



## SPITZER ULTRA FAINT SURVEY PROGRAM (SURFS UP). II. IRAC-DETECTED LYMAN-BREAK GALAXIES AT $6 \lesssim z \lesssim 10$ BEHIND STRONG-LENSING CLUSTERS

KUANG-HAN HUANG<sup>1</sup>, MARUŠA BRADAC<sup>1</sup>, BRIAN C. LEMAUX<sup>2</sup>, R. E. RYAN, JR.<sup>3</sup>, AUSTIN HOAG<sup>1</sup>, MARCO CASTELLANO<sup>4</sup>, RICARDO AMORÍN<sup>4</sup>, ADRIANO FONTANA<sup>4</sup>, GABRIEL B. BRAMMER<sup>3</sup>, BENJAMIN CAIN<sup>1</sup>, L. M. LUBIN<sup>1</sup>, EMILIANO MERLIN<sup>4</sup>, KASPER B. SCHMIDT<sup>5</sup>, TIM SCHRABBACK<sup>6</sup>, TOMMASO TREU<sup>7</sup>, ANTHONY H. GONZALEZ<sup>8</sup>, ANJA VON DER LINDEN<sup>9,10,11</sup>, AND ROBERT I. KNIGHT<sup>1</sup>

<sup>1</sup> University of California Davis, 1 Shields Avenue, Davis, CA 95616, USA; khhuang@ucdavis.edu, astrokuang@gmail.com

<sup>2</sup> Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

<sup>3</sup> Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, F-13388 Marseille, France

<sup>4</sup> INAF—Osservatorio Astronomico di Roma Via Frascati 33, I-00040 Monte Porzio Catone, Italy

<sup>5</sup> Department of Physics, University of California, Santa Barbara, CA 93106, USA

<sup>6</sup> Argelander-Institut für Astronomie, Auf Dem Hügel 71, D-53121 Bonn, Germany

<sup>7</sup> Department of Physics and Astronomy, UCLA, Los Angeles, CA 90095, USA

<sup>8</sup> Department of Astronomy, University of Florida, 211 Bryant Space Science Center, Gainesville, FL 32611, USA

<sup>9</sup> Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA

<sup>10</sup> Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen Ø, Denmark

<sup>11</sup> Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305-4060, USA

Received 2015 April 7; accepted 2015 November 21; published 2016 January 19

### ABSTRACT

We study the stellar population properties of the IRAC-detected  $6 \lesssim z \lesssim 10$  galaxy candidates from the *Spitzer* Ultra Faint Survey Program. Using the Lyman Break selection technique, we find a total of 17 galaxy candidates at  $6 \lesssim z \lesssim 10$  from *Hubble Space Telescope* images (including the full-depth images from the Hubble Frontier Fields program for MACS 1149 and MACS 0717) that have detections at signal-to-noise ratios  $\geq 3$  in at least one of the IRAC 3.6 and 4.5  $\mu\text{m}$  channels. According to the best mass models available for the surveyed galaxy clusters, these IRAC-detected galaxy candidates are magnified by factors of  $\sim 1.2$ – $5.5$ . Due to the magnification of the foreground galaxy clusters, the rest-frame UV absolute magnitudes  $M_{1600}$  are between  $-21.2$  and  $-18.9$  mag, while their *intrinsic* stellar masses are between  $2 \times 10^8 M_\odot$  and  $2.9 \times 10^9 M_\odot$ . We identify two Ly $\alpha$  emitters in our sample from the Keck DEIMOS spectra, one at  $z_{\text{Ly}\alpha} = 6.76$  (in RXJ 1347) and one at  $z_{\text{Ly}\alpha} = 6.32$  (in MACS 0454). We find that 4 out of 17  $z \gtrsim 6$  galaxy candidates are favored by  $z \lesssim 1$  solutions when IRAC fluxes are included in photometric redshift fitting. We also show that IRAC [3.6]–[4.5] color, when combined with photometric redshift, can be used to identify galaxies which likely have strong nebular emission lines or obscured active galactic nucleus contributions within certain redshift windows.

*Key words:* galaxies: evolution – galaxies: high-redshift – gravitational lensing: strong – methods: data analysis

### 1. INTRODUCTION

Galaxies at  $6 \lesssim z \lesssim 10$  are one of the frontiers in observational astronomy because they are a key player in the reionization process. It is widely postulated that galaxies provided the bulk of ionization photons, but low-level active galactic nucleus (AGN) activity (with their likely very high escape fractions of ionizing photons) is still a possibility (e.g., Giallongo et al. 2015). To improve our knowledge about the importance of galaxies on reionization, we should measure their ionizing photon production rate (through their star formation rate (SFR) density) and their ionizing photon escape fraction (Robertson et al. 2010). We should also measure their stellar mass and, under reasonable assumptions about their star formation history, infer how many ionizing photons they produced in the past.

Rest-frame optical stellar emission from  $z \gtrsim 6$  galaxies is crucial for stellar mass measurement; the rest-frame 4000 Å break shifts to  $\gtrsim 3 \mu\text{m}$  in the observed frame and requires deep *Spitzer* observations at the moment. Over a thousand  $z \gtrsim 6$  galaxy candidates have been identified in deep *Hubble Space Telescope* (*HST*) extragalactic blank fields like CANDELS (Grogin et al. 2011; Koekemoer et al. 2011), HUDF/XDF (Beckwith et al. 2006; Koekemoer et al. 2013; Illingworth et al. 2013; Bouwens et al. 2015), BoRG (Trenti et al. 2011, 2012; Bradley et al. 2012), and HIPPIES (Yan et al. 2011).

Among all the  $z \gtrsim 6$  galaxy candidates, more than 100 of them have individual *Spitzer*/IRAC detections (Yan et al. 2006; Eyles et al. 2007; Stark et al. 2009; Labbé et al. 2010, 2013, 2015; Capak et al. 2011; Roberts-Borsani et al. 2015); their IRAC fluxes enable more robust constraints on their stellar masses. The inferred stellar masses of these  $z \gtrsim 6$  galaxy candidates range from  $\sim 10^9$  to  $\sim 10^{11} M_\odot$ , surprisingly large for a universe younger than 1 Gyr old. It is likely that the majority of IRAC-detected  $z \gtrsim 6$  galaxies are at the high-mass end of the stellar mass function, although some of these galaxies likely have their IRAC fluxes boosted by strong nebular emission lines like [O III], H $\alpha$ , and H $\beta$  (e.g., Finkelstein et al. 2013; De Barros et al. 2014; Smit et al. 2014).

Using the strong gravitational lensing power of rich galaxy clusters is a novel avenue to explore high-redshift galaxies (Soucail 1990). Galaxy candidates at  $z \gtrsim 6$  that are magnified by foreground clusters were starting to be identified from *HST* images more than a decade ago (e.g., Ellis et al. 2001; Hu et al. 2002; Kneib et al. 2004); these observations provide an alternative way to probe the faint end of the luminosity function with shorter exposure time than in blank fields. Recently, the Cluster Lensing And Supernova survey with *Hubble* (CLASH; Postman et al. 2012) program and HST-GO-11591 (PI: Kneib) program observed 34 galaxy clusters. The ongoing Hubble

Frontier Fields (HFF; PI: Lotz<sup>12</sup>) program, upon its complete execution, will obtain deep *HST* Advanced Camera for Surveys (ACS)/WFC3-IR images in six galaxy cluster fields (four of them are in the CLASH sample). Bradley et al. (2014) recently reported 262  $6 \lesssim z \lesssim 8$  galaxy candidates across 18 clusters in the CLASH sample based on photometric redshift selection, and they demonstrated the power of using strong gravitational lensing to identify high- $z$  galaxies, especially at the bright end of the luminosity function. The even deeper HFF data, despite being  $\gtrsim 0.7$  mag shallower than the HUDF data, can probe galaxies *intrinsically* fainter than can be probed in the HUDF due to the power of gravitational lensing. Other coordinated campaigns are also underway to complement the deep *HST* images in those targeted galaxy cluster fields (e.g., the Grism Lens-Amplified Survey from Space program; Schmidt et al. 2014; Treu et al. 2015).

Here we use the deep *Spitzer*/IRAC images obtained from the *Spitzer* UltraFaint Survey Program (SURFS UP; Bradač et al. 2014; hereafter Paper I) to probe the rest-frame optical emission from  $z \gtrsim 6$  galaxy candidates. SURFS UP surveys 10 strong-lensing cluster fields (our sample partially overlaps with both CLASH and HFF) with *Spitzer* IRAC images in the 3.6  $\mu\text{m}$  and 4.5  $\mu\text{m}$  channels, with exposure times of  $\gtrsim 28$  hr per channel per cluster. Paper I summarizes the science motivations and observational strategies of SURFS UP, and Ryan et al. (2014) presented the  $z \sim 7$  galaxy candidates in the Bullet Cluster, one of which is detected in both IRAC channels. In this work, we explore the  $6 \lesssim z \lesssim 10$  galaxy candidates with IRAC detections in eight additional clusters in our sample<sup>13</sup> and present their physical properties inferred from their broadband fluxes. We also make all the IRAC imaging data available for the community on our webpage.<sup>14</sup>

The structure of this paper is as follows: Section 2 describes our *HST* and *Spitzer* imaging data and photometry; Section 3 describes our  $6 \lesssim z \lesssim 10$  galaxy sample selection procedure; Section 4 describes the identification of Ly $\alpha$  emission from spectroscopy for two galaxy candidates with IRAC detection at  $z = 6.76$  (RXJ 1347-1216) and  $z = 6.32$  (MACS 0454-1251); Section 5 presents our spectral energy distribution (SED) modeling procedure and results, and Section 6 explores the idea of using IRAC [3.6]–[4.5] color to identify galaxies with strong nebular emission lines. Finally, Section 7 summarizes our findings. Throughout the paper, we assume a  $\Lambda$ CDM concordance cosmology with  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ , and the Hubble constant  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . Coordinates are given for the epoch J2000.0, and all magnitudes are in the AB system.

## 2. IMAGING DATA AND PHOTOMETRY

### 2.1. *HST* Data and Photometry

We list the eight galaxy clusters analyzed in this work in Table 1. Among the eight clusters, six (MACS 0717, MACS 0744, MACS 1149, RXJ 1347, MACS 1423, and MACS 2129) are in the Cluster Lensing And Supernova Survey (CLASH; Postman et al. 2012) sample; therefore, each of them has *HST*

imaging data in at least twelve ACS/WFC and WFC3/IR filters<sup>15</sup> from the ACS and WFC3/IR cameras. The typical  $5\sigma$  depths for the CLASH clusters reported by Postman et al. (2012) are between 27.0 and 27.5 mag (within 0".4 diameter apertures) for each filter, and the large number of filters provides unique constraints for high- $z$  galaxy searches among the *HST* deep fields.

In addition to the CLASH data, full-depth images for MACS 0717 and MACS 1149 from the HFF program were also released in 2015 June. In these two clusters, *HST* spent a total  $\sim 140$  orbits that are roughly split between the ACS and WFC3/IR filters, and these images achieve  $\approx 28.7$ – $29$  mag<sup>16</sup> in the optical (ACS) and NIR (WFC3). We use the deepest HFF images for MACS 1149 and MACS 0717 for photometry in the filters where such images are available.<sup>17</sup> The six CLASH clusters in our sample are also observed by the Grism Lens-Amplified Survey from Space (GLASS; PI: Treu; Schmidt et al. 2014; Treu et al. 2015) program (MACS 0717, MACS 1149, MACS 1423, RXJ 1347, MACS 0744, and MACS 2129) and have deep grism spectra available.

For the remaining two clusters being analyzed in this work, we have data for RCS 2327 as part of the SURFS UP *HST* observations (HST-GO13177 PI Bradač; Hoag et al. 2015) and previous archival data (HST-GO10846 PI Gladders; see also Sharon et al. 2015). For MACS 0454, we use the archival observations from HST-GO11591 (PI: Kneib), GO-9836 (PI: Ellis), and GO-9722 (PI: Ebeling). We list the limiting magnitudes of point sources within a 0".4 aperture for MACS 0454 and RCS 2327 in Table 2.

We will use the template-fitting software T-PHOT (Merlin et al. 2015)—the successor to TFIT (Laidler et al. 2007)—to measure the colors between *HST* and *Spitzer* IRAC images (see Section 2.2). To prepare the *HST* images for T-PHOT, we use the public 0".03 pixel<sup>-1</sup> CLASH images and block-sum the images to make 0".06 pixel<sup>-1</sup> images. We also edit the astrometric image header values (CRVALs and CRPIXs) to conform to T-PHOT's astrometric requirements<sup>18</sup> and make sure that *HST* and *Spitzer* images are aligned to well within 0".1 (Paper I).

We also match the point-spread functions (PSFs) among all *HST* filters to get consistent colors. To do so, we identify isolated point sources in each cluster field, and we use the `psfmatch` task in IRAF to match all *HST* images to have the same PSF as the reddest band, F160W. In practice, because of the small field of view of each cluster and the crowded environment, we can only select  $\sim 5$  isolated point sources in each cluster for PSF-matching. However, we measure the curves of growth of each point source after PSF-matching and find that in most filters, the curves match to within  $\sim 20\%$  of that of F160W.

After pre-processing, we extract photometry on *HST* images using SExtractor (Bertin & Arnouts 1996; version 2.8.6). We use the combined IR images as the detection image, and the SExtractor detection/deblending settings similar to (but slightly more conservative than) the values adopted by CLASH for their high- $z$  galaxy search (Postman et al. 2012). Because our focus in this work is on identifying IRAC-detected high- $z$  sources, the slightly more conservative settings do not reject

<sup>12</sup> <http://www.stsci.edu/hst/campaigns/frontier-fields>

<sup>13</sup> The *HST* WFC3/IR imaging data for the tenth cluster, MACS 2214, will be obtained in late 2015 (HST-GO 13666; PI: Bradač).

<sup>14</sup> <http://bradac.physics.ucdavis.edu/SurfsUp.html>

<sup>15</sup> For the CLASH clusters, the ACS filters include F435W, F475W, F606W, F625W, F775W, F814W, and F850LP; the WFC3/IR filters include F105W, F110W, F125W, F140W, and F160W.

<sup>16</sup> The magnitude limits are from the Frontier Fields website.

<sup>17</sup> The HFF filter sets include F435W, F606W, F814W, F105W, F125W, F140W, and F160W.

<sup>18</sup> T-PHOT requires that the pixel boundaries of the high- and low-resolution images be perfectly aligned, meaning that both images should have the same CRVALs and both have half-integer CRPIXs.

**Table 1**  
SURFS UP Galaxy Cluster Sample

Cluster Name	Short Name <sup>a</sup>	R.A. (deg)	decl. (deg)	$z_{\text{cluster}}$ <sup>b</sup>	$N_{\text{LBG}}$ <sup>c</sup>	$N_{\text{LBG,IRAC}}$ <sup>d</sup>	
1	MACS J0454.1–0300	MACS 0454	73.545417	−3.018611	0.54	10	2
2	MACS J0717.5+3745 <sup>e,f</sup>	MACS 0717	109.390833	37.755556	0.55	10	0
3	MACS J0744.8+3927 <sup>e</sup>	MACS 0744	116.215833	39.459167	0.70	4	1
4	MACS J1149.5+2223 <sup>e,f</sup>	MACS 1149	177.392917	22.395000	0.54	11	3
5	RXJ 1347–1145 <sup>e</sup>	RXJ 1347	206.883333	−11.761667	0.59	9	3
6	MACS J1423.8+2404 <sup>e</sup>	MACS 1423	215.951250	24.079722	0.54	9	6
7	MACS J2129.4–0741 <sup>e</sup>	MACS 2129	322.359208	−7.690611	0.59	0	0
8	RCS 2–2327.4–0204	RCS 2327	351.867500	−2.073611	0.70	6	1
9	1E0657–56	Bullet Cluster	104.614167	−55.946389	0.30	10	1
10	MACS 2214.9–1359 <sup>g</sup>	MACS 2214	333.739208	−14.003000	0.50	N/A	N/A
Total						69	17

**Notes.**

<sup>a</sup> We will refer to each cluster by its short name.

<sup>b</sup> Cluster redshift.

<sup>c</sup> Number of  $6 \lesssim z \lesssim 10$  LBG candidates selected by their *HST* colors.

<sup>d</sup> Number of  $6 \lesssim z \lesssim 10$  LBG candidates with  $\geq 3\sigma$  detections in at least one IRAC channel.

<sup>e</sup> A CLASH cluster.

<sup>f</sup> A Hubble Frontier Fields cluster.

<sup>g</sup> The *HST* WFC3/IR data for MACS 2214 will be collected in late 2015.

**Table 2**  
*HST*  $5\sigma$  Limiting Magnitudes (Point Source,  $0.''4$  Aperture) for RCS 2327 and MACS 0454

Cluster	F435W	F555W	F775W	F814W	F850LP	F098M	F110W	F125W	F160W
MACS 0454	...	27.7	26.9	27.9	26.5	...	28.1	...	27.4
RCS 2327	27.6	...	...	27.6	...	27.3	...	27.6	27.5

many potential IRAC-detected candidates but eliminates most spurious detections. We also follow the procedure outlined in Trenti et al. (2011) to rescale the flux errors reported by SExtractor. At the end of the process, we use the resulting photometric catalogs and segmentation maps for both IRAC photometry (Section 2.2) and color selection (Section 3).

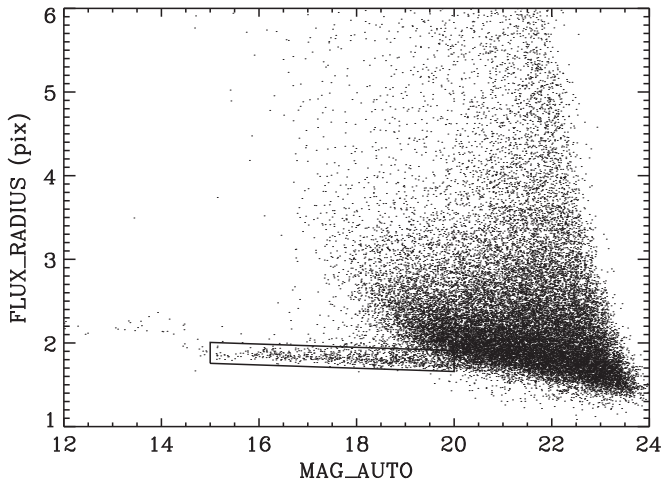
## 2.2. IRAC Data and Photometry

The IRAC imaging data set for SURFS UP was presented in Paper I; the total exposure time for each IRAC channel is about 30 hr (see Ryan et al. 2014 and Paper I for details.) The coadded IRAC mosaics are deeper within the main cluster fields covered by WFC3/IR, and the typical  $3\sigma$  limiting magnitude within  $3''$ -radius apertures is 26.6 mag in the  $3.6 \mu\text{m}$  channel (hereafter ch1) and 26.2 mag in the  $4.5 \mu\text{m}$  channel (hereafter ch2) where source blending is not severe.

