limit given by

$$12 + \log(\text{N}/\text{H}) = 12 + \log(\text{O}/\text{H}) + \log(\text{N}/\text{O})_{68},$$

where $\log(\text{N}/\text{O})_{68}$ is the 68th percentile in $\log(\text{N}/\text{O})$. Of the 195 galaxies in our sample, 36 have upper limits on N/H determined in this way.

Iron abundance can also be more difficult to estimate because of the more complicated nature of the Fe/H and Fe/H–O/H posteriors. The most common appearance is a horseshoe shape in Fe/H versus O/H space, opening downward in the lower left panels of the corner plots in Figure 2. This structure in the 2D posterior arises because there is an inherent trade-off between O and Fe for a given galaxy spectrum: as the ionizing radiation field softens with increasing Fe/H—moving toward lower O/Fe—it is necessary to have more O to produce the same line emission, but eventually gas cooling due to the increasing O/H begins to dominate, and harder ionizing radiation is required to counteract this effect—moving back toward higher O/Fe. This is closely related to the double-valued behavior of R23 with O/H, but additionally highlights the crucial interplay between gas heating and gas cooling.

Because of the prior requiring $0.0 \leq [\text{O}/\text{Fe}] \leq 0.73$, the two arms of the horseshoe are commonly separated and appear as a bimodal O/H posterior, as we have previously discussed. The 1D Fe/H posterior is also frequently bimodal, although the local maxima are less likely to be bounded. An example of this can be seen in the low-O/H panel on the left in Figure 2, where the marginalized PDF in the bottom right histogram has high posterior density even at the lowest allowed Fe/H, and the 2D posteriors involving Fe/H appear as vertical stripes. In these cases, which also occur at higher O/H, we adopt the 68th percentile in [Fe/H] as an upper limit. A smaller number of galaxies have posteriors that abut the low-O/Fe boundary at $[\text{O}/\text{Fe}] = 0.0$, but these often have clear local maxima in the univariate Fe/H posterior, and so we treat them as bounded. In total, 24 galaxies have upper limits on Fe/H.

4. The $\langle z \rangle = 2.3$ Stellar Mass–O/H Relation

Most studies of metallicity in star-forming galaxies, which are largely studied via the nebular emission from their H II regions, focus on oxygen. This is both physically-motivated—oxygen is the most abundant element (by mass) in the universe after hydrogen and helium—and practical, as emission lines of various ions of oxygen are among the brightest features observed in the spectra of galaxies at all redshifts. As a result, many studies use “metallicity” interchangeably with oxygen abundance, and the relation between galaxies’ stellar mass and gas-phase O/H is among the most common scaling relations discussed in the literature.

Figure 5 shows the observed correlation between O/H and M_* (hereafter O–MZR) for our sample of 195 $z \sim 2$ galaxies, where O/H has been measured using GalDNA, as described in Section 3. The asymmetric vertical error bars for each point represent the 68% HDI on the univariate O/H posterior, and the horizontal error bars represent the typical ≈ 0.16 dex uncertainty on the stellar masses, not taking into account any systematic uncertainties in determining M_* . A Spearman rank correlation test indicates a weak ($\rho = 0.21$) but significant

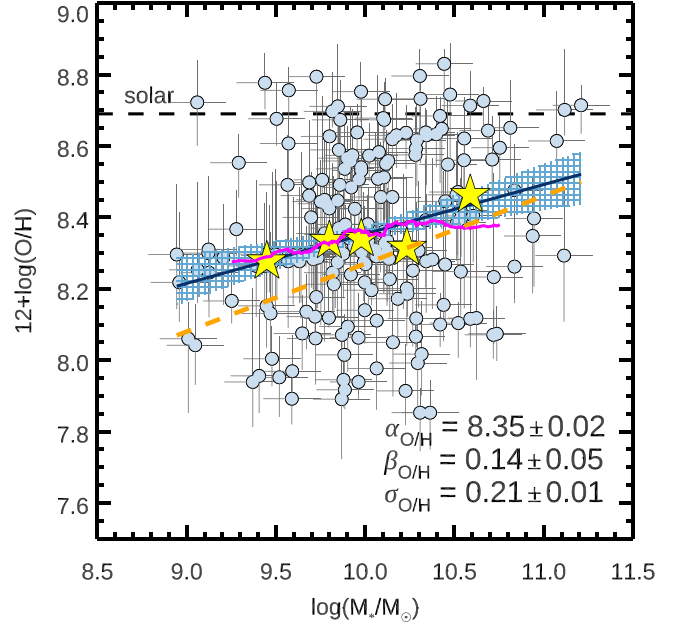


Figure 5. The correlation between stellar mass M_* and GalDNA-based gas-phase O/H for 195 galaxies at $z \sim 2$, with the medians in bins of M_* represented by the yellow stars. The solar value of $12 + \log(\text{O}/\text{H})_{\odot} = 8.69$ is represented by the horizontal dashed line. The magenta line shows the center moving average for O/H with M_* . The dark blue line is the best-fit linear relation, with the blue hatching indicating the 1σ uncertainties on the slope and normalization. Most $z \sim 2$ galaxies have subsolar O/H, and there is a marginally significant positive correlation between O/H and M_* (Spearman $\rho = 0.21$, $p = 0.003$). However, there is also substantial intrinsic scatter ($\sigma_{\text{O}/\text{H}} = 0.21 \pm 0.01$ dex), which together with the measurement uncertainties could camouflage a more significant correlation. The O3N2-based O–MZR reported by Steidel14 for a smaller sample of KBSS galaxies is shown by the dashed orange line.

($p = 0.003$) positive correlation¹⁶ (Table 1). The magenta line shows the center moving average to give a nonparametric sense of this correlation, which increases by ≈ 0.12 dex in $\log(\text{O}/\text{H})$ (or $\sim 30\%$ in O/H) over 1.5 decades in stellar mass. The yellow stars represent the median $12 + \log(\text{O}/\text{H})$ in bins of M_* , which also show an increase in oxygen enrichment with increasing mass.

The large apparent scatter (due to both measurement uncertainties and intrinsic scatter) could camouflage a more significant correlation than is indicated by the Spearman test. To account for this, we also use LINMIX_ERR (Kelly 2007) to characterize the relation between M_* and O/H; LINMIX_ERR is an IDL-based routine that performs linear regression when there are measurement errors in both variables. Using this method, we find a 99.8% probability of a positive correlation, compared to 99.7% from the Spearman test. The solid blue line shows the best-fit linear relation,

$$12 + \log(\text{O}/\text{H}) = \alpha + \beta[\log(M_*/M_{\odot}) - 10], \quad (1)$$

with $\alpha_{\text{O}/\text{H}} = 8.35 \pm 0.02$ and $\beta_{\text{O}/\text{H}} = 0.14 \pm 0.05$. These fit parameters, as well as the intrinsic scatter of this relation ($\sigma_{\text{O}/\text{H}} = 0.21 \pm 0.01$), are also reported in Table 1. The hatched

¹⁶ For the purpose of characterizing the correlations in this paper, we consider $\rho < 0.3$ to be a weak correlation, $0.3 \leq \rho \leq 0.6$ to be a moderate correlation, and $\rho > 0.6$ to be a strong correlation. We use the probability of the null result p to assess the significance of the correlation, with $p \leq 0.003$ considered significant.

Table 1
Abundance Scaling Relations with M_*

Element	Spearman Test		LINMIX_ERR			
	ρ	p	P(%) ^a	α^b	β	σ_{int} (dex)
12+log(O/H) _{GalDNA}	0.21	3×10^{-3}	99.8	8.35 ± 0.02	0.14 ± 0.05	0.21 ± 0.01
12+log(O/H) _{O3N2}	0.50	9×10^{-14}
12+log(N/H)	0.28	8×10^{-5}	100.0	7.07 ± 0.03	0.29 ± 0.07	0.33 ± 0.02
[Fe/H]	0.27	1×10^{-4}	100.0	-0.69 ± 0.02	0.17 ± 0.05	0.22 ± 0.01

Notes.

^a Probability of a positive linear relation, based on the posterior distribution for β .

^b Determined at $M_* = 10^{10} M_\odot$; c.f. Equation (1).

blue area represents the 1σ uncertainties in the slope and normalization.

The general sense of this relation is consistent with other studies of the O-MZR at $z \sim 2-3$. The majority of galaxies have subsolar oxygen abundance, with $\approx 45\%$ (O/H) _{\odot} at $M_* = 10^{10} M_\odot$ and trending toward solar at high M_* . For comparison, the first O-MZR reported for KBSS galaxies by Steidel14, which used the O3N2¹⁷ diagnostic from Pettini & Pagel (2004; hereafter PP04) to infer O/H, is shown by the dashed orange line in Figure 5 and is defined by $\alpha_{\text{Steidel14}} = 8.27 \pm 0.01$ and $\beta_{\text{Steidel14}} = 0.19 \pm 0.02$. The new GalDNA relation is shallower than the Steidel14 relation, with the slopes differing by a little more than 1σ . The normalization of the GalDNA O-MZR is also higher than the earlier O3N2 relation by 0.08 ± 0.02 dex. These differences can be attributed to changes in the sample and the sample size, as well as differences in methodology, which we discuss in more detail below.

