The Heavy Metal Survey: The Evolution of Stellar Metallicities, Abundance Ratios, and Ages of Massive Quiescent Galaxies since $z \sim 2$

Aliza G. Beverage[1] , Mariska Kriek[2] , Katherine A. Suess[3,10] , Charlie Conroy[4] , Sedona H. Price[5] , Guillermo Barro[6] ,
Rachel Bezanson[5] , Marijn Franx[2] , Brian Lorenz[1] , Yilun Ma[7] , Lamiya A. Mowla[8] , Imad Pasha[9] ,
Pieter van Dokkum[9] , and Daniel R. Weisz[1]

[1] Astronomy Department, University of California, Berkeley, CA 94720, USA; abeverage@berkeley.edu
[2] Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands
[3] Kavli Institute for Particle Astrophysics and Cosmology and Department of Physics, Stanford University, Stanford, CA 94305, USA
[4] Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA 02138, USA
[5] Department of Physics & Astronomy and PITT PACC, University of Pittsburgh, Pittsburgh, PA 15260, USA
[6] University of the Pacific, Stockton, CA 90340, USA
[7] Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA
[8] Astronomy Department, Whitin Observatory, Wellesley College, 106 Central Street, Wellesley, MA 02481, USA
[9] Department of Astronomy, Yale University, New Haven, CT 06511, USA

Abstract

We present the elemental abundances and ages of 19 massive quiescent galaxies at $z \sim 1.4$ and $z \sim 2.1$ from the Keck Heavy Metal Survey. The ultradeep LRIS and MOSFIRE spectra were modeled using a full-spectrum stellar population fitting code with variable abundance patterns. The galaxies have iron abundances between [Fe/H] $= -0.5$ and -0.1 dex, with typical values of -0.2 [-0.3] at $z \sim 1.4$ [$z \sim 2.1$]. We also find a tentative $\log \sigma_v$–[Fe/H] relation at $z \sim 1.4$. The magnesium-to-iron ratios span [Mg/Fe] $= 0.1$–0.6 dex, with typical values of 0.3 [0.5] dex at $z \sim 1.4$ [$z \sim 2.1$]. The ages imply formation redshifts of $z_{\rm form} = 2$–8. Compared to quiescent galaxies at lower redshifts, we find that [Fe/H] was ~ 0.2 dex lower at $z = 1.4$–2.1. We find no evolution in [Mg/Fe] out to $z \sim 1.4$, though the $z \sim 2.1$ galaxies are 0.2 dex enhanced compared to $z = 0$–0.7. A comparison of these results to a chemical evolution model indicates that galaxies at higher redshift form at progressively earlier epochs and over shorter star formation timescales, with the $z \sim 2.1$ galaxies forming the bulk of their stars over 150 Myr at $z_{\rm form} \sim 4$. This evolution cannot be solely attributed to an increased number of quiescent galaxies at later times; several Heavy Metal galaxies have extreme chemical properties not found in massive galaxies at $z \sim 0.0$–0.7. Thus, the chemical properties of individual galaxies must evolve over time. Minor mergers also cannot fully account for this evolution as they cannot increase [Fe/H], particularly in galaxy centers. Consequently, the buildup of massive quiescent galaxies since $z \sim 2.1$ may require further mechanisms, such as major mergers and/or central star formation.

1. Introduction

In the present-day Universe, nearly all massive galaxies have quiescent stellar populations. Archaeological studies have revealed that they formed the bulk of their stars rapidly at $z > 2$ (e.g., Gallazzi et al. 2005; Thomas et al. 2005; McDermid et al. 2015). Thus, these studies imply that quiescent galaxies should already exist when the Universe was only a fraction of its current age.

Quiescent galaxies have indeed been identified and studied out to $z \sim 5$ (e.g., Franx et al. 2003; Cimatti et al. 2004, 2008; Daddi et al. 2005; Kriek et al. 2006; Kriek et al. 2009; van de Sande et al. 2013; Belli et al. 2014; Glazebrook et al. 2017; Schreiber et al. 2018; Valentino et al. 2020; Antwi-Danso et al. 2023; Carnall et al. 2023; Nanayakkara et al. 2024) and are thought to dominate the massive galaxy population out to

$z = 2.5$ (e.g., Muzzin et al. 2013; Tomczak et al. 2014; McLeod et al. 2021). Interestingly, quiescent galaxies at higher redshifts are significantly more compact than the population at $z \sim 0$, with their half-mass radii growing by a factor of 3 since $z \sim 2$ (Daddi et al. 2005; Dokkum et al. 2008; Damjanov et al. 2009; van der Wel et al. 2014; Suess et al. 2019a, 2019b; Mowla et al. 2019; Miller et al. 2023). Distant quiescent galaxies, however, have central mass densities similar to their nearby analogs, and the size difference is primarily attributed to the larger envelopes of the lower-redshift galaxies (e.g., Bezanson et al. 2009; Hopkins et al. 2009; van Dokkum et al. 2010; van de Sande et al. 2013; Barro et al. 2017; Ji & Giavalisco 2022). Furthermore, the color gradients of quiescent galaxies become stronger with time, with the galaxy outskirts becoming progressively bluer than the centers (Suess et al. 2019a, 2019b, 2020; Miller et al. 2023).

One way to understand the structural evolution of massive quiescent galaxies is through the "inside-out" growth scenario, in which compact galaxies still grow via gas-poor minor mergers after they have become quiescent (e.g., Bezanson et al. 2009; Naab et al. 2009; van Dokkum et al. 2010; Oser et al. 2012). An alternative explanation is suggested by the fact that the stellar mass function shows a tenfold increase in the number of massive quiescent galaxies since $z \sim 2$ (McLeod et al. 2021).

---

[10] NHFP Hubble Fellow.

If galaxies that quench at lower redshift have larger sizes and stronger color gradients, the observed evolution of *average* properties could instead be explained by population growth (i.e., progenitor bias; e.g., Khochfar & Silk 2009; van Dokkum et al. 2010; Carollo et al. 2013; Poggianti et al. 2013). The relative importance of minor mergers versus progenitor bias remains a central question in massive galaxy evolution.

The evolution in elemental abundances and stellar population properties over cosmic time provides a unique insight into the assembly and star formation histories of massive quiescent galaxies (e.g., Matteucci 1994; Trager et al. 2000; Choi et al. 2014; Conroy et al. 2014; Peng et al. 2015; Spitoni et al. 2017; Maiolino & Mannucci 2019; Trussler et al. 2020; Beverage et al. 2021). Unfortunately, measuring elemental abundances beyond the low-redshift Universe poses a considerable challenge. At higher redshifts, key absorption features are faint and shifted to near-infrared (NIR) wavelengths. Recently, ultradeep spectroscopic surveys have started pushing these measurements beyond the low-redshift Universe and finding that the chemical properties of massive quiescent galaxies at $z \sim 0.7$ closely resemble today's massive early-type galaxies (Choi et al. 2014; Leethochawalit et al. 2018; Bevacqua et al. 2023; Beverage et al. 2023). These studies show that the most massive galaxies were already in place 6 billion yr ago, while the low-mass quiescent population has continued to grow since $z \sim 0.7$ (see Leethochawalit et al. 2019; Beverage et al. 2023).

While minimal evolution is found out to $z \sim 0.7$, the picture at higher redshift is less clear. The few existing stellar metallicity measurements of massive quiescent galaxies at $z \gtrsim 1.4$ show systematically lower [Fe/H] compared to $z \sim 0$, but there is large scatter depending on the methods used to derive the metallicities (Lonoce et al. 2015; Onodera et al. 2015; Kriek et al. 2016, 2019; Carnall et al. 2022; Saracco et al. 2023; Zhuang et al. 2023). At $z > 2$, the only two existing measurements show strong α-enhancement, with [Mg/Fe] $\sim$ 0.5–0.6 dex, indicative of extremely short star formation timescales of <100 Myr (Kriek et al. 2016; Jafariyazani et al. 2020). To further clarify the picture at $z > 1$ and assess the assembly of the quiescent galaxy population over the past 11 billion yr, larger galaxy samples are needed.

To that end, we have executed the Keck Heavy Metal Survey, an ultradeep spectroscopic survey with LRIS (Oke et al. 1995; Rockosi et al. 2010) and MOSFIRE (McLean et al. 2010, 2012) on the Keck I Telescope (see Kriek et al. 2024). This survey collected high signal-to-noise ratio (S/N) spectra of 21 massive quiescent galaxies at $1.4 \lesssim z \lesssim 2.2$, covering multiple Balmer and metal absorption lines.

The Heavy Metal Survey has more than tripled the number of individual galaxies at $z \gtrsim 1.4$ with measurements of stellar metallicities, ages, and elemental abundance ratios. Preliminary results of five Heavy Metal galaxies were presented in Kriek et al. (2019) and are updated in the current work. The survey design is presented in the companion survey paper (Kriek et al. 2024). In Section 2, we describe our full-spectrum modeling technique. In Section 3, we present the stellar population results. In Section 4, we assess how the metal content of quiescent galaxies evolves between $z \sim 2.2$ and $z \sim 0$ by comparing our results with the LEGA-C survey ($z \sim 0.7$) and the Sloan Digital Sky Survey (SDSS; $z \sim 0$). In Section 5, we consider the implications of our results in the buildup of the massive quiescent galaxy population. In Section 6, we summarize our

results. Throughout this work, we assume a flat ΛCDM cosmology with $\Omega_m = 0.3$ and $H_0 = 70$ km s^{-1} Mpc^{-1}.

2. Sample and Methods

2.1. The Heavy Metal Survey

The targets in this study are selected from the Keck Heavy Metal Survey, an ultradeep rest-frame optical spectroscopic survey of massive quiescent galaxies at $z = 1.3$–2.2. The 21 quiescent targets were identified using the UltraVISTA K-band selected catalog (v4.1) by Muzzin et al. (2013) and were selected to be quiescent using UVJ colors (Williams et al. 2009; Wuyts et al. 2016). The redshift intervals $1.3 < z < 1.5$ and $1.92 < z < 2.28$ were chosen such that key rest-frame NIR metal absorption features, namely, Balmer lines, Hβ, and Mg I, fall within the wavelength windows of the Keck/LRIS and Keck/MOSFIRE spectrometers. Primary targets were further required to have $J < 21.5$ and $H < 21.7$ at $z \sim 1.4$ and $z \sim 2.1$, respectively.

The $z \sim 1.4$ targets were observed for a total of $\sim$17 hr on Keck ($\sim$5 hr with LRIS using the 600/10,000 red grating and $\sim$12 hr with MOSFIRE-J), while the $z \sim 2.1$ targets were observed for $\sim$30 hr ($\sim$13 hr with MOSFIRE-J and $\sim$16 hr with MOSFIRE-H). These integration times provide a sufficient S/N for modeling the faint stellar absorption features of individual massive quiescent galaxies. The typical S/N for the $z \sim 1.4$ sample is 8 Å^{-1} and 15 Å^{-1} in the rest frame for LRIS and MOSFIRE-J, respectively, while the S/N for the $z \sim 2.1$ sample is 4 Å^{-1} and 12 Å^{-1} for MOSFIRE-J and MOSFIRE-H. For further details on the target selection, observing strategy, data reduction, galaxy sample characteristics, and stellar population properties (i.e., stellar masses), we refer to Kriek et al. (2024).

One of the 21 primary targets (ID: 55878) is removed from this study before fitting due to strong line emission, likely originating from active galactic nuclei (AGN; see Y. Ma et al. 2024, in preparation). Another galaxy is also removed (ID: 59449) because there are no clear absorption lines in the spectra.

2.2. Stellar Ages and Elemental Abundances

We derive chemical compositions and stellar ages of the 19 Heavy Metal targets using the full-spectrum absorption line fitter (alf) code (Conroy & van Dokkum 2012; Conroy et al. 2018). alf is designed to fit the individual elemental abundances and stellar ages from the optical-NIR stellar continuum of >1 Gyr stellar populations. alf combines metallicity-dependent MIST isochrones (Choi et al. 2016), the empirical MILES and IRTF spectral libraries (Sanchez-Blazquez et al. 2006; Villaume et al. 2017), and synthetic metallicity- and age-dependent elemental response spectra for 19 elements.

3. Results

In the fitting presented here, we assume a model with a single burst of star formation, approximated as a simple stellar population (SSP), and a fixed Kroupa (2001) initial mass function. We note that alf also allows for a young second SSP component to account for late-time star formation. In the end, the full-spectrum model has 26 free parameters: velocity offset, velocity dispersion, SSP age, isochrone metallicity, the

abundances of 19 individual elements (Fe, O, C, N, Na, Mg, Si, K, Ca, Ti, V, Cr, Mn, Co, Ni, Cu, Sr, Ba, and Eu), and the emission line strengths of H, [O III], and [N II].