We use T-PHOT to measure consistent colors between *HST* and *Spitzer* IRAC images with a template-fitting approach (see also Laidler et al. 2007 for the template-fitting concept employed by T-PHOT.) The template-fitting approach has been demonstrated to work well for blank-field surveys such as CANDELS (Guo et al. 2013), but it does require images with zero mean background. Most galaxy cluster fields have considerable spatial variations in local sky background, so subtracting a constant background does not generally work. Therefore, instead of fitting all sources in the field at the same time—which is the strategy for blank-field surveys—we subtract the local background and perform the fit for each high- $z$  candidate separately to get the cleanest residual possible.

As described in Merlin et al. (2015), T-PHOT is designed to measure the fluxes in the low-resolution image (in our case the IRAC images) for all the sources detected in the high-resolution image (in our case the F160W images). T-PHOT does so by constructing a template for each source; it convolves the cutout of each source in the F160W image by a PSF-transformation kernel that matches the F160W resolution to the IRAC resolution. Once the templates are available (and with their fluxes normalized to 1), T-PHOT solves the set of linear equations and finds the combination of coefficients for each template that most closely reproduce the pixel values in IRAC images; each coefficient is therefore the flux of the source in IRAC. T-PHOT also calculates the full covariance matrix and uses the diagonal terms of the covariance matrix to calculate flux errors. For each source, T-PHOT reports a “covariance index,” defined as the ratio between the maximum covariance of the source with its neighbors ( $\max(\sigma_{ij})$ ) over its flux variance ( $\sigma_{ii}$ ), which serves as an indicator of how strongly correlated the source’s flux is with its closest (or brightest) neighbor. Generally, a high covariance index ( $\gtrsim 1$ ) is associated with more severe blending and large flux errors, at least from simulations (Laidler et al. 2007; Merlin et al. 2015). Therefore, sources with high covariance indices should be treated with caution.

Obviously the PSF-transformation kernel that matches the F160W PSF to the IRAC PSF is a crucial element in this process. We generate IRAC PSFs by stacking point sources observed in the exposures from both the primary cluster field and the flanking field. We identify point sources using SExtractor with `DEBLEND_MINCONT` =  $10^{-7}$ , `MINAREA` = 9, `DETECT_THRESH` = `ANALYSIS_THRESH` = 2, and a Gaussian convolution kernel with  $\sigma = 3$  pixels (defined over



**Figure 1.** Brightness and half-light radii for all sources in IRAC ch1 of the SURFS UP clusters. The half-light radii (FLUX\_RADIUS) are in  $0''.6$  IRAC pixels. The box illustrates the crudely defined stellar loci where all point sources are expected to fall. When selecting putative point sources for each field, we first make sure the objects are near this locus for a given cluster, then place additional constraints on proximity to neighbors, axis ratio, and re-centering/alignment accuracy for the final sample per cluster. There are typically  $\gtrsim 40$  stars per cluster for PSF generation.

a 5 by 5-pixel grid). We require that all point sources have an axis ratio of  $b/a > 0.9$ , lie on the stellar locus within the box shown in Figure 1, and are sufficiently separated from neighboring objects to have reliable centroids ( $FLAGS \leq 3$ ). We recompute the PSF centroids by fitting a Gaussian profile to the inner profile ( $r < 4$  pixels) using the SExtractor barycenters as initial guesses, and align the point sources using sinc interpolation. To mask neighboring objects, we grow the segmentation maps from SExtractor by 2 pixels. At each phase we subtract the local sky (assuming there are no local gradients) and normalize the flux of the point source to unity. We sigma-clip average the masked, registered, normalized point sources and do one further background correction only to ensure the convolutions with T-PHOT are flux conserving. As discussed in Paper I, our stacked PSFs are consistent with the IRAC handbook. Each of our clusters contains at least 40 point sources per bandpass in our PSF-making process.

In practice, T-PHOT still breaks down in very crowded regions (e.g., near the cluster center or near bright cluster galaxies); in this case, we are limited mostly by our knowledge of the IRAC PSFs and our ability to subtract sky background underneath the sources. We also measure the “reduced  $\chi^2$ ” for each source in IRAC within a  $2''.4$  by  $2''.4$  box by calculating the average difference per pixel between the model pixel values and the observed pixel values:  $\chi_\nu^2 = \sum_i^{N_{\text{pix}}} (f_{i,\text{model}} - f_{i,\text{obs}})^2 / (f_{i,\text{obs}}^2 \times N_{\text{pix}})$ , where  $f_{i,\text{model}}$  and  $f_{i,\text{obs}}$  are the model (best-fit) and observed flux in pixel  $i$ , respectively. Later in this work, we only report the IRAC fluxes of the high- $z$  galaxies with reliable IRAC flux measurements, i.e.,  $\chi_\nu^2 \leq 3$ .

For the sources with nominal signal-to-noise ratios (S/Ns) above 3 but with poor T-PHOT residuals, we do not trust the T-PHOT-measured fluxes and estimate the local  $3\sigma$  flux limits via artificial source simulations. We insert artificial point sources into the F160W image (but *not* into the IRAC image) within  $5''$  of the high- $z$  candidates and run T-PHOT to measure the local sky level. We repeat this process at least 100 times near each high- $z$  candidate and use the resulting IRAC flux distribution to determine the  $3\sigma$  flux limits. In our analysis in Section 5, we use

the  $3\sigma$  flux limit for only one IRAC filter for one source (ch1 for MACS 1423-1384); for all the other sources, their IRAC flux measurements in both IRAC channels pass the  $\chi_\nu^2$  test.

We also run a separate set of simulations that independently estimate the magnitude errors in case T-PHOT underestimates the magnitude errors in crowded regions even when they pass the  $\chi_\nu^2$  test. In this set of simulations, we insert fake point sources around each high- $z$  target (within  $5''$ ) in IRAC images with the same magnitude as the T-PHOT-measured value, and measure the flux of the fake sources again with T-PHOT. We then calculate the median difference between the input and output magnitudes of the fake sources as an independent magnitude error estimate. We find that for most sources, the T-PHOT-reported magnitude error is within 0.1 mag from the simulated magnitude error, but sometimes the simulated magnitude error is much larger than the T-PHOT-reported value. In these cases, T-PHOT might underestimate the true magnitude errors, so we adopt the simulated magnitude errors in our SED modeling. We note that by adding fake sources in IRAC images, we increase the flux error due to crowding, so the simulated magnitude errors could be higher than the true magnitude errors.

### 3. SAMPLE SELECTION

We select galaxy candidates at  $z \gtrsim 6$  based on their rest-frame UV colors using the Lyman-break selection method (Steidel & Hamilton 1993; Giavalisco 2002). For the CLASH clusters, we use the published color criteria presented below for selecting  $z \sim 6-9$  Lyman-break galaxies (LBGs); for RCS 2327 and MACS 0454, we design our own color cuts. All galaxy colors are calculated using their isophotal magnitudes (MAG\_ISO) from SExtractor. After the initial color selection, we inspect the galaxy candidates to remove image artifacts and objects with problematic photometry. We then measure each LBG candidate’s fluxes in IRAC, and we only present the candidates with  $S/N \geq 3$  in at least one channel. Because of the differences in the available filters in each cluster, we explain our color-selection process in more detail below and list the sample size in Table 1. The full sample of the color-selected  $z \gtrsim 6$  galaxy candidates with IRAC detections is presented in Table 3.

#### 3.1. CLASH and HFF Clusters: MACS 0717, MACS 1149, MACS 0744, MACS 1423, MACS 2129, and RXJ 1347

For the six clusters that in the CLASH sample, we use the criteria below. To be selected as a  $z \sim 6$  LBG candidate, a source has to satisfy *all* of the following criteria from Gonzalez et al. (2011):

$$\begin{cases} F775W - F850LP > 1.3 \\ F850LP - F125W < 0.8 \\ S/N \geq 5 & \text{in F850LP and F125W} \\ S/N < 2 & \text{in filters bluer than F606W,} \end{cases} \quad (1)$$

where we calculate all S/Ns within the isophotal (ISO) aperture.<sup>19</sup> If the S/N in F775W is below one, we use the  $1\sigma$  flux limit in F775W to calculate the F775W – F850LP color. To select LBG candidates at  $z \sim 7$ , we use the color criteria

<sup>19</sup> The S/N limits in blue filters roughly correspond to magnitude limits of  $\gtrsim 28$  mag, based on the typical limiting magnitudes presented by Postman et al. (2012).

**Table 3**  
IRAC-detected  $6 \lesssim z \lesssim 10$  Galaxy Candidates

Object ID	R.A. (deg)	decl. (deg)	F160W <sup>a</sup> (mag)	[3.6] <sup>b</sup> (mag)	$R_{3.6}$ <sup>c</sup>	[4.5] <sup>d</sup> (mag)	$R_{4.5}$ <sup>c</sup>	[3.6]–[4.5] (mag)	Spectroscopy <sup>e</sup>
F125W-dropouts ( $z \sim 9$ )									
MACS 1149-JD <sup>f</sup>	177.389943	22.412719	$25.6 \pm 0.1$	$25.8 \pm 0.4$	0.27	$25.0 \pm 0.2$	0.29	$0.8 \pm 0.4$	M,G
F105W-dropouts ( $z \sim 8$ )									
MACS 1423-1384	215.942115	24.079401	$25.7 \pm 0.2$	$>23.6^h$	0.95	$24.1 \pm 0.4^i$	0.95	$>-0.5$	G
RXJ 1347-1080 <sup>g</sup>	206.891236	-11.752594	$26.3 \pm 0.2$	$25.4 \pm 0.2$	0.12	$25.3 \pm 0.2$	0.11	$0.2 \pm 0.2$	D,G
F850LP-dropouts ( $z \sim 7$ )									
MACS 0744-2088 <sup>g</sup>	116.250405	39.453011	$25.5 \pm 0.2$	$25.2 \pm 0.4^i$	0.50	$25.1 \pm 0.2$	0.50	$0.1 \pm 0.3$	G
MACS 1423-587	215.940493	24.090848	$25.3 \pm 0.1$	$24.3 \pm 0.3^i$	0.86	$26.2 \pm 0.5$	0.80	$-1.8 \pm 0.6$	G
MACS 1423-774	215.935607	24.086475	$25.9 \pm 0.2$	$25.1 \pm 0.2$	0.70	$25.5 \pm 0.3$	0.69	$-0.4 \pm 0.3$	D,G
MACS 1423-2248	215.932958	24.070875	$25.6 \pm 0.1$	$25.0 \pm 0.1$	0.42	$25.3 \pm 0.2$	0.45	$-0.2 \pm 0.2$	D,G
MACS 1423-1494	215.935871	24.078414	$26.3 \pm 0.2$	$26.1 \pm 0.4$	0.92	$25.2 \pm 0.2$	0.89	$0.9 \pm 0.4$	D,G
MACS 1423-2097 <sup>j</sup>	215.945534	24.072433	$25.8 \pm 0.2$	$24.6 \pm 0.3^i$	0.68	$24.6 \pm 0.1$	0.68	$0.0 \pm 0.2$	D,G
RXJ 1347-1216 <sup>i,k,g</sup>	206.900848	-11.754199	$26.1 \pm 0.2$	$24.3 \pm 0.1$	0.16	$25.6 \pm 0.2$	0.11	$-1.3 \pm 0.2$	D,G
RXJ 1347-1800	206.881657	-11.761483	$25.4 \pm 0.2$	$24.3 \pm 0.3$	0.63	$25.6 \pm 0.7$	0.55	$-1.3 \pm 0.8$	G
Bullet-3 <sup>l</sup>	104.667375	-55.968067	$25.0 \pm 0.2$	$23.8 \pm 0.3$	n/a	$23.8 \pm 0.3$	n/a	$0.0 \pm 0.4$	FORS2
F814W-dropouts ( $z \sim 6-7$ )									
RCS 2327-1282	351.880595	-2.076292	$24.8 \pm 0.1$	$24.4 \pm 0.1$	0.08	$24.1 \pm 0.1$	0.06	$0.3 \pm 0.1$	D,M
MACS 0454-1251 <sup>m</sup>	73.535653	-3.004116	$24.1 \pm 0.1$	$23.2 \pm 0.2$	0.60	$23.4 \pm 0.2$	0.60	$-0.2 \pm 0.2$	D
MACS 0454-1817	73.551806	-3.001018	$26.4 \pm 0.2$	$24.1 \pm 0.3^i$	0.31	$24.5 \pm 0.2^i$	0.31	$-0.4 \pm 0.1$	D
F775W-dropouts ( $z \sim 6$ )									
MACS 1149-274 <sup>g</sup>	177.412009	22.415783	$24.8 \pm 0.04$	$24.1 \pm 0.1$	0.44	$24.0 \pm 0.1$	0.36	$0.0 \pm 0.1$	G
MACS 1149-1204 <sup>g</sup>	177.378959	22.402429	$25.0 \pm 0.1$	$24.3 \pm 0.1$	0.64	$24.4 \pm 0.1$	0.67	$-0.1 \pm 0.1$	G

#### Notes.

<sup>a</sup> Lensed total magnitude (MAG\_AUTO) in F160W; the magnification factors ( $\mu$ ) are listed in Table 4.

<sup>b</sup> Isophotal lensed magnitude in IRAC channel 1 based on the isophotal aperture defined in F160W.

<sup>c</sup>  $R_{3.6}$  and  $R_{4.5}$  are the covariance indices for the [3.6] and [4.5] measurements, respectively. The covariance index of a source  $i$  is defined as the ratio between the maximum covariance among the neighbors ( $\sigma_{ij}$ ) over the flux variance of itself ( $\sigma_{ii}$ ) in the covariance matrix.

<sup>d</sup> Isophotal lensed magnitude in IRAC channel 2 based on the isophotal aperture defined in F160W.

<sup>e</sup> Instruments we used for spectroscopy: D = DEIMOS, M = MOSFIRE, G = *HST* grism from the GLASS program (Schmidt et al. 2014; Treu et al. 2015).

<sup>f</sup> First reported by Zheng et al. (2012).

<sup>g</sup> Also reported by Bradley et al. (2014).

<sup>h</sup> The IRAC residual in the 3.6  $\mu\text{m}$  channel shows that T-PHOT breaks down due to severe blending, so we report the simulated  $3\sigma$  magnitude limit.

<sup>i</sup> T-PHOT likely underestimates the magnitude errors for these sources due to crowding, so we use the simulated magnitude errors.

<sup>j</sup> Also reported by Smit et al. (2014).

<sup>k</sup> Has Ly $\alpha$  detection at  $z = 6.76$ .

<sup>l</sup> Reported by Ryan et al. (2014).

<sup>m</sup> Has tentative Ly $\alpha$  detection at  $z = 6.32$ .

from Bouwens et al. (2011):

$$\begin{cases} \text{F850LP} - \text{F105W} > 0.7 \\ \text{F105W} - \text{F125W} < 0.45 \\ \text{F850LP} - \text{F105W} > 1.4 \times (\text{F105W} - \text{F125W}) + 0.42 \\ \text{S/N} \geq 5 & \text{in F105W and F125W} \\ \text{S/N} < 5 & \text{in F814W} \\ \text{S/N} < 2 & \text{in filters bluer than F775W.} \end{cases} \quad (2)$$

To select the LBGs at  $z \sim 8$ , we use the criteria from Bouwens et al. (2011):

$$\begin{cases} \text{F105W} - \text{F125W} > 0.45 \\ \text{F125W} - \text{F160W} < 0.5 \\ \text{S/N} \geq 5 & \text{in F125W and F160W} \\ \text{S/N} < 2 & \text{in filters bluer than F850LP.} \end{cases} \quad (3)$$

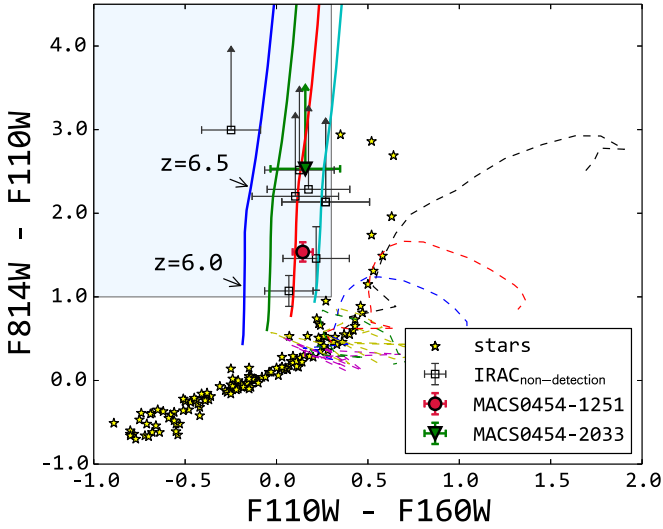
Finally, to select the LBG candidates at  $z \gtrsim 9$ , we use the criteria from Zheng et al. (2014):

$$\begin{cases} \text{F125W} - \text{F160W} > 0.8 \\ \text{S/N} \geq 5 & \text{in F160W} \\ \text{S/N} < 2 & \text{in filters bluer than F850LP.} \end{cases} \quad (4)$$

In total, we find 43  $z \gtrsim 6$  LBG candidates from the six CLASH/HFF clusters; among them, 13 have  $\geq 3\sigma$  detections in at least one IRAC channel.

### 3.2. MACS 0454

For the two galaxy clusters that are not in the CLASH sample (MACS 0454 and RCS 2327), we design our own selection criteria for  $z \gtrsim 6$  galaxy candidates. For MACS 0454, we have *HST* imaging data in F555W, F775W, F814W, F850LP, F110W, and F160W, although the images in F775W



**Figure 2.** Color-color diagram of the F814W-dropout selection from MACS 0454. The shaded region (light blue) shows where the expected F814W-dropout colors should be. We also plot (in *solid* curves) the theoretical color tracks of a 100 Myr old stellar population with constant star formation taken from Bruzual & Charlot (2003) with different amounts of dust attenuation. Sources within the shaded region that also pass the S/N cuts are shown in unfilled squares; two of them, MACS 0454-1251 and MACS 0454-1817 (shown in large filled symbols), are detected in IRAC. The color tracks of  $z \leq 3$  galaxies, calculated from the local galaxy templates of Coleman et al. (1980), are shown in *dashed* curves. We also show the expected colors of stars from Pickles (1998).

and F850LP are shallower than the other filters and we use both filters only for S/N rejection of low- $z$  interlopers. We use the following criteria to select  $6 \lesssim z \lesssim 7.5$  galaxy candidates (F814W-dropouts):

$$\begin{cases} F814W - F110W \geq 1.0 \\ F110W - F160W \leq 0.3 \\ S/N \geq 5 \text{ in } F110W \text{ and } F160W \\ S/N < 2 \text{ in } F555W. \end{cases} \quad (5)$$

We show the *HST* color-color diagram of the F814W-dropout selection in Figure 2. A total of ten sources satisfy the color and S/N cuts listed above; among them, MACS 0454-1251 and MAC 0454-1817 are detected in at least one IRAC channel.

