



COSMOS2020: Exploring the Dawn of Quenching for Massive Galaxies at $3 < z < 5$ with a New Color-selection Method

Katriona M. L. Gould^{1,2} , Gabriel Brammer^{1,2} , Francesco Valentino^{1,2,3} , Katherine E. Whitaker^{1,4} , John. R. Weaver^{1,2,4} ,
Claudia del P. Lagos^{1,5,6} , Francesca Rizzo^{1,2} , Maximilien Franco⁷ , Bau-Ching Hsieh⁸ , Olivier Ilbert⁹ ,
Shuowen Jin^{1,10} , Georgios Magdis^{1,2,10} , Henry J. McCracken¹¹ , Bahram Mobasher¹² , Marko Shuntov^{1,2} ,
Charles L. Steinhardt^{1,2} , Victoria Strait^{1,2} , and Sune Toft^{1,2}

¹ Cosmic Dawn Center (DAWN), Denmark; katriona.gould@nbi.ku.dk

² Niels Bohr Institute, University of Copenhagen, Jagtvej 128, DK-2200 Copenhagen N, Denmark

³ European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei Munchen, Germany

⁴ Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA

⁵ International Centre for Radio Astronomy Research (ICRAR), M468, University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia

⁶ ARC of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

⁷ Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA

⁸ Academia Sinica Institute of Astronomy and Astrophysics, No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan

⁹ Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France

¹⁰ DTU-Space, Technical University of Denmark, Elektrovej 327, DK-2800 Kgs. Lyngby, Denmark

¹¹ Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, F-75014 Paris, France

¹² Department of Physics and Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA

Received 2022 December 1; revised 2023 February 18; accepted 2023 February 19; published 2023 May 23

Abstract

We select and characterize a sample of massive ($\log(M_*/M_\odot) > 10.6$) quiescent galaxies (QGs) at $3 < z < 5$ in the latest Cosmological Evolution Survey catalog (COSMOS2020). QGs are selected using a new rest-frame color-selection method, based on their probability of belonging to the quiescent group defined by a Gaussian mixture model (GMM) trained on rest-frame colors ($NUV - U$, $U - V$, $V - J$) of similarly massive galaxies at $2 < z < 3$. We calculate the quiescent probability threshold above which a galaxy is classified as quiescent using simulated galaxies from the SHARK semi-analytical model. We find that, at $z \geq 3$ in SHARK, the GMM/ $NUVU - VJ$ method outperforms classical rest-frame UVJ selection and is a viable alternative. We select galaxies as quiescent based on their probability in COSMOS2020 at $3 < z < 5$, and compare the selected sample to both UVJ - and $NUVrJ$ -selected samples. We find that, although the new selection matches UVJ and $NUVrJ$ in number, the overlap between color selections is only $\sim 50\%$ – 80% , implying that rest-frame color commonly used at lower-redshift selections cannot be equivalently used at $z > 3$. We compute median rest-frame spectral energy distributions for our sample and find the median QG at $3 < z < 5$ has a strong Balmer/4000 Å break, and residual NUV flux indicating recent quenching. We find the number densities of the entire quiescent population (including post-starbursts) more than doubles from $3.5 \pm 2.2 \times 10^{-6} \text{ Mpc}^{-3}$ at $4 < z < 5$ to $1.4 \pm 0.4 \times 10^{-5} \text{ Mpc}^{-3}$ at $3 < z < 4$, confirming that the onset of massive galaxy quenching occurs as early as $3 < z < 5$.

Unified Astronomy Thesaurus concepts: [Quenched galaxies \(2016\)](#); [High-redshift galaxies \(734\)](#); [Galaxy classification systems \(582\)](#); [Gaussian mixture model \(1937\)](#)

Supporting material: machine-readable table

1. Introduction

In the past few years it has become evident that there exists an extraordinary population of massive galaxies which had already ceased their star formation and become quiescent when the universe was just 2 billion years old ($z \geq 3$). The first hints of this population were candidates selected from optical/near-infrared (NIR) surveys such as in the GOODS-South field (Fontana et al. 2009), and later the Newfirm Medium-Band Survey (Marchesini et al. 2010; Whitaker et al. 2011) and FourStar Galaxy Evolution Survey (Spitler et al. 2014; Straatman et al. 2014). Targets were selected using various indicators, either on the basis of their specific star formation rate (sSFR) or rest-frame colors. Although these candidates were tentative then, the recent influx of spectroscopically

confirmed massive galaxies at $z > 3$ puts aside any doubt that these galaxies were already passive, and rapid evolution in the first few billion years caused the onset of the dawn of quenching. There are several examples of single galaxies (e.g., Marsan et al. 2015; Forrest et al. 2020a; Saracco et al. 2020; Valentino et al. 2020; Carnall et al. 2023), as well as full samples (Marsan et al. 2017; Schreiber et al. 2018; D'Eugenio et al. 2021; Forrest et al. 2020b; McConachie et al. 2022; Nanayakkara et al. 2022) that have been spectroscopically confirmed as both massive (usually, $\log(M_*/M_\odot) \gtrsim 10$) and quiescent ($\log(\text{sSFR}) \text{ yr}^{-1} \lesssim -9.5$), as well as a number of studies which have selected and analyzed massive quenched galaxies in large photometric data sets (Merlin et al. 2018; Cecchi et al. 2019; Girelli et al. 2019; Guarnieri et al. 2019; Merlin et al. 2019; Carnall et al. 2020; Santini et al. 2021; Shahidi et al. 2020; Stevans et al. 2021; Ito et al. 2022).

Thanks to the named studies, we now know about the properties of massive quiescent galaxies (QGs) at $z > 3$. These galaxies exhibit significant Balmer/4000 Å breaks indicative of

ages between 100 Myr and ~ 1 Gyr, although the exact ages can be quite uncertain (e.g., Schreiber et al. 2018; D’Eugenio et al. 2021; Forrest et al. 2020a; Carnall et al. 2020; Nanayakkara et al. 2022; Carnall et al. 2023). A few spectroscopically confirmed galaxies exhibit weak flux bluewards of the break (a “blue bump”), indicating a stellar population comprising predominantly A-type stars, and therefore likely recent quenching (< 1 Gyr). This is distinct from the UV upturn observed in local quenched galaxies, which seems to be due to emission from post-main-sequence stars such as horizontal branch or asymptotic giant branch stars (Greggio & Renzini 1990; Yi et al. 1997; see Dantas et al. 2020 for a more recent overview). The age and mass of such early QGs implies a rapid formation period and recent quenching, much like post-starburst (PSB) or E/K + A galaxies observed at $z < 3$ (e.g., Dressler & Gunn 1983; Goto 2005; Wild et al. 2009; Ichikawa & Matsuoka 2017; Wu et al. 2018; Chen et al. 2019; Wild et al. 2020; Wu et al. 2020; Wilkinson et al. 2021; see French 2021 for a review of PSB galaxies). An onset of rapid quenching at $z \sim 4$ – 6 , as suggested by the previous works, would result in a fairly young massive quiescent population at $z \sim 3$ that had recently undergone an intense period of star formation.

Recently quenched QGs at $z > 3$ also seem to be fairly dust free (Schreiber et al. 2018; Kubo et al. 2021). This could be because dust is being destroyed in the PSB phase (e.g., Akins et al. 2022) and our selections for QGs at $z > 3$ are dominated by PSB galaxies and relatively dust-free QGs, or because our selections and surveys are blind to dusty QGs at $z > 3$ (D’Eugenio et al. 2021). The upcoming investigations into galaxies appearing in NIR observations with no optical flux (galaxies dark to the Hubble Space Telescope, HST) with the JWST will hopefully address this pertinent debate (Wang et al. 2019; Barrufet et al. 2021, 2023; Long et al. 2022).

Quiescent populations are commonly selected using rest-frame color–color diagrams, with an increasing number of studies further analyzing the correlations of physical properties relative to the sample location within these diagrams (e.g., Whitaker et al. 2012; Belli et al. 2019). The four most popular in use are the *UVJ* selection criteria (Wuyts et al. 2008; Williams et al. 2009) or variations thereof (e.g., Whitaker et al. 2012), *BzK* (Daddi et al. 2004), and *NUVrK* or *NUVrJ* (Arnouts et al. 2013; Ilbert et al. 2013; Davidzon et al. 2017; Noirot et al. 2022). Rest-frame color–color diagrams have been favored for their quick and easy use with large photometric data sets; galaxies tend to occupy distinct parameter spaces within these diagrams, making selection for quiescent populations moderately straightforward. Critically, the use of two colors allows for the separation of quiescent from star-forming galaxies, with the $U - V$ or $NUV - r$ color separating star-forming galaxies from quiescent, and the $r - J$ or $V - J$ color separating red dusty galaxies from those that are quiescent.

These selections were originally designed to separate galaxy populations at low redshift, but as newer data have emerged demonstrating that the bimodality in populations clearly extends to higher redshift (e.g., Whitaker et al. 2011; Ilbert et al. 2013; Muzzin et al. 2013; Straatman et al. 2014), the selections and variations thereof are now often adopted at $z > 2$ (e.g., Whitaker et al. 2012; Straatman et al. 2014; Hwang et al. 2021; Ji & Giavalisco 2022; Suzuki et al. 2022). In sum, rest-frame color–color diagrams are ideal to select quiescent populations as they are computationally cheap and

straightforward. They can also be model dependent, but this depends on the way in which the rest-frame colors are calculated.

However, the application of rest-frame color diagram selection boxes to higher-redshift samples (in particular, $z > 3$) should be cautioned against. Several studies (Leja et al. 2019b; Belli et al. 2019; D’Eugenio et al. 2021) find that 10%–30% of *UVJ*-selected QGs at higher redshift are either low-redshift interlopers or dusty star-forming galaxies. Lustig et al. (2023) show that this contamination occurs more frequently in the *UVJ* area populated by older QGs, at the top-right corner of the quiescent selection box. Moreover, despite sSFR decreasing with redder $U - V$ and bluer $V - J$ colors, Leja et al. (2019b) demonstrate that a *UVJ* selection cannot differentiate between moderate and low sSFRs. Such a distinction in sSFR is critical at high redshift, where recently quenched galaxies, or those in the process of quenching, are more prevalent. Furthermore, Antwi-Danso et al. (2023) show that the extrapolation of rest-frame J (which can be common at $z > 3$) leads to galaxies being wrongly classified.

While rest-frame color diagram selections remain useful, particularly for large ground-based surveys, the current methods explored above are less reliable at $z > 3$ for several reasons. First, many studies clearly show that $3 < z < 5$ is the epoch where QGs first appear (e.g., Marsan et al. 2022; Valentino et al. 2020), and so the galaxy population is not yet bimodal. Instead, as Marsan et al. (2022) find, massive galaxies in this epoch exhibit a wide spectrum of properties, with varying amounts of dust and star formation, as well as some harboring an active galactic nucleus (AGN; see also Marsan et al. 2015, 2017 and Ito et al. 2022). It therefore does not make sense anymore to apply a color–color cut that was designed to separate galaxies in a bimodal universe when that bimodality does not yet exist. Second, it takes on the order of a billion years for galaxies to evolve and become truly red (without dust), which implies that we would not expect to find the upper edge of the *UVJ* quiescent box highly populated at $z > 3$. Instead, it is more likely that galaxies found in this parameter space are instead dusty star-forming galaxies that have scattered in over the $V - J < 1.5$ line. This is seen clearly in Lustig et al. (2023), who compare observations of QGs at $z \sim 3$ with several simulations. They find that the contamination of the *UVJ* quiescent box at $z \sim 2.7$ is most significant at the reddest edge of the box, near the $V - J < 1.5$ line. To remedy this, they suggest changing the $V - J$ constraint from $V - J < 1.5$ to $V - J < 1$, which effectively cuts the selection area in half.

The third reason is that galaxies in the high-redshift universe are faint, and thus even state-of-the-art ground-based surveys will have fairly large photometric errors for most galaxies which propagate to uncertainties a factor of 2 larger in colors (see Appendix C in Whitaker et al. 2010). This is further antagonized by the fact that surveys such as the Cosmological Evolution Survey (COSMOS) are limited to the resolution of the now decommissioned Infra Red Array Camera (IRAC) on the Spitzer Space Telescope, which, besides having a large point-spread function (PSF), also barely probes the Balmer break at $z > 4$ for QGs. The use of rest-frame $V - J$ color as a separator between dusty star-forming galaxies and red QGs becomes highly uncertain at $z > 5$, where rest-frame J moves beyond $\sim 9 \mu\text{m}$ (IRAC/Channel 4) and must therefore be extrapolated. For surveys limited to $\sim 5 \mu\text{m}$ (Channel 2), this

becomes $z > 3$. All of these reasons together imply that simply drawing a dividing line over a nonbimodal, noisy population will result in both incomplete and quite significantly contaminated samples.

Given the arguments against them, one might reconsider the usefulness of rest-frame color–color selections when there are multiple options for spectral energy distribution (SED) fitting tools to calculate sSFRs which can be used to select for QGs. While this alternative is often used (e.g., Carnall et al. 2020; Carnall et al. 2022; Ito et al. 2022), more complex methods may require careful tuning, time, and sometimes vast computing resources. Finally, this still does not solve the problem because these approaches require a sSFR cut or otherwise to select QGs in a population where the lines between star-forming and quiescent are blurred. While these methods offer opportunities to study the statistical properties and their stellar assembly, the selection of QGs should still be possible with simpler methods. It is therefore important that we take care in optimizing the rest-frame color–color selections as they are fast and can easily be applied to different surveys irrespective of the filters used, therefore serving as crucial tools for the selection of candidates for spectroscopic follow-up with instruments such as JWST. This is something that the community has already begun to explore, e.g., Leja et al. (2019b; efficient selection of QGs at $z < 1$) and Antwi-Danso et al. (2023; efficient selection of QGs at $z > 3$ using synthetic ugi_s filters).

In this paper, we present a new rest-frame color–color diagram and selection method specifically designed to find and select high-confidence massive QGs at $z \sim 3$ and beyond. Using the latest COSMOS catalog, COSMOS2020, we explore the utility of this new selection technique. In Sections 2 and 3, we describe the data and a modified COSMOS2020 catalog made specific to this study. In Section 4, we describe the selection of a robust sample of massive galaxies at $3 < z < 5$. In Section 5, we introduce the new color-selection method. In Section 6, we present this new selection applied to COSMOS2020 and the main results. Finally, we summarize our conclusions and outlook in Section 10. For all calculations, we use the WMAP9 flat Lambda cold dark matter cosmology (Hinshaw et al. 2013) with $H_0 = 69.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $\Omega_m = 0.307$. All magnitudes are in the AB system defined by Oke & Gunn (1983) as $m_{AB} = -2.5 \log_{10} \left(\frac{f_\nu}{3631 \text{ Jy}} \right)$.