The fitting is done in the rest frame, and the instrumental resolution of the data is accounted for by convolving the grid of models to match the resolution of the spectrum. We measure this instrumental resolution for each spectrum by finding the average full width at half-maximum (FWHM) of 20–25 skylines in the variance spectrum. The average instrumental dispersions are $\sigma_{inst} = 62$, 36, and 33 km s^{-1} for the LRIS-Red, MOSFIRE-J, and MOSFIRE-H spectra, respectively. We note that the instrumental dispersion is lower than the template resolution in `alf`. This template resolution can be an issue if the intrinsic velocity dispersion is lower than 100 km s^{-1}. However, as we will find, the intrinsic velocity dispersions of the Heavy Metal galaxies are significantly higher than 100 km s^{-1}. The spectral continuum is also removed from the observations by fitting a high-order polynomial to the ratio of the data to the model (where $n \equiv (\lambda_{rmmax} \lambda_{min})/100$ Å). We test the impact of the polynomial on the results by assuming different orders than the default ($n = 8$ and $n = 6$). We find that the polynomial order does not make a meaningful impact on the results, to well within the 1σ uncertainties. Furthermore, Conroy et al. (2018) test the impact of the normalizing polynomial on the fitting results and find that if fluxing issues are approximately linear on <50 Å scales, the default polynomial assumption is sufficient. Once normalized, the spectrum is fit using a Markov Chain Monte Carlo (MCMC; Foreman-Mackey et al. 2013) with 1024 walkers and a burn-in of 20,000 steps. The priors of the MCMC fit are uniform, and the walkers are initialized around the solar abundance pattern.

After fitting all 19 galaxies, we visually inspect the best-fit models and the corresponding posterior distribution functions (PDFs) and remove any galaxy with poorly constrained parameters. We removed three galaxies because their PDFs do not follow normal distributions and/or they have ill-defined peaks. Two of these galaxies lack spectral coverage of key absorption features (D4000 break), leading to unreliable abundance measurements, and the third galaxy has strong [O II] and hydrogen emission. We also remove two galaxies with best-fit stellar ages <1 Gyr, as the stellar population models are not optimized for these young ages. As a sanity check, we also inspect correlations between all fitted parameters to ensure the relevant parameters are not being driven by other poorly constrained parameters. After removing these five galaxies, we are left with elemental abundances and stellar age measurements for 10 galaxies at $z \sim 1.4$ and four galaxies at $z \sim 2.1$.

Five of the original 19 galaxies in this study were originally presented in Kriek et al. (2019). In this study, we have since reextracted the 1D spectra using an improved optimal weighting procedure. We also assume a single SSP when fitting, whereas Kriek et al. (2019) assume a two-burst model. Nonetheless, the results presented here are consistent to within 1σ of what is found in Kriek et al. (2019).

Example spectra with corresponding best-fit models are shown in Figure 1, while the remaining spectra and fits can be found in Appendix A. Corner plots corresponding to the galaxies in Figure 1 are included in Appendix B. The best-fit parameters are listed in Table 1 for all 19 galaxies. Objects with poorly constrained abundances are noted and are removed from the following analysis. In Appendix D, we conduct a mock recovery test to check for systematic uncertainties in the `alf` fitting as a function of S/N. Systematic uncertainties remain <0.1 dex even at the lowest S/N (8 Å^{-1}). We refer to Kriek et al. (2024) for tabulated details of the observations and spectral energy distribution fitting results.

In addition to elemental abundances and stellar ages, the full-spectrum fitting provides measurements of the stellar velocity dispersions. We find that the velocity dispersions of the Heavy Metal galaxies agree with previous measurements of the dynamical mass–stellar mass relation at similar redshifts (e.g., Belli et al. 2016), and we explore the implications of the Heavy Metal dynamical masses in the companion survey paper (Kriek et al. 2024).

Figure 2 shows the variation in [Fe/H], [Mg/Fe], and formation time (t_{form}) as a function of velocity dispersion σ_v for the Heavy Metal galaxies at $z \sim 1.4$ (blue triangles) and $z \sim 2.1$ (green stars). For comparison, we also show trends of quiescent galaxies at $z \sim 0$ and $z \sim 0.7$ from Beverage et al. (2023) in Figure 2. The $z \sim 0$ trends are derived by refitting the stacked spectra of a sample of 5234 quiescent galaxies in SDSS from Conroy et al. (2014), while the $z \sim 0.7$ trends are based on 135 quiescent galaxies in the LEGA-C survey. Initially, Beverage et al. (2021) reported a -0.2 dex offset in [Fe/H] between LEGA-C and SDSS galaxies based on the spectra from the second LEGA-C data release (DR2). However, with the subsequent release, DR3, which doubled the sample size and improved data reduction (see van der Wel et al. 2021), Beverage et al. (2023) no longer found an offset in [Fe/H] between $z \sim 0.7$ and $z \sim 0$. Figure 2 also includes results from other studies of quiescent galaxies at $z \sim 1.5$ (Carnall et al. 2022; Zhuang et al. 2023) and $z \sim 2.1$ (Kriek et al. 2016; Jafariyazani et al. 2020).[11] We note that several other works have derived stellar metallicities of distant quiescent galaxies by fitting low-resolution grism spectra (e.g., Morishita et al. 2018; Estrada-Carpenter et al. 2019). However, such studies rely on the shape of the stellar continuum to constrain the metallicities, leading to high degeneracies with stellar age; therefore, we do not include these results in our comparison. To ensure unbiased comparisons, we restrict the analysis to studies that utilize `alf`.

In the left panel of Figure 2, we show that the Heavy Metal galaxies have lower [Fe/H] compared to the $z \sim 0$ and $z \sim 0.7$ galaxies with similar velocity dispersions. Both the $z \sim 1.4$ and $z \sim 2.1$ galaxy populations have Fe abundances between -0.5 and -0.1 dex, with typical uncertainties of 0.2 dex. There is a hint that the $z \sim 2.1$ galaxies are slightly offset to lower [Fe/H] compared to $z \sim 1.4$. We fit a σ–[Fe/H] relation to the $z \sim 1.4$ population using a simple linear regression, with confidence intervals determined by perturbing the [Fe/H] of each data point according to their uncertainties and refitting 1000 times. The resulting relation is shown in the left panel of Figure 2 (blue line with shaded confidence intervals). We find a positive slope of $0.52^{+0.38}_{-0.23}$, which is consistent with the $z \sim 0$ and $z \sim 0.7$ relations. However, there is a clear offset to lower [Fe/H]. We do not fit a relation to the $z \sim 2.1$ population because of the small sample size.

The observed evolution in [Fe/H] between $z = 1.4$–2.1 and $z < 1$ is corroborated by studies of other quiescent galaxies at similar redshifts. At $z \sim 1.4$, galaxies in the VANDELS

---

[11] Carnall et al. (2022) do not measure a velocity dispersion. We infer one using the average stellar mass of galaxies in the stacked spectrum and an empirically calibrated σ_v–M_* relation from Zahid et al. (2016).
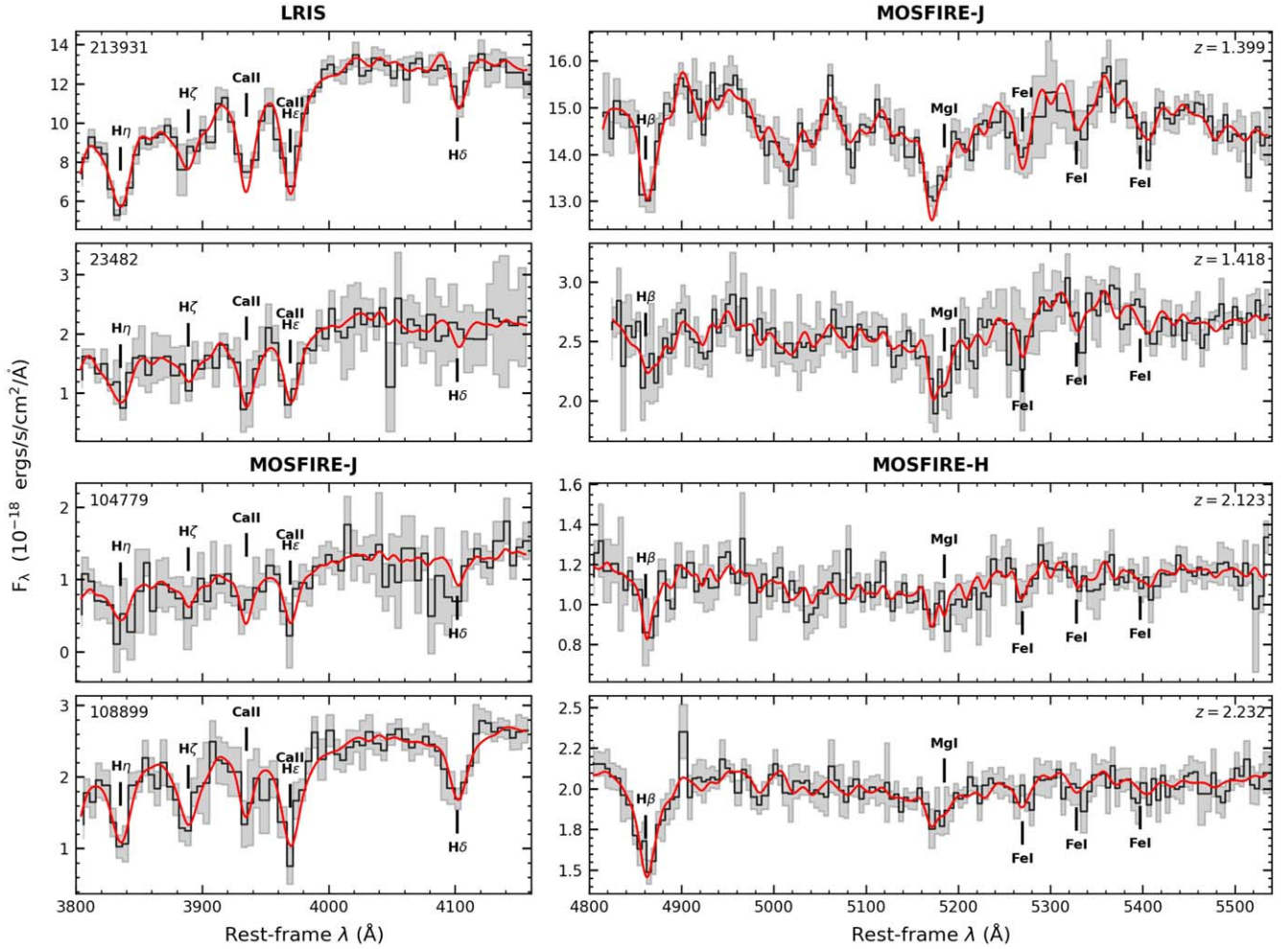
Figure 1. Spectra (black) and corresponding 1σ uncertainties on the flux (gray), along with the best-fitting `alf` models used to derive the ages and elemental abundances (red) of four example primary Heavy Metal Survey targets (two at $z \sim 1.4$ and two at $z \sim 2.2$). The other spectra can be found in Appendix A. Spectra are binned such that 1 pixel is $\simeq 5$ Å in the rest frame. The spectra were fit before binning.

+KMOS survey (Carnall et al. 2022) and two lensed galaxies in the AGEL survey (Zhuang et al. 2023) have [Fe/H] ~ -0.25, in agreement with the $z \sim 1.4$ Heavy Metal galaxies (see Figure 2). At $z = 2.1$, Kriek et al. (2016) measure [Fe/H] $= -0.25$ for a massive quiescent galaxy (COSMOS-11494), consistent with the values of the $z \sim 2.1$ Heavy Metal galaxies (refer to Figure 2). Interestingly, the only other study at $z \sim 2.1$ finds supersolar [Fe/H] for a lensed galaxy (MRG-M0138) at $z = 1.98$ (Jafariyazani et al. 2020). It was previously unclear whether this large variation in [Fe/H] is typical of the $z \sim 2.1$ quiescent galaxy population. Now, with four Heavy Metal galaxies at $z \sim 2.1$ with subsolar [Fe/H], there is additional evidence suggesting that typical massive galaxies at $z \sim 2.1$ are indeed Fe-deficient and that the Fe abundance of MRG-M0138 may be anomalous. Nevertheless, the sample size is small, and the large uncertainties on the Fe abundances prevent any definitive statement about the $z \sim 2.1$ population.

In the middle panel of Figure 2, we see that the Heavy Metal galaxies at $z \sim 1.4$ have similar [Mg/Fe] to what is found in the SDSS ($z \sim 0$) and LEGA-C ($z \sim 0.7$) samples. Meanwhile, the $z \sim 2.1$ Heavy Metal galaxies have [Mg/Fe] that is ≈ 0.2 dex higher than what is found at lower redshifts, though the typical uncertainties are large (0.2 dex) and the sample size is small. We derive a σ–[Mg/Fe] relation for the $z \sim 1.4$ Heavy Metal

galaxies using the same method as with [Fe/H], resulting in a negative slope of $-0.15^{+0.17}_{-0.29}$ (blue line with shaded confidence intervals). This slope is consistent with 0 to within 1σ, and there is no significant offset compared to the $z < 1$ relations. These results are supported by the other studies of quiescent galaxies at $z > 1$. The two studies at $z \sim 1.4$ find [Mg/Fe] consistent with the LEGA-C and SDSS, while the two at $z \sim 2$ find enhanced [Mg/Fe] consistent with the $z \sim 2.1$ Heavy Metal galaxies.