In particular, discrepancies in the abundance scale or normalization of the O-MZR determined using different methods are not unexpected. The O3N2 calibration from PP04 that is used in Steidel14 is based on a sample of nearby H II regions with known T_e -based abundances. However, the direct method has been shown to result in abundances up to ≈ 0.24 dex lower than those measured using recombination lines (Esteban et al. 2004, 2014; Blanc et al. 2015). Steidel et al. (2016) found an offset of 0.25 dex between the T_e -based O/H and the photoionization model-inferred O/H for a composite spectrum of 30 KBSS star-forming galaxies; the O/H estimated using the O3N2 index was intermediate between these two values. The offset between the new GalDNA O-MZR and the older O3N2 O-MZR from Steidel14 is less than the offset between the model O/H and O3N2 O/H reported in that work.

4.1. Slope

The GalDNA O-MZR in Figure 5 has $\beta_{\text{O/H}} = 0.14 \pm 0.05$ and is shallower than the previously reported KBSS O-MZR from Steidel14, as well as most other high- z O-MZR relations reported in the literature (Erb et al. 2006a; Maiolino et al. 2008; Yabe et al. 2014; Zahid et al. 2014; Sanders et al. 2018, 2021, among others), which range from $\beta \sim 0.2$ to $\beta \sim 0.35$. Our current understanding of galaxy assembly reasonably predicts a positive correlation between M_* and O/H, given that the more stars that are created, the larger the mass of oxygen that is produced. The slope of the O-MZR can also be affected by

other astrophysical processes, including inflows and outflows, which determine how much of the oxygen is retained or expelled and how much it is diluted by infalling low-metallicity gas. A comparatively shallow O-MZR may reflect a relationship between gas content and M_* at $z \sim 2$ that keeps O/H higher in low- M_* galaxies and/or leads to relatively lower O/H in high- M_* galaxies. For example, a shallower O-MZR at high redshift could result if low- M_* galaxies are not efficiently ejecting their enriched gas or if high- M_* galaxies retain more of their original (pristine) gas mass or accrete more low-metallicity gas from their surroundings, both of which would dilute the oxygen. The effects of these processes are explored by multiple cosmological simulations (e.g., Brooks et al. 2007; Ma et al. 2016; De Rossi et al. 2017; Torrey et al. 2019), and we briefly return to this issue in Section 6, where we discuss the O-MZR together with the scaling relations for nitrogen and iron.

However, before considering the astrophysical interpretation of the O-MZR, we must first consider the practical issue of converting the observed quantities to O/H and how that may affect the resulting relation. For example, a shallow O-MZR is a natural consequence of a relatively small dynamic range observed in line ratios like R23 in $z \sim 2$ samples. However, other line ratios (including O3N2) exhibit a stronger correlation with M_* , even in high- z samples, and these have historically been used to measure the O-MZR instead. As we have discussed, these O-MZR relations can be dramatically different from one another, but these differences should not always be interpreted as having physical meaning, because one of the primary drivers of the slope of strong-line relations is the calibration used to infer O/H.

Figure 6 shows how different O3N2 calibrations with different coefficients for the line ratio can result in very different O-MZR slopes. The black points represent the slope of the O3N2 O-MZR measured when different calibrations for O3N2 (identified by the labels) are used to infer O/H for the same sample of galaxies as shown in Figures 1 and 5; these results are also summarized in Table 2. Comparing the two most dissimilar calibrations, from Strom18 and Bian et al. (2018), shows that the slope of the O-MZR can differ by up to $\Delta\beta \approx 0.1$ *only because a different calibration was used*. One may attempt to circumvent this uncertainty by selecting the calibration deemed most appropriate for the sample being studied, but it is not always clear how that decision should be made. Bian et al. (2018) use direct-method abundances from stacks of “local analogs” to high- z galaxies in order to construct their calibration. Marino et al. (2013) also use the direct method, in combination with a multiple line-ratio technique, for H II complexes in local galaxies that were observed as part of

¹⁷ O3N2 = log([O III]/H β) – log([N II]/H α).

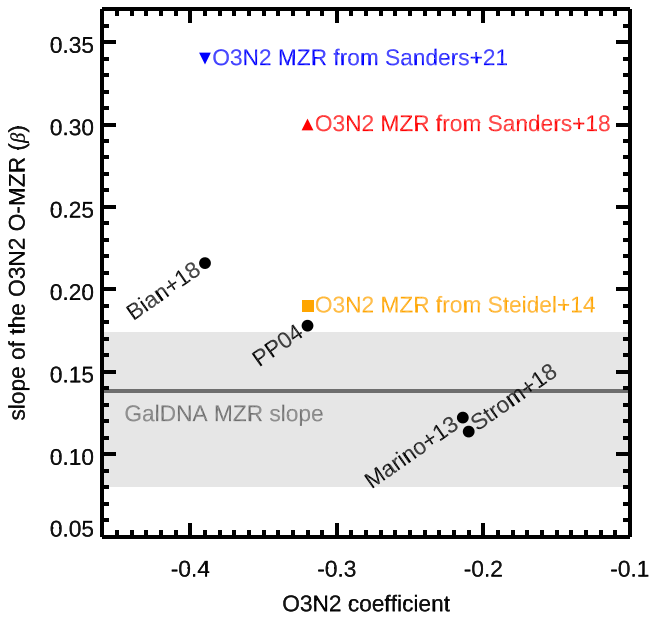


Figure 6. The slope of the O3N2 O-MZR is extremely sensitive to the choice of O3N2 calibration, which is in turn dependent on the correlations between the parameters in the original calibration sample. The round points represent the O3N2 O-MZR measured for our $z \sim 2$ galaxy sample using four different calibrations. As the slope of the O3N2 calibration becomes more negative, the slope of the O-MZR grows steeper, which can result in a steeper O-MZR than observed in Figure 5 (the horizontal gray line and shaded region). Three O3N2 O-MZRs from the literature are shown by the orange square (Steidel et al. 2014, based on a smaller sample of KBSS galaxies), the red triangle (Sanders et al. 2018, from MOSDEF), and the inverted blue triangle (Sanders et al. 2021, also from MOSDEF). These points illustrate the differences that can arise merely from the sample size and sample selection. Notably, the magnitude of the differences that result from choosing different diagnostics is similar to or even larger than the magnitude of the differences that result from sample effects.

Table 2
O3N2 O-MZR Slopes

Sample	O3N2 Calibration	Coefficient	Slope
This paper	PP04	-0.32	0.18 ± 0.02
This paper	Marino+13	-0.214	0.12 ± 0.01
This paper	Bian+18	-0.39	0.22 ± 0.02
This paper	Strom+18	-0.21	0.11 ± 0.01
Steidel+14	PP04	-0.32	0.19 ± 0.02
Sanders+18	PP04	-0.32	0.30
Sanders+21	Bian+18	-0.39	0.34

the Calar Alto Legacy Integral Field Area (CALIFA) survey (Sánchez et al. 2012). In Strom18, we constructed an O3N2 calibration based on the photoionization model oxygen abundances for our $z \sim 2$ sample (albeit only the subsample with single peaks), which resulted in an equation with a coefficient very similar to one based on H II regions from Pilyugin et al. (2012). In the absence of a sufficient sample of direct-method abundances in high-redshift galaxies, any or all of these calibrations might be good choices, depending on the sample under consideration.

The difference between the GalDNA O-MZR and the PP04 O3N2 O-MZR for our current sample is $\Delta_\beta = 0.04$. This is larger than the difference in slope when the same PP04 O3N2 calibration is used for both the current sample and the sample from Steidel14 ($\Delta_\beta = 0.01$). We therefore conclude that, in this

case, the principal explanation for the difference in slope between our current analysis and Steidel14 as seen in Figure 5 is the choice of abundance inference technique, with sample effects playing a more minor role.

At the same time, it is worth noting that the slope of the O3N2 O-MZR reported for KBSS by Steidel14 (the orange square in Figure 6) and the slope reported for the MOSFIRE Deep Evolution Field survey (MOSDEF; Kriek et al. 2015) by Sanders et al. (2018; the red triangle in Figure 6) is still quite large ($\Delta_\beta = 0.11$), even though both are based on the PP04 calibration for O3N2. This indicates that differences in sample selection rather than sample size can also play a role. Nonetheless, because it is impossible to know a priori the specific origin(s) of any reported differences in the O-MZR, it is likely more useful to compare the scaling relations between M_* and line ratio—rather than between M_* and O/H—if O/H has been inferred using a strong-line diagnostic.

4.2. Intrinsic Scatter

A notable feature of the GalDNA O-MZR presented in Figure 5 is the substantial scatter, with the intrinsic scatter (after accounting for measurement error) estimated to be $\sigma_{\text{O/H}} = 0.21 \pm 0.05$ dex. This is reasonably consistent with the 0.15 dex intrinsic scatter reported for the $z \sim 0$ O-MZR for galaxies with similar M_* (Tremonti et al. 2004), but more visually apparent here because of the comparatively larger O/H measurement uncertainties for the high- z galaxies.

The origin of the scatter in the O-MZR is still a matter of some debate, although there is a general consensus that deviations from the $z \sim 0$ relation correlate with SFR, with higher-SFR galaxies exhibiting lower O/H at fixed M_* (e.g., Ellison et al. 2008; Lara-López et al. 2010). This “fundamental metallicity relation” (Mannucci et al. 2010) is likely the result of differences in gas fraction (Bothwell et al. 2013), in the sense that higher-SFR galaxies have larger gas fractions and, thus, have diluted the O formed by their stellar populations. The scatter may also be attributed to the characteristic timescale for oxygen enrichment (Matthee & Schaye 2018; see also Section 6), gas flows (van Loon et al. 2021), or observational effects like inclination angle (Tremonti et al. 2004).