2. Data

2.1. COSMOS2020

The Cosmological Evolution Survey is currently the deepest NIR wide-field multiwavelength survey that provides data over 2 square degrees of the sky (Koekemoer et al. 2007; Scoville et al. 2007). The latest version of the COSMOS photometric catalog, COSMOS2020 (Weaver et al. 2022), improves on previous versions (Laigle et al. 2016) in many ways. First, with the improved depth (~ 1 mag) and uniformity in the u -band data, taken from the CFHT Large Area U -band Deep Survey2 (Sawicki et al. 2019); second, with the addition of ultra-deep optical data from Hyper-Suprime-Cam (HSC; Aihara et al. 2018; Miyazaki et al. 2018); third, the fourth data release (DR4) of Ultra-VISTA (UVISTA; McCracken et al. 2012; Moneti et al. 2022), effectively 1 mag deeper in K_s over the entire field; and, finally, the inclusion of all Spitzer/IRAC data ever taken in the COSMOS field (Weaver et al. 2022;

Moneti et al. 2022). COSMOS2020 contains four catalogs, combining two source-extraction methods and two photometric redshift codes. The CLASSIC catalog is produced using a classical aperture-based photometry method using SOURCE EXTRACTOR (Bertin & Arnouts 1996), while THE FARMER catalog is produced using a profile-fitting photometry method (J. Weaver et al. 2023, in preparation), which is a wrapper written for the source profile-fitting code THE TRACTOR (Lang et al. 2016). THE FARMER has a distinct advantage over classical photometry methods due to its ability to correctly model and deblend overlapping sources. This is particularly beneficial in deep fields where source crowding and overlapping is common. Both photometry catalogs are fit with two different photometric redshift codes, Le PHARE (Arnouts & Ilbert 2011) and `eazy-py` (Brammer et al. 2008; Brammer 2021). Weaver et al. (2022) show that THE FARMER performs equivalently to SOURCE EXTRACTOR and excels at $i < 24$, independent of the redshift code used. For this analysis, we use THE FARMER photometric catalog combined with redshift and physical parameter estimation using `eazy-py`.

Although THE FARMER photometry and its associated uncertainties are suitable for general use SED fitting, Weaver et al. (2022) noted that it is likely the photometric uncertainties are underestimated. To test this, we ran the custom code GOLFIR (Kokorev et al. 2022) on a small subset of the COSMOS2020 IRAC images. GOLFIR uses a prior based on the highest-available-resolution imaging to model the flux in the IRAC photometry. Overall, the agreement of both photometric catalogs is high, however the IRAC errors derived with THE FARMER are smaller by a factor of ~ 5 . This is similar to the approach and conclusions in Valentino et al. (2022). At $z > 3$, the IRAC photometry plays a crucial role in the identification of massive galaxies. This is due to the fact that the light emitted from the bulk of the stellar mass is measured at observed frame $\lambda_{\text{obs}} > 2.5 \mu\text{m}$, which is measured predominantly by the IRAC bands. Although the depths probed by the first two IRAC bands, Channels 1 and 2, are comparable to those in UVISTA bands JHK_s (~ 25.5 mag, 3σ), IRAC Channels 3 and 4 are shallower by ~ 3 mags. For massive QGs at $z > 3$ this may result in the IRAC bands approaching the detection limit of the survey. For this reason and the investigations explained above, for this work we conservatively boost the IRAC photometric errors by a factor of 5 and refit the COSMOS2020 FARMER photometry with `eazy-py` (see Section 3). As with Weaver et al. (2022), we do not include the Subaru Suprime-Cam broadband photometry or the GALEX FUV and NUV photometry, due to their shallow depth and broad PSF, respectively. In total, we use 27 photometric bands for SED fitting, probing an observed frame $\sim 0.3\text{--}8 \mu\text{m}$.

2.2. 3D-HST

In Section 3, we will introduce the latest python implementation of `eazy`, `eazy-py`, and its newer features. To demonstrate their validity, in particular the computation of physical properties such as stellar mass and star formation rate (SFR), we use an observational catalog with similar properties as COSMOS2020 to benchmark the performance. We consider the PROSPECTOR version of the 3D-HST catalog from Leja et al. (2019d). The 3D-HST survey (Skelton et al. 2014) is a 248-orbit Hubble Treasury Program completed in 2015, providing Wide Field Camera 3 and Advanced Camera for

Surveys (ACS) spectroscopy for galaxies in five extragalactic deep fields, including COSMOS, covering 900 square arcminutes (199 square arcminutes in COSMOS). The catalog includes photometry in up to 44 bands in COSMOS covering 0.3–8 μm in the observed frame, with the inclusion of Spitzer Multi-Band Imaging Photometer 24 μm photometry from Whitaker et al. (2014). The version that we considered contains a representative subsample of 58,461 galaxies from the 3D-HST survey (roughly 33%), which spans the observed SFR density and mass density at $0.5 < z < 2.5$. These galaxies are fit with the Bayesian SED-fitting code PROSPECTOR (Leja et al. 2017; Johnson et al. 2021). PROSPECTOR uses galaxy models generated using python-Flexible Stellar Population Synthesis (FSPS; Conroy et al. 2009; Conroy & Gunn 2010) code, nonparametric star formation histories, variable stellar metallicities, and a two-component dust-attenuation model, under the assumption of energy balance. For this paper, we cross-match our catalog and the 3D-HST PROSPECTOR catalog using a $0''6$ matching radius, resulting in 8964 galaxies in common.

2.3. Simulated Data

To test the robustness of our QG selection method, we use simulated galaxies from the SHARK semi-analytical model (SAM; Lagos et al. 2018). SHARK models the formation and evolution of galaxies following many physical processes, including star formation, chemical enrichment, gas cooling, and feedback from stars and AGNs, among others. Critically for this work, SHARK includes a two-phase dust model for light attenuation that is based on radiative-transfer results from hydrodynamical simulations (Trayford et al. 2020). This model was combined with the stellar population synthesis model of Bruzual & Charlot (2003), adopting a Chabrier (2003) initial mass function (IMF), and a model for reemission in the IR in Lagos et al. (2019). SHARK does this using the generative SED mode of the PROSPECT SED code of Robotham et al. (2020). The predicted SED of each galaxy is then convolved with a series of filters to get their broadband photometry. Lagos et al. (2019) showed that SHARK is very successful in reproducing the far-UV to far-IR emission of galaxies in observations at a wide redshift range, including luminosity functions, number counts, and cosmic SED emission. Lagos et al. (2020) used this model to show that the contamination fraction of dusty, star-forming galaxies (those with 870 μm flux > 0.1 mJy) in the *UVJ* diagram increases with increasing redshift, reaching $\approx 47\%$ at $z=3$, showing that the commonly adopted *UVJ* color selection could be improved to reduce contamination. Here, we use a light cone presented in Lagos et al. (2019, their Section 5) that has an area of $\approx 107 \text{ deg}^2$, and includes all galaxies in the redshift range $0 \leq z \leq 8$ that have a dummy magnitude, computed assuming a stellar mass-to-light ratio of 1, < 32 . The method used to construct light cones is described in Chauhan et al. (2019).

3. Spectral Energy Distribution Fitting with *eazy-py*

3.1. *eazy-py*

In this study, we use photometric redshifts, stellar masses, SFRs, and rest-frame colors derived with *eazy-py*¹³ (Brammer 2021). Here, we detail this process, including an updated description and presentation of the *eazy-py* code.

eazy-py is a pythonic photometric redshift code adapted from the original *eazy* code written in C+ (Brammer et al. 2008). *eazy-py* estimates redshifts by fitting nonnegative linear combinations of galaxy templates to broadband photometry to produce a best-fit model. The flexibility of the code means that, in theory, any kind of template can be used, e.g., quasar or AGN templates. The templates used in this paper are comprised of 13 stellar population synthesis templates from FSPS (Conroy et al. 2009; Conroy & Gunn 2010) spanning a wide range in age, dust attenuation, and log-normal star formation histories. This template set is designed to encompass the expected distribution of galaxies in *UVJ* color space at $0 < z < 3$, and in theory beyond. We use a Chabrier IMF (Chabrier 2003) and the Kriek & Conroy (2013) two-component dust-attenuation law, which allows for a variable UV slope (dust index $\delta = -0.1$, $RV = 4.05$).

The latest version of *eazy-py* (version 0.5.2) has been adapted to be a general use SED-fitting code, providing estimates of physical parameters such as stellar mass, SFR, and dust attenuation. These physical properties are assigned to each template and “propagate through” the fitting process with the same nonnegative linear combination coefficients that return the best-fit SED model. This results in star formation histories that are nonnegative linear combinations of a number of basis star formation histories, and therefore the star formation history of a galaxy best-fit model will mimic nonparametric methods such as those used in Pacifici et al. (2016), Leja et al. (2019a), Iyer et al. (2019), and other works. This means that *eazy-py* is more similar to codes that implement nonparametric star formation histories than it is to traditional SED-fitting codes, and can account for the mass that is often missed as a result of using parametric star formation histories (Lower et al. 2020). We demonstrate this in Section 3.1.2. Errors on physical parameters are calculated by drawing 100 fits from the best-fit template error function and using the 16th and 84th percentiles as the extremes of the 1σ errors. It should be noted that the physical parameter errors are not marginalized over the redshift error, therefore it is likely that the errors from *eazy-py* are underestimated. Figure 1 shows a comparison of photometric and spectroscopic redshifts for all galaxies at $0 < z < 6$ in our catalog. We find similar outlier fractions and bias for both star-forming and QGs compared to the official COSMOS2020 release (Weaver et al. 2022).

3.1.1. Rest-frame Flux Densities

The rest-frame fluxes are calculated based on the approach of Brammer et al. (2011), which uses the best-fit model as a guide to interpolate between the observed photometry. This interpolation itself is weighted by the photometric errors, and therefore the rest-frame flux errors reflect those of the photometry. This means that the rest-frame fluxes are almost entirely derived from the photometry with only partial guidance from the best-fit template. The results also account for the filter shapes and relative depths, which is advantageous particularly for multi-instrument photometric catalogs. This essentially ensures a “model-free” approach, which is crucial for our method, as rest-frame colors are used in the selection process, and should ideally reflect the observed universe and not our models. For a full description of calculation of the rest-frame fluxes, we refer the reader to Appendix A. The rest-frame filters used in this paper are the GALEX *NUV* band ($\lambda = 2800\text{\AA}$; Martin et al. 2005), the recalibrated *U* and *V* filters by Maíz

¹³ <https://github.com/gbrammer/eazy-py>

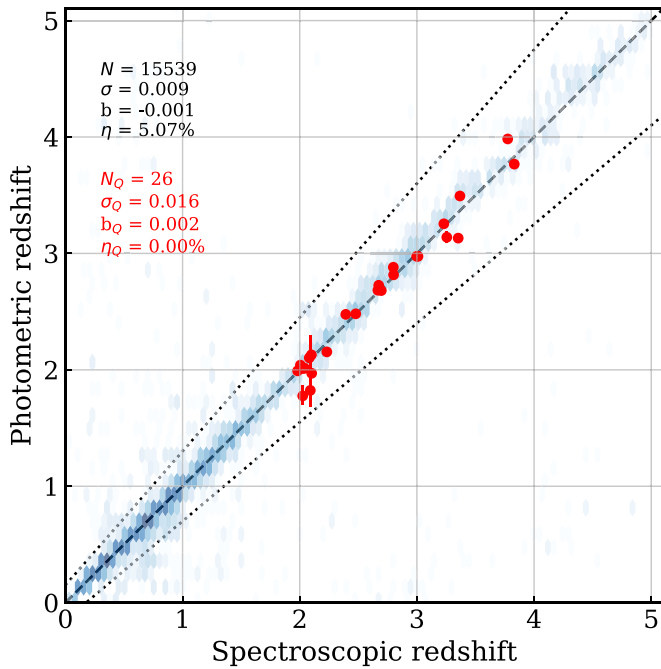


Figure 1. Photometric redshift vs. spectroscopic redshift for all galaxies at $0 < z < 5$ in the combined UVISTA and HSC area with a K_s S/N > 5 . Red points are spectroscopically confirmed QGs at $z > 2$ from Schreiber et al. (2018), Stockmann et al. (2019), D’Eugenio et al. (2021), Forrest et al. (2020b), and Valentino et al. (2020).

Apellániz (2006), and the Two Micron All Sky Survey J band (Skrutskie et al. 2006).

3.1.2. Stellar Masses and Star Formation Rates

In this section, we explore the stellar masses (M_*) and SFRs derived with `eazy-py`. We note that it has been shown that SFRs derived from broadband photometric SED fitting assuming parametric star formation histories can underestimate the SFR of a galaxy by several orders of magnitude (Lower et al. 2020). Sherman et al. (2020), who used `eazy-py` with default FSPS templates using a variety of realistic star formation histories (rising, bursty), find that M_* are overestimated on average by 0.085 dex, and SFRs are underestimated on average by ~ 0.46 dex. As we fit with a different, newer set of templates using log-normal star formation histories, we cannot assume the derived masses and SFRs have from the same offsets.

To benchmark the `eazy-py` M_* and SFRs in this study, we use the 3D-HST catalog described in Section 2.2. From the PROSPECTOR 3D-HST catalog we use redshift (which are a compilation of spectroscopic redshifts, grism redshifts, and redshifts derived using `eazy`), stellar masses, and SFRs, which are averaged over the past 100 million years. Here, we use the masses and SFRs derived with PROSPECTOR as the “ground truth.” We select all galaxies with a photometric redshift agreeing within 15% of $\Delta z/(1+z)$ to the PROSPECTOR redshift, which results in 8134 galaxies spanning $0.5 < z < 3.0$. We then compute the difference (in dex) between our parameters (SFR, M_*) and the PROSPECTOR parameters. Due to differences in method, the K_s -band magnitudes for COSMOS2020 are marginally brighter (~ 0.09 dex) than those in the Skelton et al. (2014) catalog that was used in Leja et al. (2019a). To facilitate a fairer comparison, we scale the PROSPECTOR masses and SFRs accordingly. In practice, this

results in a very small difference. We show both the SFR and mass comparisons in Figure 2 as a function of the estimates derived with PROSPECTOR. We calculate the median offset and scatter, and find that there is excellent agreement between the two catalogs, with only a minor overestimation of 0.04 dex for masses and 0.07 dex for SFRs by `eazy-py` with a scatter of 0.21 dex for mass and 0.55 dex for SFRs. The difference between the results of this investigation compared to Sherman et al. (2020) is notably that of the SFRs: while they find an underestimation of ~ 0.5 dex, we report a slight overestimation. This is likely due to the different template sets used. In conclusion, given the agreement and no catastrophic bias in stellar masses or SFRs derived with `eazy-py`, we deem them acceptable for use.

4. Selecting a Robust Sample of Massive Galaxies

We begin by selecting all objects within the area covered by both HSC and UVISTA (i.e., FLAG COMBINED = 0, area $\approx 1.27^{\circ 2}$), classified as galaxies according to the star-galaxy separation criteria (see Appendix B), and with $K_s < 25.6$ mag, which is the 3σ limiting magnitude. From these, we select all galaxies with photometric redshifts $3.0 < z_{\text{phot}} < 5.0$, with a constraint that the lower 16% of the photometric redshift probability distribution, $p(z)$, should be contained above $z = 3$, and the upper 84% below $z = 5$. If a spectroscopic redshift is available in the catalog we require $3 < z_{\text{spec}} < 5$. At $4.5 < z < 5$, the rest-frame J band is covered by IRAC Channels 3 and 4, which means that it is still observationally constrained and not extrapolated (see Appendix C). This is vital to ensure the robustness of the sample (Antwi-Danso et al. 2023). For this reason, we do not extend beyond $z = 5.0$ in our sample selection. We next make a cut on stellar masses, $10.6 < \log(M_*/M_\odot) < 12$, which is the knee of the mass function at this redshift range and well above the stellar mass completeness (Weaver et al. 2022). However, it is important to note that at $4.5 < z < 5$ the Balmer/4000 Å break moves through the K_s band, which is the reddest band in the chi-mean $izYJHK_s$ detection image used in the photometric modeling. Therefore, our selection is likely not sensitive to older QGs at this redshift range, if they exist. To our sample, we then add the following requirements to ensure the confidence in our massive galaxy sample is high. We choose to conservatively exclude anything with a reduced χ^2 equal to $\chi^2/(N_{\text{fit}})$ greater than 10. We also require a K_s signal-to-noise ratio (S/N) > 5 . In summary, we require the following:

1. To be located inside the combined HSC/UVISTA area.
2. To be classified as a galaxy.
3. K_s mag < 25.6 .
4. $3.0 < z_{\text{phot}} < 5.0$ (or $3.0 < z_{\text{spec}} < 5.0$ if available).
5. $z_{\text{phot},16\%} > 3.0$ and $z_{\text{phot},84\%} < 5.0$.
6. $10.6 < \log(M_*/M_\odot) < 12$.
7. Reduced $\chi^2 < 10$.
8. K_s S/N > 5 .

