CANDELS VISUAL CLASSIFICATIONS: SCHEME, DATA RELEASE, AND FIRST RESULTS

Jeyhan S. Kartaltepe[1,2], Mark Mozena[3], Dale Kocevski[4], Daniel H. McIntosh[5], Jennifer Lotz[6], Eric F. Bell[7], Sandy Faber[3], Harry Ferguson[6], David Koo[3], Robert Bassett[8], Maksym Bernyk[8], Kirsten Blancato[2,9], Frederic Bournaud[10], Paolo Cassata[11], Marco Castellano[12], Edmond Cheung[3], Christopher J. Conselice[13], Darren Croton[8], Tomas Dahlen[6], Duilia F. de Mello[14], Laura DeGroot[15], Jennifer Donley[16], Javiera Guedes[17], Norman Grogin[6], Nimish Hathi[15], Matt Hilton[18], Brett Hollon[5], Anton Koekemoer[6], Nick Liu[6,19], Ray A. Lucas[6], Marie Martig[8], Elizabeth McGrath[20], Conor McPartland[21], Bahram Mobasher[15], Alice Morlock[13], Erin O'Leary[2,22], Mike Peth[23], Janine Pforr[2], Annalisa Pillepich[24], David Rosario[25], Emmaris Soto[14], Amber Straughn[26], Olivia Telford[27], Ben Sunnquist[5], Jonathan Trump[28], Benjamin Weiner[29], Stijn Wuyts[30], Hanae Inami[2], Susan Kassin[6], Caterina Lani[13], Gregory B. Poole[31], and Zachary Rizer[5]

[1] School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623, USA; jeyhan@astro.rit.edu
[2] National Optical Astronomy Observatory, 950 N. Cherry Ave., Tucson, AZ, 85719, USA
[3] University of California Observatories/Lick Observatory, University of California, Santa Cruz, CA 95064, USA
[4] Department of Physics and Astronomy, University of Kentucky, Lexington, KY 40506, USA
[5] Department of Physics, University of Missouri-Kansas City, 5110 Rockhill Road, Kansas City, MO 64110, USA
[6] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[7] Department of Astronomy, University of Michigan, 500 Church Street, Ann Arbor, MI 48109, USA
[8] Centre for Astrophysics & Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia
[9] Astronomy Department, Wellesley College 106 Central Street, Wellesley, MA 02481, USA
[10] Laboratoire AIM-Paris-Saclay, CEA/DSM/Irfu—CNRS—Université Paris Diderot, CE-Saclay, F-91191 Gif-sur-Yvette, France
[11] Aix Marseille Université, CNRS, LAM (Laboratoire dAstrophysique de Marseille) UMR 7326, F-13388, Marseille, France
[12] IINAFOsservatorio Astronomico di Roma, Via Frascati 33, I-00040, Monteporzio, Italy
[13] School of Physics & Astronomy, University of Nottingham, Nottingham NG7 2RD, UK
[14] Catholic University of America, Washington, DC, USA
[15] Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA
[16] Los Alamos National Laboratory, Los Alamos, NM, USA
[17] Institute for Astronomy, ETH Zurich, Wolgang-Pauli-Strasse 27, 8093 Zurich, Switzerland
[18] Astrophysics & Cosmology Research Unit, School of Mathematics, Statistics & Computer Science, University of KwaZulu-Natal, Durban 4041, South Africa
[19] Pioneer High School, 601 W. Stadium Boulevard, Ann Arbor, MI 48103, USA
[20] Department of Physics and Astronomy, Colby College, Waterville, ME 04901, USA
[21] Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
[22] Department of Physics and Astronomy, Macalester College, 1600 Grand Avenue, Saint Paul, MN 55105, USA
[23] Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA
[24] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
[25] Max-Planck Institut für Astronomie, Konigstuhl 17, D-69117, Heidelberg, Germany
[26] NASA Goddard Space Flight Center, USA
[27] Astronomy Department, University of Washington, 3910 15th Ave NE, Seattle, WA 98195, USA
[28] Penn State, State College, PA, USA
[29] Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA
[30] Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse 1, D-85748 Garching bei München, Germany
[31] School of Physics, University of Melbourne, Parksville, VIC 3010, Australia