Accurately characterizing the intrinsic scatter in chemical scaling relations like the O-MZR is critical, as it may provide insight regarding the importance of these other processes and/or effects. However, as with the slope of such relations, measuring intrinsic scatter is very sensitive to the method used to infer abundances (Tremonti et al. 2004). For example, the intrinsic scatter we report for the GalDNA O-MZR is significantly larger than that reported for the O-MZR when strong-line diagnostics are used to infer O/H, which is usually $\sigma_{\text{O/H}} \lesssim 0.08$ dex. However, given the similar (or even larger) scatter in the measurements originally used to establish the diagnostics, such a small amount of scatter is unexpected—and may instead indicate that the O-MZR determined using strong-lines is a projection of a higher-dimensional surface linking M_* , O/H, and other parameters.

Figure 7 shows that ionization parameter U is a likely candidate for a third parameter that would account for the diminution in the observed scatter. The left panel shows the GalDNA O-MZR from Figure 5, and the right panel shows the O3N2 O-MZR for the same galaxies using the O3N2 calibration from Strom18. In both panels, the points have been color-coded based on the value of U for that galaxy. The trend

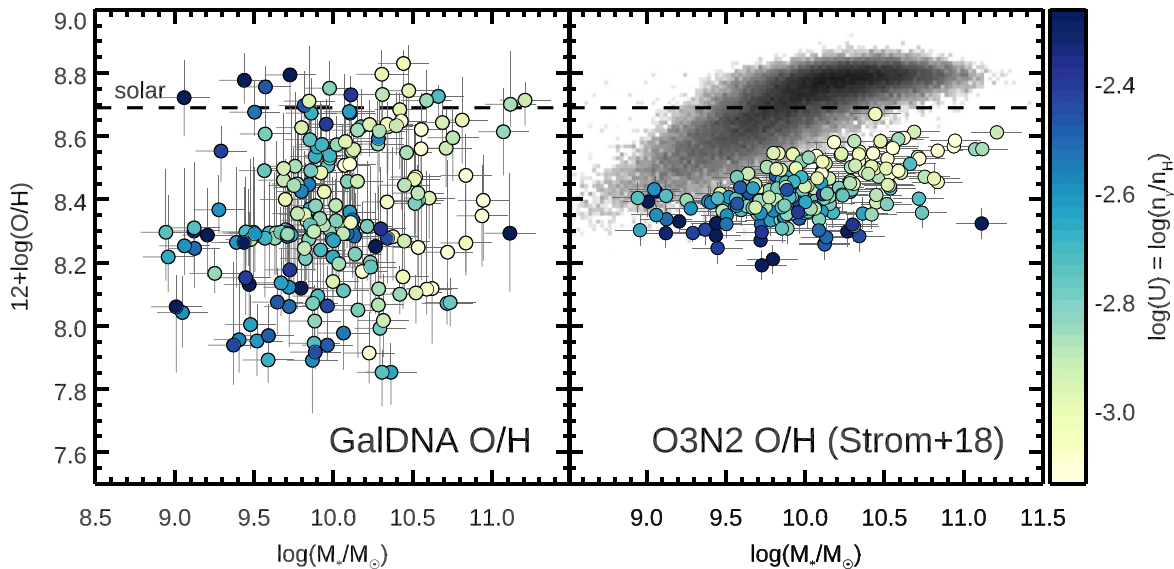


Figure 7. The O-MZR for the $z \sim 2$ galaxy sample presented in this paper, color-coded by ionization parameter (yellow corresponds to low U , and blue corresponds to high U). The oxygen abundances were determined using GalDNA in the left panel and using the O3N2 diagnostic from Strom18 in the right panel (the distribution of $12 + \log(\text{O}/\text{H})_{\text{O3N2}}$ for $z \sim 0$ SDSS galaxies is shown in grayscale). The correlation between O/H and M_* is statistically significant in both cases, but a difference in the intrinsic scatter leads to the remarkably different appearances of the two panels. A strong correspondence between O3N2 and U , which manifests as horizontal isocontours in the right panel, is fundamental to the construction of strong-line diagnostics at $z \sim 0$. Consequently, the tighter O3N2 O-MZR is the result of effectively measuring a combination of two parameters— U and O/H—and not due to low intrinsic scatter in the O-MZR itself. This trend is not observed in the GalDNA O-MZR in the left panel, as O/H and U are only weakly anticorrelated using the model method.

with U is remarkably different between the panels: the isocontours of U are nearly horizontal in the right panel, while there is considerable scatter in U in the left panel. Using the Pearson correlation coefficients, we can compare the strength of the correlation between the model parameters from GalDNA and O3N2: we find $r_{\text{O/H-O3N2}} = -0.40$, indicating a moderate anticorrelation between O3N2 and $\log(\text{O}/\text{H})$, whereas there is a much stronger correlation between O3N2 and $\log(U)$, $r_{U-\text{O3N2}} = 0.74$. The correlation between O3N2 and both $12 + \log(\text{O}/\text{H})$ and $\log(U)$ is $r_{\text{O/H}+U} = 0.78$. This reflects the fact that O3N2 calibrations depend on a strong underlying (anti)correlation between O/H and U . We did not find strong evidence for such an anticorrelation in our $z \sim 2.3$ sample in Strom18, but our new results for galaxies with bimodal posteriors indicate that the galaxies with higher U tend to have lower O/H and vice versa, although it is not clear whether this is consistent with the $z \sim 0$ U -O/H relation underpinning all locally calibrated diagnostics. Even if the U -O/H relation is the same at $z \sim 2$ as at $z \sim 0$, the strong correlation between O3N2 and U in most calibration samples means that the scatter in *any* O3N2 O-MZR is not accurately tracing the scatter in the underlying O/H scaling relation. The additional dependence on U will artificially suppress the apparent scatter and thus the inferred intrinsic scatter as well (c.f. Section 7.2 of Steidel et al. 2014).

The trends with ionization parameter could be the result of underlying differences in the ionizing photon flux density, in the H II region gas density and pressure, or in both. However, we find no significant correlation between M_* and n_e in our sample (Spearman $\rho = -0.06$, $p = 0.43$), suggesting that differences in n_e may be more important. The role of ionization parameter, its relationship to other quantities, such as SFR, and how it should be physically interpreted in high- z galaxies will be addressed in future work. Here, we emphasize that methods like GalDNA that explicitly account for other key parameters—including U , but also multiple abundances—offer a promising

opportunity to investigate scaling relations in greater detail and build more astrophysical intuition from the same observations.

4.3. Measuring Shallow Relations with Large Scatter

Above, we have highlighted scenarios that might result in a shallow O-MZR with relatively large scatter. However, it is much more challenging to confidently characterize such relations than it is to characterize steeper or tighter correlations.

To understand our results in the context of other O-MZRs measured for similar samples, we simulated the correlation between O/H and M_* for a sample of 195 galaxies with the same distribution of M_* , as shown by the blue histogram in Figure 1. We consider a range of slopes $\beta_{\text{O/H}} = 0.05$ – 0.16 and intrinsic scatter $\sigma_{\text{O/H}} = 0.10$ – 0.23 dex. The Spearman rank correlation coefficient and the probability of the null result are then calculated for $n = 500$ realizations for each $\beta_{\text{O/H}}$ and $\sigma_{\text{O/H}}$.

The top panel of Figure 8 shows the median correlation significance ($1 - p_{\text{Spearman}}$) for the O-MZR as a function of slope $\beta_{\text{O/H}}$ and intrinsic scatter $\sigma_{\text{O/H}}$, with the darker shading indicating a more significant correlation. The dashed and dotted white lines denote the 95%, 99%, and 99.9% confidence isocontours. The location of the GalDNA O-MZR in this parameter space is marked by the white star (with $p = 0.997$). This experiment demonstrates why the strong-line-determined O-MZRs from the literature and the O3N2 O-MZR in Figure 7 appear more obvious visually and tend to be more statistically significant. Such relations are generally steeper and have smaller scatter, which would position them to the upper left in comparison to the corresponding confidence isocontours. As we have noted, however, it is not obvious that the observed scatter in the O3N2 O-MZR reflects the true intrinsic scatter.