In total, this selection results in 1568 galaxies with a median redshift of $z_{\text{phot}} = 3.55$ and a median stellar mass of $\log(M_*/M_\odot) = 10.94$. In addition to this, we select a sample at $2 < z < 3$ in a similar manner with a lower mass cut of $10 < \log(M_*/M_\odot) < 12$, in order to fit the Gaussian mixture model (GMM) described in the following section. This gives a fitting sample of 12,985 galaxies at $2 < z < 3$ with a median

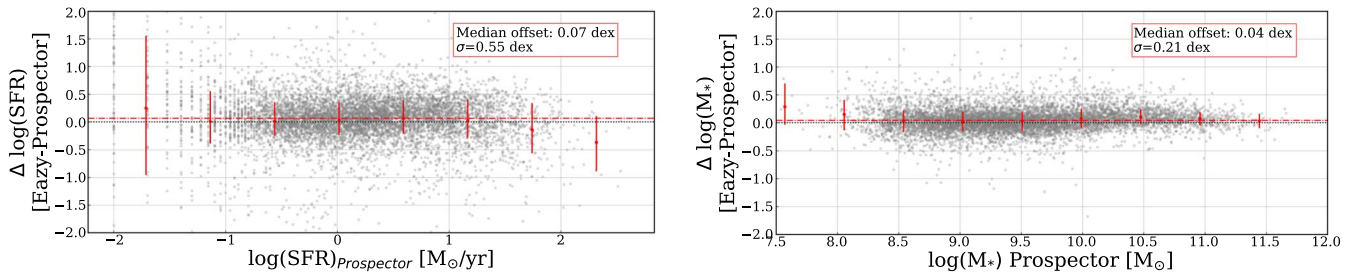


Figure 2. Left: $\Delta \log(\text{SFR})$ (EAZY-PROSPECTOR) as a function of PROSPECTOR SFR for 8134 galaxies at $0.5 < z < 3.0$ matched to our catalog with that of the 3D-HST survey in COSMOS. The median $\Delta \log(\text{SFR})$ is 0.07 dex (shown as red dashed line) and the scatter is 0.55 dex. The binned median offset and the associated 16th and 84th percentiles are shown as red points with error bars. Right: $\Delta \log(M_*)$ (EAZY-PROSPECTOR) as a function of PROSPECTOR stellar mass for the same sample. The median $\Delta \log(M_*)$ is 0.04 dex and the scatter is 0.21 dex. The binned mean offset and the associated 16th and 84th percentiles are shown as red points with error bars.

photometric redshift of $z_{\text{phot}} = 2.46$ and median stellar mass of $\log(M_*/M_\odot) = 10.52$.

5. Quiescent Galaxy Selection Method

5.1. Introducing the Gaussian Mixture Model

We present here a new color selection designed to make finding QGs at $z > 3$ easier. Our new color-selection method combines the $NUV - U$, $U - V$, and $V - J$ colors, using the filters defined in Section 3.1.1. This color-selection method is similar to both UVJ and $NUVrJ$, with two major differences. The first is that four bands (three colors) are used instead of three, and the second is that the separation itself is not defined by a traditional “box” but rather using simple machine-learning methods. Both are introduced to make the separation of QGs from (dusty) star-forming more pronounced, and furthermore to allow an easier separation between recently quenched and older QGs. The $U - V$ and $V - J$ colors are included to separate QGs from star-forming, as originally designed. The $NUV - U$ color is added to measure the light emitted from recent star formation in the form of A-type stars and is similar to $NUV - r$, except that it probes a shorter baseline, specifically the flux directly bluewards of the 4000 \AA /Balmer break. Previous works have suggested $NUV - U$ is a viable indicator for time since quenching, e.g., Schawinski et al. (2014). Similarly, Phillipps et al. (2020) use $NUV - u$ as a means to remove passive galaxies with some residual star formation from a sample of truly passive galaxies. The reason for the use of four bands/three colors is twofold. First, the increase in dimensions allows us to extract more information while still only requiring four data points. While this is similar to dimensionality reduction, it has the advantage of using information that is likely already available or easy to compute, and does not require imposing a high-S/N cut to the data. Second, $NUV - U$ does not have a codependent variable with $V - J$, which allows for the spread of colors in $NUV - U/V - J$ to become more obvious, making it easier to both separate the quenched population from star-forming, as well as to explore the properties of galaxies within the quenched population with different amounts of recent star formation. This is particularly relevant at $z > 3$, where the fraction of PSB galaxies is high (e.g., D’Eugenio et al. 2020 and Lustig et al. 2021 report a PSB fraction of 60%–70% for photometrically selected QGs at $\log(M_*/M_\odot) > 11$). If we consider the 2D projection ($NUV - U$, $V - J$), it is clear that the star-forming and quiescent populations separate more clearly due to the divergent alignment of the dust vector relative to the evolution tracks (see Figure 3). As an example, we plot the

massive $3 < z < 5$ sample in this color space with tracks showing the color evolution of a single stellar population aged to 1 Gyr at both $A_V = 0$ and $A_V = 1$ generated using the Python-FSPS package. Both tracks move directly up through the quiescent area as they evolve over time. The lower portion of this area preferentially selects quenching or PSB galaxies, while the upper half encompasses older QGs. This age trend has been observed in the UVJ diagram (e.g., Whitaker et al. 2012, 2013; Belli et al. 2019; Carnall et al. 2019), thus it is not entirely surprising that it appears in this color diagram too—and in fact more pronounced due to the addition of the $NUV - U$ color. We note that the direction of the age tracks diverge from the direction of the dust vector for ($NUV - U$, $V - J$), resulting in red dusty star-forming galaxies being pushed downwards and to the left relative to old, red QGs, whereas for UVJ they follow the same direction, resulting in old, red QGs populating the same space as red dusty star-forming galaxies.

5.2. Defining the Separator

Most selections in rest-frame color diagrams have been made by drawing a box empirically. We choose to separate QGs from star-forming by using a GMM. This is a probabilistic model that operates under the assumption that the data can be fit by a finite number of Gaussian curves, for which the parameters are not known. GMMs have already been employed successfully in multiple research areas in astrophysics, including to select and explore various galaxy populations such as quiescent and PSB galaxies (e.g., Ardila et al. 2018; Black & Evrard 2022). Probabilistic selection of QGs at $z > 3$ has also been carried out with great success by Shahidi et al. (2020) and Santini et al. (2021). We use the GMM package supplied by scikit-learn (Pedregosa et al. 2011). Briefly, the GMM algorithm uses expectation maximization (EM), which is an iterative process used for classification when there are no “correct” labels (as is our case). EM works by choosing random components and iterating toward the best fit by computing the likelihood of each data point being drawn from the current model, and then adjusting the parameters to maximize that likelihood. In this approach, we choose to use all three colors: $NUV - U$, $U - V$, and $V - J$. We fit the model using the $2 < z < 3$ sample, which is first cleaned by requiring all colors have $0 < \text{color} < 4$. To ascertain the optimal number of components, we fit multiple times using 1–10 Gaussian components and calculate the Bayesian information criterion (BIC) for each number of

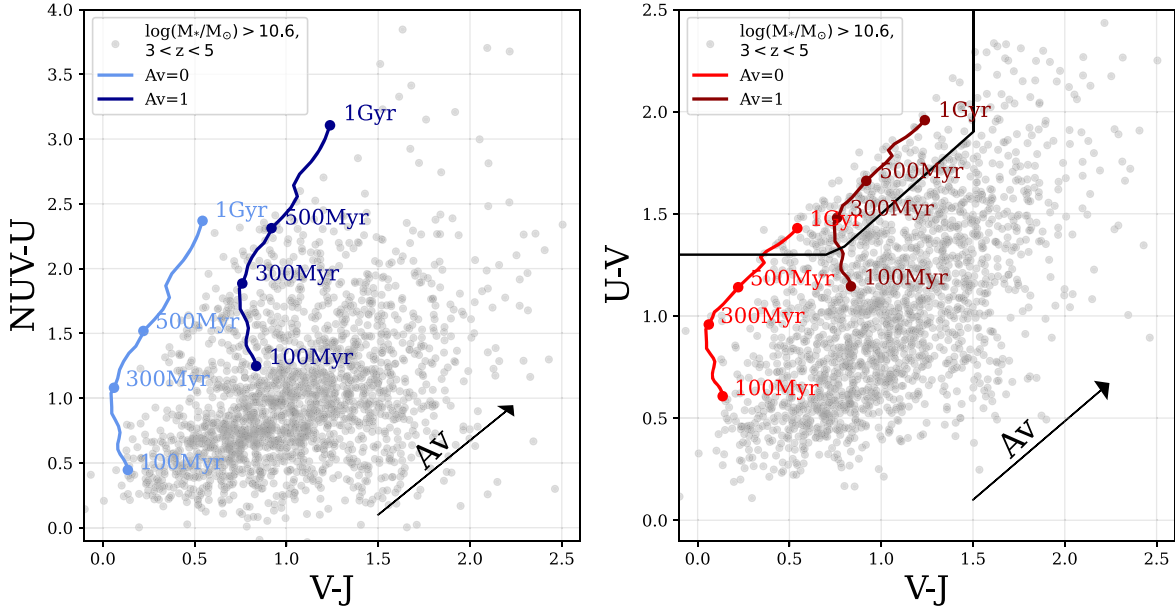


Figure 3. Left: $NUV - U$, $V - J$ colors of our massive ($10.6 < \log(M_*/M_\odot) < 12$) galaxy sample at $3 < z < 5$ (gray) and evolution tracks for an SSP with $A_V = 0$ (blue line on the left of the figure) and $A_V = 1$ (dark blue line) and Chabrier IMF aged to 1 Gyr. The dust vector representing the movement of a galaxy in this space by an attenuation of $A_V = 1$ is shown by the arrow at the bottom right. Right: same but for UVJ with the quiescent box defined by Whitaker et al. (2012; defined as $U - V > 0.8 \times (V - J) + 0.7$, $U - V > 1.3$ and $V - J < 1.5$).

components, which is defined as

$$\text{BIC} = k \ln(n) - 2 \ln(\mathcal{L}_{\max}), \quad (1)$$

where n is the number of model parameters, k is the number of data points fit, and \mathcal{L}_{\max} is the maximum likelihood of the fit. The BIC discriminates between models by penalizing the number of parameters, which avoids overfitting. We randomly select 60% of the sample, fit and compute the BIC for 1–10 components, and repeat 1000 times. We find that a six-component model is best fit in $\sim 75\%$ of the repeated fits and adopt this as our baseline model. The GMM returns for each galaxy a likelihood of belonging to each group, which we convert to a probability. We could simply use this value to decide what to classify as quiescent, however this does not account for the errors in a galaxy’s colors. To assign each galaxy a probability of belonging to the quiescent group, which is identified by eye as the component in the quiescent area of UVJ space, we instead take the following approach, inspired by Hwang et al. (2021). Each galaxy has rest-frame fluxes in the NUV , U , V , and J bands, and associated errors. Assuming these are Gaussian distributed, we bootstrap the fluxes 1000 times and compute the $NUV - U$, $V - J$, and $U - V$ colors for each set of rest-frame fluxes. Each set of 1000 colors is then fit with the GMM and the percentiles (5, 16, 50, 84, 95) of the probabilities are calculated. This results in a quiescent probability distribution $p(q)$ for each galaxy, which is marginalized over its rest-frame flux errors. Therefore, the final distribution of $p(q)$ for each galaxy accounts for flux errors. We refer mainly to $p(q)_{50}$ (abbreviated to $p(q)$) throughout the text, which is the median quiescent probability calculated in this way. Here, we use the GMM purely as a tool to classify QGs and do not explore the other five groups defined by the model. In Figure 4, we show the $2 < z < 3$ fitting sample and the $3 < z < 5$ sample in the $NUV - U$ vs $V - J$ plane, colored by their

quiescent probability $p(q)$. It is evident that the model correctly finds the boundary between quiescent and star-forming, but with a boundary that is a smooth gradient rather than a binary separation.

5.2.1. The $p(q)$ Threshold for Quiescence

To calculate the $p(q)$ threshold above which a galaxy is classed as quiescent and to assess the performance of the method, we use galaxies in the SHARK simulation (Section 2.3). At redshift snapshots of $z = 2$, $z = 3$, and $z = 4$, we select a massive galaxy sample ($\log(M_*/M_\odot) > 10$). We then define QGs in the sample as those with specific star formation rates $\log(\text{sSFR } t_{\text{Hyr}}^{-1}) < 0.2$, where t_{H} is the age of the universe at that redshift snapshot and the sSFR is averaged over the last ~ 50 Myr. We fit the GMM to the $NUV - U$, $V - J$, and $U - V$ colors at each redshift snapshot. We then calculate the receiver operator characteristic (ROC) curve, which is simply the false-positive rate (FPR) as a function of true-positive rate (TPR). The area under curve (AUC) can be used as a measure of effectiveness of the classification method: an AUC very close to unity shows that the method correctly classifies objects with a high TPR and low FPR. We also compute the same numbers for the Whitaker et al. (2012) UVJ selection method (defined as $U - V > 0.8 \times (V - J) + 0.7$, $U - V > 1.3$, and $V - J < 1.5$). Figure 5 shows both the $NUV - U$ vs $V - J$ and UVJ diagrams and the ROC curves for each redshift snapshot. The AUC increases from $z = 2$ to $z = 4$, meaning that the GMM has increasing chance of correctly identifying QGs (88%, 93%, and 97% at $z = 2, 3$, and 4, respectively). It is not possible to calculate the AUC for UVJ because there is only one selection “threshold,” which is whether or not a galaxy falls inside the UVJ quiescent area. We can instead calculate the TPR and FPR for UVJ . The optimal $p(q)$ threshold is defined as the maximum of the geometric mean, which finds the threshold at which the TPR and FPR are perfectly balanced (or the top-left point on the ROC curve), and is defined as

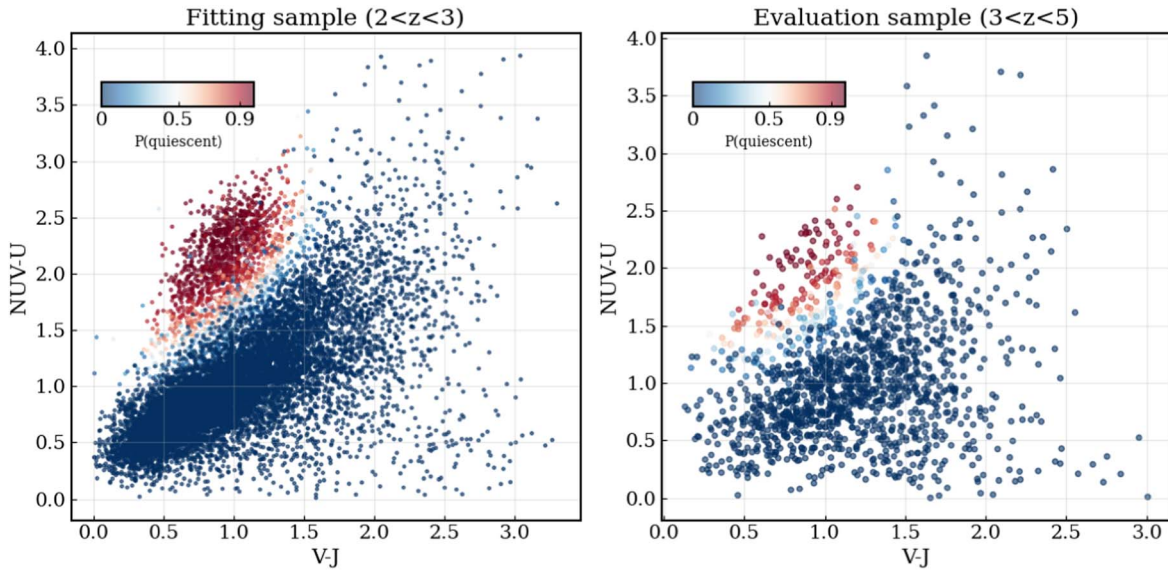


Figure 4. Left: $NUV - VJ$ diagram with the $2 < z < 3$ fitting sample colored by $p(q)_{50\%}$, which is the median of the bootstrapped quiescent probability distribution based on the GMM. Right: same but for the $3 < z < 5$ sample.