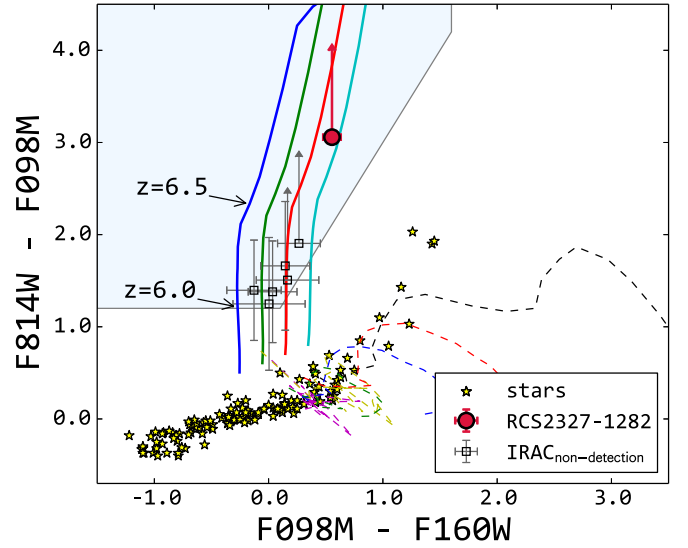
Because we have only one filter redward of F110W, we refrain from searching for F110W-dropouts in MACS 0454 as it would yield objects detected in only one *HST* filter.

### 3.3. RCS 2327

For RCS 2327, we have deep *HST* images in F435W, F814W, F098M, F125W, and F160W, so we use the following criteria for  $6 \lesssim z \lesssim 7.5$  galaxy candidates (F814W-dropouts):

$$\begin{cases} F814W - F098M \geq 2.2 \\ F098M - F160W \leq 1.6 \\ F814W - F098M \geq 2.0 \times (F098M - F160W) + 1.0 \\ S/N \geq 5 \text{ in } F098M \text{ and } F160W \\ S/N < 3 \text{ in } F435W. \end{cases} \quad (6)$$

When the S/N in F814W is less than unity, we use its  $1\sigma$  flux limit to calculate colors. We demonstrate the F814W-dropout selection from RCS 2327 in Figure 3, and six sources



**Figure 3.** Color-color diagram of the F814W-dropout selection from RCS 2327. The plot style and model assumptions are the same as Figure 2. A total of six sources satisfy the F814W-dropout color criteria, and one of them (RCS 2327-1282; shown as a red circle) has IRAC detections.

pass the above color and signal-to-noise cuts. One of the six sources, RCS 2327-1218, is detected in both IRAC channels. In addition to the above criteria, we also use the criteria similar to the BoRG survey (Trenti et al. 2011) to search for  $z \gtrsim 7.5$  galaxy candidates:

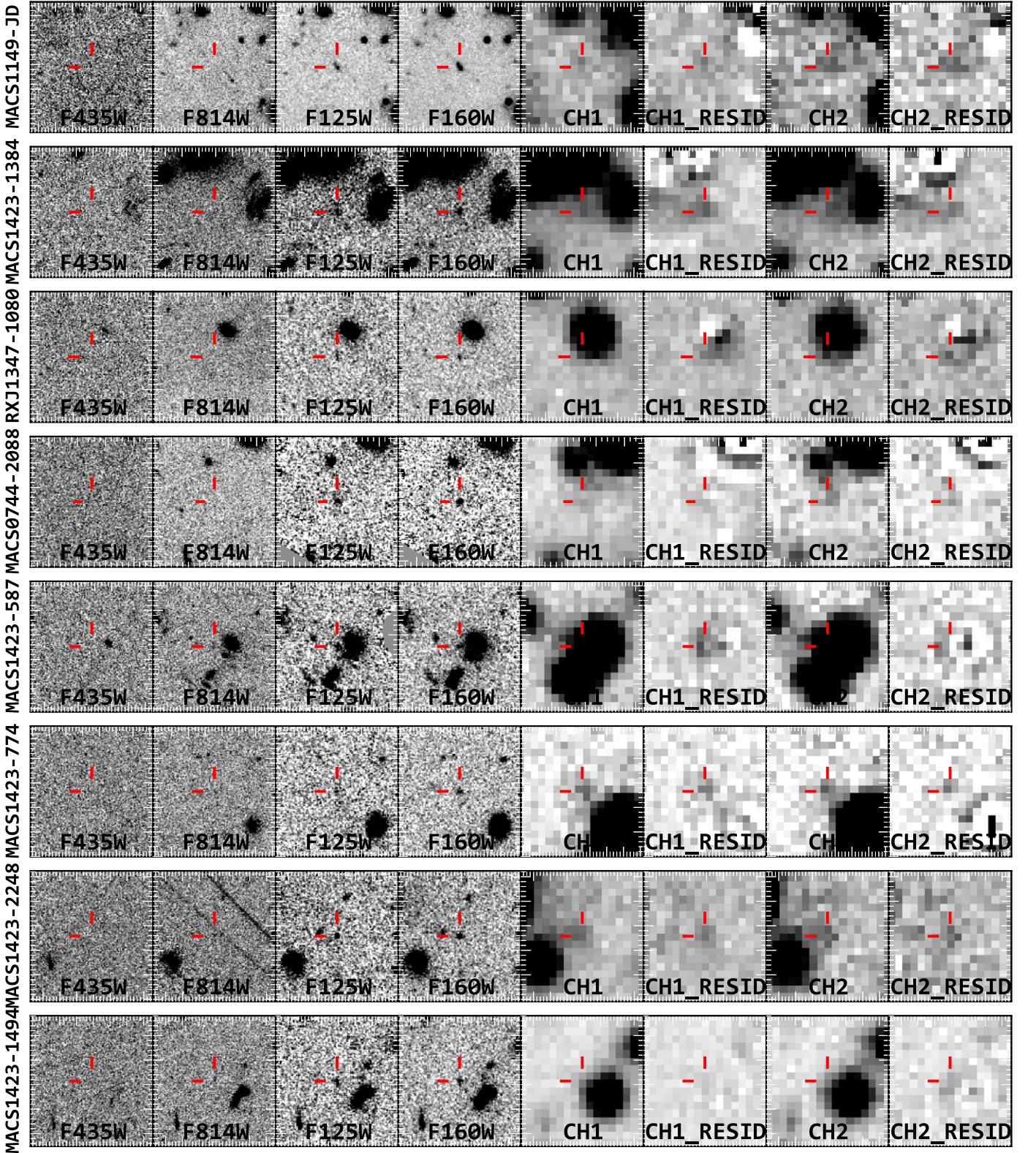
$$\begin{cases} F098M - F125W \geq 1.75 \\ S/N \geq 5 \text{ in } F125W \text{ and } F160W \\ S/N < 3 \text{ in } F814W \text{ and } F435W. \end{cases} \quad (7)$$

and when the S/N in F098M is less than unity, we use its  $1\sigma$  flux limit to calculate colors. The F098M-dropout search yields no galaxy candidate, so we find a total of one galaxy candidate with IRAC detections at  $z \gtrsim 6$  from RCS 2327 (see also Hoag et al. 2015 for more details on the dropout search in RCS 2327).

To summarize, we find a total of 69 LBG candidates at  $z \gtrsim 6$  from 9 clusters in SURFS UP; 17 of them have IRAC detections in at least one channel. Figures 4 and 5 show the cutouts of the 16 IRAC-detected LBG candidate in *HST* and *Spitzer* images (one candidate was reported by Ryan et al. 2014). We also report the IRAC photometry for the entire sample in Table 3. We use the simulated IRAC magnitude errors for MACS 1423-1384, MACS 1423-587, MACS 0744-2088, MACS 1423-2097, and MACS 0454-1817 because we find that T-PHOT likely underestimates their IRAC magnitude errors from the simulations. We keep MACS 1423-1384 in our sample because it has a nominal  $5.9\sigma$  detection in ch2 from T-PHOT, although the simulated magnitude error suggests that the additional flux error due to crowding reduces it to a  $2.2\sigma$  detection in ch2.

## 4. SPECTROSCOPIC OBSERVATIONS

We report the detection of two likely Ly $\alpha$  emitters among our sample with DEIMOS (Faber et al. 2003) on the Keck II telescope. The DEIMOS observation is part of a larger campaign to systematically target lensed high- $z$  galaxies behind strong-lensing galaxy clusters with DEIMOS and MOSFIRE (McLean et al. 2010, 2012) on Keck. We observed six cluster



**Figure 4.** Cutouts of the first eight IRAC-detected,  $z \gtrsim 6$  LBG candidates in SURFS UP (excluding the one candidate in the Bullet Cluster reported by Ryan et al. 2014). Each row shows the cutout in two *HST* ACS filters (F435W and F814W in all clusters except for MACS 0454, where we show F435W and F555W), two *HST* WFC3/IR filters (F125W and F160W), and two IRAC channels. We also show the neighbor-subtracted cutouts around each LBG candidate in both IRAC channels (designated by CH1\_RESID and CH2\_RESID, respectively). The LBG candidate ID is to the left of each row. Each panel is centered on the LBG candidate (marked by the red lines), and each panel spans  $10''$  by  $10''$  on the sky.

fields between 2013 April and 2014 August and targeted 9 out of 17 high- $z$  galaxy candidates in Table 3 with DEIMOS. Which galaxy candidates were observed with DEIMOS and

MOSFIRE are indicated in Table 3. The DEIMOS data were reduced using the standard DEEP2 spec2d pipeline slightly modified to reduce the data observed also with  $600 \text{ l mm}^{-1}$  and

8301 mm<sup>-1</sup> gratings (Lemaux et al. 2009; Newman et al. 2013). We focus here on the two galaxies (RXJ 1347-1216 and MACS 0454-1251) that have line detections; we will present the full spectroscopic survey (with both DEIMOS and MOS-FIRE) and the line flux limits for the non-detections in a future work. In addition to the Keck observations, 13 of the 17 galaxy candidates in Table 3 are also observed by the GLASS (HST GO-13459; PI: Treu) program; the spectroscopic constraints on Ly $\alpha$  emission from the *HST* grism data will be presented by Schmidt et al. (2015).

Below we discuss the two galaxy candidates with robust line detections and the likelihood that they are Ly $\alpha$  emitters at  $z = 6.76$  and  $z = 6.32$ .

#### 4.1. RXJ 1347-1216

We selected this object as a  $z \sim 7$  LBG candidate for spectroscopic follow-up on 2013 April 3 and 2014 May 26. This source was also selected by both Smit et al. (2014) and Bradley et al. (2014) as a  $z_{\text{phot}} \sim 6.7$  LBG with high [O III]+H $\beta$  equivalent widths ( $>1300 \text{ \AA}$  rest frame). We used the 8301 mm<sup>-1</sup> grating in the 2013 run and the 12001 mm<sup>-1</sup> grating in the 2014 run, and the total integration times are roughly 6000 and 7200 s, respectively. We had good (but not photometric) conditions with  $\sim 1''$  seeing in the 2013 run, but the conditions were highly variable in the 2014 run. Therefore, we only present the line flux measurements from the 2013 run, though the line was detected significantly in both runs.

In Figure 6 we show the two-dimensional spectra of RXJ 1347-1216 from both observation runs (in the top and middle panels) and the combined one-dimensional spectrum (in the bottom panel). We detect an extended emission feature with  $\text{FWHM}_{\text{obs}} \sim 16.5 \text{ \AA}$  in the 2013 run, and although the blue side of the feature is severely contaminated by sky line residual, its asymmetric profile with a tail to the red side of the spectrum strongly suggests that it is Ly $\alpha$ . Using the centroid of the sky line residual at 9439  $\text{\AA}$  as the peak of the line profile, we determine its Ly $\alpha$  redshift as  $z_{\text{Ly}\alpha} = 6.76 \pm 0.003$  (the uncertainty corresponds to the width of the sky line residual). The Ly $\alpha$  redshift is in excellent agreement with its photometric redshift  $z_{\text{phot}} = 6.8 \pm 0.1$ , lending additional support to the identification of the Ly $\alpha$  feature.

The emission feature is also independently detected at  $4\sigma$  in the GLASS grism data at  $\sim 9440 \text{ \AA}$ . With the grism spectra in both G102 and G141, one can test the possibility of this feature being an [O II] doublet ( $\lambda 3727, \lambda 3729$ ) at  $z \sim 1.5$  by looking for the [O III]  $\lambda 5007$  line at  $\sim 1.3 \mu\text{m}$ . For a typical star-forming galaxy, the line ratio [O III]/[O II] should be at least  $\sim 0.3$  (Jones et al. 2015), and the [O III]/[O II] ratio is even higher for low-metallicity galaxies (e.g., Maiolino et al. 2008). Assuming the detected line in DEIMOS is [O II] at  $z = 1.53$ , HST G141 grism data imply a  $2\sigma$  upper limit on [O III]/[O II]  $\lesssim 0.3$ , which is highly unlikely for a star-forming galaxy (Schmidt et al. 2015). Therefore, we conclude that the *HST* grism data also strongly support the  $z = 6.76$  Ly $\alpha$  interpretation of this emission line.

We perform line flux measurements from the DEIMOS data obtained during the 2013 run (with the 8301 mm<sup>-1</sup> grating) following the procedure outlined in Section 2.4 of Lemaux et al. (2009). In short, we place two filters of width 20  $\text{\AA}$  on both sides of the emission line that are free of spectral features and sky line residuals to measure the background, and a central filter encompassing the emission line to measure the integrated flux. The background, which is fit to a linear function, is then

subtracted from the integrated line flux. We perform spectrophotometric calibration of the DEIMOS data using two other compact sources (with half-light radii  $\sim 0''.3$  as measured from the F775W image) on the same mask with continuum detection in the following manner: for each object, the combination of slit loss and loss due to clouds was determined by calculating a spectral magnitude, done by correcting each DEIMOS spectrum for spectral response and atmospheric extinction and convolving the F775W filter curve with the resulting spectra, and comparing this value with the magnitude measured in the *HST* image. The ratio of the flux densities for each of the two sources is calculated and averaged to estimate the total spectral loss for this mask, which is then applied to RXJ 1347-1216 assuming a similar half-light radius for this object. The reason for this assumption is, though the half-light radius of RXJ 1347-1216 is smaller ( $0''.2$ ), which would, in principle, mean less slit loss, the size of the Ly $\alpha$  nebula is known to far exceed the size of the UV continuum region (e.g., Wisotzki et al. 2015). For the two sources on the mask with which we performed the flux calibration, the total measured throughput of the slit was  $\sim 40\%$ , lower than the  $\sim 60\%$  expected for sources of this size (if they are symmetric), suggesting at least some departure from a photometric night.

Using the above procedure, we measure the line flux from the 8301 mm<sup>-1</sup> data to be  $7.8 \pm 0.7 \times 10^{-18} \text{ erg cm}^{-2} \text{ s}^{-1}$ , which translates into a Ly $\alpha$  luminosity of  $4.1 \pm 0.4 \times 10^{42} \text{ erg s}^{-1}$ . We do not detect continuum in the spectrum, so we estimate the rest-frame Ly $\alpha$  equivalent width using the object's broadband magnitude in F105W (on the red side of Ly $\alpha$ ) to be  $26 \pm 4 \text{ \AA}$ . The equivalent width uncertainty include the Poisson noise in the central filter encompassing the sky line residual, the uncertainty in the continuum, and the uncertainty in DEIMOS absolute flux calibration.

We note that our measured Ly $\alpha$  line flux from DEIMOS data is roughly a factor of 3 lower than that measured from the *HST* grism data, which is  $2.6 \pm 0.5 \times 10^{-17} \text{ erg cm}^{-2} \text{ s}^{-1}$  (Schmidt et al. 2015). The difference in the line flux measurements could be due to the following factors: (1) our DEIMOS data are shallower than the *HST* grism data ( $\sim 2$  hr of total integration time in the 2013 mask, versus  $\sim 5$  orbits of *HST* integration time in G102 grism), with the difference in depths leading to differences in the spatial and extent of the emission detected above the noise in each observation, which impacts the total integrated line flux; (2) we might still underestimate the Ly $\alpha$  slit loss because the true extent of the Ly $\alpha$ -emitting region is much larger than the continuum-emitting region, while the *HST* slitless grism recovers more of the Ly $\alpha$  flux; (3) there is a sky line coincident with the peak wavelength of the Ly $\alpha$  emission in the DEIMOS data which appears slightly over-subtracted that serves to slightly lower the measured line flux; and (4) there could also be issues with contamination subtraction in the *HST* grism data, although it is unclear which direction it biases line flux measurement. We do expect the ground-based line flux measurement to be a lower limit to the space-based measurement, which is consistent with the measured values from DEIMOS and from *HST* grism data.

We also measure the inverse line asymmetry  $1/a_\lambda$  as defined in Lemaux et al. (2009), and the line's inverse asymmetry value ( $1/a_\lambda = 0.22$ ) is well within the range ( $1/a_\lambda < 0.75$ ) typical for convincing Ly $\alpha$  emission.