ABSTRACT

We have undertaken an ambitious program to visually classify all galaxies in the five CANDELS fields down to $H < 24.5$ involving the dedicated efforts of over 65 individual classifiers. Once completed, we expect to have detailed morphological classifications for over 50,000 galaxies spanning $0 < z < 4$ over all the fields, with classifications from 3 to 5 independent classifiers for each galaxy. Here, we present our detailed visual classification scheme, which was designed to cover a wide range of CANDELS science goals. This scheme includes the basic Hubble sequence types, but also includes a detailed look at mergers and interactions, the clumpiness of galaxies, k-corrections, and a variety of other structural properties. In this paper, we focus on the first field to be completed—GOODS-S, which has been classified at various depths. The wide area coverage spanning the full field (wide+deep+ERS) includes 7634 galaxies that have been classified by at least three different people. In the deep area of the field, 2534 galaxies have been classified by at least five different people at three different depths. With this paper, we release to the public all of the visual classifications in GOODS-S along with the Perl/Tk GUI that we developed to classify galaxies. We present our initial results here, including an analysis of our internal consistency and comparisons among multiple classifiers as well as a comparison to the Sérsic index. We find that the level of agreement among classifiers is quite good (>70% across the full magnitude range) and depends on both the galaxy magnitude and the galaxy type, with disks showing the highest level of agreement (>50%) and irregulars the lowest (<10%). A comparison of our classifications with the Sérsic index and rest-frame colors shows a clear separation between disk and spheroid populations. Finally, we explore morphological k-corrections between the V-band and H-band observations and find that a small fraction (84 galaxies in total) are classified as being very different between these two bands. These galaxies typically have very clumpy and extended morphology or are very faint in the V-band.

1

1. INTRODUCTION

Since galaxies were first discovered in the early 20th century, astronomers have used information about their structure and morphology to understand galaxy properties in a larger context. We know that in the local universe, massive galaxies follow the Hubble sequence and their placement along this sequence (i.e., spiral, elliptical, etc.) is closely correlated with many other galaxy properties, such as stellar mass, color, star formation rate, and the relative dominance of a central bulge (e.g., Roberts & Haynes 1994). Using this basic separation of galaxy morphology, along with the various correlated properties, astronomers have been able to piece together a basic picture of galaxy evolution, connecting blue star forming disk-like galaxies to red and dead elliptical galaxies.

While this picture of galaxy formation and evolution is supported by galaxy studies in the nearby universe, we would like to understand how it holds up against observations of galaxies in the early universe. When did the Hubble Sequence, as we know it, form and how have galaxies changed over time? Were major mergers more important in the early universe and how did they affect the morphology of progenitor galaxies? Are the highly irregular, clumpy galaxies we see at high redshift the result of a higher gas fraction in the early universe? Many of these open questions in galaxy evolution can be addressed by understanding the structure of galaxies at high redshift and how galaxy morphology relates to other properties, such as stellar mass, star formation rate, and active galactic nucleus (AGN) content.

Visual classifications have long been used to study galaxy morphology and structure in the local universe. Large surveys such as the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2003) and the CFHTLS-Deep Survey have enabled the visual classification of thousands of galaxies in the nearby universe (e.g., Schawinski et al. 2007; Bridge et al. 2010; Nair & Abraham 2010; Atkinson 2013) and studies utilizing these classifications have led to a greater understanding of the various correlations described above. The citizen science project *Galaxy Zoo* has extended morphological classifications to the general public and has produced catalogs of nearly one million galaxies from the SDSS (Lintott et al. 2008, 2011; Darg et al. 2010; Willett et al. 2013, 2015 Simmons et al. 2014). Thanks to the various *Hubble Space Telescope* (*HST*) deep field surveys (such as the Hubble Deep Field, GOODS, COSMOS, AEGIS, and GEMS), we have been able to probe galaxy morphology to ever increasing redshift. The large optical cameras on *HST* (WFPC2 and Advanced Camera for Surveys (ACS)) have enabled detailed morphological studies based on visual classifications for increasing numbers of objects (e.g., Brinchmann et al. 1998; Brinchmann & Ellis 2000; Bell et al. 2005; Bundy et al. 2005; Jogee et al. 2009; Bridge et al. 2010; Kartaltepe et al. 2010) at intermediate redshifts. However, beyond $z \sim 1$, these optical surveys start to probe the rest-frame UV morphologies of galaxies, which are sensitive to the regions of most active star formation. In order to study the overall structure of galaxies, and provide the best comparison to morphologies in the local

universe, we wish to trace the structures where the older stellar populations live. This can best be done in the rest-frame optical using near-infrared imaging at $z > 1$.