Finally, in the right panel of Figure 2, we show galaxy σ_v versus galaxy formation time. This age is determined using the best-fit stellar population age. We find that the formation times of the Heavy Metal galaxies range from 1 Gyr ($z_{form} = 9$) to 4 Gyr ($z_{form} = 2$) after the Big Bang, with an average formation time of 2.5 Gyr ($z_{form} = 3$) and 1.5 Gyr ($z_{form} = 4$) for the $z \sim 1.4$ and $z \sim 2.1$ samples, respectively. We note that the galaxy formation times are calculated using an SSP-equivalent age, and more complex star formation histories would have likely resulted in slightly earlier formation times. The assumption of an SSP, however, does not bias the stellar metallicities or elemental abundance measurements (Serra & Trager 2007). In the next section, we consider the results in Figure 2 in the context of galaxy evolution.
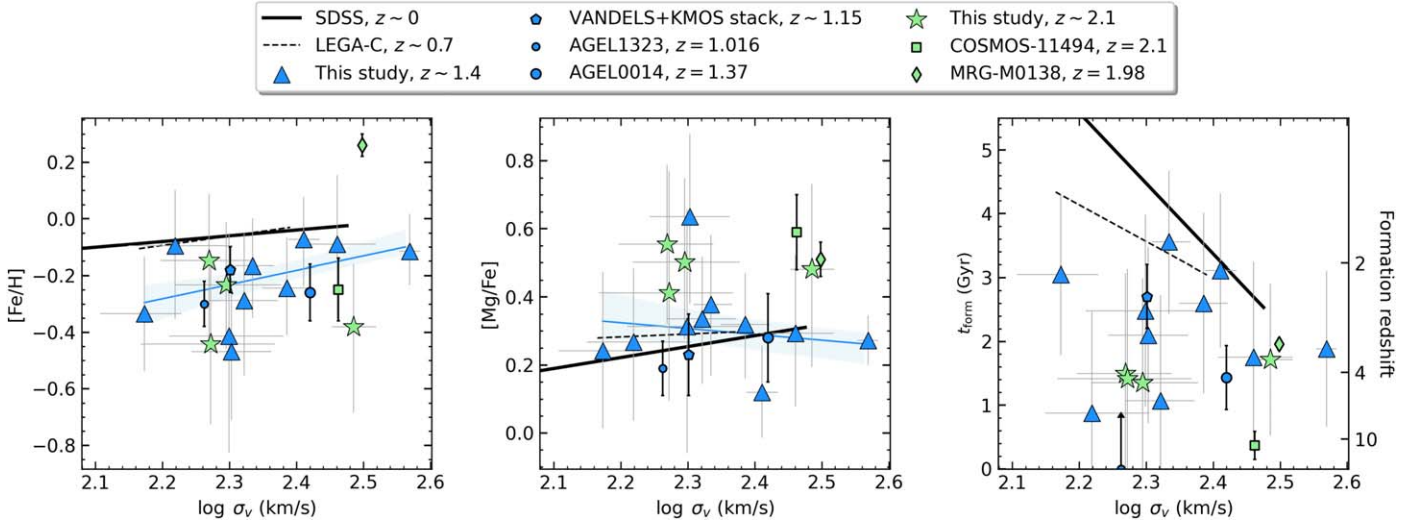
Figure 2. [Fe/H] (left), [Mg/Fe] (middle), and formation redshift (right) results for the $z \sim 1.4$ (blue triangles) and $z \sim 2.1$ (green stars) Heavy Metal targets as a function of stellar velocity dispersion. The black solid line shows the σ_v abundance relations of massive quiescent galaxies at $z \sim 0$ (Conroy et al. 2014; Beverage et al. 2021), while the black dashed lines denote relations at $z \sim 0.7$ (Beverage et al. 2023). We include a quiescent galaxy at $z = 2.1$ (green square; Kriek et al. 2016), a lensed quiescent galaxy at $z = 1.98$ (green diamond; Jafariyazani et al. 2020), a stack of quiescent galaxies at $1.0 < z < 1.3$ (blue pentagon; Carnall et al. 2022), and two lensed quiescent galaxies at $z = 1.016$ and $z = 1.37$ (blue circles; Zhuang et al. 2023). The solid blue line in the left panel represents the best-fit σ_v–[Fe/H] with the 16th and 84th percentile uncertainties on the fit shown in light blue. The results from all studies were derived using `alf`. Quiescent galaxies at $z > 0.7$ have lower [Fe/H], enhanced [Mg/Fe], and earlier formation times than galaxies at lower redshifts.

Table 1
Heavy Metal Quiescent Galaxy Parameters

ID[a]	z	σ (km s^{-1})	Age (Gyr)	[Fe/H]	[Mg/Fe]
23621	1.358	199^{+38}_{-41}	$2.1^{+1.0}_{-0.7}$	$-0.41^{+0.36}_{-0.41}$	$0.31^{+0.34}_{-0.37}$
25407	1.359	201^{+27}_{-28}	$2.5^{+0.9}_{-0.6}$	$-0.47^{+0.26}_{-0.24}$	$0.64^{+0.24}_{-0.26}$
26888	1.359	166^{+33}_{-27}	$3.7^{+2.3}_{-1.1}$	$-0.10^{+0.20}_{-0.26}$	$0.27^{+0.22}_{-0.23}$
217249[b, c]	1.377	185^{+21}_{-24}	$0.8^{+0.2}_{-0.1}$	$-0.38^{+0.21}_{-0.26}$	$0.34^{+0.25}_{-0.25}$
214695[b]	1.396	210^{+24}_{-25}	$3.4^{+2.2}_{-0.9}$	$-0.29^{+0.24}_{-0.27}$	$0.34^{+0.18}_{-0.19}$
213947[b]	1.397	216^{+16}_{-13}	$0.9^{+0.1}_{-0.1}$	$-0.17^{+0.17}_{-0.18}$	$0.38^{+0.20}_{-0.21}$
213931[b]	1.399	371^{+13}_{-13}	$2.6^{+0.6}_{-0.4}$	$-0.12^{+0.13}_{-0.12}$	$0.27^{+0.07}_{-0.07}$
214340[b]	1.418	149^{+19}_{-23}	$1.4^{+0.4}_{-0.3}$	$-0.34^{+0.20}_{-0.20}$	$0.24^{+0.23}_{-0.23}$
23482	1.419	289^{+38}_{-35}	$2.6^{+1.3}_{-0.7}$	$-0.09^{+0.24}_{-0.26}$	$0.29^{+0.23}_{-0.23}$
25702	1.419	243^{+21}_{-20}	$1.8^{+0.7}_{-0.4}$	$-0.24^{+0.18}_{-0.20}$	$0.32^{+0.15}_{-0.16}$
23351	1.420	257^{+14}_{-13}	$1.3^{+0.3}_{-0.2}$	$-0.07^{+0.15}_{-0.13}$	$0.12^{+0.14}_{-0.13}$
59375[c]	1.550	333^{+42}_{-43}	$9.5^{+3.4}_{-4.9}$	$0.21^{+0.30}_{-0.30}$	$0.36^{+0.28}_{-0.29}$
60736[c]	1.861	270^{+30}_{-27}	$2.1^{+1.7}_{-0.6}$	$-0.49^{+0.23}_{-0.25}$	$0.60^{+0.26}_{-0.30}$
104779	2.123	186^{+30}_{-31}	$1.5^{+0.9}_{-0.4}$	$-0.15^{+0.24}_{-0.28}$	$0.55^{+0.23}_{-0.23}$
56163[c]	2.160	213^{+54}_{-48}	$2.9^{+1.1}_{-1.1}$	$-0.43^{+0.32}_{-0.32}$	$0.29^{+0.32}_{-0.31}$
106812	2.230	187^{+41}_{-45}	$1.5^{+1.1}_{-0.4}$	$-0.44^{+0.28}_{-0.28}$	$0.41^{+0.36}_{-0.32}$
108899	2.233	306^{+24}_{-23}	$1.2^{+0.2}_{-0.1}$	$-0.38^{+0.22}_{-0.30}$	$0.48^{+0.25}_{-0.29}$
107590[c]	2.234	173^{+25}_{-23}	$0.8^{+0.1}_{-0.1}$	$-0.42^{+0.18}_{-0.22}$	$0.44^{+0.26}_{-0.30}$
103236	2.243	197^{+37}_{-35}	$1.5^{+1.0}_{-0.4}$	$-0.23^{+0.22}_{-0.23}$	$0.50^{+0.25}_{-0.23}$

Notes.
[a] See Kriek et al. (2024) for coordinates, magnitudes, and stellar population and structural properties.
[b] Appears in Kriek et al. (2019).
[c] Poorly constrained, removed from analysis.

4. The Evolution of the Massive Quiescent Galaxy Population since $z = 2.1$

In this section, we investigate the changes of *average* stellar metallicities, abundance ratios, and ages of massive quiescent galaxies as a function of redshift and discuss implications for their evolution over the past 11 billion yr. In Figure 3, we show the average [Fe/H], [Mg/H], $t_{\rm form}$, and [Mg/Fe] for quiescent galaxies in the SDSS, LEGA-C survey, and Heavy Metal Survey as a function of redshift. To allow for a fair comparison between surveys, we only consider the SDSS and LEGA-C galaxies that fall in bins with similar σ_v as the Heavy Metal galaxies. Specifically, we remove three SDSS bins and one LEGA-C bin with $\sigma < 170$ km s^{-1}. Thus, the average velocity dispersions of the SDSS and LEGA-C galaxies used in Figure 3 are $\sigma_v = 240$ km s^{-1}, with the number of galaxies in each respective sample now being 2990 (SDSS) and 101 (LEGA-C).

Examining the top left panel of Figure 3, we observe an increase in the average [Fe/H] from $z = 2.1$ to $z = 0$, with distinct offsets of -0.2 and -0.3 dex at $z \sim 1.4$ and $z \sim 2.1$, respectively, between the Heavy Metal galaxies and the lower-redshift samples. Moving to the upper middle panel, we find less evolution in [Mg/H] compared to [Fe/H]; the $z = 1.4$ sample is 0.1 dex lower compared to the $z = 0.7$ and $z = 0$ galaxies. The $z = 2.1$ galaxies have no [Mg/H] offset compared to lower-redshift galaxies, although the uncertainties remain substantial. Primarily due to the increase in [Fe/H], the average [Mg/Fe] decreases toward lower redshift, as revealed in the upper right panel. The $z = 2.1$ sample exhibits [Mg/Fe] elevated by 0.25 dex, while the $z = 1.4$ sample is only elevated by 0.08 dex. In the bottom left panel, we show that the average $t_{\rm form}$ changes from $t_{\rm form} = 1.5$ to $t_{\rm form} = 4$ Gyr from $z = 2.1$ to $z = 0$.

In the bottom middle panel of Figure 3, we compare the average [Mg/Fe] as a function of [Fe/H] with a simple closed-box chemical evolution model from Kriek et al. (2016; represented by a black dashed line). The model provides mass-weighted abundances for various different star formation timescales (indicated in Gyr next to the black boxes). The behavior of the model is driven by the differential enrichment timescales of Fe and Mg. While Mg is predominantly synthesized in massive stars and promptly returned to the interstellar medium through core-collapse supernovae (SNe),
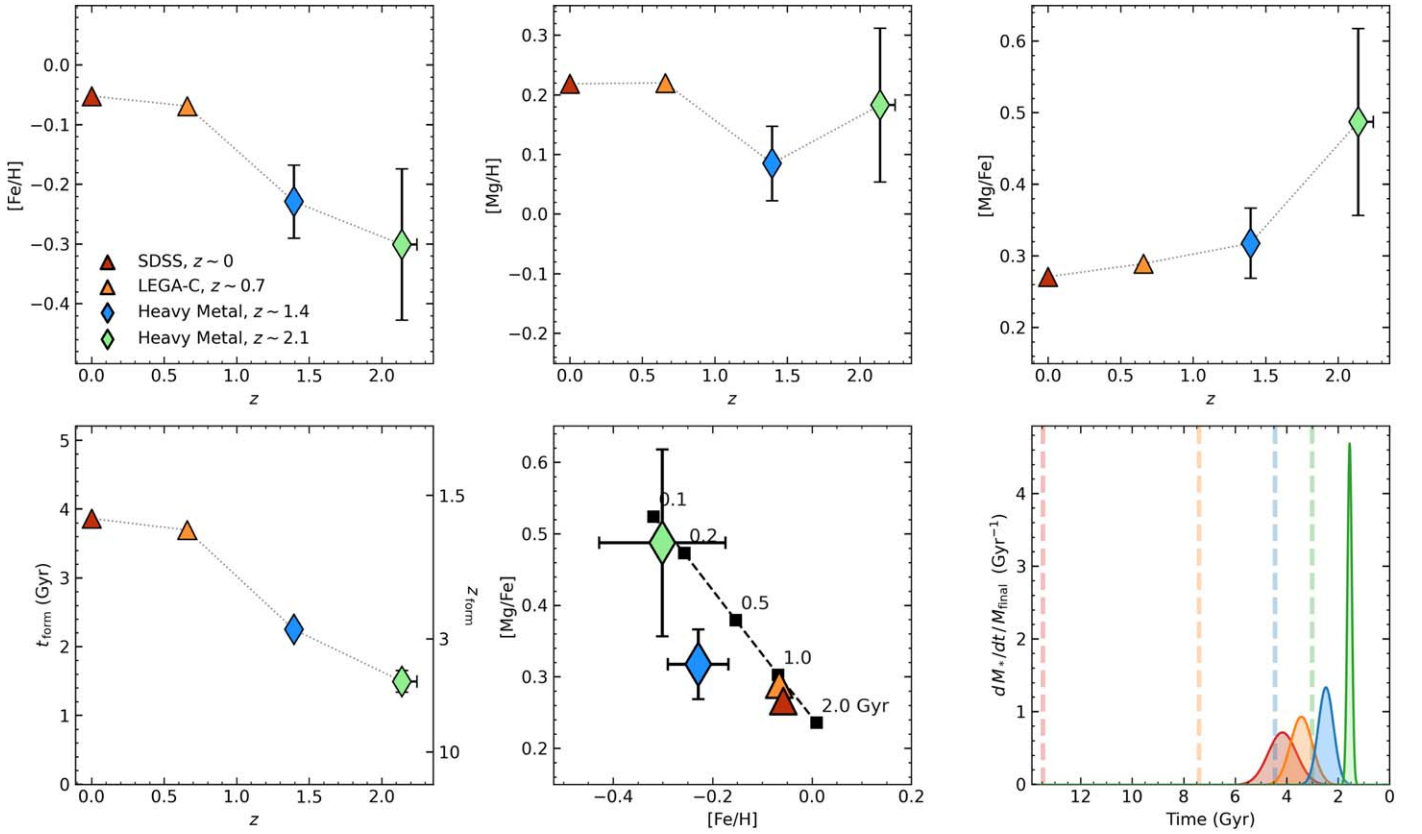
Figure 3. Average elemental abundance and stellar age results for the Heavy Metal galaxies at $z \sim 1.4$ and $z \sim 2.1$ and for stacks of massive quiescent galaxies at $z \sim 0$ and $z \sim 0.7$ across various axes. The error bars on the SDSS and LEGA-C data points are smaller than the symbol size. Top row: the [Fe/H], [Mg/H], and [Mg/Fe] abundances as a function of redshift. Bottom row: the left panel shows the formation time as a function of redshift. In the middle panel, we compare the average abundances on the [Fe/H]–[Mg/Fe] plane to a simple closed-box chemical evolution model (black dashed line). Black squares represent predicted elemental abundances for various star formation timescales, denoted in Gyr. The right panel shows the inferred star formation history at each redshift interval. The age of the Universe for each sample is marked with a dashed line. Typical quiescent galaxies at higher redshift formed earlier and over shorter timescales than massive quiescent galaxies today.