Using Ly $\alpha$  line flux, one can also infer a SFR following Lemaux et al. (2009). The inferred SFR is strictly a lower limit,

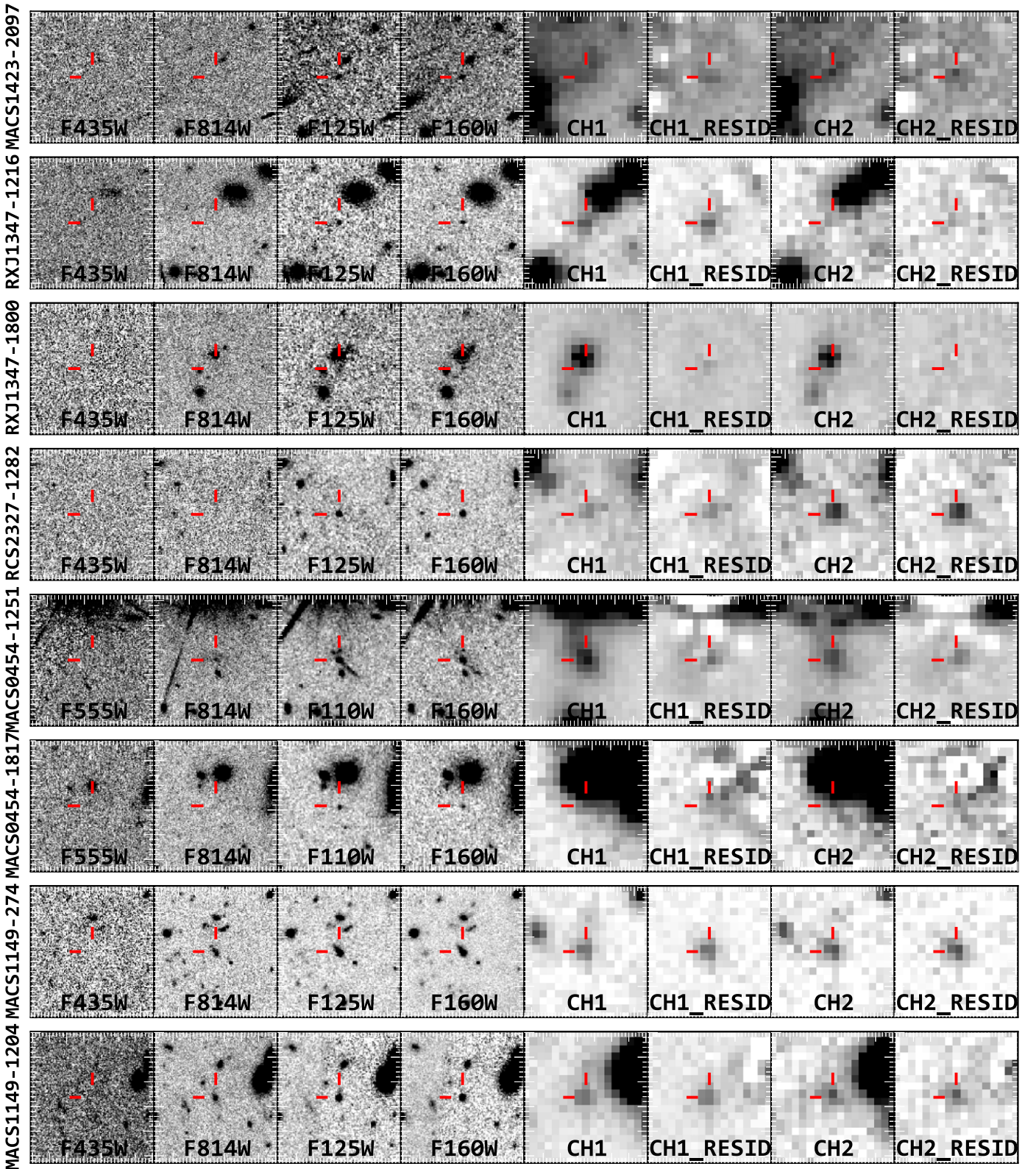
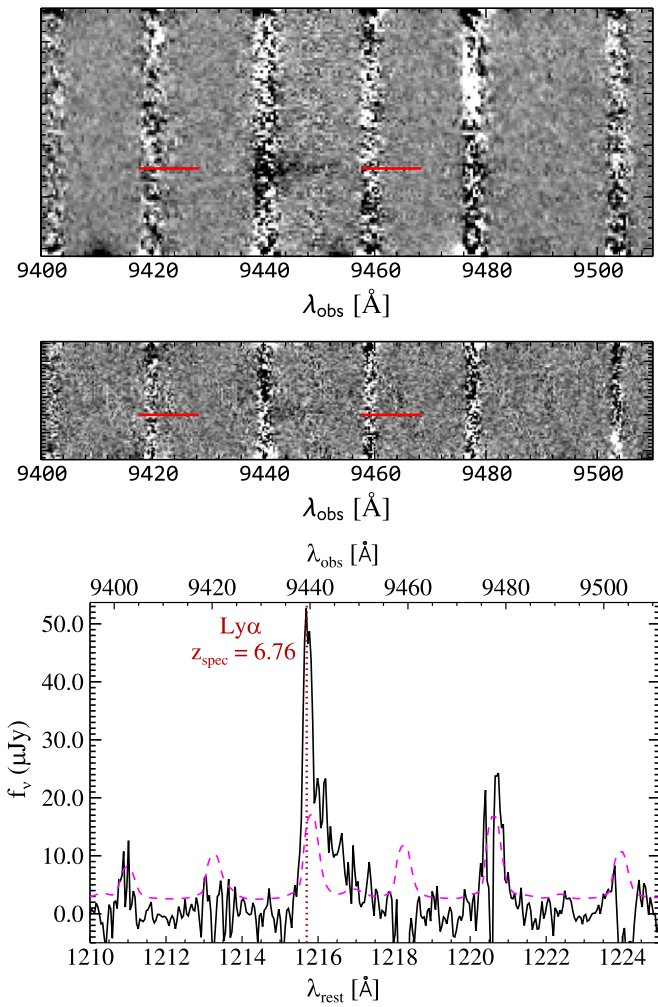


Figure 5. Same as Figure 4, but for the remaining seven LBG candidates.

because our conversion assumes no attenuation of Ly $\alpha$  photons by dust or neutral hydrogen. The inferred SFR is  $1.6 \pm 0.1 M_{\odot} \text{ yr}^{-1}$ , consistent with being a lower limit to the value we derive from SED fitting in Section 5,  $17.0 \pm 0.5 M_{\odot} \text{ yr}^{-1}$ . The Ly $\alpha$ -inferred SFR roughly yields a Ly $\alpha$  escape fraction of  $\sim 10\%$ .

#### 4.2. MACS 0454-1251

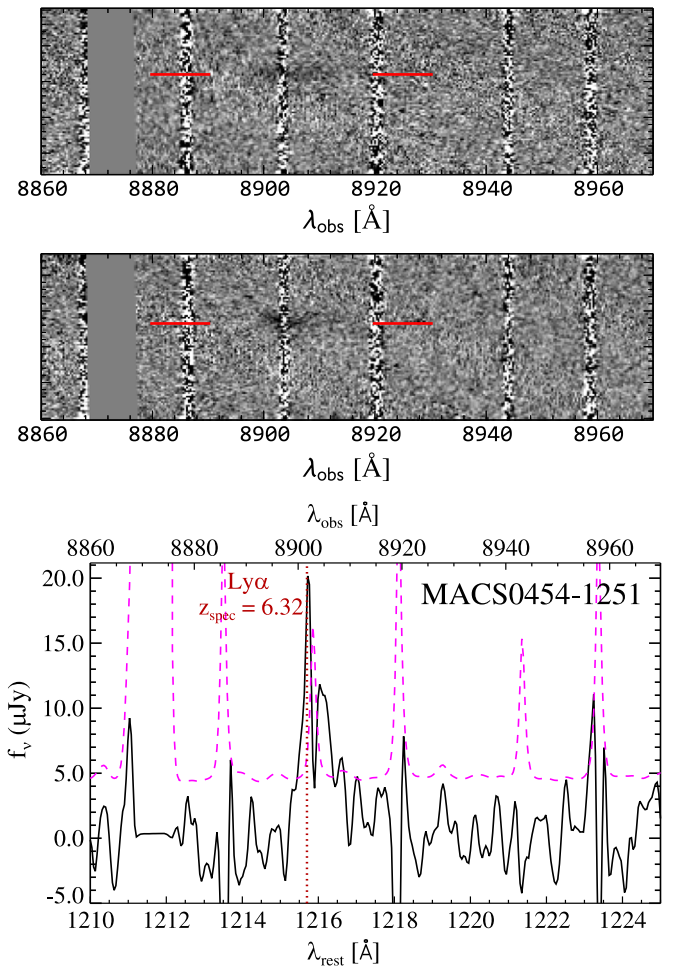
We observed MACS 0454 with DEIMOS on 2014 November 28 and 29, using the  $1200 \text{ l mm}^{-1}$  grating on both nights. The total exposure time for this mask is 7200 s. We also reduce the DEIMOS data and extract one-dimensional



**Figure 6.** Reduced two-dimensional (top and middle panels) and one-dimensional (bottom panel) spectrum of RXJ 1347-1216. The top panel shows the data taken with the  $8301 \text{ mm}^{-1}$  grating on 2013 April 3, and the middle panel shows the data taken with the  $12001 \text{ mm}^{-1}$  grating on 2014 May 26. The one-dimensional spectrum was extracted from the  $8301 \text{ mm}^{-1}$  spectrum. We also plot the rms spectra in dashed lines and mark the emission line redshift if the line is to be Ly $\alpha$ . The flux density values given on the ordinate are calculated in the rest-frame assuming the line to be Ly $\alpha$ .

spectrum following the procedure in Lemaux et al. (2009), in the same way as RXJ 1347-1216.

We show the reduced two-dimensional spectra of MACS 0454-1251 in Figure 7, and an extended emission feature is clearly detected on both nights. However, a bright sky line residual cuts through the middle of the emission feature in the spectral direction and makes the line interpretation ambiguous. Given the width of the emission line, it could be either Ly $\alpha$  at  $z = 6.32$  or [O II] at  $z = 1.39$ , but the sky line residual makes it difficult to either confirm or rule out Ly $\alpha$  or [O II] based on line shape alone. However, as we will show in Section 5.2, this line is more likely to be Ly $\alpha$  than [O II] because its *HST* fluxes (and upper limits) are much better fit by a galaxy template at  $z = 6.32$  than at  $z = 1.39$  (Section 5.2), and its photometric redshift probability density function  $P(z)$  has very a low probability at  $z = 1.39$  (see the  $P(z)$  curve in Figure 10). We measure the line fluxes from both nights to be  $1.2 \pm 0.2 \times 10^{-17} \text{ erg cm}^{-2} \text{ s}^{-1}$  (first night) and  $8.0 \pm 1.5 \times 10^{-18} \text{ erg cm}^{-2} \text{ s}^{-1}$  (second night), and these translate to Ly $\alpha$  luminosities of  $5.5 \pm 0.9 \times 10^{42} \text{ erg s}^{-1}$  (first night) and



**Figure 7.** Reduced two-dimensional (top and middle panels) and one-dimensional (bottom panel) spectrum of MACS 0454-1251. The top panel shows the data taken on 2014 November 27, and the middle panel shows the data taken on 2014 November 28; we observed with the  $12001 \text{ mm}^{-1}$  grating on both nights. The one-dimensional spectrum (bottom panel) was extracted from the data in the first night (top panel). We also plot the rms spectra in dashed lines and mark the emission lines redshift if the line is to be Ly $\alpha$ . The flux density values given on the ordinate are calculated in the rest-frame assuming the line to be Ly $\alpha$ .

$3.6 \pm 0.3 \times 10^{42} \text{ erg s}^{-1}$  (second night). From the line fluxes, we estimate its rest-frame equivalent widths (assuming Ly $\alpha$ ) using the continuum on the red side of Ly $\alpha$  (estimated from its F110W flux density) from both nights to be  $8.2 \pm 1.4$  and  $5.4 \pm 1.0 \text{ \AA}$ . We also infer SFR lower limits to be  $2.3 \pm 0.4 M_{\odot} \text{ yr}^{-1}$  (first night) and  $1.5 \pm 0.3 M_{\odot} \text{ yr}^{-1}$  (second night) based on Ly $\alpha$  fluxes, also fully consistent with being lower limits to the value derived from SED fitting in Section 5,  $17.0^{+18.0}_{-4.1} M_{\odot} \text{ yr}^{-1}$ . The Ly $\alpha$ -inferred SFR also roughly corresponds to an escape fraction of  $\sim 10\%$  for Ly $\alpha$  photons.

The inverse asymmetry value measured for the one-dimensional emission feature is  $\sim 0.5$  for both masks, consistent with the values typical for Ly $\alpha$ . However, the line asymmetry estimate is less reliable than that of RXJ 1347-1216 because we have to mask the over-subtracted skyline near the central wavelength of the emission feature. Based on the object's photometric information and line asymmetry measurements, we identify this source as a Ly $\alpha$  emitter at  $z = 6.32$ , although we are less confident with this Ly $\alpha$  interpretation than we are with RXJ 1347-1216. MACS 0454 is not in the GLASS

sample, so we are unable to cross check our measurements with *HST* grism data.

## 5. REDSHIFT AND SED MODELING

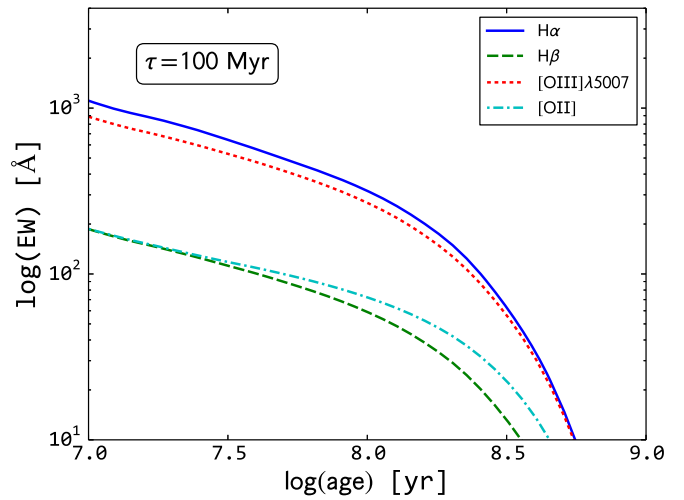
In this section, we present our stellar population modeling of the IRAC-detected,  $z \gtrsim 6$  galaxy candidates using broadband photometry. We present the modeling procedure in Section 5.1, the modeling results in Section 5.2, and we discuss the sources of bias and uncertainty in Section 5.3.

### 5.1. Modeling Procedure

For photometric redshift and stellar population modeling, we use the photometric redshift code EAZY (Brammer et al. 2008) with stellar population templates from Bruzual & Charlot (2003, BC03) with Chabrier (2003) initial mass function (IMF) between 0.1 and  $100 M_{\odot}$  and a metallicity of  $0.2 Z_{\odot}$ . There is very little direct observational evidence for galaxy metallicity at  $z \geq 3$ , but limited results so far suggest that the majority of them have sub-solar metallicity (Maiolino et al. 2008); we choose  $0.2 Z_{\odot}$  for easy comparison with other works. The galaxy templates are generated assuming an exponentially declining star formation history with  $e$ -folding time  $\tau$  ranging from 0.1 and 30 Gyr, and ages of the stellar population range from 10 Myr to 13 Gyr. For each combination of age and  $\tau$ , we implement galaxy internal dust attenuation using the Calzetti et al. (2000) prescription, with the reddening parameter  $E(B - V)$  ranging from 0 to 1 mag to include potential low- $z$  dusty galaxy solutions. The  $E(B - V)$  grid we use have a step size of  $\Delta E(B - V) = 0.02$  mag from  $E(B - V) = 0$  to 0.5 mag and a step size of  $\Delta E(B - V) = 0.1$  mag from  $E(B - V) = 0.5$ –1 mag. We also use the stellar templates from the photometric code *Le Phare* (Ilbert et al. 2006)<sup>20</sup> in the fitting to check if stellar templates provide significantly better fits.

Recent studies have shown that for some galaxy candidates, strong nebular emission lines contribute significantly to broadband fluxes and therefore influence the inferred galaxy properties (e.g., Schaerer & De Barros 2010; Smit et al. 2014). Therefore, we use galaxy templates that include nebular emission lines in the modeling. In order to calculate the expected line fluxes for a given BC03 galaxy template, we calculate the integrated Lyman continuum flux (before dust attenuation) and use the relation from Leitherer & Heckman (1995) to calculate the expected fluxes from hydrogen recombination lines (mainly  $H\alpha$ ,  $H\beta$ ,  $Pa\beta$ , and  $Br\gamma$ ) while assuming the Lyman continuum escape fraction to be zero. Non-zero Lyman continuum escape fraction will reduce the strength of optical nebular emission lines (see Inoue 2011 and Salmon et al. 2015). We then use the tabulated line ratios between  $H\beta$  and the metal lines from Anders & Alvensleben (2003) to calculate the metal line fluxes for a metallicity of  $0.2Z_{\odot}$ . For templates with dust attenuation, we include the dust attenuation effects *after* adding nebular emission lines. The resultant equivalent widths as a function of galaxy age, for  $\tau = 100$  Myr and  $E(B - V) = 0$ , are shown in Figure 8, in agreement with Leitherer & Heckman (1995).

In addition to nebular emission lines, we also include nebular *continuum* emission that account for the bound-free emission of  $H\text{I}$  and  $\text{HeI}$  as well as the two-photon emission of hydrogen



**Figure 8.** Equivalent widths of  $H\alpha$ ,  $H\beta$ ,  $[\text{O III}] \lambda\lambda 3726, 3729$ , and  $[\text{O II}] \lambda 5007$  that we add to the BC03 models as a function of stellar population age, assuming that all Lyman continuum photons are converted into nebular emission. The BC03 galaxy templates used in this figure have a metallicity of  $0.2Z_{\odot}$ , no dust attenuation, and a star formation rate  $e$ -folding time of 100 Myr. Some galaxy candidates (e.g., RXJ 1347-1216) require strong nebular emission lines to explain their observed IRAC colors.

from the 2 s level. We follow the prescription in Krueger et al. (1995, their Equations (7) and (8)) to calculate the nebular continuum flux as a function of Lyman continuum photon density, and we calculate the emission coefficients using the methods in Brown & Mathews (1970) and Nussbaumer & Schmutz (1984).<sup>21</sup> Nebular continuum emission could be an important component for very young ( $\sim 10$  Myr) starbursts and can contribute up to  $\sim 1/3$  of the total continuum just blueward of the rest-frame 4000 Å break. Nebular continuum emission also makes the rest-frame UV slope redder than expected from stars alone (Schaerer & De Barros 2010).

We do not include  $\text{Ly}\alpha$  in our galaxy templates. Strong  $\text{Ly}\alpha$  emission could affect the LBG color selection by changing the rest-frame UV broadband colors and could affect the derived physical properties from SED fitting (Schaerer et al. 2011; De Barros et al. 2014). But  $\text{Ly}\alpha$  photons suffer from complicated radiative transfer processes and does not show tight correlations with the stellar population properties, so for simplicity we do not include  $\text{Ly}\alpha$  emission in our modeling.

Strong gravitational lensing boosts galaxy fluxes and increases the *apparent* SFR and stellar mass. To calculate the *unlensed* SFR and stellar mass, we use the magnification factor  $\mu_{\text{best}}$  estimated from the cluster mass models for each galaxy candidate at its redshift. We generate our own models for MACS 1149, MACS 0717, MACS 0454, RXJ 1347, and RCS 2327, following the procedures outlined in Bradač et al. (2005, 2009). In short, we constrain the gravitational potential on a mesh grid within a galaxy cluster field via  $\chi^2$  minimization, and we adaptively use denser pixel grids near the core(s) of the cluster and around multiple images. We find the minimum  $\chi^2$  values by iteratively solving a set of linearized equations that satisfy  $\partial\chi^2/\partial\psi_k = 0$ , where  $\psi_k$  is the gravitational potential in the  $k$ th dimension. We then produce the magnification ( $\mu$ ) map from the best-fit gravitational potential map. For the rest of the

<sup>20</sup> The fit using stellar templates is still done using EAZY, only the templates are from *Le Phare*.

<sup>21</sup> The fitting formula in Nussbaumer & Schmutz (1984) is crucial to calculate the two-photon continuum emission between rest-frame 1216 and 2431 Å, where two-photon emission dominates the nebular continuum.

clusters in our sample (MACS 1423, MACS 2129, and MACS 0744), we use the public PIEMD-eNFW<sup>22</sup> models by Zitrin et al. (2013).<sup>23</sup> The magnification factors (and their errors) are estimated at the galaxy candidate positions from the  $z = 9$  magnification maps except for MACS 1423-587, MACS 1423-774, MACS 1423-2248, and RXJ 1347-1800 (which have  $z_{\text{best}} \sim 1$  so we estimate their  $\mu_{\text{best}}$  from the  $z = 1$  magnification map).