Until the installation of Wide Field Camera 3 (WFC3) on *HST* during Servicing Mission 4 (SM4) in 2009, near-infrared imaging surveys were limited to ground-based telescopes (either with wide field cameras or with adaptive optics (AO) for small numbers of objects) or with NICMOS on *HST* (e.g., McGrath et al. 2008; Kriek et al. 2009; Conselice et al. 2011b). The small field of view of NICMOS and AO observations placed practical constraints on the total area coverage and thus numbers of galaxies that these surveys were able to study. WFC3 has enabled detailed investigations of large samples of high redshift galaxies for the first time. Since the first WFC3 surveys began, a number of studies have focused on the morphological properties of samples of tens to hundreds of galaxies at high redshift (e.g., Cassata et al. 2010; Cameron et al. 2011; Conselice et al. 2011a; Szomoru et al. 2011; Law et al. 2012; Morishita et al. 2104; Glikman et al. 2015).

We can greatly expand on these morphological studies with larger samples from the Cosmic Assembly Near-infrared Dark Energy Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011). CANDELS provides deep, high resolution near-infrared imaging from WFC3 across five of the most commonly studied deep fields. A key goal of the CANDELS survey is to study the structure and morphology of galaxies at $z = 1$–3, a key period of galaxy assembly. A number of studies on the morphological properties of galaxies in the CANDELS fields have already been published (e.g., Bell et al. 2012; Bruce et al. 2012; Kartaltepe et al. 2012; Kocevski et al. 2012; Wuyts et al. 2012; Lee et al. 2013; Mortlock et al. 2013; Targett et al. 2013; Trump et al. 2014; Villforth et al. 2014; Guo et al. 2015; Rosario et al. 2015). Our team has undertaken an ambitious effort to visually classify all CANDELS galaxies brighter than $H = 24.5$. Once complete, this will result in detailed morphological classifications for over 50,000 galaxies across all five fields, spanning a wide range in redshift ($0 < z < 4$)—the largest such sample of classifications at these redshifts. Two publications (Kartaltepe et al. 2012; Kocevski et al. 2012) have already made use of these classifications, investigating the role of mergers among Ultraluminous Infrared Galaxies and X-ray selected AGN, respectively, at $z \sim 2$. In this paper, we present our classification scheme and initial results based on the first field that we classified (GOODS-S). With this publication, we are also releasing the full set of classifications for GOODS-S (covering over 7000 galaxies) to the public. At the time of writing, we have completed the classifications for two other fields (UDS and COSMOS), which will be released with future publications.

This paper is organized as follows. Section 2 introduces the CANDELS survey and the data sets discussed in this paper. In Section 3 we present our visual classification scheme. In Section 4 we discuss our results and various comparisons to test for consistency and in Section 5 we summarize our findings. Throughout this paper we assume a ΛCDM

cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\Lambda = 0.7$, and $\Omega_m = 0.3$. All magnitudes are in the AB system unless otherwise stated.

2. OBSERVATIONS AND DATA SETS

2.1. CANDELS

CANDELS (PIs Faber & Ferguson; see Grogin et al. 2011; Koekemoer et al. 2011) is an *HST* Multi-Cycle Treasury Program to image portions of five of the most commonly studied legacy fields (GOODS-N: Giavalisco et al. 2004; GOODS-S, COSMOS: Scoville et al. 2007; UDS: Lawrence et al. 2007; and EGS: Davis et al. 2007) with WFC3 in the NIR. The survey has observed all five fields to 2-orbit depth in F125W (*J*-band, 2/3 orbit) and F160W (*H*-band, 4/3 orbit) and the central regions of GOODS-N and GOODS-S to 10 orbit depth in these bands as well as F105W (*Y*-band). ACS parallel imaging has also been obtained for all of these fields in F814W and F606W. For details on the full CANDELS survey, see Grogin et al. (2011). In addition to the CANDELS observations, a portion of GOODS-S was also observed as a part of the WFC3 Early Release Science (ERS; Windhorst et al. 2011) campaign in *Y*, *J*, and *H*. While we are classifying galaxies in all of the five CANDELS fields, for this paper, we focus on GOODS-S and include the ERS coverage along with CANDELS for full coverage across the entire field.

The CANDELS observations began in 2010 October and were completed in 2013 August. For this paper, we use mosaics at three different depths for comparison. First, we use a uniform 2-orbit depth ($J + H$) mosaic across the full field. This mosaic represents the wide coverage that has been obtained for all five CANDELS fields. For the deep region of GOODS-S, we also use a 4-orbit (available at the beginning of the visual classifications) and the final 10-orbit depth mosaic in order to test the dependence of our classifications on image depth. The original images were reduced and drizzled to a $0\farcs06$ pixel scale to create each of the mosaics. The details of the data reduction pipeline are described in Koekemoer et al. (2011). The WFC3 photometry in both the *J* and *H* bands were measured using SExtractor version 2.5.0 (Bertin & Arnouts 1996) in a "cold mode" setup found to work best for extracting $z \sim 2$ galaxies by detecting faint sources as well as optimally deblending and resolving multiple source issues (see, for example, Caldwell et al. 2008).