Fe is produced through the explosions of intermediate-mass stars (Type Ia SNe; SNe Ia) and is released on delayed timescales. Therefore, galaxies with shorter star formation timescales experience less Fe enrichment from SNe Ia prior to quenching, resulting in stellar populations with lower [Fe/H] and higher [Mg/Fe].

The simple chemical evolution model used in this paper is illustrative and is by no means meant to capture the complex feedback processes and star formation histories likely present in the galaxies at hand. More detailed chemical evolution models, including inflows and outflows, along with varied star formation histories, will help decipher the various factors shaping the elemental abundances and stellar metallicities of integrated stellar populations, especially as we move to higher redshifts (e.g., see N. Gountanis 2024, in preparation).

A visual comparison between the observations and the simple chemical evolution model highlights that the inferred star formation timescales of the $z \sim 0$ and $z \sim 0.7$ samples are longer than the extreme timescales observed for the Heavy Metal galaxies, which are approximately 0.15 and 0.7 Gyr for the $z \sim 2.1$ and $z \sim 1.4$ samples, respectively. These inferred star formation timescales derived from [Mg/Fe], along with the results for $t_{\rm form}$, paint a picture of galaxy formation in which typical massive quiescent galaxies at higher redshift formed earlier and underwent more rapid star formation compared to their counterparts at $z \sim 0$. To visually depict this scenario, we present the implied average star formation histories in the lower

right panel of Figure 3. We adopt a Gaussian parameterization for the star formation histories, with the mean centered at $t_{\rm form}$ and the FWHM corresponding to the star formation timescale, which is determined by comparing the abundances to the chemical evolution model. It is important to note that this panel serves as an illustrative tool, and the choice of a Gaussian star formation history is driven by its simplicity rather than a specific physical motivation. Nevertheless, it is evident that the star formation histories of the Heavy Metal galaxies, particularly at $z \sim 2.1$, are extremely short compared to the lower-redshift samples. We also note that this panel does not take into account the evolution of the stellar mass function or the spread in formation redshifts and star formation timescales within the sample. Instead, it shows the *typical* star formation history at each redshift interval.

Before considering how one should interpret the observed evolution, it is important to understand how selection biases influence the results. Unlike the Heavy Metal Survey, the $z \sim 0$ and $z \sim 0.7$ samples are mass-complete; therefore, the average stellar population properties should reflect the general galaxy population at each redshift. The Heavy Metal sample, however, has a complex selection function (see Kriek et al. 2024). The selection favors quiescent galaxies with lower mass-to-light ratios, indicative of younger ages. As a result, the ages of the Heavy Metal galaxies are likely biased young compared to the true population of quiescent galaxies at $z > 1$. Accounting for this bias, however, would strengthen the observed evolution in

Figure 3; galaxies at $z > 1$ could have even earlier formation redshifts, and therefore more extreme chemical properties, than what is found in this sample.

One possible explanation for the observed evolution is the growth of the quiescent galaxy population due to the continuous quenching of star-forming galaxies between $z = 2.1$ and $z = 0$ (i.e., progenitor bias; van Dokkum & Franx 2001; Carollo et al. 2013; Lilly et al. 2013). Galaxies that quench at later epochs presumably form their stars over longer timescales, resulting in later t_{form} and more Fe enrichment from SNe Ia (lower [Mg/Fe]) before quenching. Hence, the addition of these galaxies to the quiescent population steadily increases the *average* t_{form} and [Fe/H], as observed. In fact, the number of quiescent galaxies with $M_* > 10^{10.5} M_\odot$ has increased tenfold since $z = 2$, as evident from the evolution of the stellar mass function (e.g., McLeod et al. 2021; Conselice et al. 2022). Furthermore, most of this evolution occurs at $z > 0.5$ (McLeod et al. 2021), which could explain the pronounced jump in [Fe/H] and t_{form} from $z \sim 1.4$ to $z \sim 0.7$.

An alternate explanation is mergers and late-time star formation, which impact the elemental abundances of *individual* galaxies; when galaxies accrete satellites or form new stars, their integrated elemental abundances are altered by the newly added stars. Indeed, the importance of (dry) minor mergers in the evolution of quiescent galaxies has been posited by observations that they were more compact and had flatter color gradients at earlier times (e.g., van Dokkum et al. 2008; Naab et al. 2009; Trujillo et al. 2009; Taylor et al. 2010; van Dokkum et al. 2010; Suess et al. 2019b, 2020; Miller et al. 2023), as well as by direct pair fraction measurements (Newman et al. 2012). Furthermore, Suess et al. (2023) show that tiny companions with stellar mass ratios of 1:900 compromise some 30% of their total stellar mass budget.

The extent to which mergers, late-time star formation, and progenitor bias contribute to the evolution of the massive quiescent galaxy population remains a topic of ongoing debate (e.g., Bezanson et al. 2009; Hopkins et al. 2009; Carollo et al. 2013; Poggianti et al. 2013; van de Sande et al. 2013; Williams et al. 2017; Damjanov et al. 2019). In the following section, we examine the elemental abundances and stellar ages of individual galaxies to help differentiate between the various evolutionary scenarios; if mergers and/or late-time star formation play only a small role in the evolution of massive quiescent galaxies, then the extreme galaxies observed at $z \sim 2$ should persist as rare remnants in the $z \sim 0$ population.

5. Implications for the Buildup of Massive Quiescent Galaxies

In the previous section, we observed significant evolution in the average elemental abundances of massive quiescent galaxies from $z \sim 2.1$ to $z \sim 0$, suggesting that galaxies at higher redshift are typically metal-poor, have shorter star formation timescales, and formed earlier. In this section, we examine the abundance ratios of individual galaxies over cosmic time to help differentiate between various evolutionary channels.
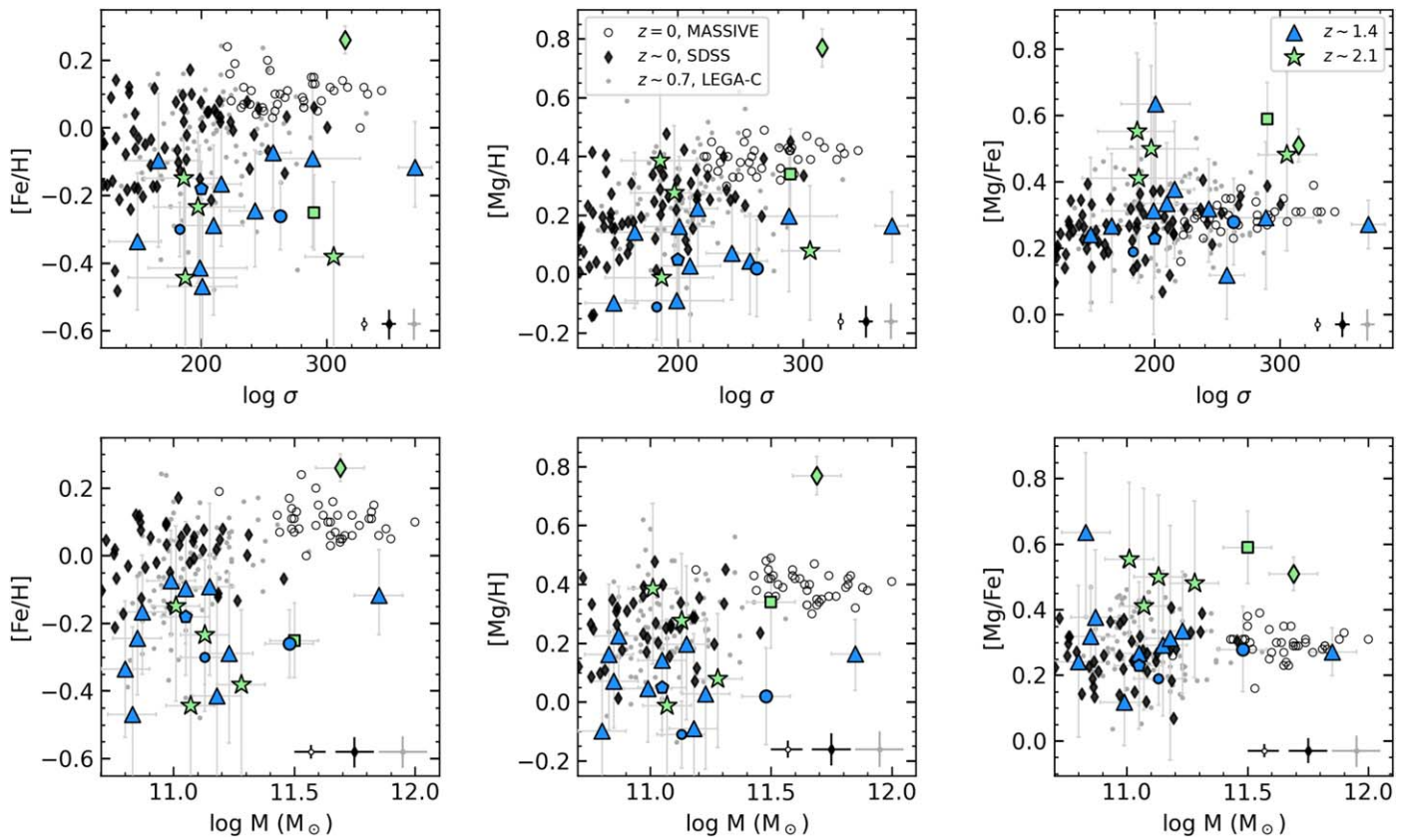
Figure 4 shows [Fe/H], [Mg/H], and [Mg/Fe] as a function of σ_v (top row) and M_* (bottom row). The stellar masses of the Heavy Metal galaxies are derived in Kriek et al. (2024). We compare the Heavy Metal galaxies to the abundances of individual LEGA-C galaxies (Beverage et al. 2023; gray

circles), and to galaxies in SDSS (black diamonds) presented by Zhuang et al. (2023). Additionally, we compare to alf-derived abundances of 41 galaxies from the MASSIVE survey, a volume-limited survey of the most massive galaxies in the local Universe (open circles; Ma et al. 2014; Gu et al. 2022). We also include abundance results from other studies of quiescent galaxies at $z \sim 1.4$ and $z \sim 2.1$, with the same markers as Figure 2. We note that the SDSS fiber and the LEGA-C and Heavy Metal slits cover similar physical extents within the galaxies (3–4 kpc). However, as galaxies were smaller at earlier times, the Heavy Metal and LEGA-C slits cover 1–1.2 R_e, while SDSS only covers 0.4–0.8 R_e. The MASSIVE spectra were extracted from the inner $R_e/8$ and therefore represent only the cores of nearby massive early-type galaxies.

In the previous section, we proposed two theories to explain the observed abundance differences between the Heavy Metal Survey and $z \sim 0$. In the first scenario, the quiescent galaxy population is continuously growing by the addition of newly quenched galaxies, driving the evolution of the observed average abundances ("progenitor bias," e.g., van Dokkum & Franx 2001; Carollo et al. 2013; Poggianti et al. 2013). In this scenario, we expect to find galaxies with properties similar to the Heavy Metal galaxies in the lower-redshift Universe, and thus the populations at higher and lower redshift should overlap. In other words, the Heavy Metal galaxies should constitute the lower-metallicity and high-[Mg/Fe] tail of the $z \sim 0$ samples. Indeed, the top panels of Figure 4 show that this is mostly the case for the Heavy Metal galaxies, though there are a few exceptions; most of the Heavy Metal galaxies with high σ_v have significantly lower [Fe/H] and [Mg/H] than any galaxy found at $z \sim 0$ and $z \sim 0.7$. Furthermore, in the rightmost panel, we find that galaxies at $z \sim 2.1$ are *all* [Mg/Fe]-enhanced compared to the galaxies at lower redshifts.