$\max(\sqrt{\text{TPR} \cdot (1 - \text{FPR})})$. The resulting $p(q)$ thresholds at $z = 2$, $z = 3$, and $z = 4$ are calculated using the geometric mean and are shown in Table 1.

5.2.2. UVJ versus Gaussian Mixture Model in SHARK

At $z = 2$, UVJ performs similarly to $NUV - VJ$ at $p(q) \gtrsim 0.1$, with UVJ having $\text{FPR} = 30\%$, $\text{TPR} = 90\%$ and $NUV - VJ$ with $\text{FPR} = 33\%$, $\text{TPR} = 89\%$. However, the optimal threshold for $NUV - VJ$ performs substantially better at this redshift, with a TPR only a few percent lower (81%), while the FPR is more than halved, reducing from 33% to 15%. If we consider instead the contamination, which is defined as the number of galaxies defined as quiescent compared to the total number of UVJ QGs, the conclusion changes. The contamination fraction of UVJ in SHARK at $z = 2$, 3, and 4 is 80, 86, and 91%, further highlighting the issues with UVJ . Looking to higher redshift, it is evident that $NUV - VJ$ /GMM classification not only outperforms UVJ classification but also substantially increases in effectiveness with increasing redshift, highlighting the usefulness of the method, especially at $z > 3$.

6. Application to COSMOS2020

We apply the GMM to the massive galaxy sample in COSMOS2020 at $2 < z < 5$ selected in Section 5. While SHARK produces galaxies at redshift snapshots of $z = 2$, 3, and 4, observed galaxies are continuously distributed in redshift. We therefore choose not to assume that the $p(q)$ threshold is a step function, but rather a smoothly varying function of redshift. We fit a second-order polynomial to the three points ($z_{\text{phot}}, p(q)$) = (2, 0.56; 3, 0.18; 4, 0.01) and use this function to determine if a galaxy is chosen as quiescent or not. This naturally incorporates the assumption that as we move to higher redshift, in order to be defined as quiescent, galaxies may look less like QGs at $z = 2$ and more like PSBs. We require that at $z = 2-4$ the median of the $p(q)$ distribution is above the threshold. At $z \geq 4$, we instead require that 95% of the $p(q)$ distribution is above $p(q) = 0.01$, in order to only select the highest-confidence candidates. The application of redshift-dependent $p(q)$ thresholds selects 1455 QGs at $2 < z < 3$ and

230 QGs at $3 < z < 5$. We visually inspected cutouts of the $3 < z < 5$ sample in the optical and NIR bands, and removed nine QG candidates which were spurious, reducing the total $3 < z < 5$ sample to 221 in number. In the following sections, we present the properties of our sample of QGs, compare its statistics with those of galaxies selected using traditional color diagrams, and discuss the differences between methods.

6.1. Comparing GMM/ $NUV - VJ$ to UVJ and $NUVrJ$ Color Selections

We apply both the Whitaker et al. (2012) UVJ selection as well as the Ilbert et al. (2013) $NUVrJ$ selection for quiescence to our massive galaxy samples at $2 < z < 3$ and $3 < z < 5$, and compare these traditional color selections to the GMM color-selection method. For each panel of Figure 6, we calculate the overlap between selections, which is defined as the sample size divided by the sample size of the “main” selection for that row, which is highlighted by the colored panel. At $2 < z < 3$, the overlap between all three selection methods is high: UVJ and $NUVrJ$ selections have an $\sim 85\%$ overlap, while the GMM-selected sample also agrees well, sharing $\sim 80\%$ – 90% with UVJ - and $NUVrJ$ -selected QGs. This would be expected if the rest-frame fluxes were measured directly from the template, but in this case they are weighted by the observed fluxes, and so the conclusion is that the different selections are generally measuring the same observed SED shape. Hwang et al. (2021), who compared $NUVrJ$ - and UVJ -selected samples at $0 < z < 3$ in COSMOS2015, found a similar overlap ($\sim 85\%$). This suggests that the choice of rest-frame color diagram selection at $z < 3$ is not crucial, although modifications to both UVJ and $NUVrJ$ color selections may be in order.

At $3 < z < 5$ the agreement between color selections is less clear: UVJ and $NUVrJ$ share 58%–77% of the same sample, while GMM shares 52%–79% with $NUVrJ$ and 66%–75% with UVJ . This difference is highlighted clearly in Figure 7. This is likely due to two reasons: first, that neither UVJ nor $NUVrJ$ selections include the region where PSB galaxies are most often found, which is at the lower edge beyond the $U - V < 1.3$ boundary. Second, the contamination fraction of UVJ and

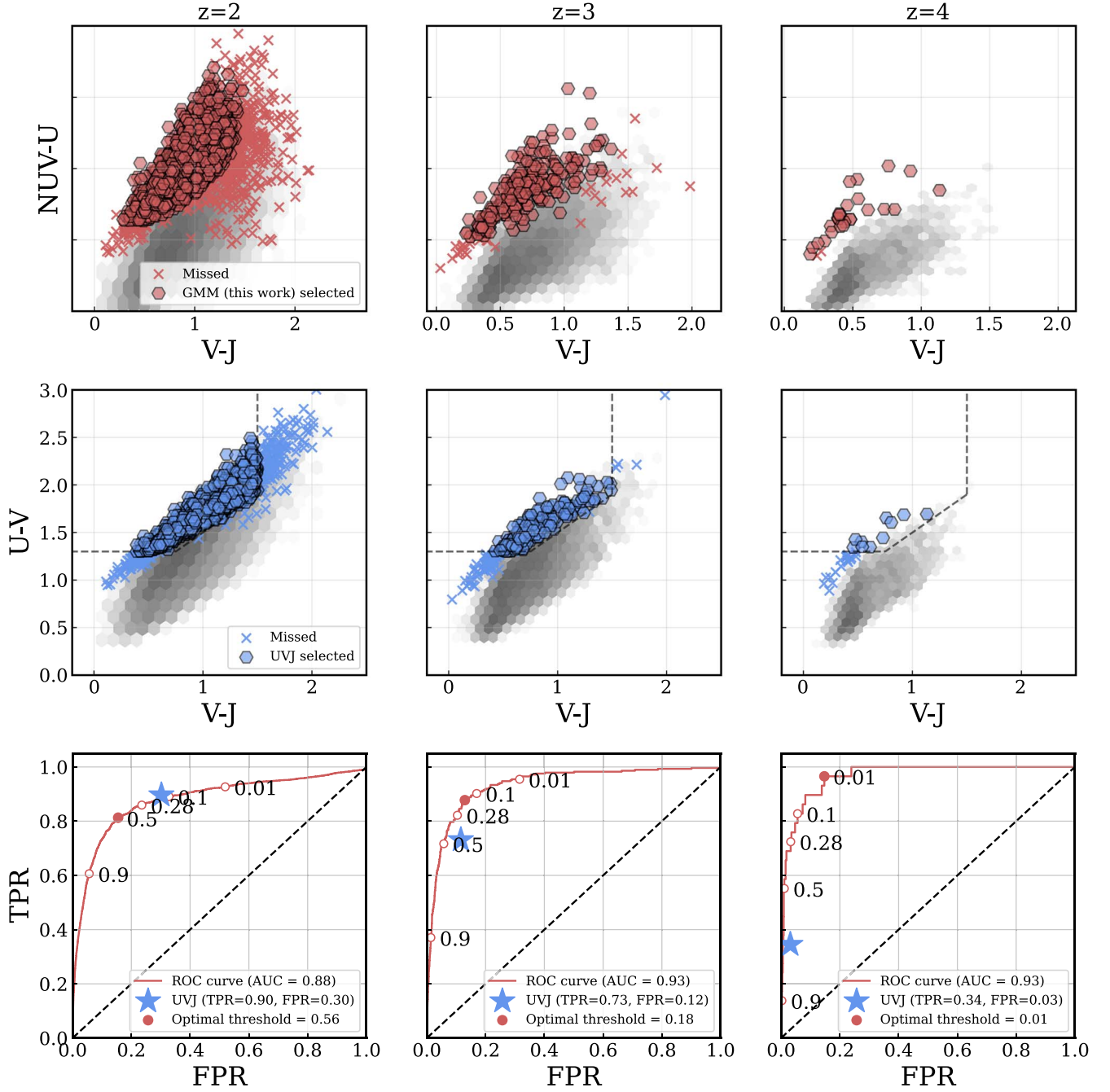


Figure 5. Left column: SHARK-simulated galaxies in different color planes. GMM/ $NUV - VJ$ (this work, top) and UVJ (middle) color diagrams for massive ($\log(M_*/M_\odot) > 10$) galaxies in the SHARK simulation (gray), with QGs found by each selection method as red (blue) hexagons and galaxies missed by both selection methods as red (blue) crosses. The bottom panel shows the receiver operator characteristic (ROC) curve for the trained GMM as a function of quiescent probability ($p(q)$) threshold. The area under the curve (AUC), which represents the percentage probability of correctly classifying QGs, is reported, as well as the $p(q)$ threshold, which is calculated as the geometric mean (shown on the figure as the colored red circle on the red curve). The false-positive rate (FPR) and true-positive rate (TPR) are also presented for the Whitaker et al. (2012) UVJ selection (blue star). Middle column: same as left column but at $z = 3$. Right column: same as left column but at $z = 4$.

Table 1

Quiescent Probability ($p(q)_{50}$) Thresholds for Redshift Ranges, Defined by the Maximum of the Geometric Mean of the ROC Curve Trained on SHARK-simulated Data

Redshift	$p(q)_{50} >$
$2.0 < z < 3.0$	0.56
$3.0 < z < 4.0$	0.18
$4.0 < z < 5.0$	0.01

$NUVrJ$ is likely much higher than the GMM selection method. While it is easy to remedy the first issue in both cases by simply lowering or removing the dividing line in $U - V/NUV - r$, as discussed above, this may also increase the contamination of the sample by introducing $U - V/NUV - r$ blue star-forming galaxies. This is something that Schreiber et al. (2018) confirmed in their analysis. For completeness, we report the performance of an extended UVJ selection benchmarked against SHARK in Appendix D. The issue of contamination is more difficult to solve due to the nature of these color

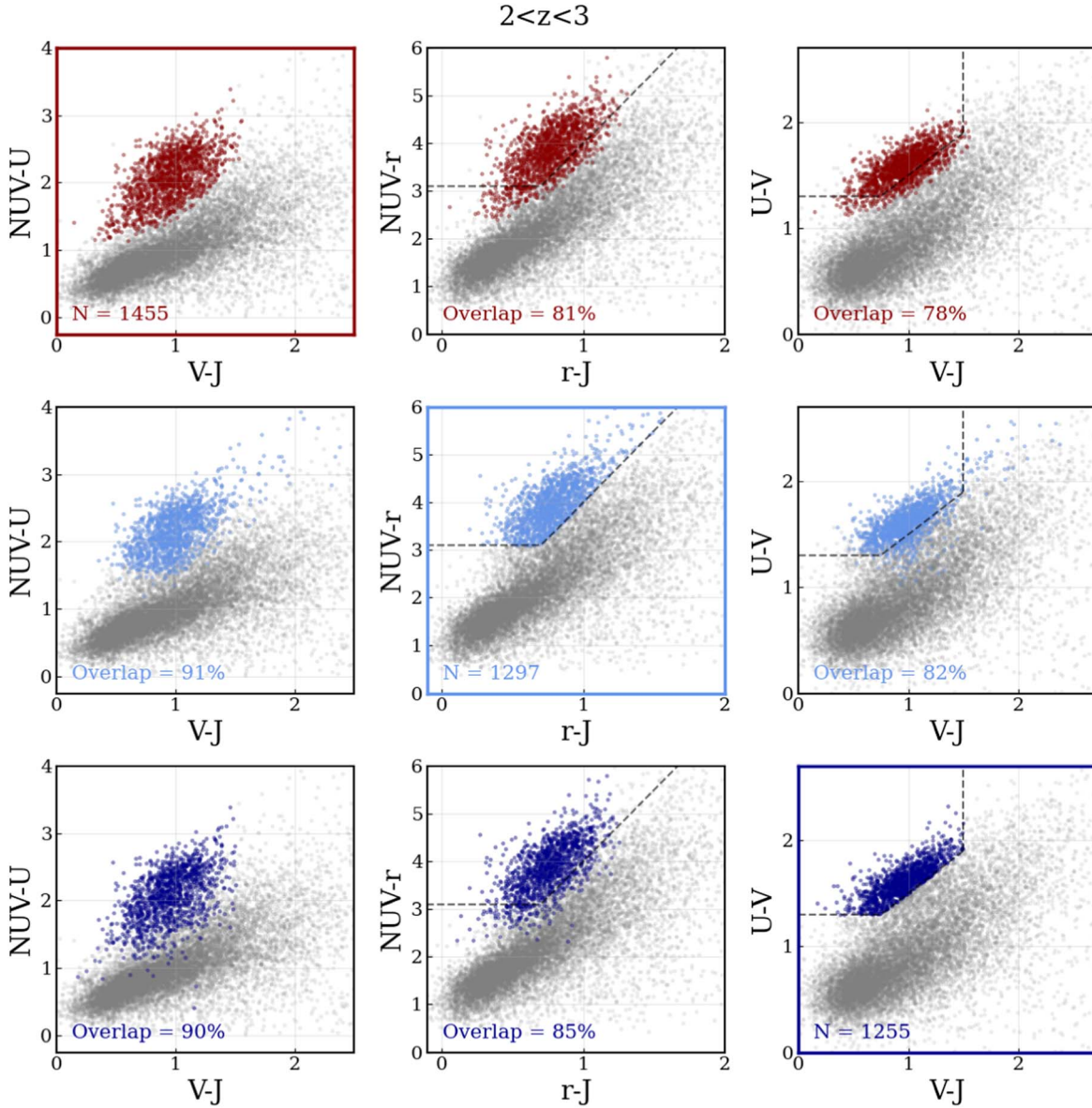


Figure 6. Top row: $NUV - VJ$ -selected QGs at $2 < z < 3$ (dark red, left figure) shown in $NUV - VJ$ space, UVJ space, and $NUVrJ$ space. Middle row: $NUVrJ$ -selected QGs (middle figure, light blue) shown in $NUV - VJ$ and UVJ color diagrams. Bottom row: UVJ -selected QGs (right figure, dark blue) shown in $NUV - VJ$ and $NUVrJ$ color diagrams. For each panel the overlap is given, defined as the number of galaxies selected by that selection method divided by the number of galaxies selected by the main selection method for that row. In the first row, we show the GMM-selected sample (dark red) at $2 < z < 3$ in the $NUV - VJ$ diagram, $NUVrJ$ diagram, and UVJ diagram. The second row shows the $NUVrJ$ -selected sample in $NUVrJ$ (light blue) as well as the other two color-color diagrams, and the third row shows the UVJ -selected sample in UVJ as well as in the other two color-color diagrams (dark blue).

selections. However, our method aims to alleviate this problem by both the introduction of the third $NUV - U$ color, and the option of choosing galaxies based on a probability distribution ideally makes it easier to sharpen the boundaries between different populations.