2.2. Ancillary Data and Data Products

In addition to the CANDELS NIR images and SExtractor catalogs, we also use the optical *HST*-ACS F606W and F850LP mosaics. For GOODS-S, these data were already publicly available (Giavalisco et al. 2004). We use the CANDELS consensus photometric redshift catalog (Dahlen et al. 2013). These photometric redshifts were computed based on the photometry in 14 bands: *U* (VLT/VIMOS), *BViz* (*HST*/ACS), F098M, F105W, F125W, F160W (*HST*/WFC3/IR), *Ks* (VLT/ISAAC), and 3.6, 4.5, 5.8, 8.0 μm (*Spitzer*/IRAC). The point-spread function-matched photometry catalog (Guo et al. 2013) is based on *H*-band detected objects with photometry for all ACS and WFC3 data measured using SExtractor and for the rest of the optical-NIR bands using TFIT (Laidler et al. 2007). These photometric redshifts were used to compute rest-frame magnitudes with the code EAZY (Brammer et al. 2008) and the and the templates of Muzzin et al. (2013).
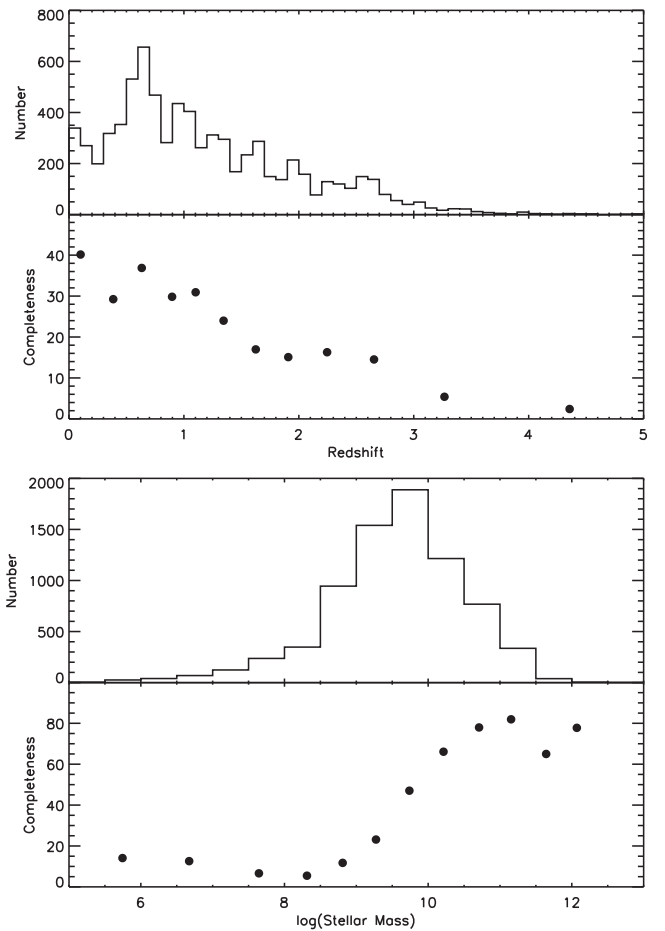

Figure 1. Redshift (top) and stellar mass (bottom) distributions and completeness. The completeness indicates the percentage of galaxies in each bin that satisfy the magnitude cut ($H < 24.5$) and are observed to at least the 2-epoch depth in the *H*-band.

Stellar masses were also computed at these photometric redshifts using BC03 templates (Bruzual & Charlot 2003), an exponentially declining star formation history, and a Chabrier (2003) initial mass function. All of these data can be found in Table 4 of Santini et al. (2015).