## 5.2. Modeling Results

We list the best-fit galaxy properties in Table 4 and show the best-fit templates and the photometric redshift probability density function  $P(z)$  (while allowing redshift to float) in Figures 9 and 10. For each galaxy candidate, we estimate the *statistical* uncertainties of stellar population properties using Monte Carlo simulations: we perturb the photometry within the errors (assuming Gaussian flux errors), re-fit with the same set of galaxy templates, and collected the distributions of each best-fit property. We only perturb the fluxes where  $S/N \geq 1$ . For upper limits, we do not perturb the fluxes in our simulations. The systematic errors related to assumptions in IMF, galaxy metallicity, and the functional form of star formation history are not represented by the error bars. We show the distributions from Monte Carlo simulations for stellar mass, SFR, and stellar population age in Appendix. From these distributions, we derive the confidence intervals that bracket 68% of the total probability in Table 4. The error bars do *not* include uncertainties in  $\mu_{\text{best}}$ .

We also show the best-fit galaxy properties for RXJ 1347-1216 and MACS 0454-1251, the two galaxies for which we have line detections from DEIMOS data (see Section 4), when we fix their redshifts at their Ly $\alpha$  redshifts in Figure 11 (assuming both lines are Ly $\alpha$ ). For RXJ 1347-1216, its photometric redshift is already sharply peaked at  $z = 6.7$ , so the best-fit template and physical properties do not change after fixing its redshift at the Ly $\alpha$  redshift. On the other hand, MACS 0454-1251 has slightly different best-fit photometric redshift and Ly $\alpha$  redshift, and we also list its best-fit properties at  $z_{\text{Ly}\alpha} = 6.32$  in Table 4. We also show the best-fit template at  $z = 1.39$  in Figure 11 if the detected emission line is [O II] instead of Ly $\alpha$  and see that the  $z = 1.39$  solution has a higher  $\chi^2_\nu$  ( $\chi^2_\nu = 4.02$ ) than the  $z = 6.32$  solution ( $\chi^2_\nu = 2.90$ ). The likelihood ratio of these two fits, calculated using the total  $\chi^2$  as  $e^{-\Delta\chi^2}$ , suggests that the  $z = 6.32$  model is  $\sim 8000$  times more likely than the  $z = 1.36$  model.

The best-fit stellar masses in our sample range from  $2 \times 10^8 M_\odot$  to  $2.9 \times 10^9 M_\odot$  after correcting for magnification by the foreground clusters and excluding the  $z \sim 1$  interlopers. The stellar masses inferred from SED fitting have smaller statistical errors when *HST* fluxes are combined with IRAC fluxes because of the constraints on rest-frame optical emission from IRAC. We show the range of stellar mass from our Monte Carlo simulations in the Appendix (Figure 15), and we find that including IRAC fluxes tighten up the possible range of stellar mass for every object. We also see that the range of stellar mass spanned by our IRAC-detected sample are not necessarily at the high-mass end of the observed (Gonzalez et al. 2011) and

simulated (Katsianis et al. 2015) stellar mass functions at  $z \gtrsim 6$ . In fact, the IRAC [3.6]–[4.5] color for several of our galaxy candidates (e.g., RXJ 1347-1216) suggest extremely young stellar population ages ( $\sim 10$  Myr) and large equivalent widths from nebular emission lines. For these sources, stellar continuum emission might not dominate the observed IRAC fluxes, hence their true stellar masses depend sensitively on the equivalent widths of nebular emission lines. This demonstrates the combined power of strong gravitational lensing and deep IRAC images that allows one to measure the stellar mass of  $z \gtrsim 6$  galaxies further down the stellar mass function.

On the other hand, the SFRs and stellar population ages are not necessarily well constrained by SED fitting even when IRAC fluxes are included (see Figures 16 and 17 in the Appendix). The SFRs of high- $z$  galaxies are often calculated from their rest-frame UV fluxes (after correcting for dust attenuation), and these are often the only constraints available from observations. However, the UV-derived SFR depends critically on the amount of dust attenuation inside each galaxy, and the effect of dust on the rest-frame UV color is degenerate with the effect of stellar population age. Furthermore, UV-derived SFRs probe the star formation activity over the past  $\sim 100$  Myr, so it could underestimate the *instantaneous* SFR if the stellar population is younger than  $\sim 100$  Myr; for these systems, nebular emission line fluxes (e.g., H $\alpha$  or [O II]) are better proxies for SFRs (Kennicutt 1998). We consider SFRs and stellar population ages more poorly constrained compared with stellar mass, and we will discuss the degeneracies in SED fitting in Section 5.3.

IRAC fluxes also reveal four  $z \sim 1$  interlopers from our sample—MACS 1423-587, MACS 1423-774, MACS 1423-2248, and RXJ 1347-1800—as shown in the three bottom-right panels in Figure 10. All four sources have significant integrated probabilities at  $z \geq 6$  when only *HST* photometry is used in the fitting, but the addition of their IRAC fluxes pushes their photometric redshifts down to  $z \sim 1$ , suggesting that the observed breaks between F850LP and F105W are more likely the rest-frame 4000 Å break instead of the Ly $\alpha$  break. This demonstrates the value of IRAC detections in discriminating between genuine  $z \gtrsim 6$  galaxies and lower- $z$  interlopers.

In Figures 9 and 10 we also show the best-fit stellar templates from *Le Phare*. Most of the sources are better fit by galaxy templates than by stellar templates, although there is only one case, RXJ 1347-1800, where both templates provide similarly good fits ( $\chi^2_\nu = 0.57$  for galaxy templates and  $\chi^2_\nu = 0.62$  for stellar templates). For all of the galaxy candidates, their best-fit stellar templates are either brown dwarfs or low-mass stars from Chabrier & Baraffe (2000). We also check if our sample contains X-ray detected sources that could have significant contributions from AGNs and find no evidence that any of our galaxy candidates have AGNs. However, low-level AGN activity is still possible, and the stellar population properties we infer from SED modeling will depend on how much (if any) AGN contribution there is to their broadband fluxes.

## 5.3. Modeling Biases and Uncertainties

The biases and uncertainties of SED fitting are well documented in the literature (e.g., Papovich et al. 2001; Lee et al. 2009), and most sources of systematic errors come from model assumptions. In general, stellar mass has the smallest systematic errors ( $\sim 0.3$  dex), but uncertainties in galaxy star

<sup>22</sup> PIEMD-eNFW models use pseudo-isothermal elliptical mass distributions for galaxies and elliptical NFW profiles for dark matter.

<sup>23</sup> These models are made public as high-end science products of the CLASH program; <http://archive.stsci.edu/prepds/clash/>.

**Table 4**  
Photometric Redshift and Stellar Population Modeling Results

Object ID	$z_{\text{best}}^a$	$\mu_{\text{best}}^b$	$M_{\text{stellar}} \times f_{\mu}^c$ ( $10^9 M_{\odot}$ )	$\text{SFR} \times f_{\mu}^c$ ( $M_{\odot} \text{ yr}^{-1}$ )	Age <sup>d</sup> (Myr)	sSFR <sup>e</sup> ( $\text{Gyr}^{-1}$ )	$E(B - V)_{\text{fit}}^f$ (mag)	$\beta^g$	$M_{1600} - 2.5 \log(f_{\mu})^h$ (mag)
F125W-dropouts ( $z \sim 9$ ; $M_{\text{UV}}^* = -20.63 \pm 0.36$ at $z \sim 8$ from Bouwens et al. 2015)									
MACS 1149-JD	$9.3^{+0.1}_{-0.1}$	$5.5^{+0.3}_{-0.3}$	$0.9^{+0.5}_{-0.4}$	$1.9^{+0.6}_{-0.5}$	$200^{+250}_{-60}$	$2.1^{+3.1}_{-0.7}$	0.00	...	$-19.9 \pm 0.1$
F105W-dropouts ( $z \sim 8$ ; $M_{\text{UV}}^* = -20.63 \pm 0.36$ at $z \sim 8$ from Bouwens et al. 2015)									
RXJ 1347-1080	$7.3^{+0.3}_{-0.4}$	$6.0^{+0.6}_{-0.6}$	$0.6^{+0.2}_{-0.4}$	$0.5^{+1.0}_{-0.5}$	$290^{+350}_{-260}$	$0.9^{+1.9}_{-0.9}$	0.00	$-2.5^{+1.2}_{-1.1}$	$-18.9 \pm 0.2$
MACS 1423-1384	$6.9^{+0.9}_{-0.1}$	$4.0^{+1.2}_{-1.2}$	$1.1^{+3.2}_{-0.6}$	$115.6^{+16.1}_{-114.6}$	$\leq 130$	$105.0^{+0.1}_{-98.9}$	0.38	$-0.5^{+1.6}_{-0.6}$	$-19.4 \pm 0.2$
F850LP-dropouts ( $z \sim 7$ ; $M_{\text{UV}}^* = -20.87 \pm 0.26$ from Bouwens et al. 2015)									
MACS 1423-1494	$7.1^{+0.3}_{-0.5}$	$1.7^{+0.1}_{-0.1}$	$0.2^{+1.2}_{-0.1}$	$22.1^{+4.3}_{-19.3}$	$\leq 10$	$105.1^{+0.1}_{-9.2}$	0.14	$-2.1^{+1.1}_{-1.4}$	$-20.1 \pm 0.2$
MACS 0744-2088	$7.0^{+0.2}_{-0.1}$	$1.5^{+0.1}_{-0.1}$	$0.7^{+1.2}_{-0.4}$	$34.0^{+4.0}_{-28.1}$	$20^{+180}_{-0}$	$45.7^{+16.8}_{-42.3}$	0.16	$-0.8^{+0.7}_{-1.2}$	$-20.7 \pm 0.1$
RXJ 1347-1216 <sup>i</sup>	6.76	$5.0^{+0.5}_{-0.5}$	$0.2^{+0.1}_{-0.1}$	$17.0^{+2.6}_{-2.7}$	$\leq 10$	$105.0^{+0.1}_{-0.1}$	0.20	$-2.5^{+0.7}_{-1.0}$	$-19.3 \pm 0.1$
MACS 1423-2097	$6.8^{+0.1}_{-0.2}$	$3.5^{+0.1}_{-0.1}$	$2.9^{+0.3}_{-2.5}$	$1.5^{+20.0}_{-0.1}$	$720^{+90}_{-490}$	$0.5^{+90.9}_{-0.1}$	0.00	$-0.6^{+0.6}_{-1.1}$	$-19.7 \pm 0.1$
Bullet-3 <sup>j</sup>	$6.8^{+0.1}_{-0.1}$	$12^{+4.0}_{-0.1}$	$2.0^{+0.6}_{-0.8}$	$1.3^{+1.4}_{-0.6}$	$630^{+160}_{-230}$	$0.7^{+0.5}_{-0.6}$	0.00	...	$-18.9 \pm 0.4$
MACS 1423-587	$0.1^{+6.7}_{-0.1}$	$1.5^{+0.1}_{-0.1}$	$0.1^{+0.2}_{-0.1}$	$\leq 0.1$	$11500^{+1250}_{-11490}$	$\leq 0.1$	0.70	...	...
RXJ 1347-1800	$0.8^{+0.7}_{-0.5}$	$4.1^{+0.1}_{-0.1}$	$0.1^{+0.1}_{-0.1}$	$\leq 1.2$	$2600^{+7830}_{-2570}$	$\leq 36.0$	0.00	...	...
MACS 1423-774	$1.2^{+5.5}_{-0.3}$	$1.2^{+0.1}_{-0.1}$	$0.3^{+0.5}_{-0.2}$	$\leq 20.6$	$1430^{+180}_{-1420}$	$\leq 100.1$	0.06	...	...
MACS 1423-2248	$1.2^{+0.3}_{-1.0}$	$1.2^{+0.1}_{-0.1}$	$1.0^{+2.7}_{-1.0}$	$\leq 5.2$	$5000^{+3640}_{-4980}$	$\leq 52.9$	0.00	...	...
F814W-dropouts ( $z \sim 6-7$ ; $M_{\text{UV}}^* = -20.87 \pm 0.26$ from Bouwens et al. 2015)									
RCS 2327-1282	$7.7^{+0.1}_{-0.4}$	$4.1^{+0.5}_{-0.4}$	$1.9^{+0.4}_{-0.5}$	$6.1^{+0.5}_{-0.2}$	$400^{+100}_{-120}$	$3.2^{+1.6}_{-0.4}$	0.00	$-3.0^{+0.2}_{-0.3}$	$-20.9 \pm 0.1$
MACS 0454-1817	$6.5^{+0.2}_{-0.1}$	$2.6^{+0.3}_{-0.3}$	$0.4^{+3.8}_{-0.1}$	$38.6^{+28.9}_{-36.3}$	$\leq 30$	$105.0^{+0.1}_{-67.6}$	0.28	$-1.4^{+0.5}_{-0.8}$	$-19.3 \pm 0.1$
MACS 0454-1251	$6.1^{+0.1}_{-0.1}$	$4.4^{+0.4}_{-0.4}$	$2.1^{+2.8}_{-0.8}$	$17.9^{+14.1}_{-5.8}$	$90^{+40}_{-10}$	$8.7^{+4.8}_{-2.5}$	0.12	$-1.6^{+0.2}_{-0.2}$	$-20.8 \pm 0.1$
MACS 0454-1251 <sup>k</sup>	6.32	$4.4^{+0.4}_{-0.4}$	$0.5^{+2.9}_{-0.2}$	$19.0^{+7.9}_{-5.4}$	$30^{+0}_{-0}$	$39.6^{+65.5}_{-35.3}$	0.08	$-1.6^{+0.2}_{-0.2}$	$-20.9 \pm 0.1$
F775W-dropouts ( $z \sim 6$ ; $M_{\text{UV}}^* = -20.94 \pm 0.20$ from Bouwens et al. 2015)									
MACS 1149-274	$5.8^{+0.1}_{-0.1}$	$1.6^{+0.1}_{-0.1}$	$2.4^{+0.5}_{-0.5}$	$17.7^{+5.2}_{-0.1}$	$100^{+10}_{-20}$	$7.3^{+3.2}_{-1.2}$	0.08	$-1.6^{+0.1}_{-0.1}$	$-21.2 \pm 0.1$
MACS 1149-1204	$5.7^{+0.1}_{-0.1}$	$1.8^{+0.1}_{-0.1}$	$1.3^{+0.3}_{-0.8}$	$12.9^{+4.7}_{-2.0}$	$80^{+30}_{-50}$	$10.2^{+29.2}_{-2.8}$	0.08	$-1.5^{+0.4}_{-0.5}$	$-20.7 \pm 0.2$

#### Notes.

<sup>a</sup>  $z_{\text{best}}$  is the photometric redshift using BC03 galaxy templates except for RXJ 1347-1216 and MACS 0454-1251, for which we identify Ly $\alpha$  emission at  $z_{\text{spec}} = 6.76$  and  $z_{\text{spec}} = 6.32$ , respectively.

<sup>b</sup> Lensing magnification factor estimated from the galaxy cluster mass models mentioned in Section 5.1. For MACS 1423-587, MACS 1423-774, MACS 1423-2248, and RXJ 1347-1800, we estimate their  $\mu_{\text{best}}$  from the magnification map at  $z = 1$ .

<sup>c</sup> The intrinsic stellar mass and SFR assuming  $\mu = \mu_{\text{best}}$ . To use a different magnification factor  $\mu$ , simply use  $f_{\mu} \equiv \mu/\mu_{\text{best}}$ , where  $\mu_{\text{best}}$  is the best magnification factor we adopt for each object. When the best-fit SFR is zero, we report the 68% upper limit.

<sup>d</sup> Time since the onset of star formation. For the sources with best-fit age equal to 10 Myr, the youngest template allowed in our models, we report the 68% upper limit of the age from Monte Carlo simulations.

<sup>e</sup> Specific star formation rate  $\equiv$  SFR / stellar mass. When the best-fit sSFR is zero, we report the 68% upper limit.

<sup>f</sup> The best-fit color excess  $E(B - V)$  of the stellar emission from our SED modeling. The dust attenuation at rest-frame 1600 Å can be calculated using the dust attenuation curve from Calzetti et al. (2000) as  $A_{1600} = 9.97 \times E(B - V)$ .

<sup>g</sup> Measured rest-frame UV slope  $\beta$  from HST/WFC3 broadband fluxes, assuming the candidates are at  $z = z_{\text{best}}$ . We do not measure  $\beta$  for MACS 1423-587, MACS 1423-774, MACS 1423-2248, and RXJ 1347-1800 because they have  $z_{\text{best}} \sim 1$ . We also do not measure  $\beta$  for MACS 1149-JD because it has only 1 filter (F160W) that samples the rest-frame UV continuum.

<sup>h</sup> Rest-frame 1600 Å absolute magnitude assuming  $z = z_{\text{best}}$  and  $\mu = \mu_{\text{best}}$ .

<sup>i</sup> All fits are performed at  $z = 6.76$ , the Ly $\alpha$  redshift. The confidence intervals reflect the maximal range returned from the simulations because the distributions for this object are highly skewed.

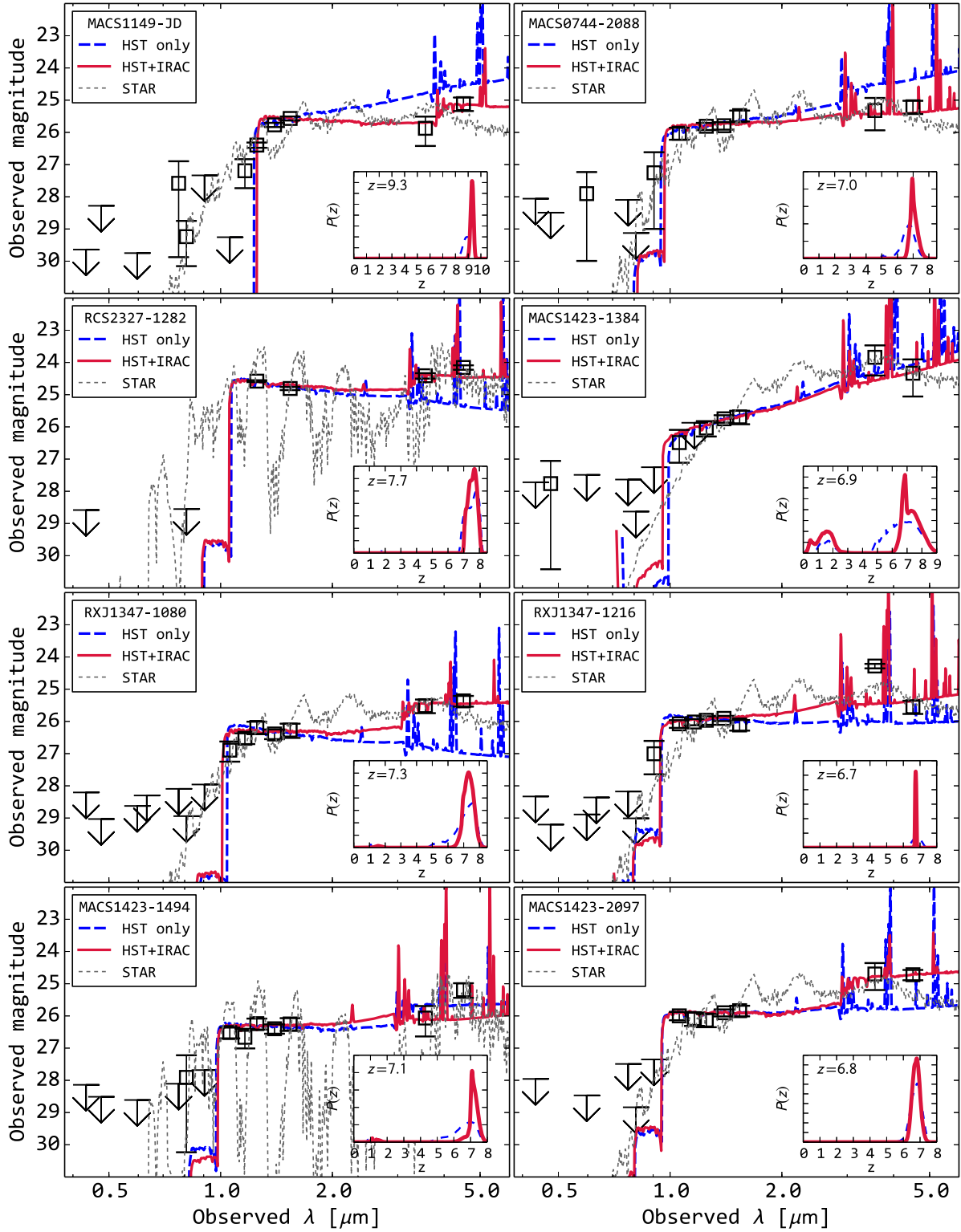
<sup>j</sup> First reported by Ryan et al. (2014); included here for completeness.