The two most straightforward ways of increasing the significance of an observed correlation with given intrinsic β and σ are to increase the sample size or to decrease the

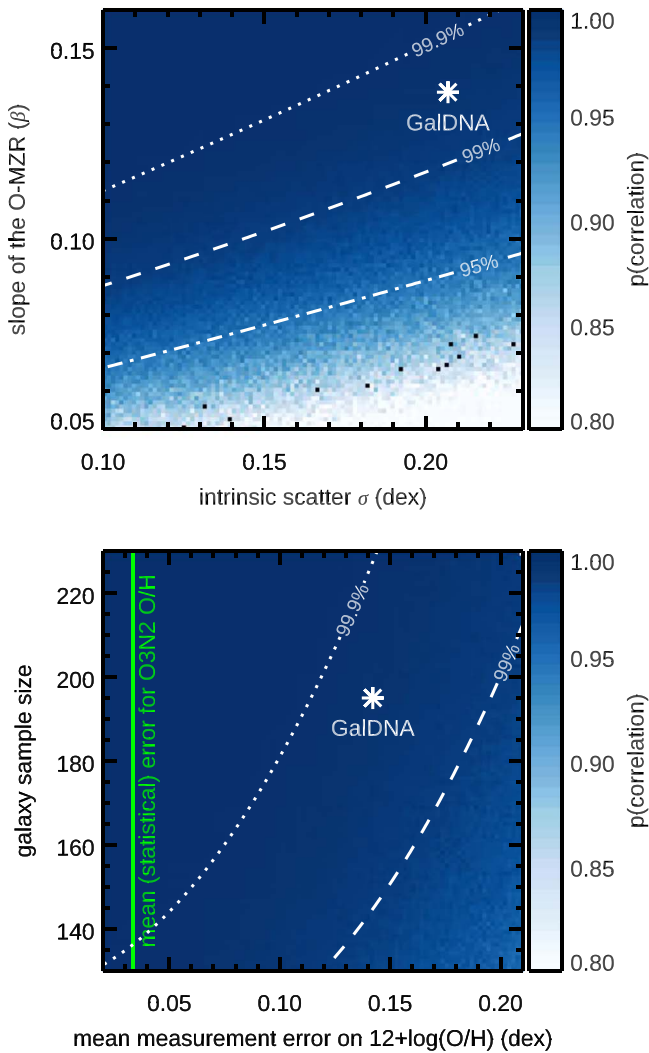


Figure 8. The significance of a monotonic correlation between O/H and M_* , assessed using a Spearman rank correlation test, as a function of the intrinsic scatter and slope of the underlying O-MZR (top), and as a function of the mean measurement error and galaxy sample size (bottom). The darker colors indicate a more significant correlation, and the 95%, 99%, and 99.9% confidence isocontours are denoted by the dashed and dotted white lines. As seen in the top panel, the intrinsic slope of the O-MZR has a stronger effect on the significance of the observed relation than differences in intrinsic scatter (for a fixed sample size and mean measurement error). Assuming the intrinsic O-MZR has the comparatively shallow slope and moderate intrinsic scatter measured from Figure 5, improving the precision of the abundance measurements could dramatically increase the significance of the observed scaling relation. The mean 1σ error on $12 + \log(\text{O}/\text{H})_{\text{O3N2}}$ (the vertical green line in the bottom panel) is small enough to ensure a highly significant observed correlation even for much smaller samples—but this does not include systematic uncertainties stemming from the scatter in the original calibration, and is likely to be an underestimate of the true uncertainty on O3N2-based O/H abundances.

statistical measurement uncertainties. The bottom panel shows the results for a similar set of simulated O-MZRs, but now varying the galaxy sample size (where the stellar masses are drawn from a normal distribution with the same mean and variance as the paper sample) and the mean measurement error on the abundance. For reference, the mean measurement error on $\log(\text{O}/\text{H})_{\text{O3N2}}$ is shown by the vertical green line, although this does not capture any systematic uncertainties. A modest increase in sample size (to ~ 230 galaxies) or decrease in measurement error (to ~ 0.11 dex) would increase our confidence in a positive correlation to 99.9%+.

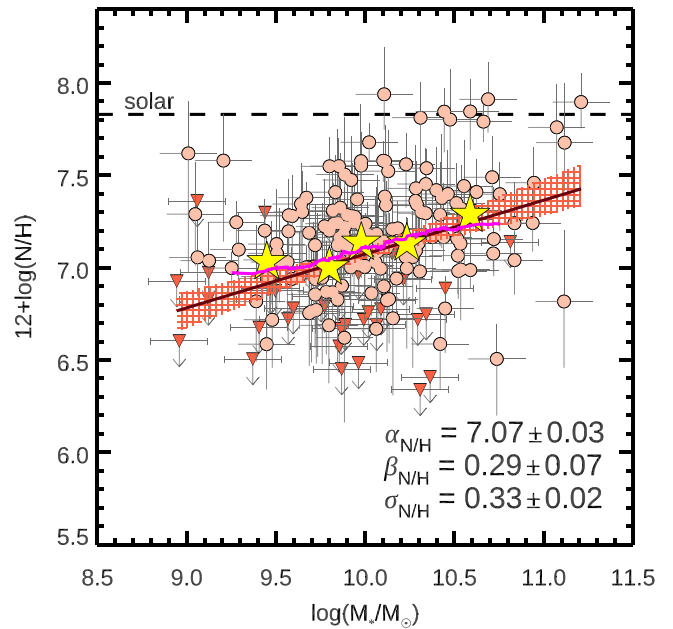


Figure 9. The correlation between stellar mass M_* and gas-phase N/H for our $z \sim 2$ galaxy sample, shown in the same manner as Figure 5. Upper limits on N/H are shown as darker red triangles. The magenta line follows the center moving average, and the solid red line indicates the best-fit linear relation (with the uncertainties illustrated by the hatched region). The medians in bins of M_* are shown by the yellow stars. The observed correlation is stronger than that for the O-MZR, with a rank correlation coefficient of $\rho = 0.28$ ($p = 8 \times 10^{-5}$) and a slope of $\beta_{\text{N/H}} = 0.29 \pm 0.07$ for the linear relation. The intrinsic scatter in this relation is also larger than that measured for the O-MZR ($\sigma_{\text{N/H}} = 0.33 \pm 0.02$ dex).

In coming years, we will continue to see progress on both fronts: samples of high-quality rest-optical spectra of $z \gtrsim 2$ galaxies are increasing in size, and advances in models of massive star evolution and independent constraints on Fe/H in high- z galaxies will both lead to narrower posterior distributions for O/H. In the meantime, it is important to continue to invest resources in developing methods of abundance inference that can accurately characterize both the slope and the scatter in the O-MZR. It is also important to study galaxies' abundance patterns, rather than a single proxy for their bulk metallicity. GalDNA is an example of this kind of method, and in the next section, we present the corresponding scaling relations for nitrogen (N-MZR) and iron (Fe-MZR) that result from our analysis.

5. Relations for Other Elements

Here we present the N-MZR and Fe-MZR for the same galaxies that define the O-MZR in Figure 5. These results comprise the largest study to date of multiple elemental abundances in individual high- z galaxies.

5.1. The N-MZR

The correlation between M_* and N/H for our sample of 195 $z \sim 2$ galaxies is shown in Figure 9. As for the O-MZR in Figure 5, individual galaxies are represented by the colored points, with the darker triangles indicating upper limits on N/H (c.f. Section 3.2.2). The center moving average (in magenta) provides a nonparametric view of the N-MZR, and the yellow stars represent the median $12 + \log(\text{N}/\text{H})$ in bins of M_* . The Spearman rank correlation coefficient for the N-MZR is

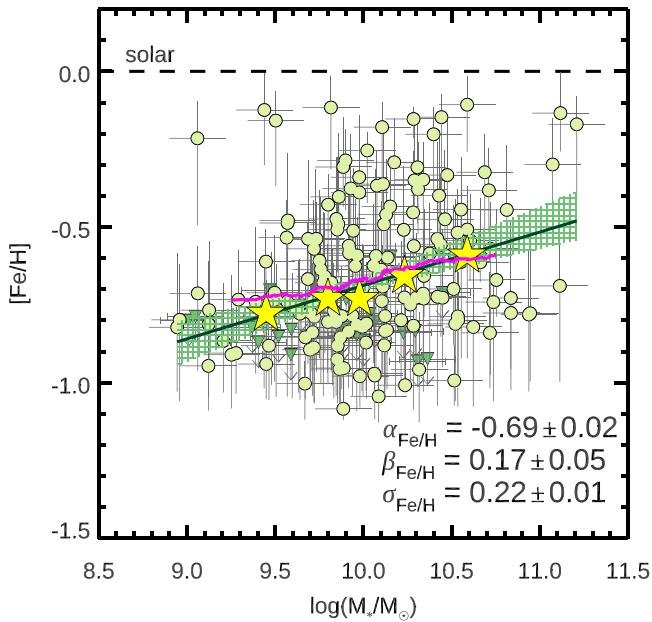


Figure 10. The correlation between stellar mass M_* and Fe/H for our $z \sim 2$ galaxy sample. As in Figures 5 and 9, the magenta line follows the center moving average, the solid green line indicates the best-fit linear relation (with the uncertainties illustrated by the hatched region), and the medians in bins of M_* are shown by the yellow stars. The rank correlation coefficient is $\rho = 0.27$ ($p = 1 \times 10^{-4}$), and the linear relation has $\beta_{\text{Fe/H}} = 0.17 \pm 0.05$ and $\sigma_{\text{Fe/H}} = 0.22 \pm 0.01$ dex. Unlike the O-MZR and N-MZR, the Fe-MZR remains significantly subsolar even at high M_* .

$\rho_{\text{N/H}} = 0.28$ ($p = 8 \times 10^{-5}$), indicating a significant moderate correlation. The correspondence with M_* is stronger for N/H than for O/H, although it also remains subsolar across the entire sampled mass range, with $12 + \log(\text{N}/\text{H}) = 7.07 \pm 0.03$ at $10^{10} M_\odot$, or $\approx 17\%$ (N/H) $_\odot$.

The best-fit linear relation determined using `LINMIX_ERR` is shown by the dark red line, with the hatched region illustrating the 1σ uncertainties on the slope and normalization; the fit parameters can be found in Table 1. The N-MZR is significantly steeper than the O-MZR, with $\beta_{\text{N/H}} = 0.29 \pm 0.07$, and has significantly larger intrinsic scatter, with $\sigma_{\text{N/H}} = 0.33 \pm 0.02$. The mean measurement error for $12 + \log(\text{N}/\text{H})$ is also larger, with $s_{\text{N/H}} = 0.23$, in large part due to the fact that N/H is determined almost entirely by observations of a single feature [N II] $\lambda\lambda 6549, 6585$ that tracks the number of nitrogen atoms in the gas. This is in contrast to O/H, which dominates the cooling in H II regions and thus influences all collisionally-excited transitions. As a result, O/H can be inferred from an ensemble of lines. The larger scatter in the N-MZR may have a physical origin as well, as nitrogen is thought to be produced by intermediate-mass stars and returned to the ISM on somewhat longer timescales than oxygen (Pettini et al. 2002).