2.3. Sample Selection and Postage Stamps

In order to maximize the amount of time spent looking at galaxies that are bright enough to be effectively classified, we settled on a magnitude cut of $H < 24.5$ based on classifying a test sample of 100 randomly selected galaxies chosen to sample the full magnitude range of our data. Five people classified all of these objects and we chose $H < 24.5$ as our cutoff because we found that at fainter magnitudes, many of the objects were difficult to classify. For GOODS-S, this magnitude cut results in a final sample of 7634 galaxies across the entire field at 2-epoch depth and 2534 galaxies in the deep region of the field. We have made no cuts based on any other galaxy properties as the values for these properties (such as photometric redshifts and stellar masses) are likely to change in future iterations of the catalogs. We also chose not to make an a priori cut based on redshift so that our final catalogs would cover all redshifts, allowing each user to select the range they are most interested in. As such, our final sample of galaxies spans a wide range in redshift (based on the consensus photometric redshifts
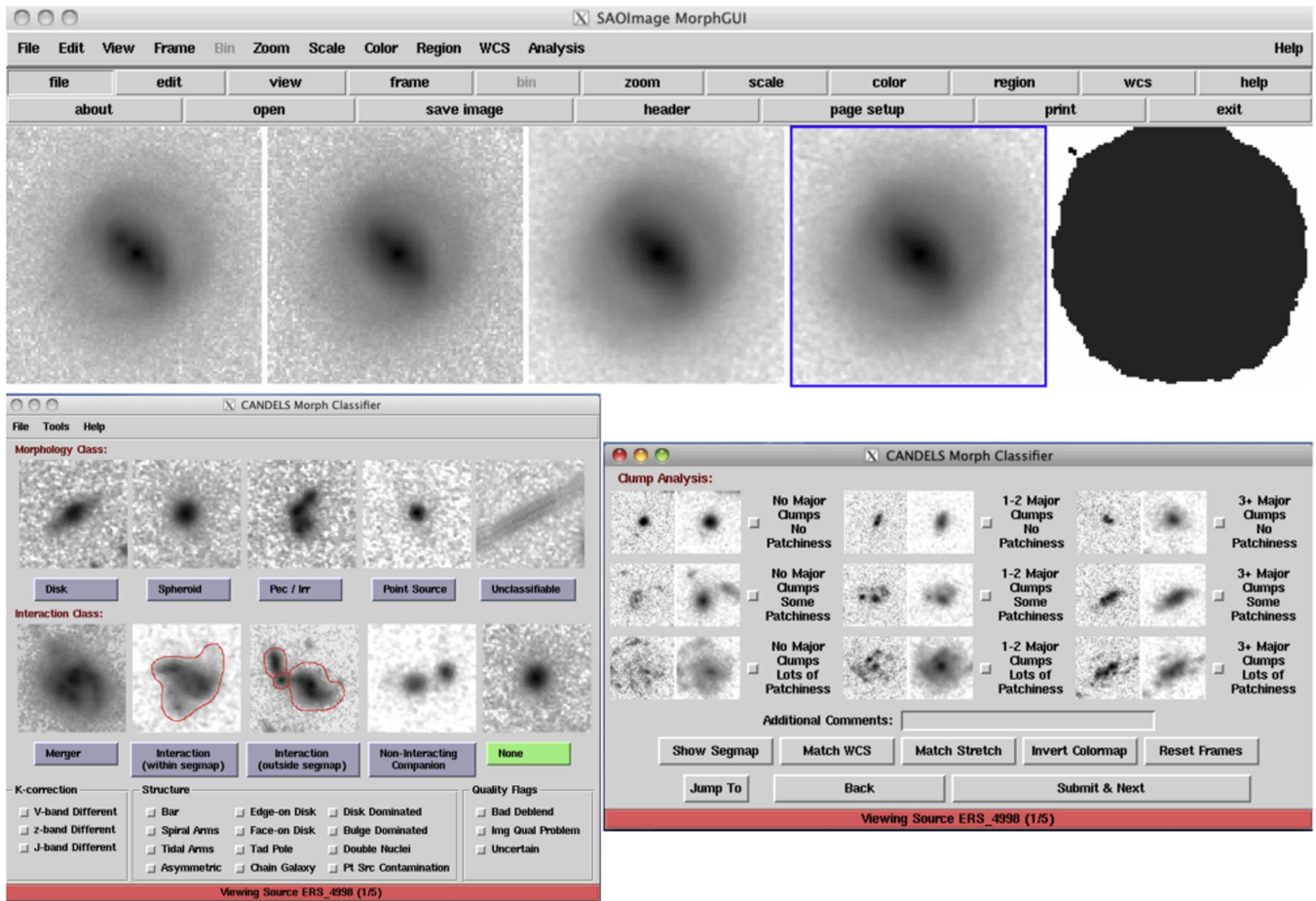
Figure 2. Screenshot of the Perl/ds9 GUI used for visual classification of CANDELS galaxies. Top: ds9 window with F606W, F850l, F125W, F160W, and segmentation map images for a sample galaxy. Bottom: GUI window showing visual classification scheme examples and check boxes or the user to select while classifying.

described above), from $z = 0$ to $z \sim 4$ with $\langle z \rangle = 0.98$ and 3511 galaxies at $z = 1-3$ where H-band CANDELS observations probe the rest frame optical. The redshift and stellar mass distributions of this sample are shown in Figure 1 along with the completeness as a function of mass and redshift. The incompleteness is a result of the magnitude cut used as well as the exclusion of objects around the edges of the mosaic with shallower coverage.