When considered as a function of stellar mass, the metallicity differences between the lower- and higher-redshift populations become more apparent. At all stellar masses, and particularly at the highest stellar masses, the Heavy Metal galaxies have [Fe/H] and [Mg/H] ≈ 0.2 dex lower than any individual galaxy in SDSS or LEGA-C. This result poses a problem for the progenitor bias scenario; if these Heavy Metal galaxies passively evolve to $z \sim 0$, we would expect to find Fe- and Mg-deficient remnants in the SDSS and MASSIVE populations. Specifically, in this scenario, 10% of the SDSS and MASSIVE galaxies should be direct descendants of $z > 2$ quiescent galaxies given the rapid evolution of the stellar mass function (McLeod et al. 2021). Thus, the $z \sim 0$ galaxies are likely not the direct descendants of the $z \sim 2$ Heavy Metal galaxies. It is important to acknowledge, however, that uncertainties are substantial and sample sizes are limited. Thus, the possibility remains that we may have simply missed the rare metal-poor remnants in the $z \sim 0$ samples. We also note that the Heavy Metal galaxies were selected to maximize the number of bright quiescent galaxies in one field of view. As such, the environments of these galaxies may not be representative of the broader massive quiescent galaxy population at higher redshift.

As previously mentioned, minor mergers provide an alternate way to explain the observed evolution. In this picture, the $z \sim 1.4$–2.1 quiescent galaxies may grow into the cores of today's massive early-type galaxies by accreting low-mass galaxies (e.g., Bezanson et al. 2009; Hopkins et al. 2009; van

Figure 4. Individual measurements of [Fe/H], [Mg/H], and [Mg/Fe] as a function of σ_v (top row) and M_* (bottom row). The Heavy Metal galaxies are split into $z \sim 1.4$ (blue triangles) and $z \sim 2.1$ (green stars) samples. We include results from other studies of massive quiescent galaxies at similar redshifts (see legend in Figure 2). We compare these results to individual abundance measurements from the MASSIVE survey ($z = 0$; Gu et al. 2022; open circles), SDSS ($z \sim 0$; Zhuang et al. 2023; black diamonds), and LEGA-C survey ($z \sim 0.7$; Beverage et al. 2023; gray dots), with their typical uncertainties provided at the bottom of each panel. At constant M_* and σ_v, galaxies at $z > 1$ have different chemical properties than those at lower redshifts.

de Sande et al. 2013). This scenario can successfully explain the observed increase in stellar metallicity and decrease in [Mg/Fe], but only if the accreted galaxies form over long timescales (i.e., lower [Mg/Fe]) and are more metal-rich. There are, however, two problems with this scenario. First, our observations show there is evolution within the inner $1\,R_e$, while minor mergers are typically accreted onto galaxy outskirts (Naab et al. 2009; Oser et al. 2010). Second, given the mass–metallicity relation, the accreted low-mass galaxies will likely be metal-poor and therefore cannot explain the increase in metallicity (Kirby et al. 2013). Major mergers with galaxies that have longer star formation timescales (and thus lower [Mg/Fe]) solve both of these problems, as more massive galaxies tend to be more metal-rich *and* can more easily disrupt galaxy cores. However, the observed major merger rate is likely too low to fully account for the observed evolution (Newman et al. 2012; Nipoti et al. 2012; Oser et al. 2012), and major mergers alone cannot explain the size evolution (e.g., Hopkins et al. 2009; van Dokkum et al. 2010; Hilz et al. 2013; Belli et al. 2014). Therefore, a combination of major and minor mergers may be necessary to explain the observed evolution.

Finally, late-time star formation, likely in combination with or triggered by mergers, can increase the metallicities of individual galaxies while decreasing [Mg/Fe]; if the star formation material is enriched in SN Ia products by previous epochs of star formation, then the newly formed stars will be younger, more metal-rich, and more α-enhanced. It should be noted that a burst of central star formation is expected in a wet

merger scenario (e.g., Hopkins et al. 2006); thus, the two scenarios together are effective at explaining the absence of metal-poor and [Mg/Fe]-enriched galaxies by $z \sim 0$ as observed in Figure 4.

Crucially, all of the above scenarios rely on the assumption that at the same stellar masses, younger galaxies have higher metallicity and lower [Mg/Fe]. For example, the progenitor bias scenario only works if the galaxies that quench at later times are more metal-rich and form over longer timescales. In Figure 5, we search for such trends in [Fe/H], [Mg/H], and [Mg/Fe] as a function of $t_{\rm form}$ for the SDSS, LEGA-C, and Heavy Metal quiescent galaxies. In each panel, we remove the first-order mass dependence by subtracting the SDSS M_*–abundance relations from all galaxies and then scaling them to the value of the M_*–abundance relations at $M_* = 10^{11}\,M_\odot$ (e.g., Leethochawalit et al. 2019; Zhuang et al. 2023). To assess the level of correlation between quenching time (as traced by $t_{\rm form}$) and stellar metallicities, we use the Pearson correlation coefficient. A correlation is deemed "marginal" if the Pearson correlation coefficient is $0.2 \leqslant r < 0.4$, "moderate" if $0.4 \leqslant r < 0.6$, and "strong" if $r \geqslant 0.6$. We list the correlation coefficients along with the associated p-value in Figure 5. At fixed stellar mass, we do in fact find that younger SDSS and LEGA-C galaxies exhibit higher [Fe/H] ($r = 0.53$), marginally higher [Mg/H] ($r = 0.26$), and moderately lower [Mg/Fe] ($r = -0.40$), as previously found by Beverage et al. (2023) and Zhuang et al. (2023), respectively (see also Leethochawalit et al. 2018).
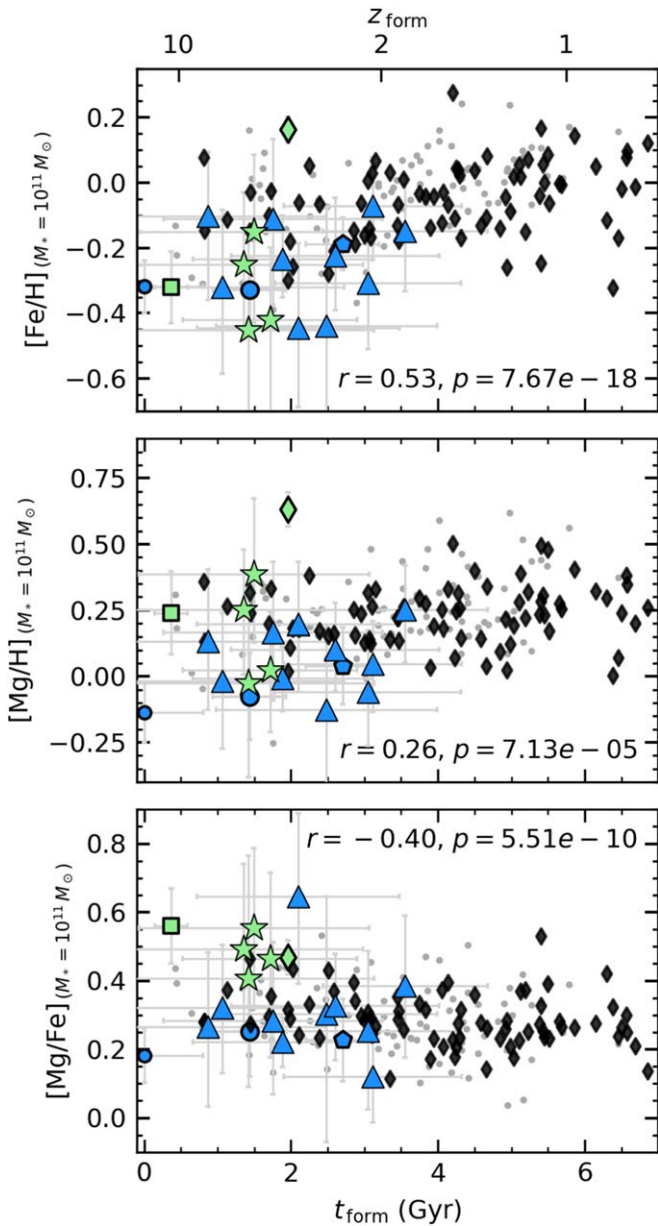
Figure 5. [Fe/H], [Mg/H], and [Mg/Fe] at a fixed stellar mass ($M_* = 10^{11}\ M_\odot$) as a function of formation time for the LEGA-C (gray circles), Heavy Metal (blue triangles and green stars), SDSS (black diamonds), and various other $z > 1$ studies (refer to the legend in Figure 2). Abundances at fixed stellar mass are calculated from the M_*–abundance relations at $z \sim 0$ shown in Figure 4. We assess the strength of the correlation between the elemental abundances and formation time using the Pearson correlation test (results listed in each panel). Galaxies that form at later cosmic epochs have higher [Fe/H] and [Mg/H] and slightly lower [Mg/Fe]. At a fixed t_{form}, the Heavy Metal galaxies exhibit similar abundances to the SDSS and LEGA-C populations.

The trends with [Fe/H] and [Mg/Fe] can be naturally explained by galaxies joining the quiescent galaxy population at later times (higher t_{form}) having more time to form stars from gas that is rich in SN Ia products (namely, Fe). However, the marginal trend with [Mg/H] is more challenging to explain, as Mg enrichment is independent of the star formation timescale and instead primarily depends on the size of the gas reservoir at the time of star formation quenching. Beverage et al. (2021) propose a model to explain the marginal trend with t_{form} by invoking rapid gas removal during the quenching process. In

summary, if galaxies at earlier epochs, which form earlier and over shorter timescales, quench more abruptly and have larger gas fractions at the time of quenching (e.g., effective AGN outflows), then [Mg/H] is truncated to lower values. Thus, the marginal trend with [Mg/H] may suggest that galaxies at earlier times experience more effective outflows (see also the outflow model proposed by Leethochawalit et al. 2019).

Furthermore, Figure 5 shows that high-z galaxies have a slight systematic offset to lower metallicities compared to local galaxies even when the dependence on M_* and stellar age is removed. This same result was also found by Zhuang et al. (2023) but for a smaller galaxy sample. Thus, this result confirms the finding by Zhuang et al. (2023) and reinforces the need for mergers or late-time star formation to alter the chemical compositions and stellar ages of these extreme individual galaxies by $z \sim 0$. Again, we note that comparisons at "fixed formation times" between galaxies at different redshift regimes should be approached with caution; the SSP-equivalent ages can be biased young. As a result, a more complicated star formation history may shift the Heavy Metal galaxies in Figure 5 toward even earlier formation times. As a result, the ages of the Heavy Metal galaxies are likely biased young compared to the true population of quiescent galaxies at $z > 1$. Nonetheless, correcting for such a bias would not change our interpretation.

Figures 4 and 5 together demonstrate that progenitor bias, minor and major mergers, and late-time star formation are likely all responsible, to some extent, for the observed evolution in the elemental abundances of massive quiescent galaxies over cosmic time. This analysis further shows that the existing story of massive quiescent galaxy assembly—namely, that the $z > 1$ quiescent galaxies are the progenitor cores of today's massive early-type galaxies—is more complicated than initially anticipated. We arrive at a similar conclusion based on the dynamical masses of the Heavy Metal galaxies in Kriek et al. (2024).

6. Summary and Conclusion

In this paper, we present the stellar metallicities, elemental abundance ratios, and stellar ages of 19 massive quiescent galaxies from the Keck Heavy Metal Survey. The galaxies are divided over two redshift intervals around $z \sim 1.4$ and $z \sim 2.1$. We fit the ultradeep LRIS+MOSFIRE spectra using a full-spectrum modeling code with variable elemental abundance ratios. Combined, these measurements represent the largest sample of individual quiescent galaxies at $z \gtrsim 1.4$ with stellar metallicity and elemental abundance measurements.

We find that massive quiescent galaxies at $1.4 \lesssim z \lesssim 2.2$ have subsolar Fe abundances, ranging from [Fe/H] $= -0.5$ to -0.1 dex, with typical values of -0.2 and -0.3 dex at $z \sim 1.4$ and $z \sim 2.1$, respectively. We also find a trend between σ_v and [Fe/H], with a best-fit slope consistent with what is found at lower redshifts. We find a high [Mg/Fe] of ≈ 0.5 for the $z \sim 2.1$ galaxies, while those at $z \sim 1.4$ are lower with [Mg/Fe] ~ 0.3. The ages range from 0.9 to 3.7 Gyr, implying formation redshifts $z_{form} = 2$–8.

We examine the evolution of the massive quiescent galaxy population over cosmic time by comparing the average Heavy Metal elemental abundances and ages to those in the LEGA-C survey ($z \sim 0.7$) and SDSS ($z \sim 0$). The $z \sim 1.4$ and $z \sim 2.1$ galaxies have [Fe/H] that is 0.15 and 0.23 dex lower than what is observed at $z \lesssim 0.7$. Interestingly, we find no evolution in

[Mg/Fe] at $z \lesssim 1.4$, though the $z \sim 2.1$ galaxies have 0.2 dex higher [Mg/Fe]. By comparing these results with a simple chemical evolution model, we find that galaxies at higher redshifts formed at progressively earlier epochs and shorter timescales, with the $z \sim 2.1$ galaxies forming in only ≈ 150 Myr at $z = 4$.