5.2. The Fe-MZR

Figure 10 shows the correlation between M_* and inferred [Fe/H] for our $z \sim 2$ galaxy sample in the same manner as the O-MZR in Figure 5 and the N-MZR in Figure 9. Rather than present the iron abundances at face value, however, we use bracket notation, which indicates the enrichment relative to solar and is commonly used in studies of stellar enrichment. The Spearman rank correlation coefficient for the Fe-MZR is $\rho_{\text{Fe/H}} = 0.27$ ($p = 1 \times 10^{-4}$), which indicates a stronger and

more significant correlation than the O-MZR for the same sample.

The Fe-MZR also has a comparatively low normalization, with $[\text{Fe}/\text{H}] = -0.69$ ($\approx 20\%$ solar) at $M_* = 10^{10} M_\odot$. This is consistent with the stellar metallicity relation (read: Fe-MZR) reported by Cullen et al. (2019) for a sample of $2.5 \leq z \leq 5$ galaxies, which had $[\text{Fe}/\text{H}] \approx -0.7$ based on measurements from rest-UV spectra. This relatively low degree of enrichment is expected, given that the typical ages of star-forming galaxies at $z \sim 2$ ($\lesssim 300$ Myr) are lower than the characteristic timescale for substantial iron enrichment due to Type Ia SNe (~ 1 Gyr).

As with the O-MZR, the Fe-MZR in Figure 10 is shallower than reported by other recent studies at $z \gtrsim 2$ (e.g., Cullen et al. 2019, 2021), which found $\beta_{\text{Fe/H}} \sim 0.3$ using galaxies’ rest-UV spectra to estimate stellar metallicity. Theios et al. (2019) have also shown a somewhat steeper increase in stellar metallicity as a function of M_* ($\beta_{\text{Fe/H}} \approx 0.29$) for KBSS galaxies, based on stacks of rest-UV spectra in bins of stellar mass (see Figure 16 in that paper). Using rest-UV photospheric spectra is a more direct method of inferring stellar metallicity than `GalDNA`, and we have shown in Steidel et al. (2016) that stellar metallicity inferred this way is relatively insensitive to the choice of stellar model. However, Cullen et al. (2021) found that stellar metallicities inferred using BPASS were systematically lower than the metallicities inferred using Starburst99 models (Leitherer et al. 2010). There is also substantial diversity in the predicted ionizing spectra of the same massive stars, which is what `GalDNA` relies on to indirectly estimate Fe/H. Both methods are ultimately model-dependent—albeit in different ways—and it is currently difficult to directly compare them without larger representative samples of individual galaxies where stellar metallicity can be measured using rest-UV spectra.

6. Astrophysical Insight from Abundance Patterns

The principal advantage of using photoionization model-based methods like `GalDNA` is the ability to explicitly account for additional factors that influence the physics of star-forming regions. Photoionization models are also the best way to constrain abundance ratios for the same galaxies using common observables. Routinely acquiring sufficiently deep rest-UV spectra in order to measure Fe/H for individual high- z galaxies remains prohibitively expensive, although this will hopefully change in the future with the construction of “extremely large telescopes.”

In this section, we revisit aspects of the discussion in Sections 4 and 5, but now consider how the normalization, slope, and scatter of the separate scaling relations can be used together to build physical intuition regarding the chemical evolution of high- z galaxies. Figure 11 shows the best-fit O-MZR (blue; the dotted white line), N-MZR (red; the dashed white line), and Fe-MZR (green; the dotted-dashed white line) described above on the same log scale relative to solar in order to facilitate comparison. The shading around the best-fit linear relations indicates the intrinsic scatter in each case.

6.1. Alpha-enhancement in High- z Galaxies

The offset between the O-MZR and Fe-MZR yields $[\text{O}/\text{Fe}] = (\alpha_{\text{O/H}} - 8.69) - \alpha_{\text{Fe/H}} = 0.35 \pm 0.03$ at $M_* = 10^{10} M_\odot$, corresponding to $\text{O}/\text{Fe} \approx 2.2 \times (\text{O}/\text{Fe})_\odot$. This is lower than the median ratio reported in Strom18, $[\text{O}/\text{Fe}] = 0.42$ or

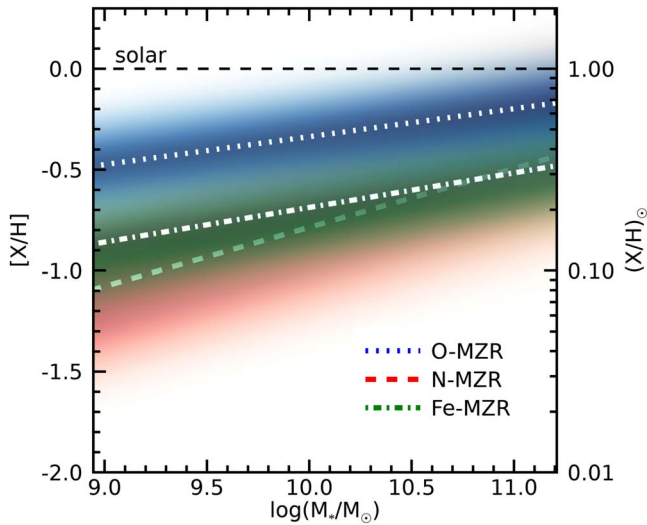


Figure 11. A comparison of the O-MZR (blue), N-MZR (red), and Fe-MZR (green) for our $z \sim 2$ galaxy sample, plotted on the same log scale relative to solar. The shading illustrates the intrinsic scatter measured for each relation. The O-MZR and Fe-MZR have very similar slopes and are offset from one another by 0.35 ± 0.03 dex at $M_* = 10^{10} M_\odot$, which translates to $O/Fe \approx 2.2 \times (O/Fe)_\odot$. This stays relatively constant across the sampled mass range, implying that young stars even in massive galaxies at $z \sim 2$ are significantly α -enhanced. The N-MZR rises much more quickly than the other relations, starting near the primary plateau in N/O at $M_* \sim 10^9 M_\odot$ and surpassing the Fe-MZR near $M_* \sim 10^{11} M_\odot$.

$O/Fe \approx 2.6 \times (O/Fe)_\odot$, but still consistent given the typical measurement uncertainties (~ 0.2 dex). Figure 12 shows the updated distribution of $[O/Fe]$ for the 63 individual galaxies in the current sample with bounded posterior PDFs for O/Fe . These galaxies have a median $[O/Fe] = 0.39$ or $O/Fe \approx 2.5 \times (O/Fe)_\odot$ and an interquartile range of $[O/Fe] = 0.28$ – 0.47 or $O/Fe \approx 1.9$ – $3.0 \times (O/Fe)_\odot$. Both estimates of the typical O/Fe in our $z \sim 2$ – 3 galaxy sample are in line with the α -enhancement reported by Cullen et al. (2019) and Sanders et al. (2020a). These works focus on different spectroscopic features from one another and from our analysis, and the agreement between all of these methods should be seen as strong confirmation of elevated α/Fe in high- z galaxies.

Although the slope of the Fe-MZR is slightly higher than the slope of the O-MZR, they agree within errors, indicating a relatively constant level of α -enhancement across the entire sampled mass range. Likewise, there is no statistically significant anticorrelation between $[O/Fe]$ and M_* for the subset of galaxies in Figure 12 (Spearman $\rho = -0.29$, $p = 0.02$). Perhaps the most direct comparison of this finding to other samples is with the recent results for $z \sim 3$ galaxies from Cullen et al. (2021), who compared their rest-UV-measured Fe/H with the strong-line-based O-MZR from Sanders et al. (2021). As we do, the authors found virtually identical—albeit much steeper—slopes for their gas-phase metallicity relation (O-MZR) and stellar metallicity relation (Fe-MZR).

The constancy of supersolar O/Fe with M_* appears robust to the choice of method, and is intriguing, as it may suggest that the modes of production for O and Fe remain the same for galaxies spanning the observed range in M_* . The implication then is that the Fe in even relatively massive galaxies has been produced—along with the O—by CCSNe, rather than by Type Ia SNe. Significant Fe production by Type Ia SNe would tend to decrease O/Fe in older galaxies, and we would expect O/Fe

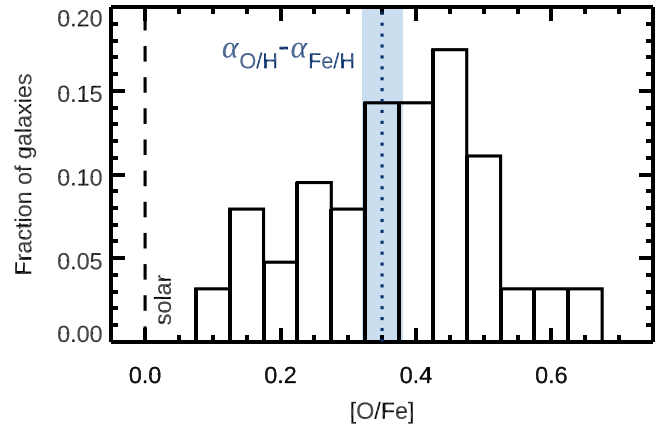


Figure 12. The distribution of $[O/Fe]$ for the 63 individual galaxies with bounded posteriors in the abundance ratio. For these galaxies, the median $[O/Fe] = 0.39$, corresponding to $O/Fe \approx 2.5 \times (O/Fe)_\odot$. This is slightly higher but still consistent with the $[O/Fe]$ inferred from the offset between the O-MZR and Fe-MZR, which is shown by the dotted blue line, with the shaded region representing the statistical uncertainty in this value.

to decrease with increasing M_* (i.e., for the Fe-MZR to be steeper than the O-MZR).