6.2. Full Spectral Energy Distribution

To offer a more complete view of the full SEDs of our selected QGs, in Figure 8 we show both the cutouts from optical/NIR bands and best-fit SEDs for three candidates at $z_{\text{phot}} = 3.12, 3.92,$ and 4.68 . We also compute the median rest-frame photometry and best-fit models for the whole GMM-selected sample normalized to $V = 1$ (Figure 9). The median values of the K_s observed magnitudes, M_* , SFRs, and sSFRs are also reported in the figure.

The galaxies in our sample show a strong Balmer/4000 Å break, indicating a dominant aged stellar population on average, as expected. However, they still have residual flux blueward of the break, indicative of a younger population and recent quenching. The median mass ($\log(M_*/M_\odot) = 10.92$) of this sample is ~ 0.3 dex higher than the mass cut for the original sample selection, confirming that only the most massive galaxies have quenched their star formation at $z > 3$. The median $\log(\text{sSFR}/\text{yr}^{-1}) = -10.8$ indicates that these galaxies are best fit mainly by templates that include little or no ongoing star formation.

6.3. Spectroscopically Confirmed Quiescent Galaxies

As a confidence check, we cross-matched our sample of candidate QGs with a literature compilation of seven

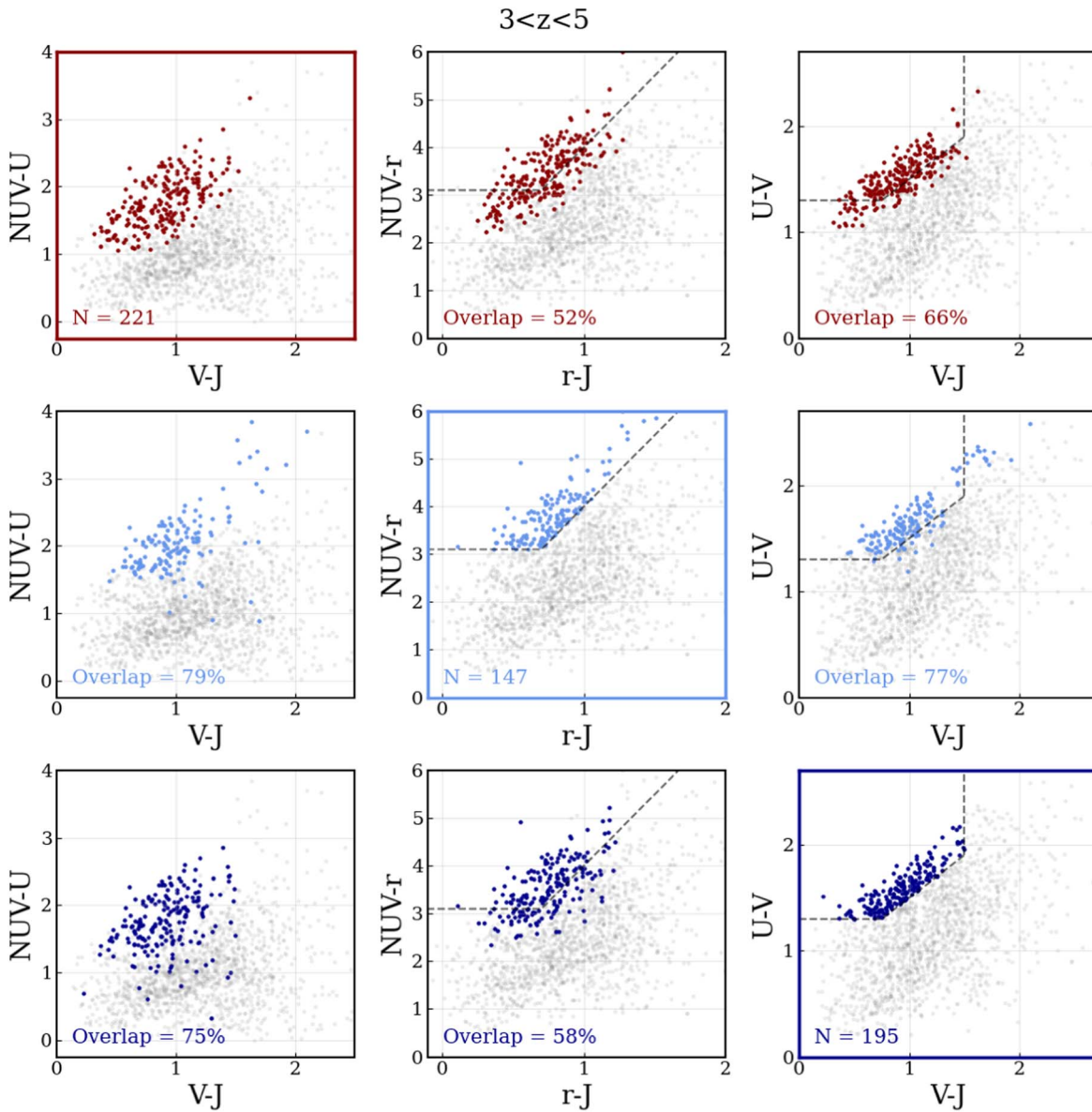


Figure 7. Same as Figure 6 but for $3 < z < 5$.

spectroscopically confirmed $z \gtrsim 3$ QGs in COSMOS from Forrest et al. (2020b, four sources), Forrest et al. (2020b)/Marsan et al. (2015)/Saracco et al. (2020, one source), Valentino et al. (2020, one source), and D’Eugenio et al. (2021, one source). For six out of seven sources we retrieve $p(q) \gtrsim 10\%$, consistent with being QG according to our selection. To the remaining galaxy at $z_{\text{spec}} = 3.352$ (Marsan et al. 2015; Forrest et al. 2020b; Saracco et al. 2020), we assign $p(q) = 0.6\%$. This source would thus not be selected using our fiducial threshold at $3 < z < 4$. We note that this galaxy has experienced rapid quenching, possibly due to an AGN. The presence of the latter is inferred from the large $[\text{OIII}]/\text{H}\beta$ emission line ratio, with the oxygen line possibly contaminating the K_s photometry (see the discussion in Forrest et al. 2020a).

7. Number Densities

The number densities of massive QGs at $z > 3$ is an important constraint on galaxy evolution and theory. The

assembly of galaxies with such high stellar masses in an evolved state within only 1.5–2 billion years not only places important constraints on the formation of the first galaxies (e.g., Steinhardt et al. 2016 and references within), but also on the cosmic SFR density (Merlin et al. 2019). As such, the number density of these galaxies provides important context to early galaxy evolution. However, the number densities of massive QGs at $z > 3$ has been an intense topic of debate, due to both the disagreement between observations and theory and between the observational studies themselves.

Number densities derived from photometric and spectroscopic observations span ~ 2 dex, but they generally remain higher than those extracted from simulations (Ilbert et al. 2013; Straatman et al. 2014; Davidzon et al. 2017; Merlin et al. 2018; Schreiber et al. 2018; Girelli et al. 2019; Merlin et al. 2019; Carnall et al. 2020; Santini et al. 2021; Shahidi et al. 2020; Valentino et al. 2020; Carnall et al. 2022; Valentino et al. 2023). The disagreement among observational works and with theoretical predictions is likely due to a mix of several factors, primarily different sample selections and quiescent criteria used

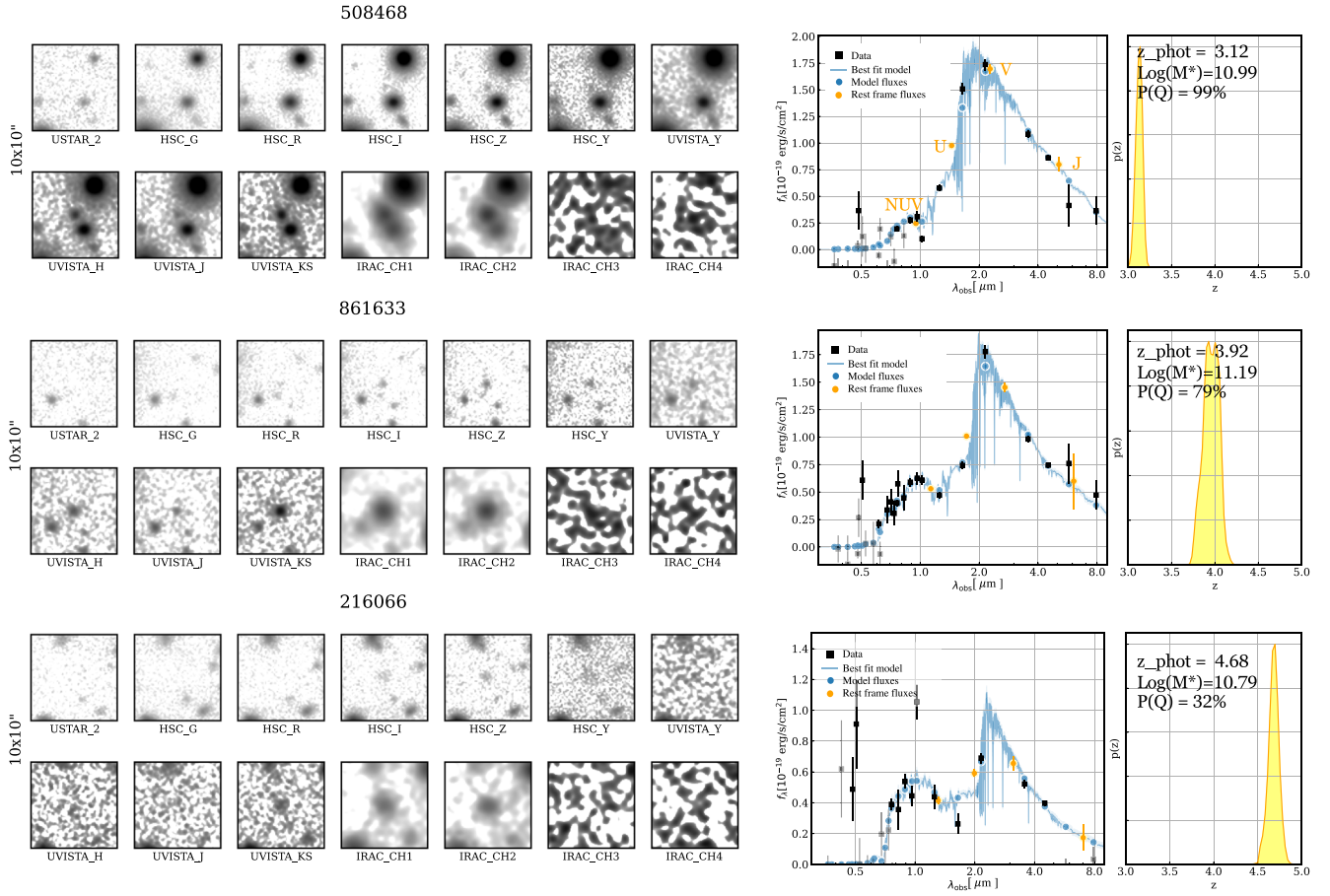


Figure 8. $10'' \times 10''$ postage-stamp cutouts from selected bands in the optical to NIR scaled by ± 3 rms (left) and corresponding best-fit SEDs (right) for three QG candidates selected using the $NUV - VJ$ method, taken from the main sample. The observed photometric points are shown as black circles (any with $S/N < 2$ in gray), and the best-fit model is shown in blue. The rest-frame NUV , U , V , J photometry is shown by the orange circles. The $p(z)$ is shown to the right of each SED in yellow, along with its best-fit photometric redshift, stellar mass, and quiescent probability median value ($P(q)_{50}$).

together with analyses done using multiple diverse data sets. Of particular note is the size of such fields, as the statistics can be affected by both the rarity of these galaxies and by cosmic variance. The latter is strongly mitigated by the large contiguous area of the COSMOS field. Here, we compute and present the number densities for our sample of QGs in three bins at $2 < z < 3$, $3 < z < 4$, and $4 < z < 5$ (Table 2). Number densities are calculated using the survey area of the combined HSC/UVISTA coverage, which corresponds to 1.27 square degrees. We calculate the fractional error due to cosmic variance using the method of Steinhardt et al. (2021), who extend the method of Moster et al. (2011) for use in the early universe (for details, see Steinhardt et al. 2021 and Weaver et al. 2022). This is combined in quadrature with the Poisson error to get the total error budget.

In Figure 10, we show our number densities along with a compilation of number densities from other observational studies (e.g., Valentino et al. 2023 and references therein). Additionally, we show QG number densities from the SHARK simulation at two mass ($\log(M_*/M_\odot) > [10, 10.6]$) and sSFR selection thresholds ($\log_{10}(\text{sSFR}/\text{Gyr}^{-1}) < [-1, -2]$), as well as both the FLARES ($z \geq 5$) and EAGLE ($z \leq 5$) simulations for QGs at ($\log(M_*/M_\odot) > 10.6$) at two different selections ($\log_{10}(\text{sSFR}/\text{Gyr}^{-1}) < [-1, -2]$) (Lovell et al. 2022). It is important to note that the literature compilation of QGs at $z > 3$ comprise selections at different mass limits ranging from

($\log(M_*/M_\odot) > 10$) (Carnall et al. 2020) upwards, while our mass limit is much higher, ($\log(M_*/M_\odot) > 10.6$). In general, our number densities at $3 < z < 4$ agree with the observational studies within 1σ errors. It is interesting to note that the number densities derived by Girelli et al. (2019) are much lower than those derived from COSMOS2020, which likely arises from the COSMOS2020 catalog including all UVISTA stripes, some of which have known overdensities (e.g., McConachie et al. 2022). Looking to higher redshift at $z > 4$, our estimates agree with those of Weaver et al. (2022) and Shahidi et al. (2020). Generally, it appears that number densities for massive QGs derived from a variety of different fields, selected with a variety of methods, are finally converging on agreement.