The observed evolution of stellar population properties provides a unique insight into the assembly histories of massive quiescent galaxies. One possible explanation is the growth of the quiescent galaxy population. If the galaxies that join the population at later times are younger, more metal-rich, and less α-enhanced, then the average properties will evolve. However, this scenario cannot fully explain the observations because we do not find the direct descendants of the chemically extreme Heavy Metal galaxies at $z \lesssim 0.7$. Minor mergers also fail to account for the observed evolution; due to the mass–metallicity relation, they cannot increase [Fe/H], in particular in the galaxy centers. Thus, major mergers and/or late-time star formation are required to explain the evolution of chemical properties across cosmic time. These results imply a more complicated assembly history than the minor-merger-driven inside-out growth scenario previously inferred from the evolution in the sizes, color gradients, and central densities of quiescent galaxies. In our accompanying paper, Kriek et al. (2024) arrive at a similar conclusion based on the dynamical masses of the Heavy Metal galaxies.

This study was enabled by ultradeep spectroscopic observations conducted with the most efficient ground-based spectrographs. With up to 16 hr of integration per galaxy per band, these spectra represent some of the deepest spectroscopic observations of individual galaxies at $z \gtrsim 1.4$ with coverage of key absorption features. As a result, the Heavy Metal Survey has tripled the number of individual stellar metallicity measurements at $z \gtrsim 1.4$. Nonetheless, sample sizes at these redshifts remain small, and the uncertainties on each measurement remain substantial.

In the future, JWST will provide even deeper spectroscopy of massive quiescent galaxies at $z \gtrsim 1.4$, extending the existing sample sizes, reaching to lower stellar masses, and enabling the measurements of more elements. Thus, JWST will revolutionize our understanding of the assembly of massive quiescent galaxies over comsic time.

Appendix A
Heavy Metal Spectra and Best-fit Models

In Figure 6, we present the LRIS+MOSFIRE spectra for all 19 galaxies from the Heavy Metal Survey (black) along with their uncertainties (gray). The spectra were binned such that each pixel represents 5 Å in the rest frame. We show the corresponding best-fit alf model for each galaxy. The fitting was accomplished prior to binning the spectra. Galaxies with (*) in front of the listed ID numbers are too young for reliable abundance measurements, galaxies with ($^\wedge$) had inaccurate photometric redshifts and therefore we miss key absorption features, and galaxies with (#) have unconstrained parameters. All (five) galaxies with symbols in front of their IDs were removed from the main analysis of this paper.
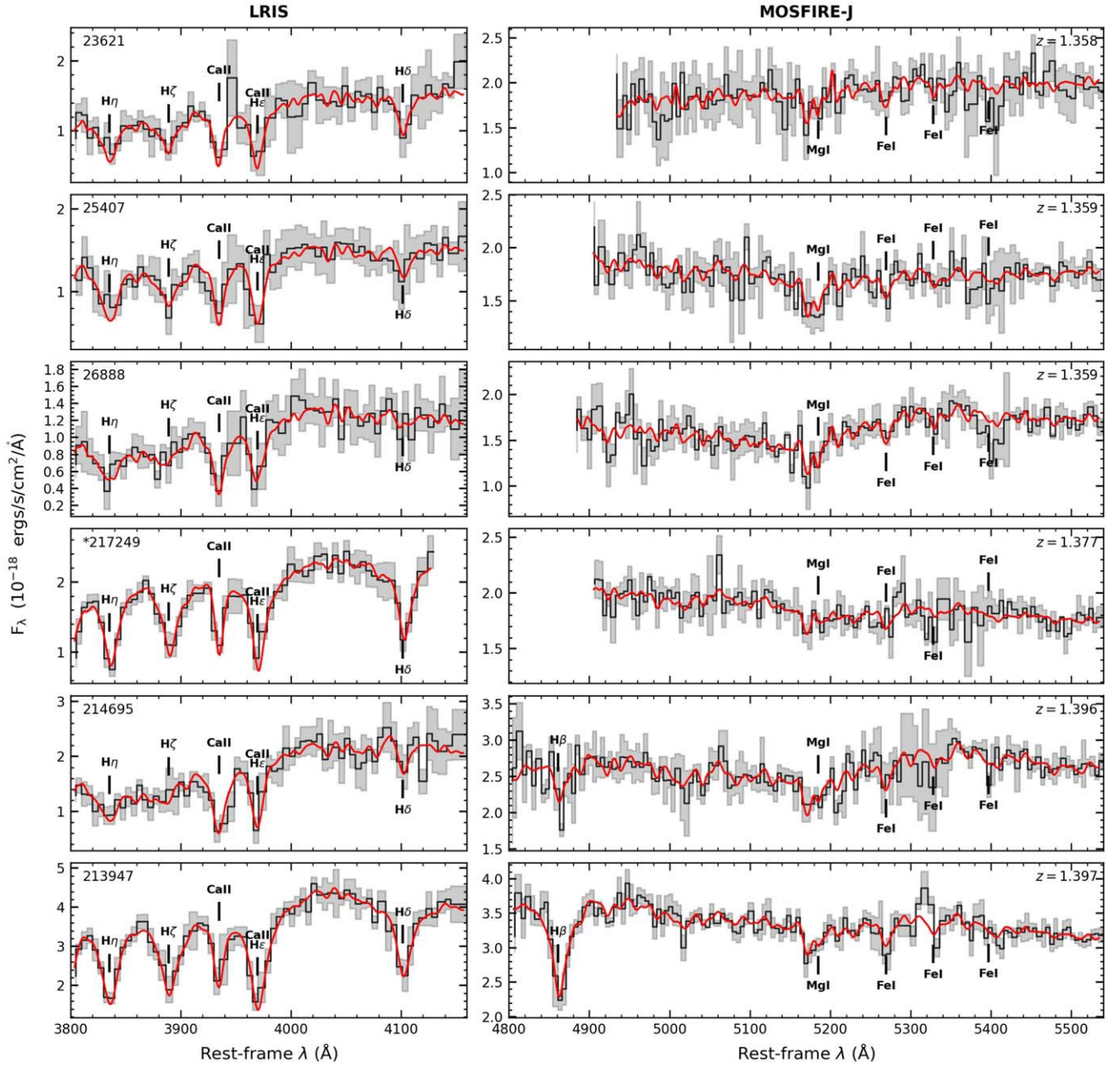
Figure 6. Spectra (black) and corresponding 1σ uncertainties on the flux (gray), along with the best-fitting `alf` models used to derive the ages and elemental abundances (red) of primary Heavy Metal Survey targets, sorted by redshift. Spectra are binned such that 1 pixel is $\simeq 5$ Å. The spectra were fit prior to binning.
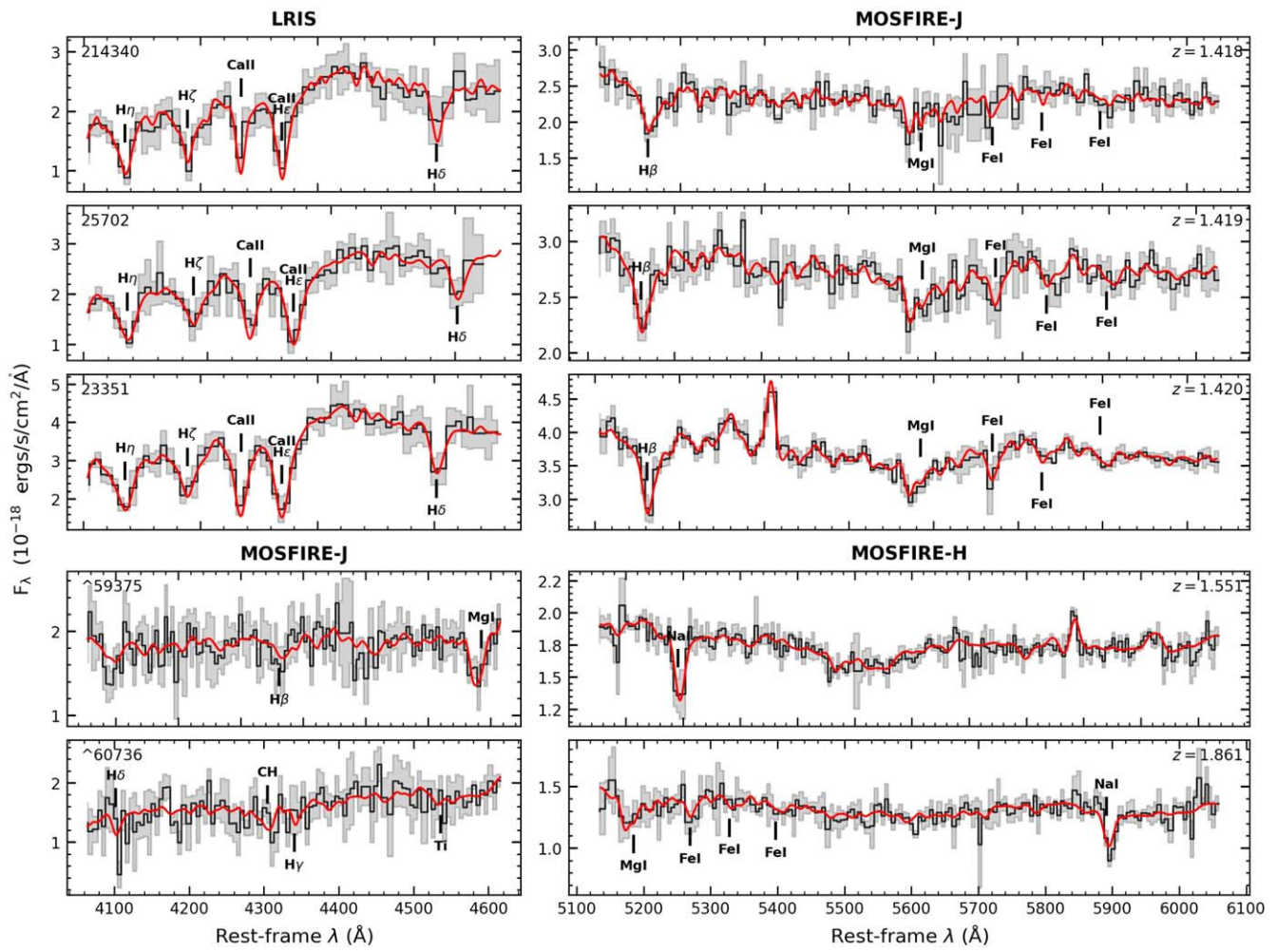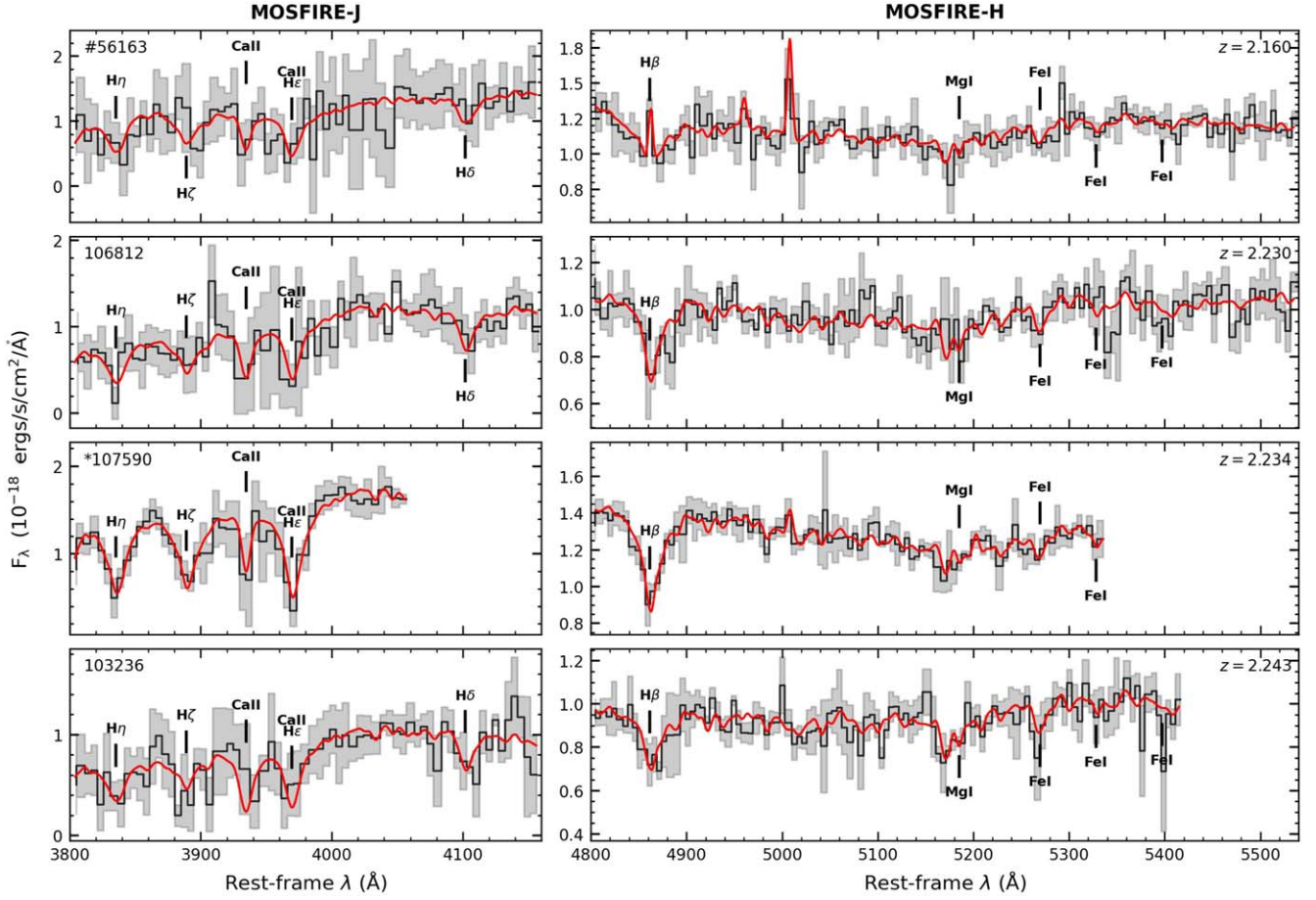
Figure 6. (Continued.)