Following the prescription of Häussler et al. (2007), we used their Equations (2) and (3) to scale the size of the postage stamps used for visual classification to the size of the galaxy. We created stamps that were square, with a size equal to the larger of *Xsize* or *Ysize* with a minimum size of 84 pixels. In addition, a larger H-band postage stamp was provided in order to identify potential nearby neighbors. While the visual classifications are conducted primarily in H-band we also used the optical *HST*-ACS mosaics to help with classifying and to study the effects of k-corrections (see Section 3). We used all four *HST* bands (F606W, F850LP, F125W, and F160W) as well as the segmentation map from SExtractor, to create cutouts of all 7634 galaxies that we classified. For the deep region of the field, we created three separate sets of cutouts, one using each of the 2-, 4-, and 10- epoch depth mosaics for the F125W and F160W bands. The cutouts for the ACS bands and segmentation map (based off of the initial mosaic of the full

field at a non-uniform 6-epoch depth) were the same for each depth.

3. VISUAL CLASSIFICATION SCHEME

All of the visual classifications are based primarily on the *H*-band WFC3 image, but the *J*-band image along with the *V*- and *I*-band ACS images are included to provide additional information and help with the classifications, since the different rest-frame wavelengths and resolutions are sensitive to different structures. In order to determine whether to classify galaxies using all of the bands at once or to classify each band separately, we conducted a test where two subsets of the co-authors, with five classifiers in each group, visually classified a sample of 100 galaxies—with one group classifying each band separately and one group classifying based on the *H*-band but using the other bands to inform the decision. We then compared the results of these two groups and found that the relative agreement between the groups was rather high but that the group classifying each band separately took much longer and had a difficult time with galaxies that were particularly faint in the optical. Since our ultimate goal is to classify galaxies in all of the CANDELS fields, we opted for the first method to minimize classifier fatigue. In order to study the effects of morphological *k*-corrections, we added a flag (see
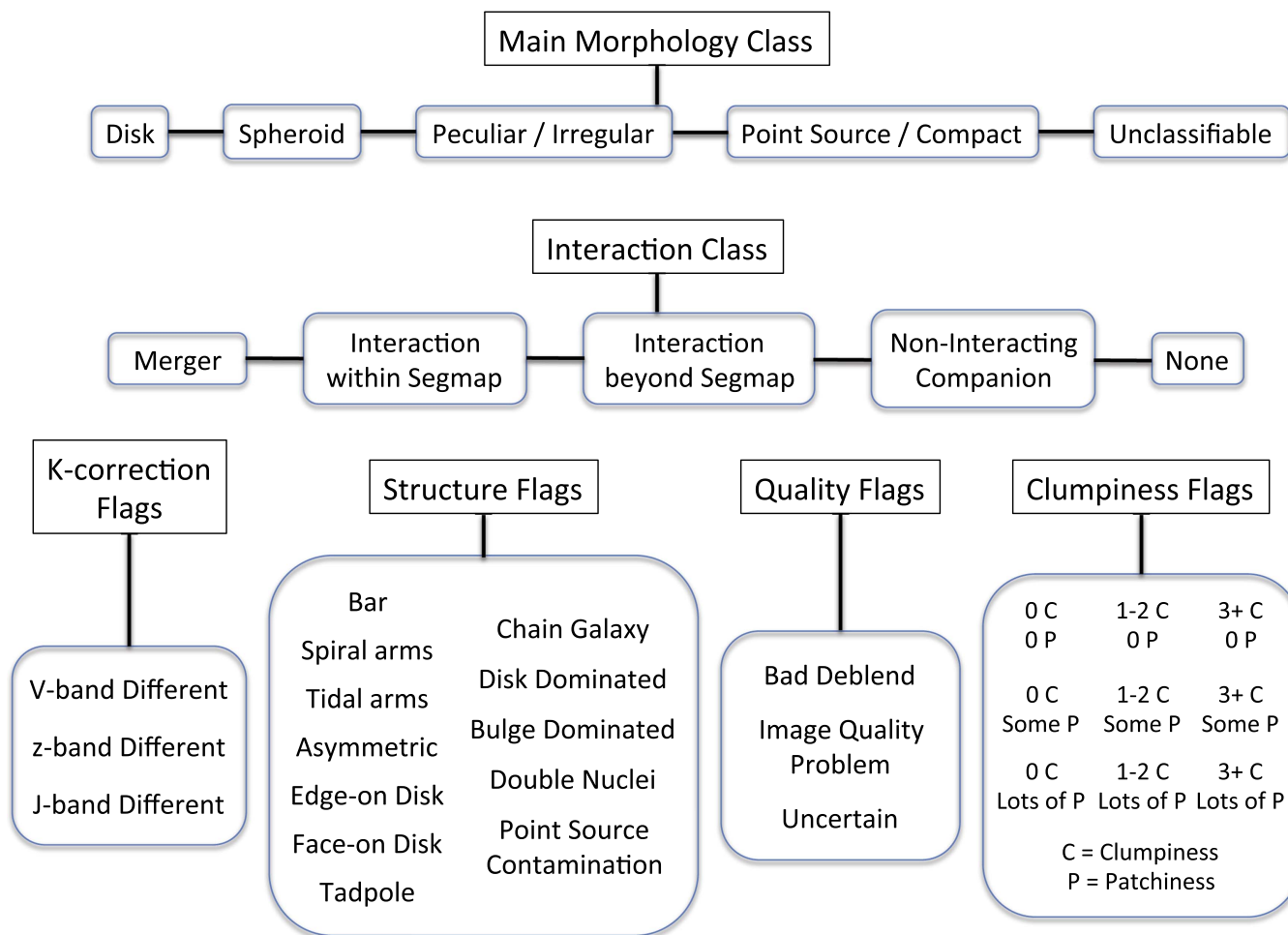
Main Morphology Class