This is contrary to our understanding of α -enhancement in nearby galaxies, which is highest in massive galaxies (e.g., Faber et al. 1992; Thomas et al. 2010; Conroy et al. 2014; Segers et al. 2016). However, our analysis pertains to the abundances in and around young massive stars, and captures the current level of enrichment in galaxies, whereas the α -enhanced stellar populations in, e.g., elliptical galaxies are composed of lower-mass stars that likely formed at earlier times, some at similar redshifts as our galaxy sample. Significant α -enhancement in $z \sim 2$ galaxies can occur if they are relatively young ($\lesssim 300$ Myr) or have rising SFHs. Considering the observed trend in cosmic SFR density at these redshifts, it is possible that this is the case, and that older galaxies and galaxies with declining SFHs—where we might expect lower O/Fe —are comparatively uncommon. However, it may also suggest that these galaxies are underrepresented in existing surveys of emission line galaxies.

Setting aside the differences in the slopes reported for the O-MZR and Fe-MZR in different studies, the slope of the Fe-MZR must effectively function as an upper limit on the slope of the O-MZR. The Fe-MZR may be steeper—if Type Ia SNe have contributed to the iron in the ISM of older, more massive galaxies—but generally not shallower than the O-MZR. This is because the limiting case is an enrichment pattern resulting from pure CCSNe ejecta at all stellar masses, which would tie together the slopes of the two relations. Curiously, the typical α -enhancement in our sample ($[O/Fe] = 0.35 \pm 0.03$) is less than half what is expected from pure Fe-poor CCSNe enrichment ($[O/Fe] \approx 0.73$; Nomoto et al. 2006). Some of this discrepancy could result from our definitions of solar enrichment for O and Fe, but it may also point to other sources of enrichment, for example “prompt” Type Ia SNe with characteristic delay times $\lesssim 400$ Myr (Scannapieco & Bildsten 2005; Matteucci et al. 2009; Yates et al. 2013).

Nitrogen production is clearly not tied to oxygen and iron production in the same way, as the N-MZR has a significantly steeper slope than both the O-MZR and Fe-MZR (Figure 11). At the low-mass end of our sample ($M_* = 10^9 M_\odot$), comparing the O-MZR and N-MZR yields $\log(N/O) \sim -1.45$, close to

the primary plateau. Based on the slope and normalization of the two relations, $\log(\text{N}/\text{O})_{\odot} = -0.86$ would occur at $M_{*} = 10^{12.8} M_{\odot}$, a much higher mass than one might expect to need to reach solar abundances. This may indicate that the *GalDNA* N-MZR relation in Figure 9 is actually too shallow.

Still, a lower N/H coupled with a more rapid increase with M_{*} relative to O/H is a robust result that persists regardless of how the final sample is constructed, including the treatment of the upper limits on N/H and N/O (c.f. Section 3.2.2). These relative abundances suggest that most high- z galaxies with $M_{*} \gtrsim 10^9 M_{\odot}$ are engaging in secondary production of nitrogen. In high- z damped Ly α systems, where nitrogen and oxygen abundances have also been studied in some detail, N/O is near or below the primary plateau at $\log(\text{N}/\text{O}) \sim -1.5$, but these systems are generally very metal-poor and have lower O/H than the galaxies in our sample (e.g., Pettini et al. 1995, 2002, 2008; Petitjean et al. 2008). It seems likely that pushing $z \sim 2$ galaxy surveys to lower masses would allow us to begin directly studying galaxies producing primary nitrogen.

6.2. Timescales of Galaxy Assembly

The behavior of the intrinsic scatter of the MZR for different elements can provide meaningful benchmarks for simulations of galaxy formation and evolution. For example, Matthee & Schaye (2018) argued that the scatter in $z=0.1$ chemical scaling relations should be expected to differ depending on the element, with the O-MZR having the least intrinsic scatter and the N-MZR and Fe-MZR having significantly larger scatter. Their interpretation of the physical origin of these differences relates to the different timescales on which elements are formed and returned to the ISM as a result of their progenitors. For α -elements like O, which is produced by CCSNe on relatively short timescales, deviations from equilibrium are short-lived and should therefore result in less observed scatter.

Figure 13 shows the posterior distributions of σ_{int} for the O-MZR (blue; the dotted line), Fe-MZR (green; the dotted-dashed line), and N-MZR (red; the dashed line). As predicted, the O-MZR exhibits the smallest intrinsic scatter, $\sigma_{\text{O/H}} = 0.21 \pm 0.01$. For our sample, however, the intrinsic scatter of the Fe-MZR is $\sigma_{\text{Fe/H}} = 0.22 \pm 0.01$ dex, statistically identical to that of the O-MZR and significantly smaller than that of the N-MZR. The relative scatter suggests that most or all of the iron enrichment in our galaxy sample occurred on a comparatively shorter timescale, like oxygen (i.e., in CCSNe). Such an interpretation is consistent with the supersolar O/Fe observed for galaxies at all M_{*} , as discussed previously.

The large intrinsic scatter of the N-MZR $\sigma_{\text{N/H}} = 0.33 \pm 0.02$ could suggest that nitrogen enrichment occurs on longer timescales than oxygen and iron. However, the observed $\sigma_{\text{N/H}}$ significantly exceeds the predictions of Matthee & Schaye (2018), and although the predictions are for scaling relations at $z=0.1$, it is not obvious why one would observe much higher scatter at $z \sim 2$. If this increase is real, it could imply differences in the production of nitrogen in lower-metallicity galaxies, such as those at high redshift. Unfortunately, there are not significantly better observational constraints that can be placed on N/H, as [N II] $\lambda\lambda 6549, 6585$ is the only nitrogen transition commonly observed in high- z spectra. In the coming years, it may be possible to measure N/H directly in high- z galaxies using the observations of the auroral [N II] $\lambda 5755$ line with JWST; however, such samples will be

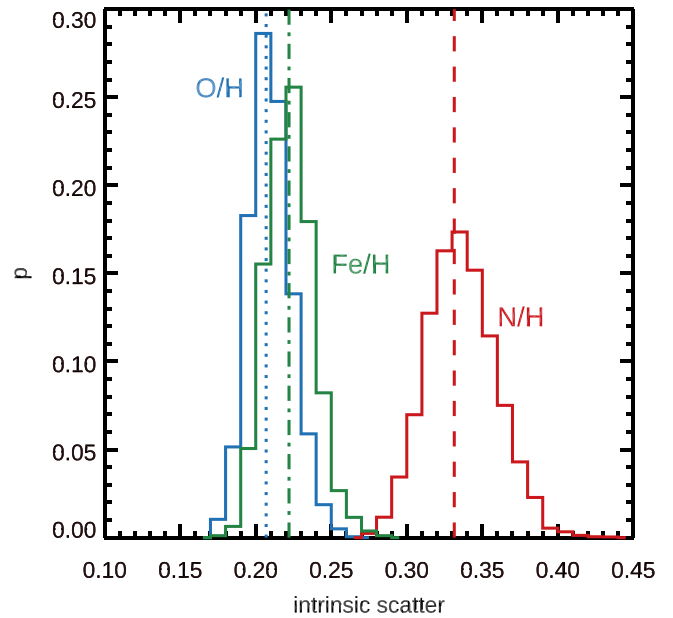


Figure 13. The intrinsic scatter in the correlation between M_{*} and abundance varies between elements. The histograms show the posterior distributions of σ_{int} from fitting the relations plotted in Figures 5, 9, and 10. The intrinsic scatter in the O-MZR is the smallest ($\sigma_{\text{O/H}} = 0.21 \pm 0.01$), followed by the Fe-MZR ($\sigma_{\text{Fe/H}} = 0.22 \pm 0.01$), and the N-MZR ($\sigma_{\text{N/H}} = 0.33 \pm 0.02$). The comparatively low scatter in the Fe-MZR relative to the N-MZR, and its relative similarity to the O-MZR, may be evidence that most or all of the iron in our galaxy sample was created by CCSNe.

small, owing to the extreme faintness of the auroral line relative to the nebular line.

6.3. Constraints on Gas Flows

Many studies use the effective yield $y_{\text{eff}} \equiv Z / \ln(f_{\text{gas}}^{-1})$ to understand the relationship between inflows and outflows and the shape of the O-MZR (e.g., Edmunds 1990; Garnett 2002; Dalcanton 2007). In a closed-box model, $y_{\text{eff}} = y$, the true yield, which is assumed to be constant; differences in y_{eff} can be attributed to gas flows in and out of the galaxy. In $z \sim 0$ galaxies, y_{eff} is seen to increase significantly with increasing baryonic mass (Tremonti et al. 2004), which is interpreted as preferential metal loss from lower-mass galaxies due to their shallower potential wells. In contrast, high- z galaxies show the opposite trend (Erb et al. 2006a; Mannucci et al. 2009; Troncoso et al. 2014), with y_{eff} decreasing slightly with increasing mass. The implications of this result are that the efficiency of outflows does not decline with M_{*} in $z \gtrsim 2$ galaxies and that high- M_{*} galaxies at these redshifts may even be more efficient at ejecting metals than their lower- M_{*} counterparts (Torrey et al. 2019). Indeed, having mass outflow rates scale proportionally with SFR naturally produces the decline in y_{eff} seen at high masses. Similarly, recent work by Lian et al. (2018) suggests that the parallel slopes of the O-MZR and Fe-MZR at high- z and the relatively flatter O-MZR at $z \sim 0$ are a consequence of larger metal loading factors at early times. The ability to compare the O-MZR and Fe-MZR at multiple redshifts is increasingly used in assessing the success of cosmological simulations (Ma et al. 2016; De Rossi et al. 2017).