Figure 6. (Continued.)

Appendix B
Quality of Stellar Population Modeling: Example Corner Plots

In Figure 7, we present corner plots from the full-spectrum `alf` fitting. The four corner plots correspond to the same four spectra showcased in Figure 1. Next to each corner plot, we provide information about the object: ID, redshift, and S/N ($\mathring{A}^{-1}$). We present the four essential parameters: [Fe/H], [Mg/H], [Mg/Fe], and the age of the stellar population. The dashed lines in the plots represent the 1σ uncertainties. Though the uncertainties are large, the results follow a normal distribution. We find there to be a degeneracy between stellar age and metallicity. Target 213931 shows the strongest degeneracy, likely because the observations are a blend of two nearby systems, resulting in the superposition of stellar populations with different ages and metallicities. For the other galaxies, the age–Z degeneracy is considerably less pronounced.
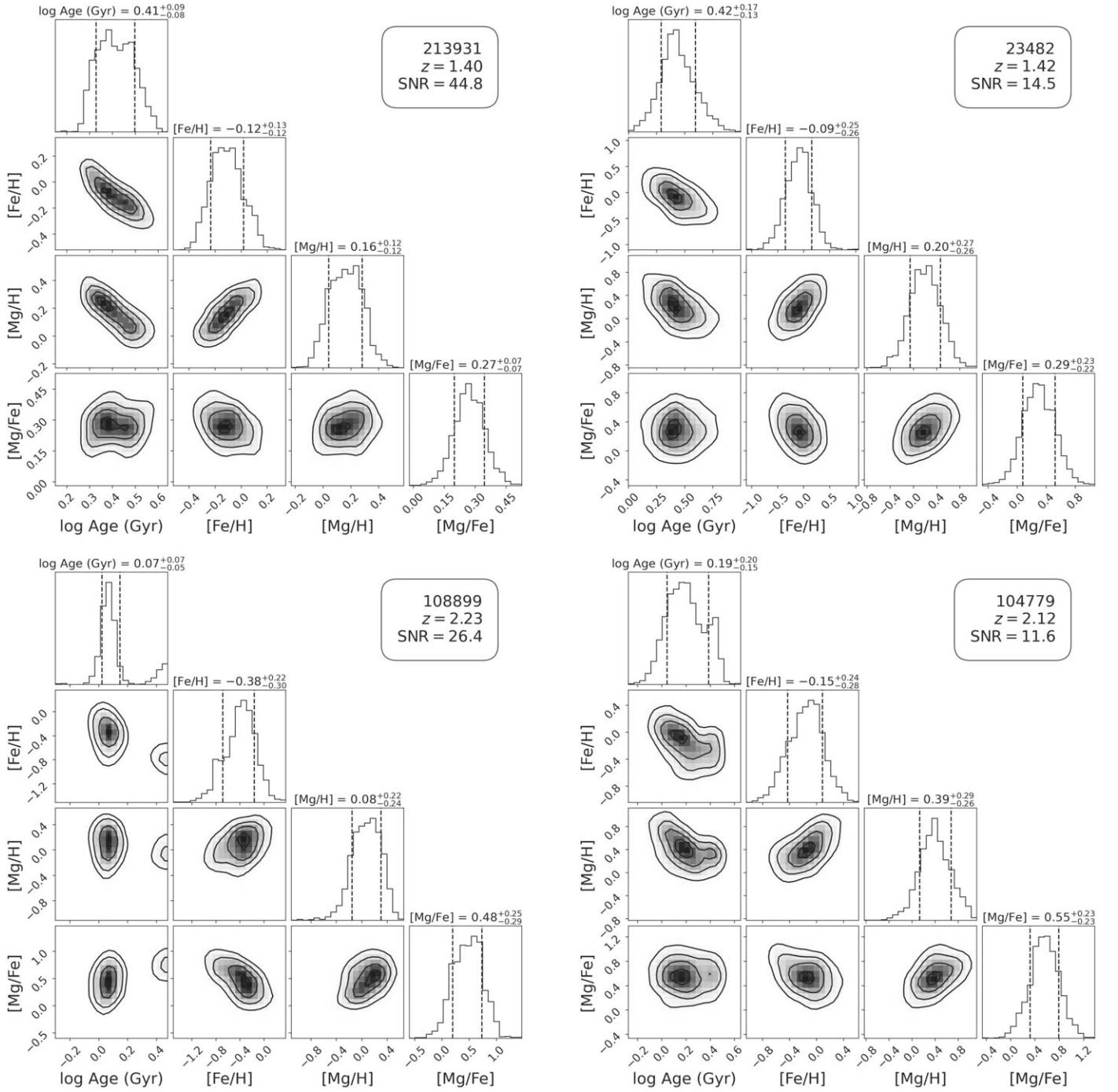
Figure 7. Four example corner plots from the full-spectrum alf modeling. In each panel, we show the probability distributions of four different physical parameters. The dashed lines in the 1D distributions show the 1σ levels. For each corner plot, we report the galaxy ID, redshift, and S/N (Å^{-1}) of the observations.

Appendix C
## Comparison with Lick Indices

In this Appendix, we compute absorption line indices for the Heavy Metal galaxies. The Lick/IDS system, first introduced by Faber et al. (1985) and updated by Worthey et al. (1994), is the most common method for measuring elemental abundances and stellar ages of nearby quiescent galaxies. However, as we continue to push observations to higher redshifts, absorption features become fainter and more contaminated by NIR skylines. As a result, Lick indices and their derived stellar population parameters become highly uncertain. Full-spectrum

modeling was, in part, developed to account for these issues. By fitting the entire stellar continuum, noise peaks can be easily downweighted, and all absorption features for a single element can be evaluated simultaneously.

In this Appendix, we check that the absorption line indices of the Heavy Metal galaxies agree with our main results from full-spectrum modeling. Specifically, we compute the $<\!\text{Fe}\!>$ and $\text{Mg}b$ indices, where $<\!\text{Fe}\!> = 0.5 \times (\text{Fe}5270 + \text{Fe}5335)$. We then compare these indices to simple spectrum models from Thomas et al. (2003).

Given the prevalence of skylines in the Heavy Metal spectra, we must account for the prominent noise peaks. However, by
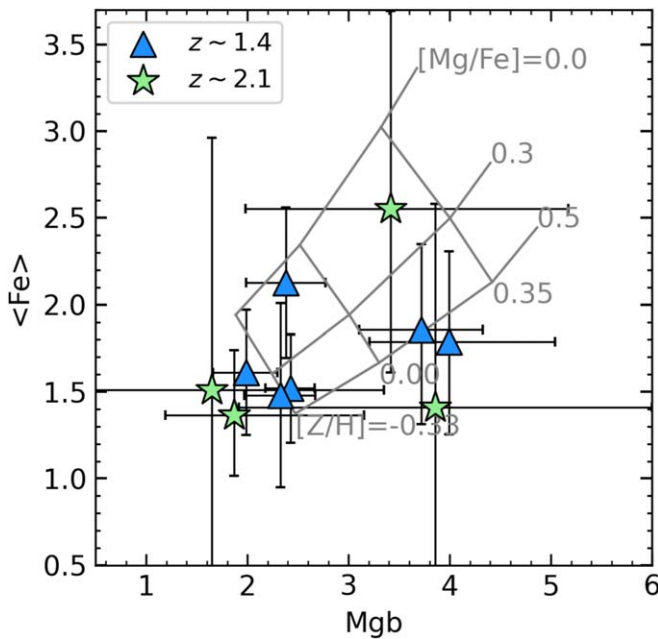
Figure 8. Lick/IDS indices Mg*b* and < Fe > for 10 primary $z \sim 1.4$ (blue triangles) and $z \sim 2.1$ (green stars) Heavy Metal targets. A grid of models from the Thomas et al. (2003) stellar population model ($t_{age} = 2$ Gyr) is superimposed in gray. The Heavy Metal galaxies tend toward low metallicities and high [Mg/Fe], in general agreement with what is found from the full-spectrum fitting. Uncertainties are large, motivating the need for full-spectrum modeling.

default, the Lick/IDS system does not use the variance spectrum because it biases the equivalent widths. Thus, to overcome skyline contamination, we follow the approach of van der Wel et al. (2021). For each spectrum, we take the best-fitting alf model and compute its χ^2. We then scale the noise to $\chi^2 = 1$. Any pixel that is more than 2σ or 20% offset from the model flux is masked. If more than 30% of the pixels in the index wavelength interval (including the pseudocontinuum intervals) are masked, the index is deemed unreliable. Four Heavy Metal galaxies are removed at this step. For the other 10 galaxies, the masked pixels are replaced by a linear interpolation using the good pixels. We note that the alf models are *only* used to help identify which pixels should be masked.

Next, we correct the absorption line indices to a constant velocity dispersion. This step is important if we want to compare indices between objects and to stellar population models. In the Lick/IDS system, indices are corrected to have zero intrinsic velocity dispersion and to have a standard wavelength-dependent resolution ranging from 8.4 Å to 11.5 Å (Worthey & Ottaviani 1997). We compute correction factors using alf models at a variety of velocity dispersions (e.g., see Davies et al. 1993; Kuntschner 2004). This correction factor is applied after measuring the equivalent widths. For the Heavy Metal galaxies, the correction ranges from $C(\sigma_v) = 1.0$ to 1.5. With the corrected Lick indices in hand, we finally estimate uncertainties by bootstrapping the spectrum and recomputing equivalent widths 10,000 times, taking the 16th and 84th percentiles as the index uncertainty.

In Figure 8, we present <Fe> as a function of Mg*b* for 10 Heavy Metal galaxies. We compare these indices with a set of 2 Gyr models from Thomas et al. (2003). Drawing conclusions from this figure is challenging, as numerous Heavy Metal galaxies display index measurements with substantial

uncertainties that span the entire model grid. This issue is particularly pronounced for galaxies at a redshift $z \sim 2.1$. The increased uncertainties at higher redshifts can be attributed to the younger ages of these galaxies, resulting in less pronounced metal lines. Additionally, significant contamination from skylines in the NIR underscores the necessity for full-spectrum modeling, as opposed to index fitting, at higher redshifts.

Appendix D
Parameter Recovery with Mock Heavy Metal Spectra

In this Appendix, we test the robustness of our results as a function of S/N. The primary goal of this test is to check for any systematic uncertainties in full-spectrum modeling, especially as we push the limits of this fitting to lower S/Ns. We begin by generating mock Heavy Metal spectra at various S/Ns, fitting them using alf, and finally comparing the best-fit values and their uncertainties to the true values. It is worth noting that Choi et al. (2014) conducted a similar test in their Appendix A. However, there are a few notable differences in our approach. First, alf has undergone significant updates since the work of Choi et al. (2014). Second, while Choi et al. (2014) used a flat noise spectrum for their simulated spectra, we use actual LRIS+MOSFIRE noise spectra. Finally, the Heavy Metal galaxies have much younger stellar ages, and therefore less pronounced absorption features, than what was tested in Choi et al. (2014).

All mock spectra are generated from a single alf model with stellar population parameters reflecting average values from Table 1. We degrade the model spectrum to the desired S/N at rest-frame 5000 Å using the Heavy Metal noise spectrum. In this way, we can replicate the observations, accounting for wavelength-dependent sensitivities. The model spectrum is also resampled to the LRIS+MOSFIRE resolution using spec-tres (Carnall 2017). We generate 45 realizations at S/Ns = 8, 12, 20, and 40 and fit them using the same method described in Section 2.2. For all realizations, we compute the difference between the best-fit parameter and the known model parameters.

We present the results in Figure 9. The black rectangles show the average parameter offset for the 45 realizations, with the blue error bars representing the 1σ scatter of these offsets. We also show the typical uncertainties on a single realization in red. The gray tick marks at the bottom of each panel represent the S/Ns of the individual Heavy Metal spectra. All but one Heavy Metal spectrum has S/N $\gtrsim 10$ Å^{-1}. σ_v is recovered extremely well with no systematic uncertainties at S/N $\gtrsim 12$. The stellar ages are also recovered well, with only a slight (~ 0.05 dex) systematic offset toward older ages at S/N $\lesssim 20$. The elemental abundance recovery also shows very little systematic uncertainty. There is a -0.1 dex offset in [Fe/H] at S/N = 8, but at the typical S/N of the Heavy Metal galaxies (S/N > 12), there is no appreciable systematic offset to within 0.05 dex. Similarly, [Mg/Fe] shows a slight systematic offset of 0.05 dex at S/N = 8 and S/N = 12, though at larger S/N, there is no appreciable systematic effect.