Disk — Spheroid — Peculiar / Irregular — Point Source / Compact — Unclassifiable

Interaction Class

Merger — Interaction within Segmap — Interaction beyond Segmap — Non-Interacting Companion — None

K-correction Flags

V-band Different

z-band Different

J-band Different

Structure Flags

Bar

Spiral arms

Tidal arms

Asymmetric

Edge-on Disk

Face-on Disk

Tadpole

Chain Galaxy

Disk Dominated

Bulge Dominated

Double Nuclei

Point Source Contamination

Quality Flags

Bad Deblend

Image Quality Problem

Uncertain

Clumpiness Flags

0 C 0 P	1-2 C 0 P	3+ C 0 P
0 C Some P	1-2 C Some P	3+ C Some P
0 C Lots of P	1-2 C Lots of P	3+ C Lots of P

C = Clumpiness
P = Patchiness

Figure 3. Chart illustrating the three different classification levels described in Section 3. At the top level is the main morphology class, where multiple class can be chosen. Next is the Interaction class, only one of which can be chosen. Finally, there are various structural flags, any of which can be chosen as they apply.

Section 3.3 below) to indicate whether the morphology is different between different bands.

To make the classifications manageable, we divided up all of the galaxies into sets of 200 objects (called a "chunk"). Each classifier was assigned one chunk at a time to classify. Once that chunk was complete, the next one was assigned. Our goal was to have a minimum of three classifiers look at each galaxy so that we could compare the different classifications and look for outliers. For the deep area of the field, we increased that minimum to five classifiers for each galaxy and assigned an independent set of five for each of the three different depths. This means that for the deep region of GOODS-S (2534 galaxies) there are 15 independent sets of classifications, five at each depth.

We developed two different GUIs to allow for a uniform implementation of the classification scheme and to make the classifications go as smoothly as possible. The first is a Perl/Tk based GUI that interacts with the image display tool ds9[32] (shown in Figure 2). The ds9 window displays all four of the *HST* bands, in order of increasing wavelength, followed by the segmentation map. This GUI allows for the user to scale each image independently while classifying. The classifier chooses their classification by checking the appropriate boxes and then moves on to the next object. When the set is complete, the GUI

writes out a text file with the classifications. We release the software for this GUI (Kocevski 2015) with this paper so that others may use it and our classification scheme for their own galaxy classifications. The second GUI is a web-based one. Every aspect of this GUI is identical except that the stamps are fixed in scale using arcsinh scaling that was determined to work the best for capturing the range of galaxy features that are present. We allowed the classifiers to choose whichever of the two GUIs they preferred and we note that when we compared the results, we could not find a significant difference based on the GUI chosen. Both of the GUIs have a comment box so that classifiers can note things that do not fall within the classification scheme, problems, or just particularly interesting objects (such as a gravitational lens, for example).