Previously, simulations were not able to reproduce the number densities of massive QGs at $z > 3$ observed in the real universe by a factor of 1–2 dex; this tension remains slightly with EAGLE, which predicts only upper limits for QGs at $\log(M_*/M_\odot) > 10.6$ that are several times lower than most observations. At $z = 5$, the results from FLARES are $\sim 50\times$ smaller than our current limits at fixed M_* threshold, resulting in a direct tension. At $z > 5$, this simulation predicts similarly low number densities, but the lack of data from observations at $z \gtrsim 5$ means that for the first time at high redshift simulations have QGs where observations have found none. SHARK performs the best at producing similar number densities of massive QGs to observations up to at least $z \sim 3$,

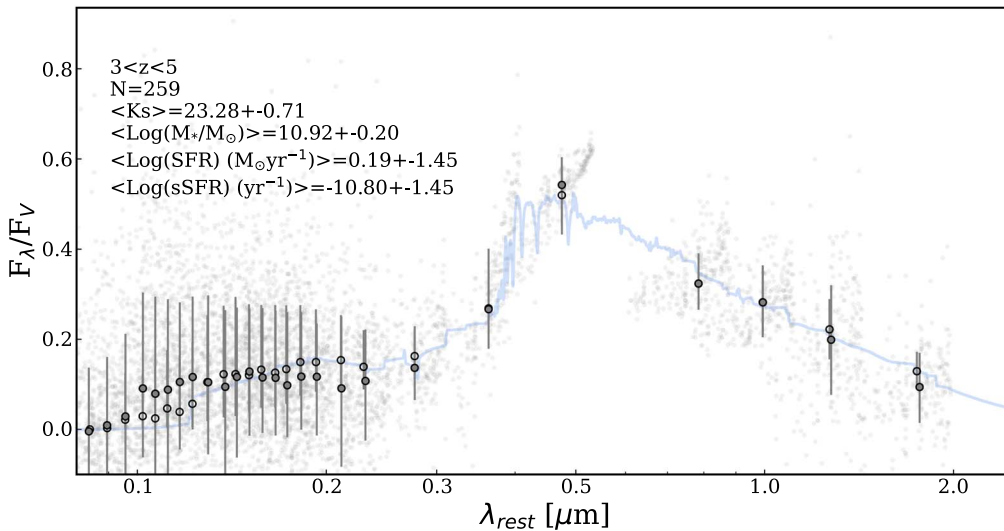


Figure 9. Top: median smoothed rest-frame SED for the GMM/ $NUV - VJ$ -selected quiescent sample at $3 < z < 5$. Rest-frame photometry for all galaxies is shown in gray, while the median per band and the associated 16th and 84th percentiles (error bars) are shown in dark gray. The median best-fit (smoothed) model is shown in red. The median model photometry per band is shown by the open face circles. At the top left the sample median properties (number, median K_s -band magnitude, median $\log(\text{stellar mass})$, median $\log(\text{SFR})$, median $\log(\text{sSFR})$) are shown, as well as the standard deviation.

Table 2
Number Densities for Galaxies in Our Sample

Redshift	$N \text{ Mpc}^{-3}$
$2.0 < z < 3.0$	$5.62 \pm 1.2 \times 10^{-5}$
$3.0 < z < 4.0$	$1.38 \pm 0.4 \times 10^{-5}$
$4.0 < z < 5.0$	$3.45 \pm 2.16 \times 10^{-6}$

Note. Errors are calculated by combining in quadrature the Poisson uncertainty and the fractional error due to cosmic variance.

but has a dearth of massive ($\log(M_*/M_\odot) > 10.6$) QGs at $z \gtrsim 4$. The only QGs at $z > 4$ in SHARK are ~ 4 times less massive than those in observations, and in fewer numbers, too, implying that those galaxies which have been able to quench are still to accumulate mass. Finding QGs at $z \gtrsim 5$ will require both the combination of wide optical/NIR surveys to find such rare galaxies and also deep NIR/mid-infrared spectroscopy to confirm them.

8. What Fraction of Quiescent Galaxies at $3 < z < 5$ are Post-starburst?

In the epoch of the universe where quenching begins, it may make more sense to instead differentiate between quenching, recently quenched, and older (or “true”) QGs. This issue confronts a more philosophical question in extragalactic astrophysics: What is the definition of a QG? There appear to be two ways to deal with this population exhibiting a wide variety of quenching states. First, one can split the entire sample into young quiescent/PSB/recently quenched, and old quiescent (e.g., Ichikawa & Matsuoka 2017; Schreiber et al. 2018; Park et al. 2022). Alternatively, the entire sample can be treated as homogeneous, which is what we have done with this analysis. The prevalence of QGs with recent star formation at $z > 3$, the so-called PSB population, has been measured by a plethora of works, both from photometric and spectroscopic data. While at low redshift the definitions of post-starburst, albeit broad, are based on measurements involving spectral indices or spectra-based principal component analysis that

encode information about strong Balmer absorption lines as well as weak or no emission lines (e.g., Wild et al. 2014; Chen et al. 2019; Wild et al. 2020; Wilkinson et al. 2021), this type of definition is not possible to apply to large photometrically selected samples. We instead consider how one might define a PSB galaxy in a sample of photometrically selected QGs based just on colors. Marsan et al. (2022) proposed a definition of PSB based on a visual SED inspection and the presence of a UV peak brighter than the emission redward of the Balmer/4000 Å break. This study estimates that PSB galaxies comprise 28% of the massive ($\log(M_*/M_\odot) > 11$) population at $3 < z < 4$ and 17% at $4 < z < 6$. Approximately twice higher fractions of PSB at $\log(M_*/M_\odot) > 11$ are reported by D’Eugenio et al. (2021) based on their sample of spectroscopically confirmed objects at $z > 3$ with measured D_{n4000} , D_B , and $H_{\delta A}$ indices. Thirty percent of the spectroscopic sample in Forrester et al. (2020b) ceased star formation less than 300 Myr prior to the epoch of the observed redshift at $3 < z < 5$. At the same epoch, Schreiber et al. (2018) report that “young quiescent” galaxies, defined as UVJ -star-forming galaxies with $V - J < 2.6$ but below some sSFR threshold, are just as numerous as classical QGs.

Here, we base a definition of PSB on the strength of the “blue bump” in the SEDs traced by the $NUV - U$ color. The distribution of $NUV - U$ colors for our QG sample of 221 galaxies peaks at $NUV - U = 1.75$. Only 28% of galaxies have $NUV - U > 2$ and only one galaxy has $NUV - U > 3$. Based on our single stellar population model from Section 5.1, a stellar population will not reach $NUV - U > 2$ until 0.5–1 Gyr has passed and $NUV - U > 3$ until well after 1 Gyr, assuming no dust attenuation. Based on this, we can define PSBs as those galaxies in our sample with $NUV - U < 2$, comprising 70% of QGs at $3 < z < 4$ and 87% at $4 < z < 5$. This is marginally higher than that previously reported (Schreiber et al. 2018). If we restrict our selection to only the most massive galaxies, i.e., those with $\log(M_*/M_\odot) > 11$, then these numbers change to 60% at $3 < z < 4$ and 86% at $4 < z < 5$. This agrees with the 60%–70% value given by D’Eugenio et al. (2021) for galaxies at $z \sim 3$. As a fraction of the entire ultra-massive galaxy

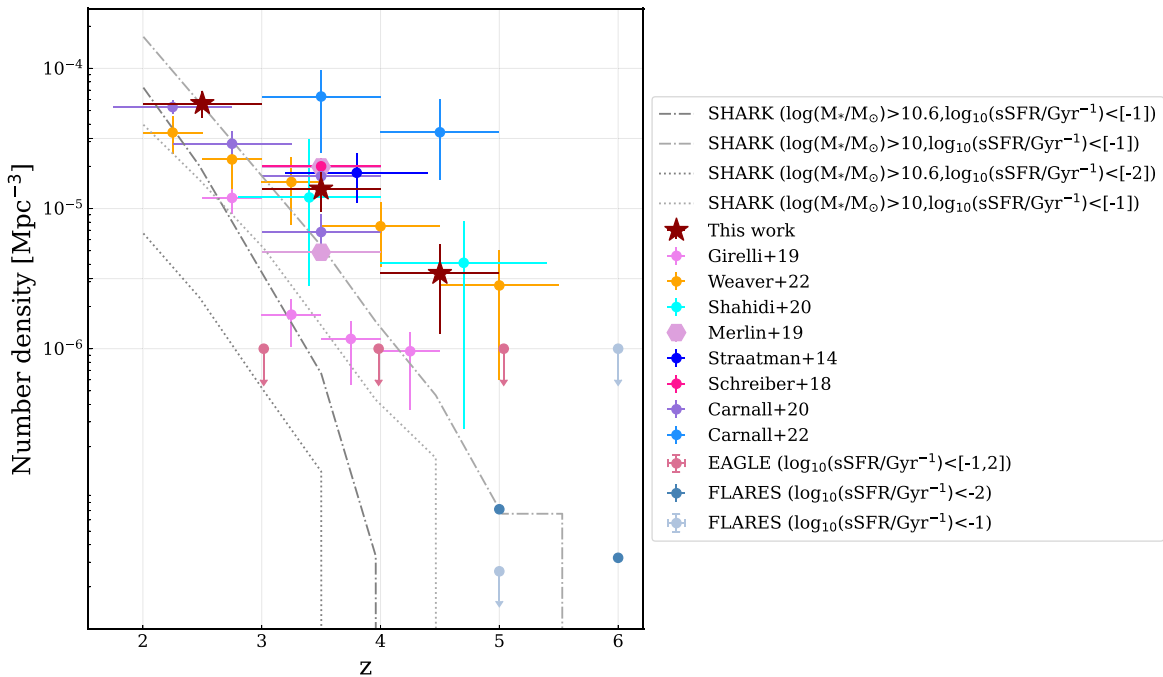


Figure 10. Number densities of massive QGs as a function of redshift at $2 < z < 5$. We report the number densities calculated for our sample (dark red stars) in redshifts bins of $2 < z < 3$ ($\log(M_*/M_\odot) > 10$), $3 < z < 4$ ($\log(M_*/M_\odot) > 10.6$), and $4 < z < 5$ ($\log(M_*/M_\odot) > 10.6$). The errors on the number densities include both Poisson noise and cosmic variance added in quadrature. We plot number densities from Straatman et al. (2014), Schreiber et al. (2018), Merlin et al. (2019), Carnall et al. (2020), Shahidi et al. (2020), Weaver et al. (2022), the extended sample from Girelli et al. (2019), and Carnall et al. (2022). Also shown are number densities for the FLARES and EAGLE simulations for massive galaxies ($\log(M_*/M_\odot) > 10.6$) at two quiescent selection thresholds ($\log_{10}(\text{sSFR}/\text{Gyr}^{-1}) < [-1, -2]$) (Lovell et al. 2022), as well as QGs from the SHARK simulation at two mass ($\log(M_*/M_\odot) > [10, 10.6]$) and sSFR selection thresholds ($\log_{10}(\text{sSFR}/\text{Gyr}^{-1}) < [-1, -2]$)

population, PSBs only comprise 7% at $3 < z < 4$ and 12% at $4 < z < 5$. Together with the results of the previous section, it is clear that $3 < z < 5$ is an active era in the history of the universe, one in which the transition to quiescence is fairly common.

9. Will We Find Old Quiescent Galaxies at $z > 4$ with JWST?

QGs have already been spectroscopically confirmed with JWST (Nanayakkara et al. 2022; Carnall et al. 2023). However, no old QGs have been found at $z > 4$ yet. In general, the blue colors of our $3 < z < 5$ QG candidates indicate a lack of very red, evolved galaxies such as the one presented in Glazebrook et al. (2017). This could be because, as an ensemble, the QG population at $z > 3$ has only recently quenched, and older QGs are much rarer. However, at $z > 4.5$, the reduced sensitivity to older/redder QGs due to the lack of data directly covering the 4000 \AA break could be biasing us to the bluest (and youngest) QGs. Although this is easily sidestepped with the JWST/Near Infrared Spectrograph (NIRSpec), it is difficult to solve even with JWST broadband Near Infrared Camera (NIRCam) imaging; at $4.5 < z < 5$, the 4000 \AA break falls at $2.2\text{--}2.4 \mu\text{m}$ and is straddled by F200W on the blue side and F277W (or redder bands such as F350W) on the red side (depending on the program), with no observations of QGs at $z > 4.5$ currently spanning the midpoint or upper edge of the break (see, e.g., Valentino et al. 2023; Carnall et al. 2022). This could easily be alleviated with the inclusion of NIRCam narrow bands covering $2.2\text{--}2.4 \mu\text{m}$, such as F210M or F250M. Currently, the only Cycle 1 program that has the combination of F200W, F277W/F350W, and at least one mediumband filter

in between is the Canadian NIRISS Unbiased Cluster Survey (CANUCS, PI: Willott; Willott et al. 2022), which has both F210M and F250M. However, the combined area covered by these filters is $\sim 50'$ squared, which would net less than one massive QG based on our number densities (see Section 7), assuming the same space distribution on the sky. In short, one would have to be incredibly lucky to find an evolved QG at $z > 4.5$ without both area and either spectroscopy or mediumband photometry.

10. Summary and Conclusions

In this work, we explore the massive QG population at $3 < z < 5$ in COSMOS using the latest photometric catalog, COSMOS2020. We create a new catalog using the COSMOS2020 photometry, with Spitzer/IRAC errors boosted, and derive redshifts and physical parameters with `eazy-py`. We motivate the need for a new rest-frame color diagram to select for QGs at $z > 3$ and present the GMM/*NUVU* – *VJ* method as a viable alternative. We use a GMM to fit the *NUV* – *U*, *U* – *V*, and *V* – *J* colors of massive galaxies at $2 < z < 3$, and galaxies are assigned a probability of being quiescent based on their bootstrapped rest-frame colors. This GMM model is then applied to the colors of massive galaxies at $3 < z < 5$. Both the GMM model and code to calculate quiescent probabilities from rest-frame flux densities are made available online.¹⁴ Our key findings can be summarized below:

1. We calculate the quiescent probability for simulated galaxies from the SHARK SAM at redshift snapshots of $z = 2, 3$, and 4 and find that GMM performs just as well

¹⁴ <https://github.com/kmlgould/GMM-quiescent>

as classical *UVJ* selection at $z=2$, with both methods returning similar true-positive rates and false-positive rates (TPR $\sim 90\%$, FPR $\sim 30\%$). However, at $z > 3$, the GMM method outperforms classical *UVJ* selection and particularly excels at $z \geq 4$, where the TPR is almost 100% compared to *UVJ*, which has less than 40% TPR. This highlights our proposed method as a viable alternative to traditional color-selection methods at $z > 3$.

2. We compare our GMM-selected QGs in COSMOS2020 to traditional rest-frame color-selected QGs (*UVJ*, *NUVrJ*) and calculate the overlap between selection methods. We find that all three color selections have high agreement at $2 < z < 3$, with methods selecting $\sim 85\%$ of the same galaxies, implying that the exact choice of rest-frame color-selection method at $2 < z < 3$ is not crucial. However, overlap between color selections at $z > 3$ reduces to $\sim 65\%$, implying that one should be careful about using *UVJ* and *NUVrJ* color selections interchangeably. Generally, the bluer colors of GMM-selected QGs at $3 < z < 5$ implies a dearth of any highly evolved QGs in this epoch.
3. We compute the number densities of QGs at $2 < z < 5$ and confirm the overall abundance of QGs presented in many different works, which our values agree with within the uncertainties. Given the conservative nature of this work (the boosting of IRAC errors, the high mass limit, and the probabilistic approach), we interpret the number densities as a lower limit.