<sup>k</sup> All fits are performed at  $z = 6.32$ , its Ly $\alpha$  redshift.

formation history can lead to large biases in galaxy age and SFR (Lee et al. 2009). In addition to the usual culprits of systematic errors (e.g., star formation history, IMF), another important source of systematic error is the nebular emission line ratios. We find that nebular emission line contributions to broadband fluxes are important for a subset of our sample, but direct observational constraints of rest-frame optical nebular emission line ratios of  $z > 6$  galaxies will not arrive until the

launch of the *James Webb Space Telescope*. Uncertainties in the amount of dust attenuation also complicates the interpretation of the best-fit parameters, as we demonstrate below.

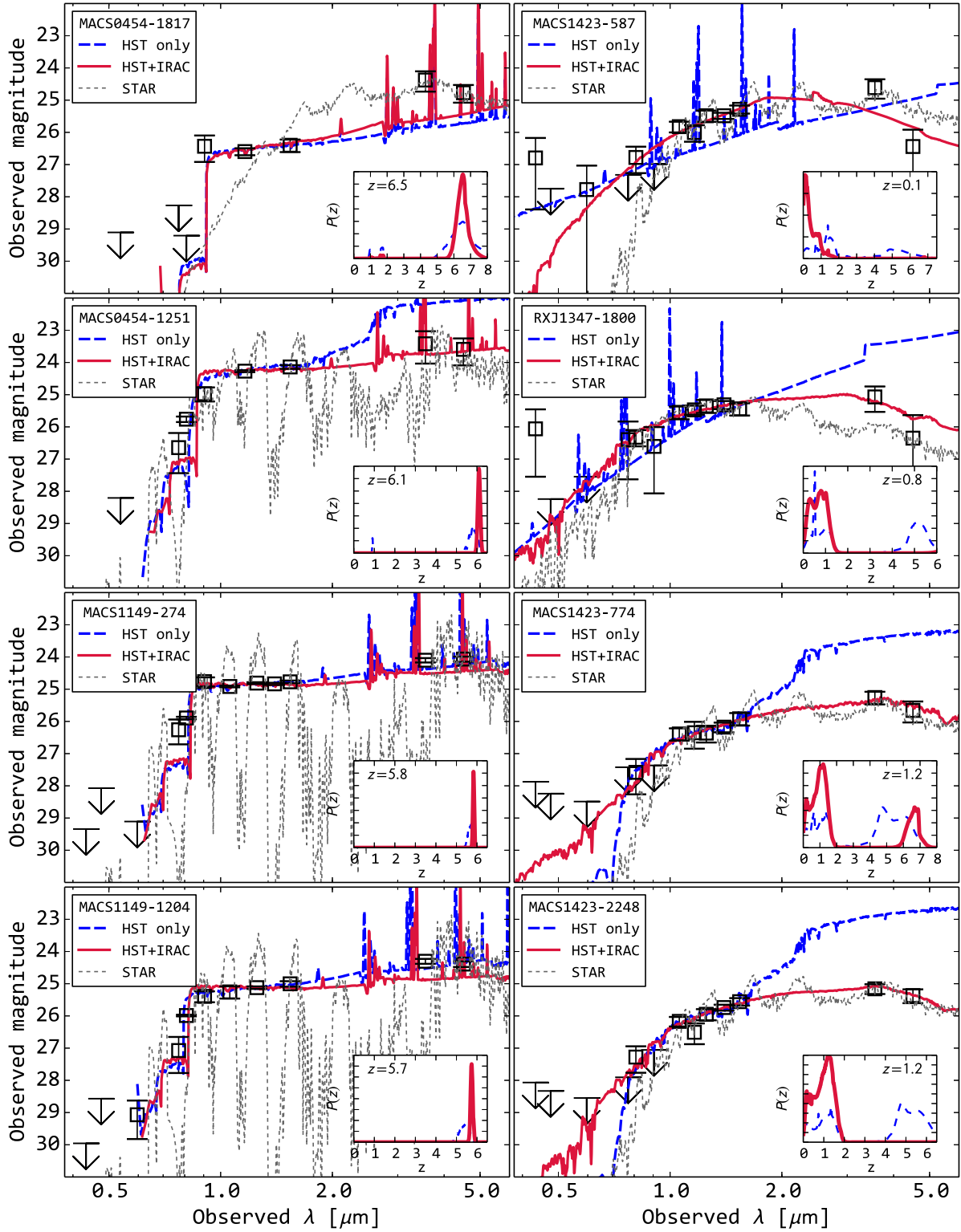
We use MACS 1149-JD to show how uncertainties in dust attenuation can lead to uncertainties in SFR in Figure 12. We estimate the number density contours from our Monte Carlo simulations (with 1000 realizations) when IRAC fluxes are included in the modeling. We show the number density contour



**Figure 9.** Best-fit SEDs for each galaxy candidate with IRAC detections. The best-fit SEDs using only *HST* photometry are shown in dashed blue lines, while the SEDs using combined *HST* and IRAC photometry are shown in solid red lines. The best-fit stellar templates (fixed at  $z = 0$ ) are shown in thin dotted lines. The photometric redshift probability density functions  $P(z)$  are shown as insets. The photometric redshifts decrease from top to bottom first, then from the left column to the right column.

of each  $E(B - V)$  value in the central panel and the marginalized histograms for galaxy age (on top) and SFR (on the right). The SFR histogram shows a single peak at  $1.9 M_{\odot} \text{ yr}^{-1}$  assuming a magnification factor of 5.5, but there is a long tail to higher SFRs that extends one order of

magnitude. The long tail in SFR corresponds to higher dust attenuation templates ( $E(B - V) > 0.1$ ; green dashed contours in the central panel) compared to the peak of the histogram ( $E(B - V) < 0.1$ ; solid blue contours in the central panel). We note that because of the high photometric redshift of

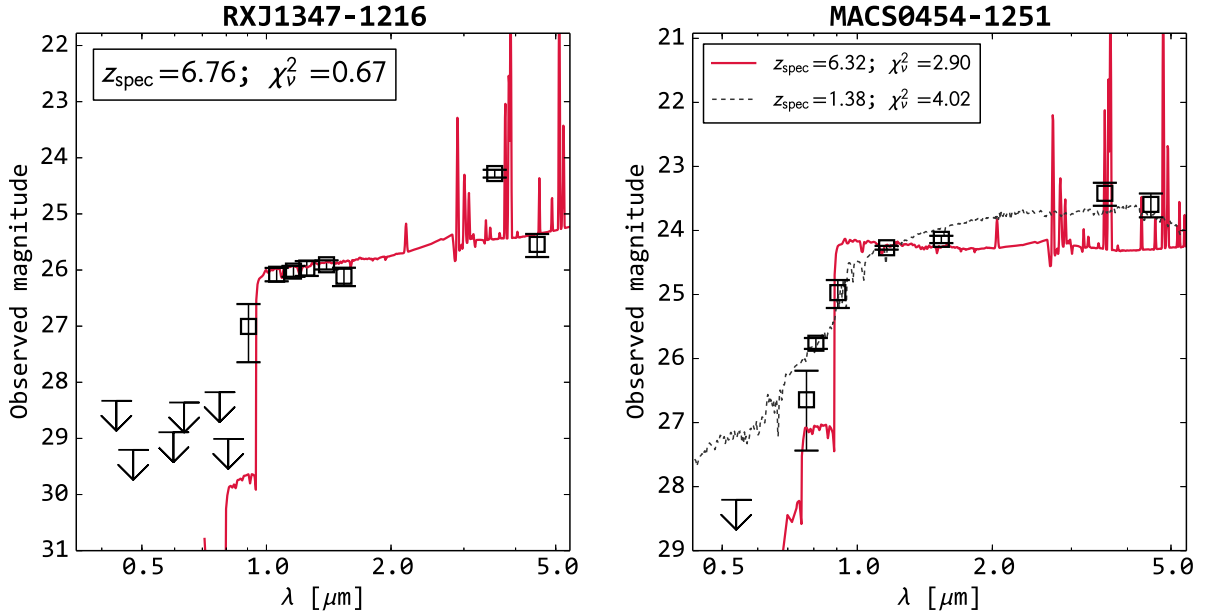


**Figure 10.** Same as Figure 9, for the remaining eight IRAC-detected  $z \gtrsim 6$  galaxy candidates. The two galaxy candidates with  $z_{\text{best}} \sim 1$  are shown here.

MACS 1149-JD ( $z_{\text{phot}} = 9.3$ ), its UV continuum between rest-frame 1250 and 2600 Å is not well sampled by *HST* and IRAC photometry; the addition of *K*-band photometry should help constrain the amount of dust attenuation inside this galaxy.

As a model-independent check on the inferred dust attenuation, we measure the rest-frame UV slope  $\beta$  for each galaxy candidate and list the values in Table 4. We can only

measure  $\beta$  when a galaxy candidate has at least two filters sampling the UV continuum (between rest-frame 1250 and 2600 Å); therefore, we do not measure  $\beta$  for MACS 1149-JD (which only has F160W that samples UV continuum) or for MACS 1423-587, RXJ 1347-1800, MACS 1423-774, and MACS 1423-2248 (which have photometric redshifts  $\sim 1$ ). We measure  $\beta$  by using a power-law spectrum  $f_{\lambda} \propto \lambda^{\beta}$ , convolving



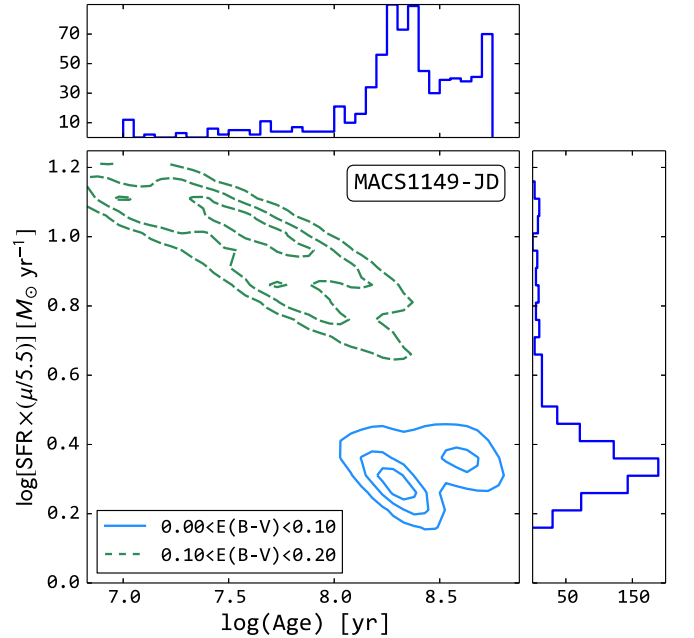
**Figure 11.** Best-fit galaxy templates for RXJ 1347-1216 and MACS 0454-1251 when their redshifts are held at the Ly $\alpha$  redshifts ( $z = 6.76$  and  $z = 6.32$ , respectively). For MACS 0454-1251, we also show the best-fit template if the DEIMOS line detection is [O II] at  $z = 1.39$  instead of Ly $\alpha$ . The best-fit template at  $z = 1.39$  has poorer fits (higher  $\chi^2_{\nu}$ ) than the best-fit template at  $z = 6.32$ , which supports our interpretation of the detection emission line as Ly $\alpha$ .

the spectrum with the filter curves that sample the UV continuum, and finding the  $\beta$  that best matches the observed fluxes. The uncertainties are quantified in bootstrap Monte Carlo simulations, and we show the  $E(B - V)$  values from SED fitting versus  $\beta$  in Figure 13.

In Figure 13, we also show the expected values of  $\beta$  given values of  $E(B - V)$  using two different empirical calibrations. To calculate the expected  $\beta$ , we use the Calzetti et al. (2000) dust attenuation law to calculate the amount of dust attenuation at rest-frame 1600 Å from  $E(B - V)$ :  $A_{1600} = 9.97 \times E(B - V)$ . Then we use the relation between  $A_{1600}$  and  $\beta$  from Meurer et al. (1999, for solar metallicity;  $A_{1600} = 4.43 + 1.99\beta$ ) and Castellano et al. (2014, for sub-solar metallicity;  $A_{1600} = 5.32^{+0.41}_{-0.37} + 1.99\beta$ ) to calculate the expected  $\beta$ . We show the Meurer et al. (1999) relation as a solid line and the Castellano et al. (2014) relation as a dashed line in Figure 13. The scatter of the measured  $\beta$  of our sample is larger than the difference between the two empirical calibrations, but the distribution is roughly consistent with both calibrations. The agreement means that the  $E(B - V)$  values derived from SED fitting is not strongly biased on average, but the large scatter also suggests that for each individual galaxy, the dust attenuation is still poorly constrained.

## 6. IRAC COLORS AND STRONG NEBULAR EMISSION LINES

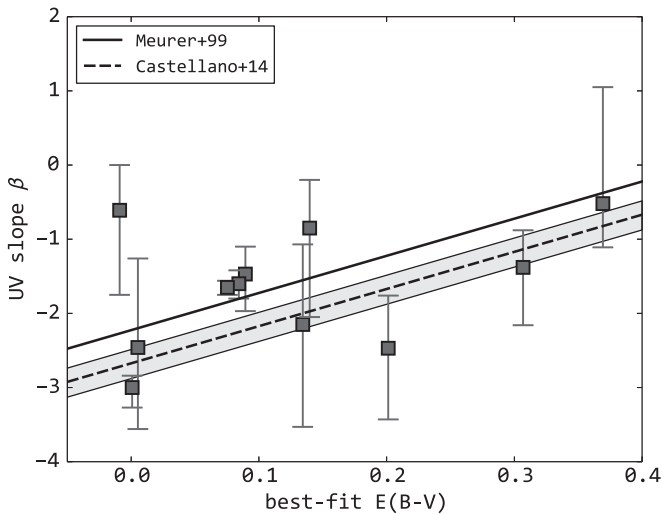
Recent works suggest that, at least in a subset of high- $z$  galaxies, strong nebular emission lines (most notably H $\alpha$ , H $\beta$ , [O III]  $\lambda 5007$ , and [O II]  $\lambda 3727$ ) with rest-frame equivalent widths  $\sim 200$  Å or higher contribute significantly to their IRAC fluxes (e.g., Schaerer & De Barros 2009; Shim et al. 2011; De Barros et al. 2014; Smit et al. 2014). The galaxies with extreme nebular emission line strengths are most likely starbursts younger than 100 Myr; such galaxies are also being found in increasing numbers at  $z \sim 2-3$  (e.g., Atek et al. 2011; van der Wel et al. 2011). If a large number of such galaxies exist at



**Figure 12.** Distribution of galaxy age vs. star formation rate for MACS 1149-JD ( $z_{\text{phot}} = 9.3$ ) from our Monte Carlo simulations. We plot the estimated number density contours for Monte Carlo realizations with different ranges of  $E(B - V)$  values in the central panel to show the correlation between galaxy age, dust attenuation, and star formation rate. We also show the marginalized histograms of star formation rate on the right, and we show the marginalized histograms of galaxy age on the top. The star formation rate is the *intrinsic* value assuming a magnification factor of 5.5 for MACS 1149-JD.

$z \gtrsim 6$ , strong nebular emission lines in the rest-frame UV/optical wavelengths need to be included in the stellar population modeling.

Within certain redshift ranges, unusual IRAC [3.6]–[4.5] colors can be tell-tale signs of strong nebular emission lines. Shim et al. (2011) identified 47 galaxies at  $z \sim 4$  that have

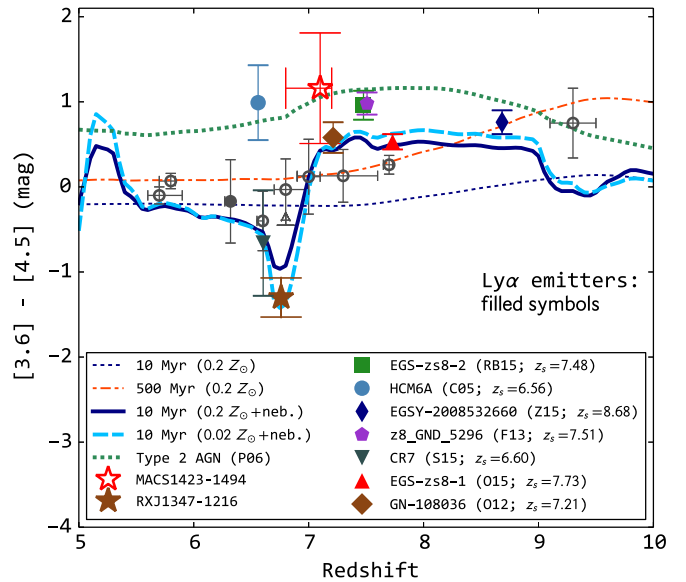


**Figure 13.** Best-fit  $E(B - V)$  vs.  $\beta$  for the IRAC-detected  $6 \lesssim z \lesssim 10$  sample. We also show the expected  $\beta$  from a given  $E(B - V)$  value assuming a dust attenuation law from Calzetti et al. (2000) and the empirical relations between  $A_{1600}$  to  $\beta$  from Meurer et al. (1999, for solar metallicity) and from Castellano et al. (2014, for sub-solar metallicity). We randomly shift the best-fit  $E(B - V)$  for each galaxy candidate around its best-fit value by no more than 0.01 for clarity. As a whole sample, the measured  $\beta$  values are consistent with the expected  $\beta$  values from  $E(B - V)$ , although the scatter is large and the dust attenuation for each galaxy candidate is poorly constrained.

bluer [3.6]–[4.5] colors than those expected from stellar continuum alone, and they categorized these galaxies as  $H\alpha$  emitters because the SED bumps at  $3.6 \mu\text{m}$  are likely due to strong  $H\alpha$  emission. More recently, Smit et al. (2014, 2015) presented  $z \sim 6.6$ – $7.0$  galaxy candidates with unusually blue [3.6]–[4.5] colors as evidence for strong contributions from [O III] and  $H\beta$  to the  $3.6 \mu\text{m}$  fluxes. The colors of these peculiar objects usually can only be reproduced with model SEDs that include strong nebular emission lines.