This parameter recovery test confirms that the abundance results corresponding to the Heavy Metal spectra (S/N $\gtrsim 12$ Å^{-1}) have systematic uncertainties of $\lesssim 0.05$ dex and can be constrained to within at least 0.15 dex. Furthermore, the results are mostly consistent with Choi et al. (2014), despite major updates to alf and mock spectra generated using real LRIS+MOSFIRE noise spectra.
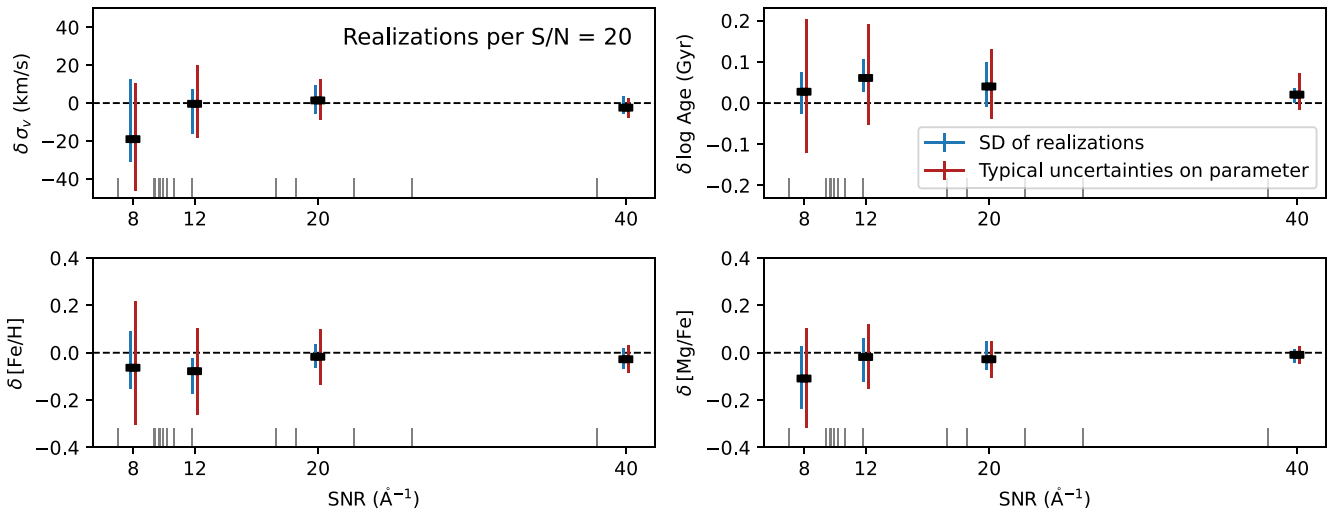
Figure 9. Results from the parameter recovery of mock Heavy Metal spectra. We show systematic offsets in the stellar population fitting code as a function of S/N for four key parameters. At each S/N, we generate 20 mock spectra and fit their elemental abundances and ages following the procedure in Section 2. The black rectangles represent the median offset between the simulated and best-fit values for the 20 realizations. The blue error bars represent the 16th and 84th percentiles of the best-fit values for the 20 realizations. The red error bars show the typical uncertainties on a single realization. The gray tick marks at the bottom of each panel show the true S/N of the individual Heavy Metal spectra. For the typical Heavy Metal galaxy, the systematic offsets are ≈ -0.05 dex for [Fe/H] and [Mg/Fe] and ≈ 0.05 dex for log Age.

ORCID iDs

Aliza G. Beverage ⓘ https://orcid.org/0000-0002-9861-4515
Mariska Kriek ⓘ https://orcid.org/0000-0002-7613-9872
Katherine A. Suess ⓘ https://orcid.org/0000-0002-1714-1905
Charlie Conroy ⓘ https://orcid.org/0000-0002-1590-8551
Sedona H. Price ⓘ https://orcid.org/0000-0002-0108-4176
Guillermo Barro ⓘ https://orcid.org/0000-0001-6813-875X
Rachel Bezanson ⓘ https://orcid.org/0000-0001-5063-8254
Marijn Franx ⓘ https://orcid.org/0000-0002-8871-3026
Brian Lorenz ⓘ https://orcid.org/0000-0002-5337-5856
Yilun Ma ⓘ https://orcid.org/0000-0002-0463-9528
Lamiya A. Mowla ⓘ https://orcid.org/0000-0002-8530-9765
Imad Pasha ⓘ https://orcid.org/0000-0002-7075-9931
Pieter van Dokkum ⓘ https://orcid.org/0000-0002-8282-9888
Daniel R. Weisz ⓘ https://orcid.org/0000-0002-6442-6030

References

Antwi-Danso, J., Papovich, C., Esdaile, J., et al. 2023, arXiv:2307.09590
Barro, G., Faber, S. M., Koo, D. C., et al. 2017, ApJ, 840, 47
Belli, S., Newman, A. B., & Ellis, R. S. 2014, ApJ, 783, 117
Belli, S., Newman, A. B., & Ellis, R. S. 2016, ApJ, 834, 18
Bevacqua, D., Saracco, P., La?Barbera, F., et al. 2023, MNRAS, 525, 4219
Beverage, A. G., Kriek, M., Conroy, C., et al. 2021, ApJL, 917, L1
Beverage, A. G., Kriek, M., Conroy, C., et al. 2023, ApJ, 948, 140
Bezanson, R., van Dokkum, P. G., Tal, T., et al. 2009, ApJ, 697, 1290
Carnall, A. C. 2017, arXiv:1705.05165
Carnall, A. C., McLure, R. J., Dunlop, J. S., et al. 2022, ApJ, 929, 131
Carnall, A. C., McLure, R. J., Dunlop, J. S., et al. 2023, Natur, 619, 716
Carollo, C. M., Bschorr, T. J., Renzini, A., et al. 2013, ApJ, 773, 112
Choi, J., Conroy, C., Moustakas, J., et al. 2014, ApJ, 792, 95
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Cimatti, A., Cassata, P., Pozzetti, L., et al. 2008, A&A, 482, 21
Cimatti, A., Daddi, E., Renzini, A., et al. 2004, Natur, 430, 184
Conroy, C., Graves, G. J., & van Dokkum, P. G. 2014, ApJ, 780, 33
Conroy, C., & van Dokkum, P. 2012, ApJ, 747, 69
Conroy, C., Villaume, A., van Dokkum, P., & Lind, K. 2018, ApJ, 854, 139
Conselice, C. J., Mundy, C. J., Ferreira, L., & Duncan, K. 2022, ApJ, 940, 168
Daddi, E., Renzini, A., Pirzkal, N., et al. 2005, ApJ, 626, 680
Damjanov, I., McCarthy, P. J., Abraham, R. G., et al. 2009, ApJ, 695, 101
Damjanov, I., Zahid, H. J., Geller, M. J., et al. 2019, ApJ, 872, 91
Davies, R. L., Sadler, E. M., & Peletier, R. F. 1993, MNRAS, 262, 650
Dokkum, P. G. v., Franx, M., Kriek, M., et al. 2008, ApJ, 677, L5

Estrada-Carpenter, V., Papovich, C., Momcheva, I., et al. 2019, ApJ, 870, 133
Faber, S. M., Friel, E. D., Burstein, D., & Gaskell, C. M. 1985, ApJS, 57, 711
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Franx, M., Labbé, I., Rudnick, G., et al. 2003, ApJ, 587, L79
Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, MNRAS, 362, 41
Glazebrook, K., Schreiber, C., Labbé, I., et al. 2017, Natur, 544, 71
Gu, M., Greene, J. E., Newman, A. B., et al. 2022, ApJ, 932, 103
Hilz, M., Naab, T., & Ostriker, J. P. 2013, MNRAS, 429, 2924
Hopkins, P. F., Bundy, K., Murray, N., et al. 2009, MNRAS, 398, 898
Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, ApJS, 163, 1
Jafariyazani, M., Newman, A. B., Mobasher, B., et al. 2020, ApJ, 897, L42
Ji, Z., & Giavalisco, M. 2022, ApJ, 935, 120
Khochfar, S., & Silk, J. 2009, MNRAS, 397, 506
Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, ApJ, 779, 102
Kriek, M., Beverage, A. G., Price, S. H., et al. 2024, ApJ, 966, 36
Kriek, M., Conroy, C., van Dokkum, P. G., et al. 2016, Natur, 540, 248
Kriek, M., Dokkum, P. G. v., Franx, M., et al. 2006, ApJ, 649, L71
Kriek, M., Price, S. H., Conroy, C., et al. 2019, ApJ, 880, L31
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Kroupa, P. 2001, MNRAS, 322, 231
Kuntschner, H. 2004, A&A, 426, 737
Leethochawalit, N., Kirby, E. N., Ellis, R. S., Moran, S. M., & Treu, T. 2019, ApJ, 885, 100
Leethochawalit, N., Kirby, E. N., Moran, S. M., Ellis, R. S., & Treu, T. 2018, ApJ, 856, 15
Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119
Lonoce, I., Longhetti, M., Maraston, C., et al. 2015, MNRAS, 454, 3912
Ma, C.-P., Greene, J. E., McConnell, N., et al. 2014, ApJ, 795, 158
Maiolino, R., & Mannucci, F. 2019, A&ARv, 27, 3
Matteucci, F. 1994, A&A, 288, 57
McDermid, R. M., Alatalo, K., Blitz, L., et al. 2015, MNRAS, 448, 3484
McLean, I. S., Steidel, C. C., Epps, H., et al. 2010, Proc. SPIE, 7735, 77351E
McLean, I. S., Steidel, C. C., Epps, H. W., et al. 2012, Proc. SPIE, 8446, 84460J
McLeod, D. J., McLure, R. J., Dunlop, J. S., et al. 2021, MNRAS, 503, 4413
Miller, T. B., Dokkum, P. v., & Mowla, L. 2023, ApJ, 945, 155
Morishita, T., Abramson, L. E., Treu, T., et al. 2018, ApJ, 856, L4
Mowla, L. A., Dokkum, P. v., Brammer, G. B., et al. 2019, ApJ, 880, 57
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18
Naab, T., Johansson, P. H., & Ostriker, J. P. 2009, ApJ, 699, L178
Nanayakkara, T., Glazebrook, K., Jacobs, C., et al. 2024, NatSR, 14, 3724
Newman, A. B., Ellis, R. S., Bundy, K., & Treu, T. 2012, ApJ, 746, 162
Nipoti, C., Treu, T., Leauthaud, A., et al. 2012, MNRAS, 422, 1714
Oke, J. B., Cohen, J. G., Carr, M., et al. 1995, PASP, 107, 375

Onodera, M., Carollo, C. M., Renzini, A., et al. 2015, ApJ, 808, 161
Oser, L., Naab, T., Ostriker, J. P., & Johansson, P. H. 2012, ApJ, 744, 63
Oser, L., Ostriker, J. P., Naab, T., Johansson, P. H., & Burkert, A. 2010, ApJ, 725, 2312
Peng, Y., Maiolino, R., & Cochrane, R. 2015, Natur, 521, 192
Poggianti, B. M., Moretti, A., Calvi, R., et al. 2013, ApJ, 777, 125
Rockosi, C., Stover, R., Kibrick, R., et al. 2010, Proc. SPIE, 7735, 77350R
Sanchez-Blazquez, P., Peletier, R. F., Jimenez-Vicente, J., et al. 2006, MNRAS, 371, 703
Saracco, P., Barbera, F. L., De Propris, R., et al. 2023, MNRAS, 520, 3027
Schreiber, C., Glazebrook, K., Nanayakkara, T., et al. 2018, A&A, 618, A85
Serra, P., & Trager, S. C. 2007, MNRAS, 374, 769
Spitoni, E., Vincenzo, F., & Matteucci, F. 2017, A&A, 599, A6
Suess, K. A., Kriek, M., Price, S. H., & Barro, G. 2019a, ApJ, 877, 103
Suess, K. A., Kriek, M., Price, S. H., & Barro, G. 2019b, ApJL, 885, L22
Suess, K. A., Kriek, M., Price, S. H., & Barro, G. 2020, ApJL, 899, L26
Suess, K. A., Williams, C. C., Robertson, B., et al. 2023, ApJL, 956, L42
Taylor, E. N., Franx, M., Brinchmann, J., van der Wel, A., & van Dokkum, P. G. 2010, ApJ, 722, 1
Thomas, D., Maraston, C., & Bender, R. 2003, MNRAS, 339, 897
Thomas, D., Maraston, C., & Bender, R. 2005, ApJ, 621, 673

Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2014, ApJ, 783, 85
Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000, AJ, 120, 165
Trujillo, I., Cenarro, A. J., de Lorenzo-Cáceres, A., et al. 2009, ApJ, 692, L118
Trussler, J., Maiolino, R., Maraston, C., et al. 2020, MNRAS, 491, 5406
Valentino, F., Tanaka, M., Davidzon, I., et al. 2020, ApJ, 889, 93
van de Sande, J., Kriek, M., Franx, M., et al. 2013, ApJ, 771, 85
van der Wel, A., Bezanson, R., D'Eugenio, F., et al. 2021, ApJS, 256, 44
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28
van Dokkum, P. G., & Franx, M. 2001, ApJ, 553, 90
van Dokkum, P. G., Franx, M., Kriek, M., et al. 2008, ApJ, 677, L5
van Dokkum, P. G., Whitaker, K. E., Brammer, G., et al. 2010, ApJ, 709, 1018
Villaume, A., Conroy, C., Johnson, B., et al. 2017, ApJS, 230, 23
Williams, C. C., Giavalisco, M., Bezanson, R., et al. 2017, ApJ, 838, 94
Williams, R. J., Quadri, R. F., Franx, M., Dokkum, P. v., & Labbé, I. 2009, ApJ, 691, 1879
Worthey, G. 1994, ApJS, 95, 107
Worthey, G., & Ottaviani, D. L. 1997, ApJS, 111, 377
Wuyts, E., Wisnioski, E., Fossati, M., et al. 2016, ApJ, 827, 74
Zahid, H. J., Geller, M. J., Fabricant, D. G., & Hwang, H. S. 2016, ApJ, 832, 203
Zhuang, Z., Leethochawalit, N., Kirby, E. N., et al. 2023, ApJ, 948, 132