The next few years will undoubtedly see our knowledge of QGs at $z > 3$ grow vastly with the recent launch and successful commissioning of JWST (Rigby et al. 2023), for which QGs at $z > 3$ have already been reported (Carnall et al. 2022; Cheng et al. 2023; Pérez-González et al. 2023; Rodighiero et al. 2023; Valentino et al. 2023) and even confirmed (Nanayakkara et al. 2022; Carnall et al. 2023), as well as ongoing large, ground-based surveys such as the Cosmic Dawn Survey (Moneti et al. 2022; C. J. R. McPartland et al. 2023, in preparation). Spectroscopic confirmation of multiple QG candidates with, for example, NIRSpec can be done in a matter of minutes (Glazebrook et al. 2021; Nanayakkara et al. 2022), and while ~ 8 hr with the Very Large Telescope/X-shooter or Keck/MOSFIRE are only sufficient to obtain a redshift and stellar populations, the same time (or less) with JWST could provide the required S/N to study detailed physics such as velocity dispersions and metallicities (Nanayakkara et al. 2022; Carnall et al. 2023). While JWST will be crucial for confirming QG candidates and studying their physical properties in detail, wide-area photometric surveys still have an important role to play both for the selection and statistical study of these galaxies.

The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant No. 140. S.J. acknowledges financial support from the European Union’s Horizon research and innovation program under the Marie Skłodowska-Curie grant agreement No. 101060888.

Software: *eazy-py* (Brammer 2021), *scipy* (Virtanen et al. 2020), *numpy* (Harris et al. 2020), *matplotlib* (Hunter 2007), *scikit learn* (Pedregosa et al. 2011), *astropy* (Astropy Collaboration et al. 2013, 2018), *python-FSPS* (Conroy et al. 2009; Conroy & Gunn 2010), *GMM-quiescent* (Gould 2023).

Appendix A Rest-frame Flux Densities

Consider two methods for inferring the flux density of a given galaxy at redshift z in a rest-frame bandpass, e.g., the *V* (visual) band in the Johnson filter system (Johnson 1955; Bessell 1990), with a central wavelength $\lambda_c \sim 5500 \text{ \AA}$.¹⁴ In the first method, we identify two observed bandpasses with central wavelengths that bracket the redshifted rest-frame bandpass $\lambda_{c,A} < \lambda_c(1+z) < \lambda_{c,B}$ and perform a linear interpolation between the flux densities in the observed bandpasses *A* and *B*. In a second method, we simply adopt the *V*-band flux density of the best-fit model template that was used to estimate the photometric redshift of the galaxy. The first method has the benefit of being purely empirical, though it will be limited by noise in the two photometric measurements used for the interpolation. The second method incorporates constraints from all available measurements, but it will be more sensitive to systematic biases induced by the adopted model template library.

Here, we adopt a methodology that is a hybrid between these two extremes that first makes better use of multiple observational constraints and second somewhat relaxes the dependence on the template library. Specifically, for a galaxy at redshift z and a given rest-frame bandpass with rest-frame central wavelength λ_r , we compute a weight, w_i , for each observed bandpass with an observed-frame central wavelength $\lambda_{o,i}$:

$$\gamma_i = \exp(x_i^2/2/\log(1+\Delta)^2), \quad (\text{A1})$$

$$w_i = W \cdot \left(\frac{2}{1 + \gamma_i/\text{Max}(\gamma_i)} - 1 \right), \quad (\text{A2})$$

where $x_i = \log(\lambda_r) - \log(\lambda_{o,i}/(1+z))$, and Δ and W are each parameters with a default value of 0.5. That is, the weights prioritize (i.e., are smallest for) observed bandpasses that are closest to the redshifted rest-frame bandpass. The templates are then refit to the observed photometry with least-squares weights $\epsilon_i^2 = \sigma_i^2 + (w_i \cdot F_i)^2$, where F_i and σ_i are the original photometric measurements and their uncertainties, respectively. The adopted rest-frame flux density is that of the template in this reweighted fit integrated through the target bandpass. In short, this approach uses the templates as a guide for a weighted interpolation of observations that best constrain a targeted rest-frame bandpass that fully accounts for all bandpass shapes and relative depths of a multiband photometric catalog. Because the weights are localized for each target rest-frame bandpass, there is no implicit requirement that, for example, the inferred rest-frame *V–J* color is within the range spanned by the templates themselves. Finally, we note that this is essentially the same methodology as that in version 1.0 of the original *eazy* code¹⁵ and now implemented in *eazy-py*.

Appendix B Star–Galaxy Separation

We use the same method for star–galaxy separation as that described in Weaver et al. (2022). Briefly, the catalog is matched to the Hubble ACS morphology catalog (Leauthaud et al. 2007) to isolate galaxies in the half-light radii versus magnitude plane.

¹⁴ The pivot wavelength is a convenient source-independent definition of the central wavelength of a bandpass (Tokunaga & Vacca 2005).

¹⁵ <https://github.com/gbrammer/eazy-photoz/tree/8be1c9>

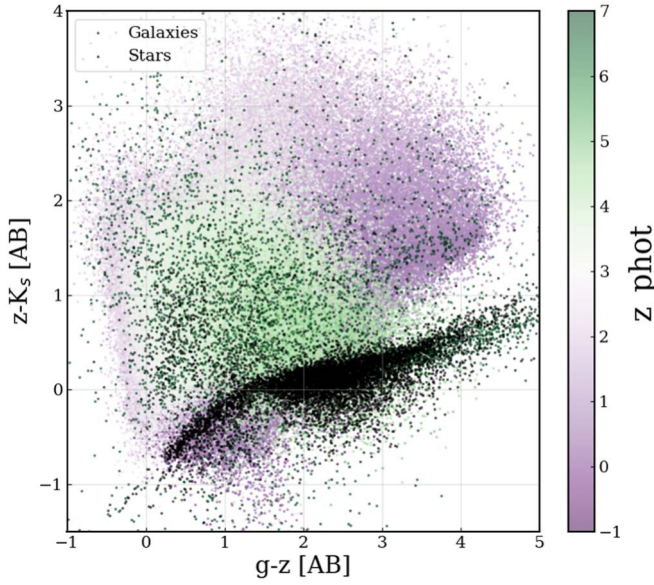


Figure 11. gzK diagram for all galaxies in the combined UVISTA and HSC area with g , z and K S/N > 3. Stars (black dots) occupy a locus across the lower part of the diagram, whereas galaxies (shown in green and purple dots) tend to occupy a wider spread of colors.

We fit each object with stellar templates from the PEGASUS library and compute the χ_{star}^2 . We calculate the reduced chi-squared χ_r^2 , where ($\chi_r^2 = \chi^2/N_{\text{filt}} - 1$), of both the galaxy best-fit template and star best template, where N_{filt} is the number of filters used. Objects with a HSC i -band magnitude <21.5 or HST/ACS F814W magnitude <23 with K_s S/N > 3 or IRAC Channel 1 S/N > 3 are classified as stars if they lie on the point-like sequence in the half-light radii versus magnitude plane or if they are fit better with a stellar template than a galaxy template (i.e., $\chi_{\text{star}}^2 < \chi_{\text{galaxy}}^2$). Additionally, objects that do not satisfy these criteria can be classified as a star if $\chi_{\text{star}}^2 + 1 < \chi_{\text{galaxy}}^2$. We find a similar fraction of stars to the number reported in the official COSMOS2020 catalog and find a similar distribution in the gzK color-color diagram (see Figure 11).

Appendix C Robustness of Rest-frame J -band Estimate

Although searches for massive QGs at $z > 3$ in the coming era will likely begin with rest-frame color-color selections, we should be critical about which rest-frame colors should be used. The choice of rest-frame colors should depend on both the desired redshift range probed and the ability of models to adequately measure rest-frame flux densities from the data. In particular, the color used to discriminate between red dusty star-forming galaxies and red QGs should be carefully chosen. Antwi-Danso et al. (2023) explored the effects of rest-frame J extrapolation for a flux-limited sample of galaxies at $0.5 < z < 6$ and find that the flux is affected significantly at $z > 2$ when extrapolation occurs. This can result in the removal of QGs in standard selections and high contamination rates.

At $z > 3$, current ground-based surveys such as COSMOS rely on data from Spitzer/IRAC Channels 3 and 4. By $z \sim 5$, the rest-frame J band is constrained entirely by Channel 4, which has a moderately shallow limiting 3σ depth of 23.1 mag in COSMOS2020. By comparison, Spitzer/IRAC Channels 1 and 2 are much deeper limited to 26.4 and 26.3 mags 3σ ,

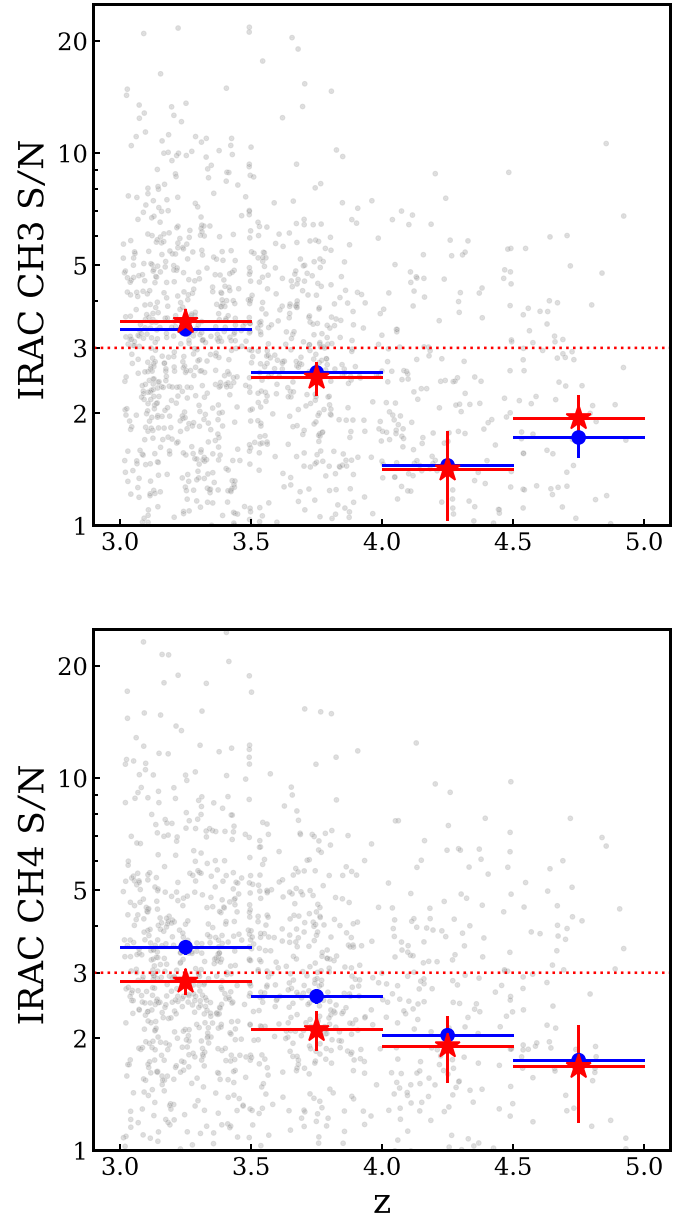


Figure 12. Signal-to-noise ratio (S/N) as a function of redshift at $3.0 < z < 5.0$ for IRAC Channel 3 (top) and IRAC Channel 4 (bottom) for the whole massive sample (gray). The average S/N for each channel is shown binned in redshift for the entire massive sample (blue dots) and the extended quiescent sample (red stars). The worst performance is at $4.5 < z < 5.0$ for both channels.

respectively. However, the fraction of galaxies in COSMOS2020 with $S/N \gtrsim 3$ in either Channel 3 or 4 at $3 < z < 5$ is only $\sim 5\%$. Making progress toward understanding the physical properties of the first quenched galaxies therefore mandates deeper data in the NIR, else inventive new methods, such as the use of a synthetic rest-frame i_s instead of J (Antwi-Danso et al. 2023). To ensure that rest-frame J is constrained for our sample, we limited the redshift upper bound at $z < 5$. Here, we compute the average S/N in bins of redshift for both Channels 3 and 4 for the entire massive sample and the quiescent sample (see Figure 12). At $3 < z < 3.5$ and for both bands, the mean S/N for QGs is $\gtrsim 3$. At $3.5 < z < 4.0$, the mean S/N in Channels 3 and 4 is 2–3, so rest-frame J is still constrained by reasonable photometry. At $z > 4$, however, both channels have mean S/N between 1–2, even after the

5× boosting. This highlights the need for deeper observations in the NIR at $>5 \mu\text{m}$. For part of the COSMOS field, this can be achieved with JWST/MIRI within the COSMOS-Web (PI: Casey; Casey et al. 2022) and PRIMER (PI: Dunlop) programs.

The issue of galaxies being classified incorrectly due to their rest-frame colors can partly be alleviated by our method of assigning a probability of quiescence to each galaxy, instead of a binary separator, which relies on the relative scatter within each population to be low—evidently not the case at $z > 3$.

Appendix D

Performance of an Extended UVJ Color Selection

At high redshift ($z \geq 2$), it may make sense to lower, or remove, the $U - V > 1.3$ boundary, allowing for the selection of “recently quenched” or PSB galaxies. This has been suggested and implemented by multiple works. For example, Schreiber et al. (2018) suggest combining the lowering of the boundary with the use of a band measuring the contribution from more recent star formation, while Marsan et al. (2022), who studied ultra-massive galaxies in the COSMOS-UVISTA field, remove the line altogether. Belli et al. (2019), Forrest et al. (2020b), Carnall et al. (2020), and Park et al. (2022) also advocated for either removing or lowering the boundary. If the $U - V > 1.3$ constraint is entirely removed, the picture changes for UVJ. At $z = 2$ the results remain roughly the same: the TPR and FPR both increase by 4%. At $z > 3$, the improvement is marked: the TPR increases by 25% to 98%, while the FPR only increases by 3% by 15%. $z = 4$ sees the most improvement, with the TPR increasing from 34% to 97% while the FPR still remains low at 6%. Considering instead the contamination, which is defined as the number of galaxies defined as quiescent compared to the total number of UVJ QGs, the conclusion changes. The contamination fraction of UVJ in SHARK at $z = 2, 3,$ and 4 is 81, 86, and 87%, which is similar to the contamination including the border (80, 86, 91%). Therefore, although the TPR and FPR imply the extended UVJ selection is suitable for use, the contamination fraction suggests otherwise, which is the same conclusion as for the classical UVJ selection.

Appendix E

Sample

The sample of the 230 QGs between redshifts 3 and 5 is included in Table 3.

Table 3
Sample QGs

Number	Label	Units	Explanation
1	ID	...	Identifier
2	R.A.	deg	R.A. (J2000)
3	Decl.	deg	decl. (J2000)
4	z_{phot}	...	Photometric derived redshift
5	Mass	M_{\odot}	Mass
6	SFR	$M_{\odot} \text{ yr}^{-1}$	Star formation rate
7	$p(q)_{50}$...	Median bootstrapped quiescent probability distribution
8	$p(q)_{95}$...	95% bootstrapped quiescent probability distribution

Note. Table 3 is published in its entirety in the electronic edition of the journal. A portion is shown here for guidance regarding its form and content.