In Figure 14, we compare the IRAC [3.6]–[4.5] colors of our sample with a range of model predictions. We use our fiducial SED model (BC03) to generate the redshift evolution of [3.6]–[4.5] color at 500 Myr old *without* nebular emission lines (thin dotted–dashed curve, roughly the age of the universe at  $z = 9.5$ ), 10 Myr old *without* nebular emission lines (thin dotted curve), and 10 Myr old *with* nebular emission lines (thick solid curve). The 10 Myr old model with nebular emission lines have equivalent widths  $1087 \text{ \AA}$ ,  $182 \text{ \AA}$ , and  $868 \text{ \AA}$  for  $H\alpha$ ,  $H\beta$ , and [O III] $\lambda\lambda 4959, 5007$ , respectively. Here we assume the star formation  $e$ -folding timescale  $\tau$  to be 100 Myr, but the [3.6]–[4.5] color does not change significantly when different values of  $\tau$  are used.

The most prominent feature in Figure 14 is the “dip” in [3.6]–[4.5] for a 10 Myr old starburst with nebular emission lines at  $z \sim 6.8$  due to the contributions from [O III] and  $H\beta$ —the same feature that Smit et al. (2014) utilized to identify strong nebular emission line objects within  $6.6 \lesssim z \lesssim 7$ . In our sample, only RXJ 1347-1216 has a photometric redshift  $\sim 6.8$  and a very blue [3.6]–[4.5] color. This source has a best-fit age of 10 Myr, the youngest age included in our templates. Extremely young stellar populations are expected to generate a large number of ionizing photons, so if these sources are indeed  $\sim 10$  Myr old starbursts, they might also have high  $\text{Ly}\alpha$  luminosities around star-forming regions. We already successfully identified one of the three sources (RXJ 1347-1216) as a  $z = 6.76$   $\text{Ly}\alpha$  emitter (LAE; see Section 4); we do not identify



**Figure 14.** IRAC [3.6]–[4.5] color as a function of redshift calculated from BC03 model spectra. We show the [3.6]–[4.5] colors for  $0.2Z_{\odot}$  stellar population models that are 500 Myr old *without* nebular emission lines (thin dotted–dashed curve), 10 Myr old *without* emission lines (thin dotted curve), and a 10 Myr old model *with* nebular emission lines (thick solid curve). In our implementation, the dust-free 10 Myr model with nebular emission lines has  $H\alpha$ ,  $H\beta$ , and [O III] $\lambda 5007$  equivalent widths of  $1087 \text{ \AA}$ ,  $182 \text{ \AA}$ , and  $868 \text{ \AA}$ , respectively; for the 10 Myr model, the equivalent widths change by  $<3\%$  within the range of  $e$ -folding time  $\tau$  that we adopt. In addition to the fiducial  $0.2 Z_{\odot}$  models, we also show the expected colors of a  $0.02 Z_{\odot}$ , 10 Myr old model with nebular emission lines for comparison (thick dashed curve); the more metal-poor model seems to reproduce the colors of RXJ 1347-1216. The measured [3.6]–[4.5] colors of two sources in our sample that likely have strong emission lines are shown in stars, and the rest of the sample is shown in black circles. We show the known  $\text{Ly}\alpha$  emitters (LAEs) in filled symbols, including the two LAEs reported in this work (see Section 4). We also show other published  $z > 6$  LAEs with measured [3.6]–[4.5] colors—z8\_GND\_5296 (Finkelstein et al. 2013, F13), HCM6A (Chary et al. 2005, C05), GN-108036 (Ono et al. 2012, O12), CR7 (Sobral et al. 2015, S15), EGS-zs8-1 (Oesch et al. 2015, O15), EGS-zs8-2 (Roberts-Borsani et al. 2015, RB15), and EGSY-2008532660 (Zitrin et al. 2015, Z15)—in filled symbols. The [3.6]–[4.5] color of z8\_GND\_5296, HCM6A, EGS-zs8-2, and MACS 1423-1494 are hard to reproduce by the stellar population models that we adopt, but they come close to the expected colors of a Type 2 (obscured) AGN template from (Polletta et al. 2006, P06).

other sources at  $z \sim 6.8$  with blue [3.6]–[4.5] color that could also be strong line emitters in our sample.

In Figure 14 we also show the redshift evolution of [3.6]–[4.5] color for a 10 Myr old,  $0.02 Z_{\odot}$  model (thick dashed curve), and it predicts a bluer [3.6]–[4.5] color at  $z \sim 6.8$  (as blue as  $\sim -1.4$  mag) than for the 10 Myr old,  $0.2 Z_{\odot}$  model. The IRAC colors of the  $0.02 Z_{\odot}$  model at  $z \sim 6.8$  show better agreements with the three sources mentioned above than the  $0.2 Z_{\odot}$  model, which suggests that these sources might have lower metallicities than our fiducial model. We note that the nebular emission line properties of individual galaxies are highly uncertain (and are sensitive to metallicity), and so any constraint on metallicity is preliminary.

Our fiducial galaxy SED models also predict that young starbursts with strong nebular emission lines should have red [3.6]–[4.5] colors at  $7.0 \lesssim z \lesssim 7.5$ . In this redshift range, [O II] and [O III] move into IRAC ch1 and ch2, respectively, and the combined [O III]+ $H\beta$  line flux is expected to be  $\sim 4$  times higher than [O II] in our implementation (Anders & Alvensleben 2003); the expected [3.6]–[4.5] color reaches

$\sim 0.5$  mag within  $7.0 \lesssim z \lesssim 7.5$ . MACS 1423-1494 ( $z_{\text{phot}} = 7.3$ ) has a photometric redshift and measured [3.6]–[4.5] color that are close to the model prediction, although its red [3.6]–[4.5] color is hard to reconcile even with the 10 Myr old galaxy model. Based on its photometric redshift and unusually red [3.6]–[4.5] color (it is also best fit by a 10 Myr old galaxy SED), we identify this source as another prime candidate for Ly $\alpha$  emission. We note that Ly $\alpha$  photons are subject to complicated radiation transfer effects both inside and outside of galaxies, so it is far from guaranteed that these sources will have detectable Ly $\alpha$  emission. But they are likely LAE candidates (compared with other high- $z$  galaxies) based on their photometric redshifts and IRAC colors.

We also compare our galaxy model-predicted IRAC colors with other  $z \gtrsim 6.5$  LAEs with published IRAC colors in Figure 14. The other LAEs include HCM6A from Hu et al. (2002;  $z_s = 6.56$ , where  $z_s$  is the spectroscopic redshift determined by Ly $\alpha$  emission), CR7 from Sobral et al. (2015;  $z_s = 6.60$ ), GN-108036 from Ono et al. (2012;  $z_s = 7.21$ ), EGS-zs8-2 from Roberts-Borsani et al. (2015;  $z_s = 7.48$ ), z8\_GND\_5296 from Finkelstein et al. (2013;  $z_s = 7.51$ ), EGS-zs8-1 from Oesch et al. (2015;  $z_s = 7.73$ ), and EGSY-2008532660 from Zitrin et al. (2015;  $z_s = 8.68$ ). All of these LAEs have IRAC colors that strongly suggest high nebular emission line equivalent widths (most likely [O III] and H $\beta$  at this redshift range), because they lie along the curve traced by a dust-free,  $0.2 Z_{\odot}$ , 10 Myr stellar population. For example, Finkelstein et al. (2013) argued that the red IRAC color of z8\_GND\_5296 is due to the galaxy’s strong [O III]+H $\beta$  emission lines in IRAC ch2, and they inferred the [O III]  $\lambda$  5007 equivalent width to be 560–640 Å from photometry. The IRAC colors of these  $z \gtrsim 6.5$  LAEs corroborate the recent findings that many galaxies detected at  $z \gtrsim 6$  likely have high nebular emission line equivalent widths.

Two notable cases among the group of LAEs in Figure 14 are MACS 1423-1494 and HCM6A.<sup>24</sup> HCM6A was found in the vicinity of the massive galaxy cluster Abell 370<sup>25</sup> and has a measured [3.6]–[4.5] color of  $1.0 \pm 0.4$  mag, significantly redder than the [3.6]–[4.5] color predicted by a 10 Myr stellar population model at its redshift ( $z_s = 6.56$ ). The red [3.6]–[4.5] color suggests a very high H $\alpha$ /([O III]+H $\beta$ ) ratio, which is unusual (but not impossible) for a young, low-metallicity stellar population. In order to explore other possibilities to explain the red [3.6]–[4.5] colors of both LAEs, we plot the predicted [3.6]–[4.5] colors of a Type 2 obscured AGN template from Polletta et al. (2007). This obscured AGN template includes a dust attenuation of  $A_V = 4$  mag that fits the obscured AGN SW 104409 ( $z = 2.54$ ; Polletta et al. 2006), and its color trajectory in redshift is shown as a thick dotted curve in Figure 14. Interestingly, the predicted [3.6]–[4.5] colors of an obscured AGN agrees quite well with the colors of both MACS 1423-1494 and HCM6A, and z8\_GND\_5296 and EGS-zs8-2 also have marginally consistent IRAC colors with this obscured AGN template. If these sources indeed harbor obscured AGNs (like SW 104409), the red [3.6]–[4.5] colors will be primarily due to large dust attenuation in the rest-frame optical, while the blue rest-frame UV colors come from the scattered light of the central QSO emission. Obscured AGN is an intriguing possibility to consider for these sources, although so far no

direct evidence exists that any of these sources have significant flux contributions from an obscured AGN.

To sum up, we identify three sources in our sample at  $z \sim 6.7$  and  $z \sim 7.3$  as likely young starbursts with very strong nebular emission lines based on their IRAC colors. We detect Ly $\alpha$  emission in one of them, RXJ 1347-1216, during our recent DEIMOS observations, and we plan to follow up all the other three sources for their potential Ly $\alpha$  emission.

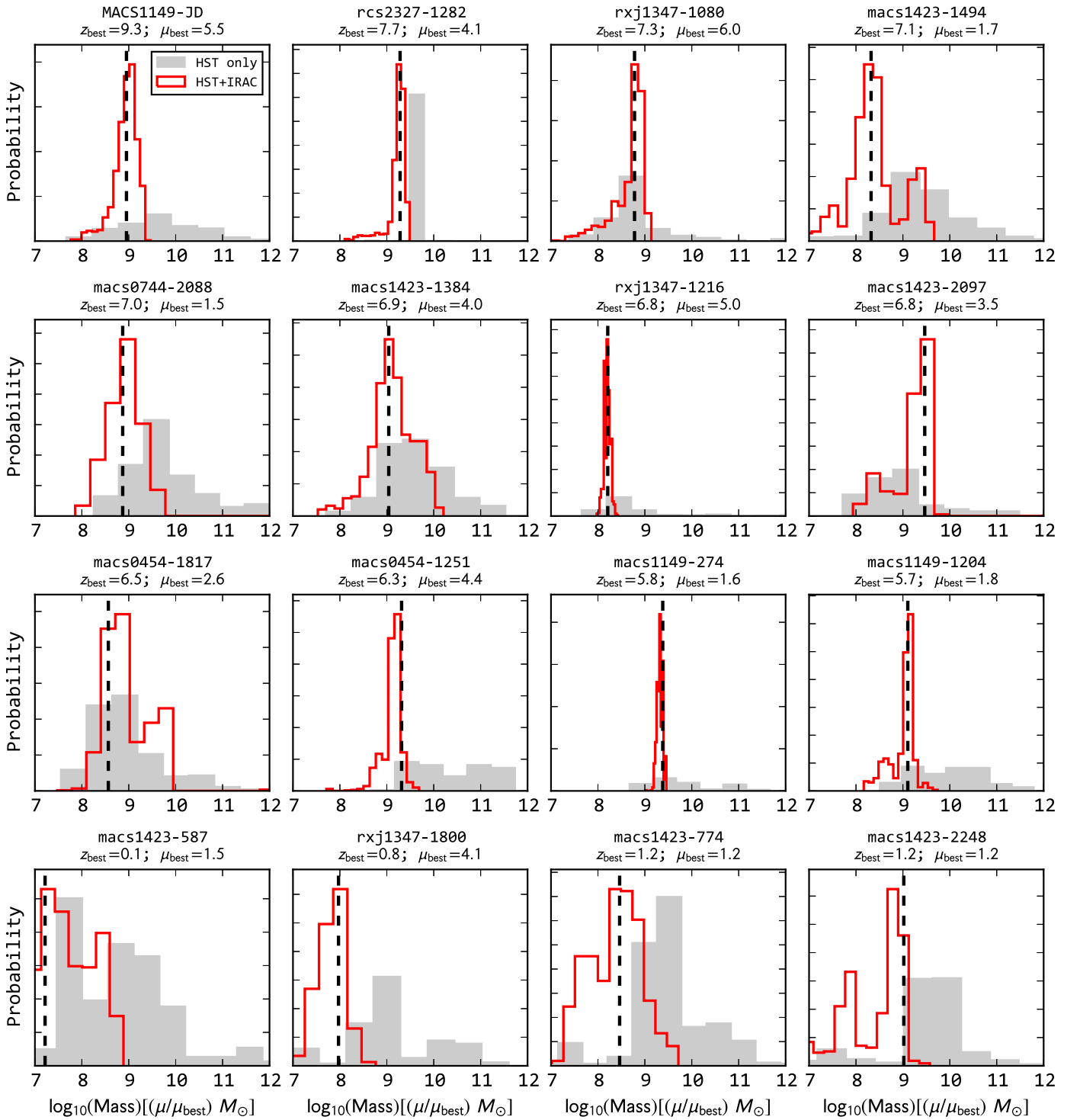
## 7. SUMMARY

In this work, we present the constraints on the  $6 \lesssim z \lesssim 10$ , IRAC-detected galaxy candidates behind eight strong-lensing galaxy clusters from SURFS UP. Six of the clusters are in the CLASH sample, and two are in the HFF sample. We summarize our findings as follows.

1. We find a total of 17 galaxy candidates using the Lyman break color selection that have  $S/N \geq 3$  in at least one IRAC channel. The photometric redshifts in our sample range from 5.7 to 9.3, and we identify four galaxy candidates (MACS 1423-587, RXJ 1347-1800, MACS 1423-774, and MACS 1423-2248) as likely  $z \sim 1$  interlopers after including their IRAC fluxes in the SED modeling. We find the largest number (6) of IRAC-detected galaxy candidates in MACS 1423.
2. From our Keck spectroscopic observations, we identify one secure Ly $\alpha$  emitter at  $z = 6.76$  (RXJ 1347-1216) and one likely Ly $\alpha$  emitter at  $z = 6.32$  (MACS 0454-1251). The line equivalent widths, assuming they are both Ly $\alpha$ , are  $26 \pm 4$  Å (RXJ 1347-1216) and  $6.8 \pm 1.7$  Å (MACS 0454-1251, averaged over two nights). We infer lower limits of their SFRs from their Ly $\alpha$  line fluxes and find them to be consistent with the SFRs from SED fitting.
3. We infer the physical properties of our sample galaxies using Bruzual & Charlot (2003) galaxy templates and add nebular emission lines to the templates. Under our SED modeling assumptions ( $0.2 Z_{\odot}$ , Chabrier IMF, exponentially decaying star formation history, and nebular line emission), the stellar masses of our sample range from  $0.2\text{--}2.9 \times 10^9 M_{\odot}$  (excluding the three likely  $z \sim 1$  interlopers) when we use the best available magnification factors for each galaxy candidate. The magnification-corrected rest-frame 1600 Å absolute magnitude ( $M_{1600}$ ; see Table 4) of our sample ranges from  $-21.2$  to  $-18.9$  mag. The range of intrinsic UV luminosity probed here is slightly fainter than the knee of UV luminosity functions at  $6 \lesssim z \lesssim 10$ , which have  $M_{UV}^*$  between  $\sim -20.6$  and  $\sim -21.6$  mag (e.g., Bouwens et al. 2015; Finkelstein et al. 2015), showing that galaxy clusters’ strong lensing power allows us to start probing the more typical UV luminosities. The range of intrinsic stellar mass probed here is also close to the knee of the stellar mass functions at this redshift range (e.g., Gonzalez et al. 2011; Katsianis et al. 2015). Some galaxies in our sample are best fit by extremely young ( $\sim 10$  Myr old) templates and others best fit by more evolved (up to  $\sim 700$  Myr old at  $z \sim 7$ ) templates, suggesting that the IRAC-detected sample contains both very young galaxies with strong nebular emission lines and more evolved and massive galaxies at  $6 \lesssim z \lesssim 10$ .
4. From the photometric redshifts and IRAC colors, we identify two galaxy candidates that likely have strong

<sup>24</sup> HCM6A was first reported by Hu et al. (2002), and later Chary et al. (2005) published its IRAC fluxes.