Although we have not measured f_{gas} and y_{eff} for our sample, the shallower slope of the *GalDNA* O-MZR should translate to a steeper decline in y_{eff} with M_{*} than has been seen in previous

high- z samples. This assumes the correlation between f_{gas} and M_* for our galaxies is similar to that reported by Erb et al. (2006a), which is a reasonable assumption, given that they share the same parent sample. A steep decline in y_{eff} requires massive galaxies to be even more efficient at ejecting metals from their ISM and implies very high mass-loading factors for outflows. This finding is more easily explained by a combination of enriched outflows *and* unenriched inflows that scale with SFR (Erb 2008), which also produces a more marked decline in y_{eff} with M_* —and thus a shallower O-MZR slope.

Although disentangling the relative importance of inflows and outflows remains challenging, the CGM provides a “fossil record” of past outflow activity in the form of the heavy elements deposited there by galactic winds. The CGM contains a significant fraction of all of the metals created in stars, and the distribution and physical state of this enriched gas can provide constraints on energetic feedback in galaxies (Tumlinson et al. 2011; Peebles et al. 2014; Oppenheimer et al. 2016; Rudie et al. 2019). Metal-bearing gaseous structures in the IGM and CGM at $z \sim 2$ –3 appear to be very small ($\lesssim 100$ pc; Rauch et al. 2001; Rudie et al. 2019), and there are significant variations in the metal content and abundance pattern of the CGM of single galaxies (Zahedy et al. 2019, 2021; Cooper et al. 2021), suggesting that metals are often poorly mixed in the gas around galaxies. Future studies comparing the composition of the CGM with the other components of galaxies, including the ISM and stars, will offer critical insight into the baryonic flows that profoundly shape the evolution of galaxies.

7. Summary

We have inferred the chemical abundances of oxygen, nitrogen, and iron in 195 individual star-forming galaxies at $z \sim 2$ –3. Self-consistent measurements of O/H, N/H, and Fe/H were made using GalDNA (Section 3), which uses BPASS stellar population synthesis models together with Cloudy photoionization models to predict the rest-optical nebular spectra of galaxies. These results comprise the largest sample of high- z galaxies with multiple elemental abundances to date and can be used to address questions relating to the chemical evolution of galaxies in the early universe. Our key findings are that:

1. The photoionized gas and massive stars in $z \sim 2$ galaxies have subsolar abundances of O, N, and Fe, which increase with increasing M_* . The fiducial abundances at $M_* = 10^{10} M_\odot$ are

$$12 + \log(\text{O}/\text{H}) = 8.35 \pm 0.02 \text{ or } 45\% (\text{O}/\text{H})_\odot$$

$$12 + \log(\text{N}/\text{H}) = 7.07 \pm 0.03 \text{ or } 17\% (\text{N}/\text{H})_\odot$$

$$[\text{Fe}/\text{H}] = -0.69 \pm 0.02 \text{ or } 20\% (\text{Fe}/\text{H})_\odot.$$
2. The O-MZR based on the GalDNA method is relatively shallow ($\beta_{\text{O}/\text{H}} = 0.14 \pm 0.05$), with moderate intrinsic scatter ($\sigma_{\text{O}/\text{H}} = 0.21 \pm 0.01$ dex; Figure 5). Although qualitatively similar to other O-MZRs in the literature, it is shallower and has larger scatter than relations based on strong-line O/H.
3. The slope of the O3N2 O-MZR (and other strong-line-based relations) depends on the coefficient(s) in the calibration (Figure 6), which in turn depend on the underlying correlations between abundances and other physical conditions, like the ionization parameter, in the

calibration sample. In other words, the correlation between U and O/H in the calibration sample affects the slope of the O-MZR for any other sample for which that calibration is used. This can introduce systematic biases in physical interpretations of the form of such scaling relations.

4. The larger intrinsic scatter observed in the GalDNA O-MZR is the result of explicitly accounting for the ionization conditions in galaxies, which are strongly correlated with many line ratios and with M_* , but not as strongly correlated with O/H. Conversely, the strong correlation between U and many commonly used line ratios for inferring O/H can artificially suppress the scatter in strong-line-based O-MZRs (Figure 7).
5. The O-MZR and Fe-MZR (Figure 10) have nearly identical slopes across two decades in M_* , indicating a high level of α -enhancement equivalent to $\text{O}/\text{Fe} \approx 2.2 \times (\text{O}/\text{Fe})_\odot$ even at relatively large M_* (Figure 11).
6. The intrinsic scatter in the O-MZR and Fe-MZR is very similar (Figure 13), which would be the case if the majority of the Fe had been produced along with the O by CCSNe, rather than by Type Ia SNe. The N-MZR has significantly larger scatter, potentially reflecting a longer timescale for nitrogen production.

Our results demonstrate the potential for measuring multiple elemental abundances in $z \sim 2$ –3 galaxies using the emission lines that are commonly observed in their rest-optical spectra. Knowledge of galaxies’ abundance patterns—rather than just a single measure of their enrichment—should prove valuable for comparing galaxies’ ISM with their CGM in order to understand the gas flows in and out of galaxies. Moreover, abundance patterns can be used to link galaxy populations observed at different redshifts with one another, by looking for commonalities in, for example, α -enhancement.


Methods like GalDNA, which are capable of accurately characterizing the slope and scatter in abundance scaling relations for multiple elements, can also be adapted for samples at other redshifts where different emission lines may be accessible (e.g., in the near-UV). Continuing to develop such methods will be key to extracting maximum scientific value from both existing galaxy surveys and the observations from future facilities, including JWST, which will provide new opportunities to study some of the earliest galaxies in the universe.

The data presented in this paper were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership between the California Institute of Technology, the University of California, and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W.M. Keck Foundation. This work has also been supported in part by the US National Science Foundation (NSF) through grants AST-0908805 and AST-1313472 (ALS and CCS). The authors wish to recognize and acknowledge the significant cultural role and reverence that the summit of Maunakea has within the indigenous Hawaiian community. We are privileged to have had the opportunity to conduct observations from this mountain.

Facilities: Keck:I (LRIS, MOSFIRE), Magellan:Baade (FourStar), Hale (WIRC), Spitzer (IRAC), HST (WFC3).

Software: GalDNA, LINMIX_ERR (Kelly 2007), BPASSv2 (Eldridge & Stanway 2016; Stanway et al. 2016), Cloudy (Ferland et al. 2013), BC03 (Bruzual & Charlot 2003).

ORCID iDs

Allison L. Strom  <https://orcid.org/0000-0001-6369-1636>
 Gwen C. Rudie  <https://orcid.org/0000-0002-8459-5413>
 Charles C. Steidel  <https://orcid.org/0000-0002-4834-7260>
 Ryan F. Trainor  <https://orcid.org/0000-0002-6967-7322>