(This table is available in its entirety in machine-readable form.)

ORCID iDs

Katriona M. L. Gould  <https://orcid.org/0000-0003-4196-5960>
Gabriel Brammer  <https://orcid.org/0000-0003-2680-005X>
Francesco Valentino  <https://orcid.org/0000-0001-6477-4011>
Katherine E. Whitaker  <https://orcid.org/0000-0001-7160-3632>
John. R. Weaver  <https://orcid.org/0000-0003-1614-196X>
Claudia del P. Lagos  <https://orcid.org/0000-0003-3021-8564>
Francesca Rizzo  <https://orcid.org/0000-0001-9705-2461>
Maximilien Franco  <https://orcid.org/0000-0002-3560-8599>
Bau-Ching Hsieh  <https://orcid.org/0000-0001-5615-4904>
Olivier Ilbert  <https://orcid.org/0000-0002-7303-4397>
Shuowen Jin  <https://orcid.org/0000-0002-8412-7951>
Georgios Magdis  <https://orcid.org/0000-0002-4872-2294>
Henry J. McCracken  <https://orcid.org/0000-0002-9489-7765>
Bahram Mobasher  <https://orcid.org/0000-0001-5846-4404>
Marko Shuntov  <https://orcid.org/0000-0002-7087-0701>
Charles L. Steinhardt  <https://orcid.org/0000-0003-3780-6801>
Victoria Strait  <https://orcid.org/0000-0002-6338-7295>
Sune Toft  <https://orcid.org/0000-0003-3631-7176>

References

- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, *PASJ*, 70, S4
Akins, H. B., Narayanan, D., Whitaker, K. E., et al. 2022, *ApJ*, 929, 94
Antwi-Danso, J., Papovich, C., Leja, J., et al. 2023, *ApJ*, 943, 166
Ardila, F., Alatalo, K., Lanz, L., et al. 2018, *ApJ*, 863, 28
Arnouts, S., Floc’h, E. L., Chevillard, J., et al. 2013, *A&A*, 558, A67
Arnouts, S., & Ilbert, O. 2011, Astrophysics Source Code Library, ascl:1108.009
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, 156, 123
Barrufet, L., Oesch, P., Fudamoto, Y., et al. 2021, JWST Proposal. Cycle, 1, 2198
Barrufet, L., Oesch, P. A., Weibel, A., et al. 2023, *MNRAS*, 522, 449
Belli, S., Newman, A. B., & Ellis, R. S. 2019, *ApJ*, 874, 17
Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
Bessell, M. S. 1990, *PASP*, 102, 1181
Black, W. K., & Evrard, A. 2022, *MNRAS*, 516, 1170
Brammer, G. 2021, eazy-py v0.6.4, Zenodo, doi:10.5281/zenodo.5012704
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, 686, 1503
Brammer, G. B., Whitaker, K. E., van Dokkum, P. G., et al. 2011, *ApJ*, 739, 24
Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
Carnall, A. C., Leja, J., Johnson, B. D., et al. 2019, *ApJ*, 873, 44
Carnall, A. C., McLeod, D. J., McLure, R. J., et al. 2023, *MNRAS*, 520, 3974
Carnall, A. C., McLure, R. J., Dunlop, J. S., et al. 2023, arXiv:2301.11413
Carnall, A. C., Walker, S., McLure, R. J., et al. 2020, *MNRAS*, 496, 695
Casey, C. M., Kartaltepe, J. S., Drakos, N. E., et al. 2022, arXiv:2211.07865
Cecchi, R., Bolzonella, M., Cimatti, A., & Girelli, G. 2019, *ApJL*, 880, L14
Chabrier, G. 2003, *PASP*, 115, 763
Chauhan, G., Lagos, C. d. P., Obreschkow, D., et al. 2019, *MNRAS*, 488, 5898
Chen, Y.-M., Shi, Y., Wild, V., et al. 2019, *MNRAS*, 489, 5709
Cheng, C., Huang, J.-S., Smail, I., et al. 2023, *ApJL*, 942, L19
Conroy, C., & Gunn, J. E. 2010, *ApJ*, 712, 833
Conroy, C., Gunn, J. E., & White, M. 2009, *ApJ*, 699, 486
Daddi, E., Cimatti, A., Renzini, A., et al. 2004, *ApJ*, 617, 746
Dantas, M. L. L., Coelho, P. R. T., & Sánchez-Blázquez, P. 2020, *MNRAS*, 500, 1870
Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, *A&A*, 605, A70
D’Eugenio, C., Daddi, E., Gobat, R., et al. 2020, *ApJL*, 892, L2
D’Eugenio, C., Daddi, E., Gobat, R., et al. 2021, *A&A*, 653, A32
Dressler, A., & Gunn, J. E. 1983, *ApJ*, 270, 7

- Fontana, A., Santini, P., Grazian, A., et al. 2009, *A&A*, 501, 15
- Forrest, B., Annunziatella, M., Wilson, G., et al. 2020a, *ApJL*, 890, L1
- Forrest, B., Marsan, Z. C., Annunziatella, M., et al. 2020b, *ApJ*, 903, 47
- French, K. D. 2021, *PASP*, 133, 072001
- Girelli, G., Bolzonella, M., & Cimatti, A. 2019, *A&A*, 632, A80
- Glazebrook, K., Nanayakkara, T., Esdaile, J., et al. 2021, JWST Proposal Cycle, 1, 2565
- Glazebrook, K., Schreiber, C., Labbé, I., et al. 2017, *Natur*, 544, 71
- Goto, T. 2005, *MNRAS*, 357, 937
- Gould, K. 2023, kmlgould/GMM-quiescent: GMM-quiescent, Zenodo, 10.5281/zenodo.7712937
- Greggio, L., & Renzini, A. 1990, *ApJ*, 364, 35
- Guarnieri, P., Maraston, C., Thomas, D., et al. 2019, *MNRAS*, 483, 3060
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, 585, 357
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *ApJS*, 208, 19
- Hunter, J. D. 2007, *CSE*, 9, 90
- Hwang, Y.-H., Wang, W.-H., Chang, Y.-Y., et al. 2021, *ApJ*, 913, 6
- Ichikawa, A., & Matsuoka, Y. 2017, *ApJL*, 843, L7
- Ilbert, O., McCracken, H. J., Fevre, O. L., et al. 2013, *A&A*, 556, A55
- Ito, K., Tanaka, M., Miyaji, T., et al. 2022, *ApJ*, 929, 53
- Iyer, K. G., Gawiser, E., Faber, S. M., et al. 2019, *ApJ*, 879, 116
- Ji, Z., & Giavalisco, M. 2022, *ApJ*, 935, 120
- Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, *ApJS*, 254, 22
- Johnson, H. L. 1955, *AnAp*, 18, 292
- Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, *ApJS*, 172, 196
- Kokorev, V., Brammer, G., Fujimoto, S., et al. 2022, *ApJS*, 263, 38
- Kriek, M., & Conroy, C. 2013, *ApJL*, 775, L16
- Kubo, M., Umehata, H., Matsuda, Y., et al. 2021, *ApJ*, 919, 6
- Lagos, C. d. P., da Cunha, E., Robotham, A. S. G., et al. 2020, *MNRAS*, 499, 1948
- Lagos, C. d. P., Robotham, A. S. G., Trayford, J. W., et al. 2019, *MNRAS*, 489, 4196
- Lagos, C. d. P., Tobar, R. J., Robotham, A. S. G., et al. 2018, *MNRAS*, 481, 3573
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24
- Lang, D., Hogg, D. W., & Myktyyn, D. 2016, ascl soft, Astrophysics Source Code Library, ascl:1604.008
- Leauthaud, A., Massey, R., Kneib, J.-P., et al. 2007, *ApJS*, 172, 219
- Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2019a, *ApJ*, 876, 3
- Leja, J., Tacchella, S., & Conroy, C. 2019b, *ApJL*, 880, L9
- Leja, J., Johnson, B. D., Conroy, C., et al. 2019d, *ApJ*, 877, 140
- Leja, J., Johnson, B. D., Conroy, C., Dokkum, P. G. V., & Byler, N. 2017, *ApJ*, 837, 170
- Long, A. S., Casey, C. M., Lagos, C. D. P., et al. 2022, arXiv:2211.02072
- Lovell, C. C., Roper, W., Vijayan, A. P., et al. 2022, arXiv:2211.07540
- Lower, S., Narayanan, D., Leja, J., et al. 2020, arXiv:2006.03599 [astro-ph]
- Lustig, P., Strazzullo, V., D'Eugenio, C., et al. 2021, *MNRAS*, 501, 2659
- Lustig, P., Strazzullo, V., Remus, R.-S., et al. 2023, *MNRAS*, 518, 5953
- Maíz Apellániz, J. 2006, *AJ*, 131, 1184
- Marchesini, D., Whitaker, K. E., Brammer, G., et al. 2010, *ApJ*, 725, 1277
- Marsan, Z. C., Marchesini, D., Brammer, G. B., et al. 2015, *ApJ*, 801, 133
- Marsan, Z. C., Marchesini, D., Brammer, G. B., et al. 2017, *ApJ*, 842, 21
- Marsan, Z. C., Muzzin, A., Marchesini, D., et al. 2022, *ApJ*, 924, 25
- Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, *ApJL*, 619, L1
- McConachie, I., Wilson, G., Forrest, B., et al. 2022, *ApJ*, 926, 37
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, 544, A156
- Merlin, E., Fontana, A., Castellano, M., et al. 2018, *MNRAS*, 473, 2098
- Merlin, E., Fortuni, F., Torelli, M., et al. 2019, *MNRAS*, 490, 3309
- Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, *PASJ*, 70, S1
- Moneti, A., McCracken, H. J., Hudelot, P., et al. 2022, VizieR On-line Data Catalog: II/373
- Moneti, A., McCracken, H. J., Shuntov, M., et al. 2022, *A&A*, 658, A126
- Moster, B. P., Somerville, R. S., Newman, J. A., & Rix, H.-W. 2011, *ApJ*, 731, 113
- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, *ApJ*, 777, 18
- Nanayakkara, T., Esdaile, J., Glazebrook, K., et al. 2022, *PASA*, 39, e002
- Nanayakkara, T., Glazebrook, K., Jacobs, C., et al. 2022, arXiv:2212.11638
- Noiro, G., Sawicki, M., Abraham, R., et al. 2022, *MNRAS*, 512, 3566
- Oke, J. B., & Gunn, J. E. 1983, *ApJ*, 266, 713
- Pacifici, C., Kassin, S. A., Weiner, B. J., et al. 2016, *ApJ*, 832, 79
- Park, M., Belli, S., Conroy, C., et al. 2022, arXiv:2210.03747
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, 12, 2825
- Pérez-González, P. G., Barro, G., Annunziatella, M., et al. 2023, *ApJL*, 946, L16
- Phillipps, S., Ali, S. S., Bremer, M. N., et al. 2020, *MNRAS*, 492, 2128
- Rigby, J., Perrin, M., McElwain, M., et al. 2023, *PASP*, 135, 048001
- Robotham, A. S. G., Bellstedt, S., Lagos, C. D. P., et al. 2020, *MNRAS*, 495, 905
- Rodighiero, G., Bisigello, L., Iani, E., et al. 2023, *MNRAS*, 518, L19
- Santini, P., Castellano, M., Merlin, E., et al. 2021, *A&A*, 652, A30
- Saracco, P., Marchesini, D., La Barbera, F., et al. 2020, *ApJ*, 905, 40
- Sawicki, M., Arnouts, S., Huang, J., et al. 2019, *MNRAS*, 489, 5202
- Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, *MNRAS*, 440, 889
- Schreiber, C., Glazebrook, K., Nanayakkara, T., et al. 2018, *A&A*, 618, A85
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Shahidi, A., Mobasher, B., Nayyeri, H., et al. 2020, *ApJ*, 897, 44
- Sherman, S., Jogee, S., Florez, J., et al. 2020, *MNRAS*, 499, 4239
- Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, *ApJS*, 214, 24
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Spitler, L. R., Straatman, C. M. S., Labbe, I., et al. 2014, *ApJL*, 787, L36
- Steinhardt, C. L., Capak, P., Masters, D., & Speagle, J. S. 2016, *ApJ*, 824, 21
- Steinhardt, C. L., Jespersen, C. K., & Linzer, N. B. 2021, *ApJ*, 923, 8
- Stevans, M. L., Finkelstein, S. L., Kawinwanichakij, L., et al. 2021, *ApJ*, 921, 58
- Stockmann, M., Toft, S., Gallazzi, A., et al. 2019, *ApJ*, 888, 4
- Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2014, *ApJL*, 783, L14
- Suzuki, T. L., Glazebrook, K., Schreiber, C., et al. 2022, *ApJ*, 936, 61
- Tokunaga, A. T., & Vacca, W. D. 2005, *PASP*, 117, 421
- Trayford, J. W., Lagos, C. d. P., Robotham, A. S. G., & Obreschkow, D. 2020, *MNRAS*, 491, 3937
- Valentino, F., Brammer, G., Fujimoto, S., et al. 2022, *ApJL*, 929, L9
- Valentino, F., Brammer, G., Gould, K. M. L., et al. 2023, *ApJ*, 947, 20
- Valentino, F., Tanaka, M., Davidzon, I., et al. 2020, *ApJ*, 889, 93
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, 17, 261
- Wang, T., Schreiber, C., Elbaz, D., et al. 2019, *Natur*, 572, 211
- Weaver, J. R., Davidzon, I., Toft, S., et al. 2022, arXiv:2212.02512
- Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, *ApJS*, 258, 11
- Whitaker, K. E., Franx, M., Leja, J., et al. 2014, *ApJ*, 795, 104
- Whitaker, K. E., Kriek, M., van Dokkum, P. G., et al. 2012, *ApJ*, 745, 179
- Whitaker, K. E., Labbé, I., van Dokkum, P. G., et al. 2011, *ApJ*, 735, 86
- Whitaker, K. E., van Dokkum, P. G., Brammer, G., et al. 2010, *ApJ*, 719, 1715
- Whitaker, K. E., van Dokkum, P. G., Brammer, G., et al. 2013, *ApJL*, 770, L39
- Wild, V., Almaini, O., Cirasuolo, M., et al. 2014, *MNRAS*, 440, 1880
- Wild, V., Taj Aldeen, L., Carnall, A., et al. 2020, *MNRAS*, 494, 529
- Wild, V., Walcher, C. J., Johansson, P. H., et al. 2009, *MNRAS*, 395, 144
- Wilkinson, A., Almaini, O., Wild, V., et al. 2021, *MNRAS*, 504, 4533
- Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbe, I. 2009, *ApJ*, 691, 1879
- Willott, C. J., Doyon, R., Albert, L., et al. 2022, *PASP*, 134, 025002
- Wu, P.-F., Wel, A. V. D., Bezanson, R., et al. 2018, *ApJ*, 868, 37
- Wu, P.-F., Wel, A. V. D., Bezanson, R., et al. 2020, *ApJ*, 888, 77
- Wuyts, S., Labbé, I., Förster Schreiber, N. M., et al. 2008, *ApJ*, 682, 985
- Yi, S., Demarque, P., & Kim, Y.-C. 1997, *ApJ*, 482, 677