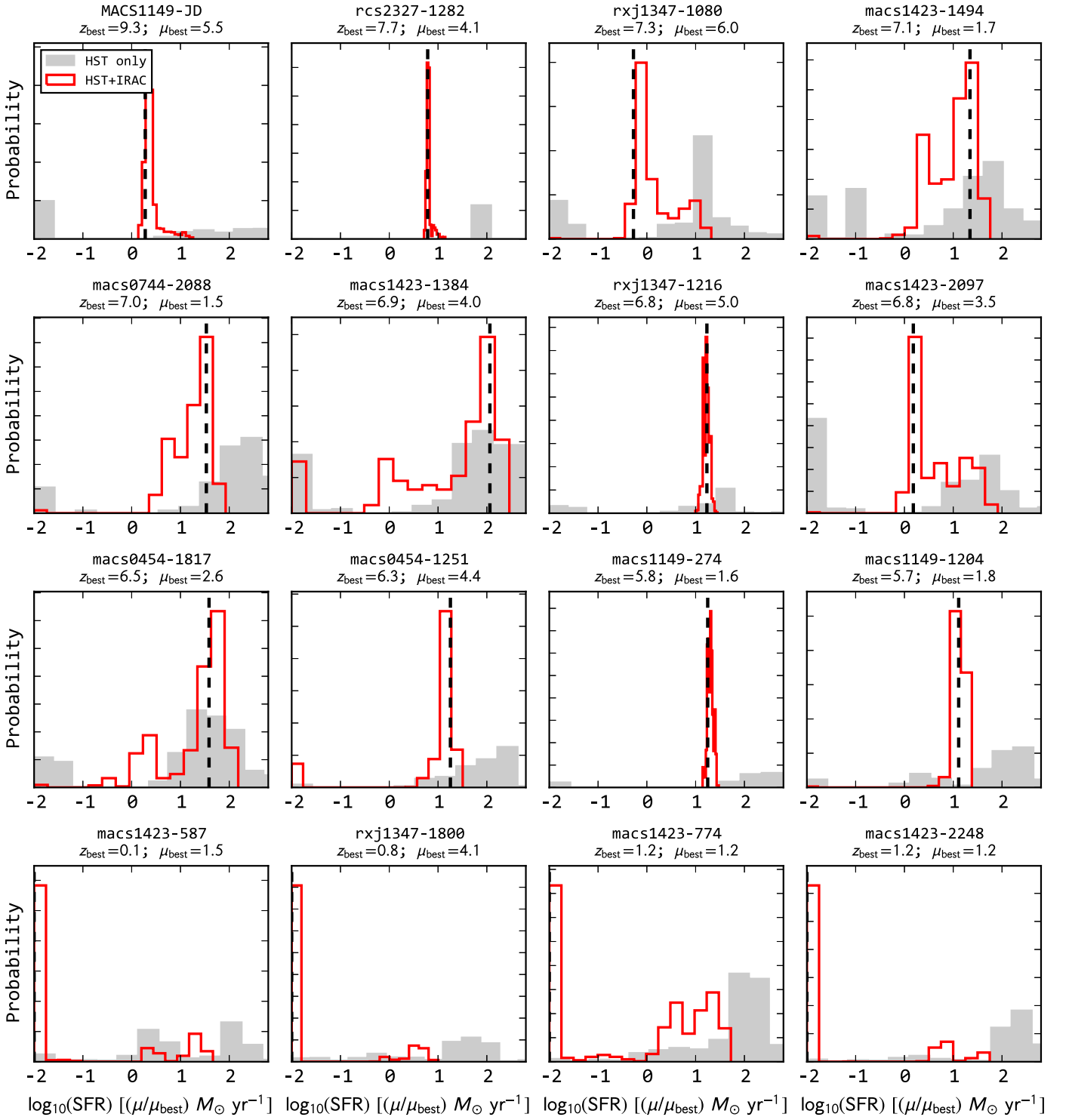
<sup>25</sup> Abell 370 is one of the Hubble Frontier Fields cluster.



**Figure 15.** Stellar mass distributions derived from Monte Carlo simulations for each IRAC-detected  $z \gtrsim 6$  galaxy candidate. The stellar mass values have been scaled by our best estimates of magnification factors  $\mu_{\text{best}}$ . The distributions from combining *HST* and IRAC photometry are shown as the unfilled red histograms; the distributions from *HST* photometry only are shown as filled gray histograms. The best-fit stellar masses from Table 4 are shown as the vertical dashed lines.

(rest-frame optical) nebular emission lines: RXJ 1347-1216 and MACS 1423-1494. Both sources are best fit by the youngest (10 Myr old) galaxy templates included in our modeling and are prime targets for spectroscopic observations. We already identified one of them (RXJ 1347-1216) as a Ly $\alpha$  emitter, and we will target the other one in our future spectroscopic observations. Other

galaxies in the sample lie in the part of the redshift-IRAC color space that makes it hard to infer their nebular emission line strengths; namely, they are within the redshift range where both IRAC bands could have contribution from strong nebular emission lines such as [O III], H $\beta$ , and H $\alpha$ , and their IRAC colors may not be very different from those of pure stellar continuum.

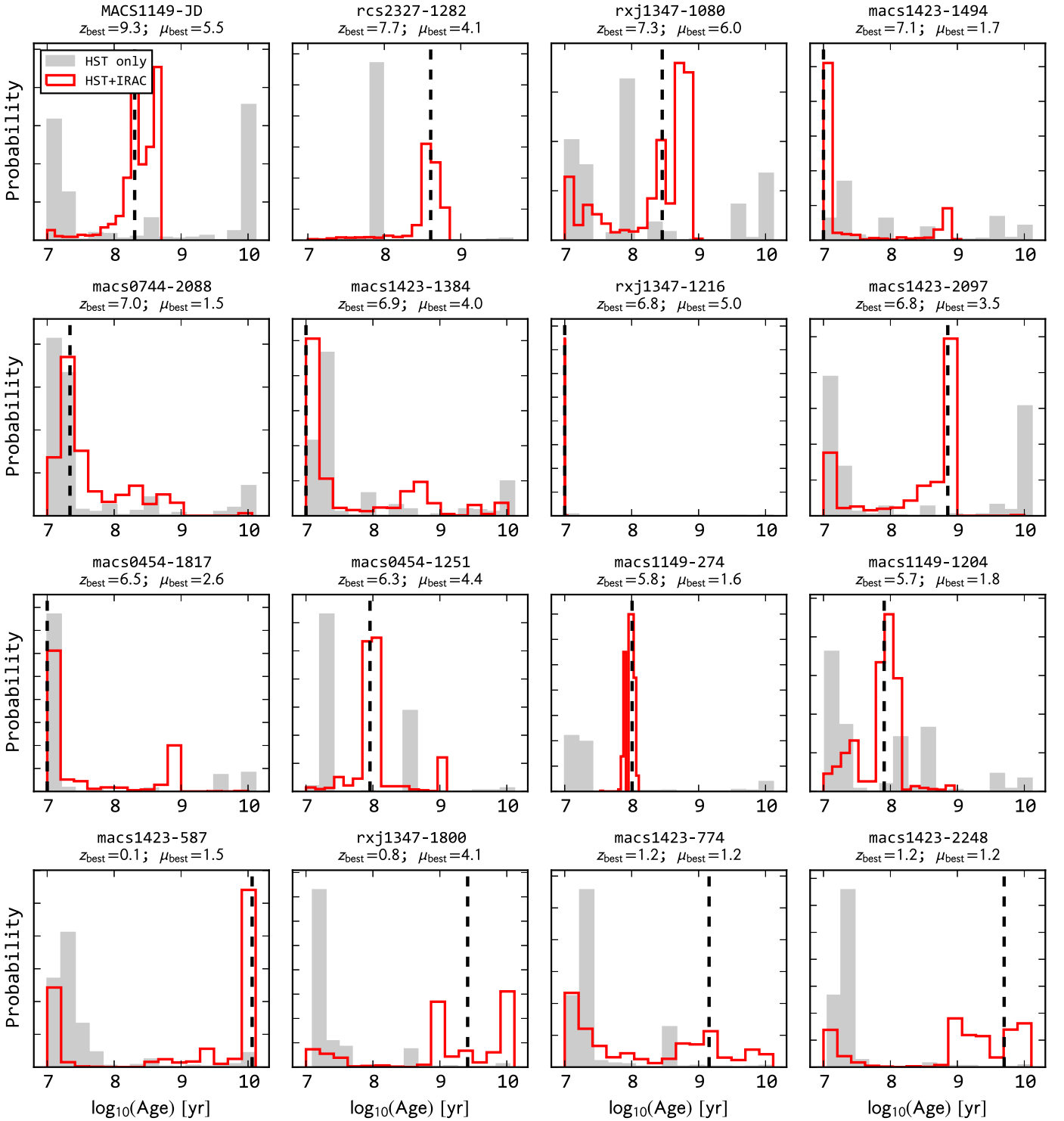


**Figure 16.** Star formation rate (SFR) distributions derived from Monte Carlo simulations for each IRAC-detected  $z \gtrsim 6$  galaxy candidate. The SFR values have been scaled by our best estimates of magnification factors  $\mu_{\text{best}}$ . All SFRs below  $0.01 M_{\odot} \text{ yr}^{-1}$  are set to  $0.01 M_{\odot} \text{ yr}^{-1}$  for clarity. The distributions from combining *HST* and IRAC photometry are shown as the unfilled red histograms; the distributions from *HST* photometry only are shown as filled gray histograms. The best-fit SFR from Table 4 are shown as the vertical dashed lines.

The IRAC fluxes provide important information about the galaxies at  $6 \lesssim z \lesssim 10$ , because it is the only probe of their rest-frame optical emission that we have at the moment. The IRAC-detected galaxies may not be representative of the entire galaxy population at  $z \gtrsim 6$ , but their IRAC colors do provide a more effective way to select spectroscopic targets for redshift confirmation. IRAC fluxes and meaningful upper limits can

also distinguish some lower-redshift galaxies from high- $z$  dropouts and are important for constructing clean  $z \gtrsim 6$  galaxy samples.

We thank the anonymous referee for constructive suggestions that improved this work. We also thank Harry Ferguson, Samuel Schmidt, Chris Fassnacht, Dennis Zaritsky, and



**Figure 17.** Model stellar population age distributions derived from Monte Carlo simulations for each color-selected, IRAC-detected  $z \gtrsim 6$  galaxy candidate. The minimum age included in the template library is 10 Myr. The distributions from combining *HST* and IRAC photometry are shown as the unfilled red histograms; the distributions from *HST* photometry only are shown as filled gray histograms. The best-fit age from Table 4 are shown as the vertical dashed lines.

Hendrik Hildebrandt for useful discussions and comments on the manuscript. Observations were carried out using *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. Also based on observations made with the NASA/ESA *Hubble Space Telescope*, obtained at the Space Telescope Science Institute, which is operated by the

Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555 and NNX08AD79G and ESO-VLT telescopes. Support for this work was provided by NASA through a *Spitzer* award issued by JPL/Caltech. This work was supported by NASA Headquarters under the NASA Earth and Space Science Fellowship Program—Grant ASTRO14F-0007. We also acknowledge support from

HST-AR-13235, HST-GO-13177, and special funding as part of the *HST* Frontier Fields program conducted by STScI. T.S. acknowledges support from the German Federal Ministry of Economics and Technology (BMWi) provided through DLR under project 50 OR 1308. T.T. acknowledges support by the Packard Fellowship. The Dark Cosmology Centre (DARK) is funded by the Danish National Research Foundation.

## APPENDIX

## DISTRIBUTIONS FROM MONTE CARLO SIMULATIONS

Here we show the distributions of stellar mass, SFR, and stellar population age (assuming an exponentially declining star formation history with  $e$ -folding time between 0.1 and 30 Gyr) in Figures 15, 16, and 17, respectively. In all panels, the distributions from using *HST* photometry only are shown as a gray filled histogram, while the distributions from combining *HST* and *Spitzer* photometry are shown as a red histogram.

## REFERENCES

- Anders, P., & Alvensleben, U. F. v. 2003, *A&A*, 401, 1063  
 Atek, H., Siana, B., Scarlata, C., et al. 2011, *ApJ*, 743, 121  
 Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, *AJ*, 132, 1729  
 Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011, *ApJ*, 737, 90  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, 803, 34  
 Bradač, M., Ryan, R., Casertano, S., et al. 2014, *ApJ*, 785, 108  
 Bradač, M., Schneider, P., Lombardi, M., & Erben, T. 2005, *A&A*, 437, 39  
 Bradač, M., Treu, T., Applegate, D., et al. 2009, *ApJ*, 706, 1201  
 Bradley, L. D., Trenti, M., Oesch, P. A., et al. 2012, *ApJ*, 760, 108  
 Bradley, L. D., Zitrin, A., Coe, D., et al. 2014, *ApJ*, 792, 76  
 Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, 686, 1503  
 Brown, R. L., & Mathews, W. G. 1970, *ApJ*, 160, 939  
 Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000  
 Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682  
 Capak, P., Mobasher, B., Scoville, N. Z., et al. 2011, *ApJ*, 730, 68  
 Castellano, M., Sommariva, V., Fontana, A., et al. 2014, *A&A*, 566, 19  
 Chabrier, G. 2003, *PASP*, 115, 763  
 Chabrier, G., & Baraffe, I. 2000, *ARA&A*, 38, 337  
 Chary, R.-R., Stern, D., & Eisenhardt, P. 2005, *ApJL*, 635, L5  
 Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, *ApJS*, 43, 393  
 De Barros, S., Schaerer, D., & Stark, D. P. 2014, *A&A*, 563, 81  
 Ellis, R., Santos, M. R., Kneib, J.-P., & Kuijken, K. 2001, *ApJL*, 560, L119  
 Eyles, L. P., Bunker, A. J., Ellis, R. S., et al. 2007, *MNRAS*, 374, 910  
 Faber, S. M., Phillips, A. C., Kibrick, R. I., et al. 2003, Proc. SPIE, 4841, 1657  
 Finkelstein, S. L., Papovich, C. J., Dickinson, M. E., et al. 2013, *Natur*, 502, 524  
 Finkelstein, S. L., Ryan, R. E., Jr., Papovich, C., et al. 2015, *ApJ*, 810, 71  
 Giallongo, E., Grazian, A., Fiore, F., et al. 2015, *A&A*, 578, 83  
 Giavalisco, M. 2002, *ARA&A*, 40, 579  
 Gonzalez, V., Labbé, I., Bouwens, R. J., et al. 2011, *ApJL*, 735, L34  
 Groggin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35  
 Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, *ApJS*, 207, 24  
 Hoag, A., Bradač, M., Huang, K.-H., et al. 2015, *ApJ*, 813, 37  
 Hu, E. M., Cowie, L. L., McMahon, R. G., et al. 2002, *ApJL*, 568, L75  
 Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, 457, 841  
 Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, *ApJS*, 209, 6  
 Inoue, A. K. 2011, *MNRAS*, 415, 2920  
 Jones, T., Martin, C., & Cooper, M. C. 2015, *ApJ*, 813, 126  
 Katsianis, A., Tescari, E., & Wyithe, J. S. B. 2015, *MNRAS*, 448, 3001  
 Kennicutt, R. C., Jr. 1998, *ARA&A*, 36, 189  
 Kneib, J.-P., Ellis, R. S., Santos, M. R., & Richard, J. 2004, *ApJ*, 607, 697  
 Koekemoer, A. M., Ellis, R. S., McLure, R. J., et al. 2013, *ApJS*, 209, 3  
 Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, 197, 36  
 Krueger, H., Fritze-v Alvensleben, U., & Loose, H. H. 1995, *A&A*, 303, 41  
 Labbé, I., González, V., Bouwens, R. J., et al. 2010, *ApJL*, 716, L103  
 Labbé, I., Oesch, P. A., Bouwens, R. J., et al. 2013, *ApJL*, 777, L19  
 Labbé, I., Oesch, P. A., Illingworth, G. D., et al. 2015, arXiv:1507.08313  
 Maitler, V. G., Papovich, C., Groggin, N. A., et al. 2007, *PASP*, 119, 1325  
 Lee, S.-K., Idzi, R., Ferguson, H. C., et al. 2009, *ApJS*, 184, 100  
 Leitherer, C., & Heckman, T. M. 1995, *ApJS*, 96, 9  
 Lemaux, B. C., Lubin, L. M., Sawicki, M., et al. 2009, *ApJ*, 700, 20  
 Maiolino, R., Nagao, T., Grazian, A., et al. 2008, *A&A*, 488, 463  
 McLean, I. S., Steidel, C. C., Epps, H., et al. 2010, Proc. SPIE, 7735, 77351  
 McLean, I. S., Steidel, C. C., Epps, H. W., et al. 2012, Proc. SPIE, 8446, 84460J  
 Merlin, E., Fontana, A., Ferguson, H. C., et al. 2015, *A&A*, 582, 15  
 Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, *ApJ*, 521, 64  
 Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *ApJS*, 208, 5  
 Nussbaumer, H., & Schmutz, W. 1984, *A&A*, 138, 495  
 Oesch, P. A., Dokkum, P. G. v., Illingworth, G. D., et al. 2015, *ApJL*, 804, L30  
 Ono, Y., Ouchi, M., Mobasher, B., et al. 2012, *ApJ*, 744, 83  
 Papovich, C. J., Dickinson, M. E., & Ferguson, H. C. 2001, *ApJ*, 559, 620  
 Pickles, A. J. 1998, *PASP*, 110, 863  
 Polletta, M., Tajer, M., Maraschi, L., et al. 2007, *ApJ*, 663, 81  
 Polletta, M. d. C., Wilkes, B. J., Siana, B., et al. 2006, *ApJ*, 642, 673  
 Postman, M., Coe, D., Benítez, N., et al. 2012, *ApJS*, 199, 25  
 Roberts-Borsani, G. W., Bouwens, R. J., Oesch, P. A., et al. 2015, arXiv:1506.00854  
 Robertson, B. E., Ellis, R. S., Dunlop, J. S., McLure, R. J., & Stark, D. P. 2010, *Natur*, 468, 49  
 Ryan, R. E., Jr., Gonzalez, A. H., Lemaux, B. C., et al. 2014, *ApJL*, 786, L4  
 Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, *ApJ*, 799, 183  
 Schaerer, D., & De Barros, S. 2009, *A&A*, 502, 423  
 Schaerer, D., & De Barros, S. 2010, *A&A*, 515, 73  
 Schaerer, D., De Barros, S., & Stark, D. P. 2011, *A&A*, 536, 72  
 Schmidt, K. B., Treu, T., Bradac, M., et al. 2015, arXiv:1511.04205  
 Schmidt, K. B., Treu, T., Brammer, G. B., et al. 2014, *ApJL*, 782, L36  
 Sharon, K., Gladders, M. D., Marrone, D. P., et al. 2015, *ApJ*, 814, 21  
 Shim, H., Chary, R.-R., Dickinson, M. E., et al. 2011, *ApJ*, 738, 69  
 Smit, R., Bouwens, R. J., Franx, M., et al. 2015, *ApJ*, 801, 122  
 Smit, R., Bouwens, R. J., Labbe, I., et al. 2014, *ApJ*, 784, 58  
 Sobral, D., Matthee, J., Darvish, B., et al. 2015, *ApJ*, 808, 139  
 Soucail, G. 1990, *Ap&SS*, 170, 283  
 Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, *ApJ*, 697, 1493  
 Steidel, C. C., & Hamilton, D. 1993, *AJ*, 105, 2017  
 Trenti, M., Bradley, L. D., Stiavelli, M., et al. 2011, *ApJL*, 727, L39  
 Trenti, M., Bradley, L. D., Stiavelli, M., et al. 2012, *ApJ*, 746, 55  
 Treu, T., Schmidt, K. B., Brammer, G. B., et al. 2015, arXiv:1509.00475  
 van der Wel, A., Straughn, A. N., Rix, H.-W., et al. 2011, *ApJ*, 742, 111  
 Wisotzki, L., Bacon, R., Blaizot, J., et al. 2015, arXiv:1509.05143  
 Yan, H., Dickinson, M. E., Giavalisco, M., et al. 2006, *ApJ*, 651, 24  
 Yan, H., Yan, L., Zamojski, M. A., et al. 2011, *ApJL*, 728, L22  
 Zheng, W., Postman, M., Zitrin, A., et al. 2012, *Natur*, 489, 406  
 Zheng, W., Shu, X., Moustakas, J., et al. 2014, *ApJ*, 795, 93  
 Zitrin, A., Ellis, R. S., Belli, S., & Stark, D. P. 2015, *ApJL*, 805, L7  
 Zitrin, A., Meneghetti, M., Umetsu, K., et al. 2013, *ApJL*, 762, L30