References

- Adelberger, K. L., Steidel, C. C., Shapley, A. E., et al. 2004, *ApJ*, 607, 226
- Aller, L. H. 1954, *ApJ*, 120, 401
- Andrews, B. H., & Martini, P. 2013, *ApJ*, 765, 140
- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, 47, 481
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, 93, 5
- Barrera-Ballesteros, J. K., Sánchez, S. F., Heckman, T., Blanc, G. A., & MaNGA Team. 2017, *ApJ*, 844, 80
- Berg, D. A., Pogge, R. W., Skillman, E. D., et al. 2020, *ApJ*, 893, 96
- Berg, D. A., Skillman, E. D., Marble, A. R., et al. 2012, *ApJ*, 754, 98
- Bian, F., Kewley, L. J., & Dopita, M. A. 2018, *ApJ*, 859, 175
- Blanc, G. A., Kewley, L., Vogt, F. P. A., & Dopita, M. A. 2015, *ApJ*, 798, 99
- Bothwell, M. S., Maiolino, R., Kennicutt, R., et al. 2013, *MNRAS*, 433, 1425
- Brooks, A. M., Governato, F., Booth, C. M., et al. 2007, *ApJL*, 655, L17
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Chabrier, G. 2003, *PASP*, 115, 763
- Conroy, C., Graves, G. J., & van Dokkum, P. G. 2014, *ApJ*, 780, 33
- Cooper, T. J., Rudie, G. C., Chen, H.-W., et al. 2021, *MNRAS*, 508, 4359
- Cullen, F., McLure, R. J., Dunlop, J. S., et al. 2019, *MNRAS*, 487, 2038
- Cullen, F., Shapley, A. E., McLure, R. J., et al. 2021, *MNRAS*, 505, 903
- Daddi, E., Cimatti, A., Renzini, A., et al. 2004, *ApJ*, 617, 746
- Dalcanton, J. J. 2007, *ApJ*, 658, 941
- De Rossi, M. E., Bower, R. G., Font, A. S., Schaye, J., & Theuns, T. 2017, *MNRAS*, 472, 3354
- Edmunds, M. G. 1990, *MNRAS*, 246, 678
- Eldridge, J. J., & Stanway, E. R. 2016, *MNRAS*, 462, 3302
- Ellison, S. L., Patton, D. R., Simard, L., & McConnachie, A. W. 2008, *AJ*, 135, 1877
- Erb, D. K. 2008, *ApJ*, 674, 151
- Erb, D. K., Pettini, M., Shapley, A. E., et al. 2010, *ApJ*, 719, 1168
- Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006a, *ApJ*, 644, 813
- Erb, D. K., Steidel, C. C., Shapley, A. E., et al. 2006b, *ApJ*, 646, 107
- Esteban, C., García-Rojas, J., Carigi, L., et al. 2014, *MNRAS*, 443, 624
- Esteban, C., Peimbert, M., García-Rojas, J., et al. 2004, *MNRAS*, 355, 229
- Faber, S. M., Worthey, G., & Gonzales, J. J. 1992, in IAU Symp. 149, The Stellar Populations of Galaxies, ed. B. Barbuy & A. Renzini (Dordrecht: Kluwer Academic), 255
- Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, *RMxAA*, 49, 137
- Franx, M., Labbé, I., Rudnick, G., et al. 2003, *ApJL*, 587, L79
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, 362, 41
- Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, *SSRv*, 123, 485
- Garnett, D. R. 2002, *ApJ*, 581, 1019
- González Delgado, R. M., Cid Fernandes, R., García-Benito, R., et al. 2014, *ApJL*, 791, L16
- Izotov, Y. I., & Thuan, T. X. 1999, *ApJ*, 511, 639
- Jafariyazani, M., Newman, A. B., Mobasher, B., et al. 2020, *ApJL*, 897, L42
- Jeong, M.-S., Shapley, A. E., Sanders, R. L., et al. 2020, *ApJL*, 902, L16
- 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
- Kaasinen, M., Kewley, L., Bian, F., et al. 2018, *MNRAS*, 477, 5568
- Kelly, B. C. 2007, *ApJ*, 665, 1489
- Kewley, L. J., & Ellison, S. L. 2008, *ApJ*, 681, 1183
- Kojima, T., Ouchi, M., Nakajima, K., et al. 2017, *PASJ*, 69, 44
- Kriek, M., Conroy, C., van Dokkum, P. G., et al. 2016, *Natur*, 540, 248
- Kriek, M., Price, S. H., Conroy, C., et al. 2019, *ApJL*, 880, L31
- Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, *ApJS*, 218, 15
- Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, *A&A*, 521, L53
- Leitherer, C., Ortiz Otálvaro, P. A., Bresolin, F., et al. 2010, *ApJS*, 189, 309
- Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, *A&A*, 500, 145
- Lian, J., Thomas, D., & Maraston, C. 2018, *MNRAS*, 481, 4000
- Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016, *MNRAS*, 456, 2140
- Maiolino, R., & Mannucci, F. 2019, *A&ARv*, 27, 3
- Maiolino, R., Nagao, T., Grazian, A., et al. 2008, *A&A*, 488, 463
- Mannucci, F., Cresci, G., Maiolino, R., et al. 2009, *MNRAS*, 398, 1915
- Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, *MNRAS*, 408, 2115
- Marino, R. A., Rosales-Ortega, F. F., Sánchez, S. F., et al. 2013, *A&A*, 559, A114
- Matteucci, F., Spitoni, E., Recchi, S., & Valiante, R. 2009, *A&A*, 501, 531
- Matthee, J., & Schaye, J. 2018, *MNRAS*, 479, L34
- McGaugh, S. S. 1991, *ApJ*, 380, 140
- McLean, I. S., Steidel, C. C., Epps, H. W., et al. 2012, *Proc. SPIE*, 8446, 84460J
- Nomoto, K., Tominaga, N., Umeda, H., Kobayashi, C., & Maeda, K. 2006, *NuPhA*, 777, 424
- Onodera, M., Carollo, C. M., Lilly, S., et al. 2016, *ApJ*, 822, 42
- Oppenheimer, B. D., Crain, R. A., Schaye, J., et al. 2016, *MNRAS*, 460, 2157
- Peeples, M. S., Werk, J. K., Tumlinson, J., et al. 2014, *ApJ*, 786, 54
- Pérez-Montero, E. 2014, *MNRAS*, 441, 2663
- Petitjean, P., Ledoux, C., & Srianand, R. 2008, *A&A*, 480, 349
- Pettini, M., Ellison, S. L., Bergeron, J., & Petitjean, P. 2002, *A&A*, 391, 21
- Pettini, M., Lipman, K., & Hunstead, R. W. 1995, *ApJ*, 451, 100
- Pettini, M., & Pagel, B. E. J. 2004, *MNRAS*, 348, L59
- Pettini, M., Zych, B. J., Steidel, C. C., & Chaffee, F. H. 2008, *MNRAS*, 385, 2011
- Pilyugin, L. S., Grebel, E. K., & Mattsson, L. 2012, *MNRAS*, 424, 2316
- Pilyugin, L. S., & Thuan, T. X. 2005, *ApJ*, 631, 231
- Rauch, M., Sargent, W. L. W., & Barlow, T. A. 2001, *ApJ*, 554, 823
- Reddy, N. A., Erb, D. K., Steidel, C. C., et al. 2005, *ApJ*, 633, 748
- Reddy, N. A., Pettini, M., Steidel, C. C., et al. 2012, *ApJ*, 754, 25
- Rudie, G. C., Steidel, C. C., Pettini, M., et al. 2019, *ApJ*, 885, 61
- Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012, *ApJ*, 750, 67
- Sánchez, S. F., Barrera-Ballesteros, J. K., López-Cobá, C., et al. 2019, *MNRAS*, 484, 3042
- Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A., et al. 2012, *A&A*, 538, A8
- Sanders, R. L., Jones, T., Shapley, A. E., et al. 2020b, *ApJL*, 888, L11
- Sanders, R. L., Shapley, A. E., Jones, T., et al. 2021, *ApJ*, 914, 19
- Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2015, *ApJ*, 799, 138
- Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2016, *ApJ*, 816, 23
- Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2018, *ApJ*, 858, 99
- Sanders, R. L., Shapley, A. E., Reddy, N. A., et al. 2020a, *MNRAS*, 491, 1427
- Sanders, R. L., Shapley, A. E., Zhang, K., & Yan, R. 2017, *ApJ*, 850, 136
- Scannapieco, E., & Bildsten, L. 2005, *ApJL*, 629, L85
- Segers, M. C., Schaye, J., Bower, R. G., et al. 2016, *MNRAS*, 461, L102
- Shapley, A. E., Steidel, C. C., Erb, D. K., et al. 2005, *ApJ*, 626, 698
- Stanway, E. R., Eldridge, J. J., & Becker, G. D. 2016, *MNRAS*, 456, 485
- Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2003, *ApJ*, 592, 728
- Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, *ApJ*, 795, 165
- Steidel, C. C., Shapley, A. E., Pettini, M., et al. 2004, *ApJ*, 604, 534
- Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, *ApJ*, 826, 159
- Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, *ApJ*, 836, 164
- Strom, A. L., Steidel, C. C., Rudie, G. C., Trainor, R. F., & Pettini, M. 2018, *ApJ*, 868, 117
- Theios, R. L., Steidel, C. C., Strom, A. L., et al. 2019, *ApJ*, 871, 128
- Thomas, A. D., Dopita, M. A., Kewley, L. J., et al. 2018, *ApJ*, 856, 89
- Thomas, D., Maraston, C., Schawinski, K., Sarzi, M., & Silk, J. 2010, *MNRAS*, 404, 1775
- Tinsley, B. M. 1979, *ApJ*, 229, 1046
- Topping, M. W., Shapley, A. E., Reddy, N. A., et al. 2020, *MNRAS*, 499, 1652
- Torrey, P., Vogelsberger, M., Marinacci, F., et al. 2019, *MNRAS*, 484, 5587
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, *ApJ*, 613, 898
- Troncoso, P., Maiolino, R., Sommariva, V., et al. 2014, *A&A*, 563, A58
- Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, *ARA&A*, 55, 389
- Tumlinson, J., Thom, C., Werk, J. K., et al. 2011, *Sci*, 334, 948
- Vale Asari, N., Stasińska, G., Morisset, C., & Cid Fernandes, R. 2016, *MNRAS*, 460, 1739
- van Loon, M. L., Mitchell, P. D., & Schaye, J. 2021, *MNRAS*, 504, 4817
- van Zee, L., Salzer, J. J., & Haynes, M. P. 1998, *ApJL*, 497, L1
- Veilleux, S., & Osterbrock, D. E. 1987, *ApJS*, 63, 295
- Yabe, K., Ohta, K., Iwamuro, F., et al. 2014, *MNRAS*, 437, 3647
- Yates, R. M., Henriques, B., Thomas, P. A., et al. 2013, *MNRAS*, 435, 3500
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, 120, 1579
- Yuan, T. T., & Kewley, L. J. 2009, *ApJL*, 699, L161
- Zahedy, F. S., Chen, H.-W., Cooper, T. M., et al. 2021, *MNRAS*, 506, 877
- Zahedy, F. S., Chen, H.-W., Johnson, S. D., et al. 2019, *MNRAS*, 484, 2257
- Zahid, H. J., Kashino, D., Silverman, J. D., et al. 2014, *ApJ*, 792, 75