Before beginning the classification process, there are several steps that each classifier must go through. The first is to familiarize themselves with the scheme as described in the following subsections. After this, and after asking any questions they might have, they then go through a training set of 25 galaxies. These 25 galaxies were chosen to reflect the wide range of different classifications that are possible and to illuminate some of the aspects of the scheme that typically create confusion. Throughout the process, we reiterate that there is no right or wrong answer to classification. As long as the scheme is being interpreted correctly, then each classifier's response is valid since they may each see different aspects of

---

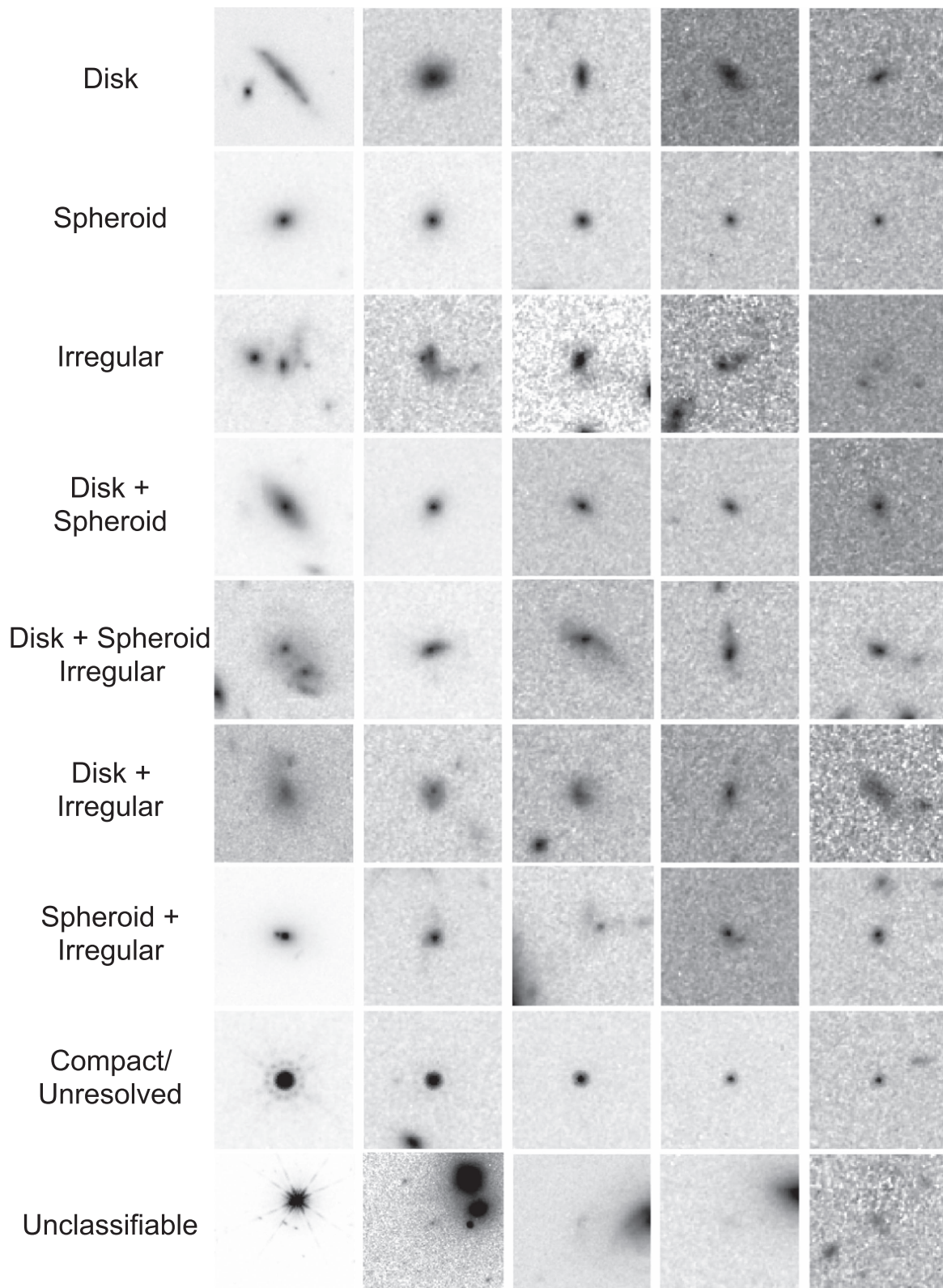[32] http://hea-www.harvard.edu/RD/ds9/site/Home.html

Figure 4. Sample *HST*/WFC3 postage stamps illustrating the main morphology class of the visual classification scheme. The stamps are ordered by *H*-band magnitude with the brightest galaxies to the left. The sizes of the stamps follow the prescription described in Section 2.3.

the same galaxy. Once each classifier has gone through these 25 and understands the scheme and why they have made each selection, we assign the first chunk. The first chunk is the same for everybody. This chunk is known as our calibration set. This set of 200 galaxies is simply the first set of objects in the deep area of GOODS-S from the 2-epoch depth mosaic and samples the full magnitude and redshift range of the full sample. By having everyone classify this set first, we can identify any
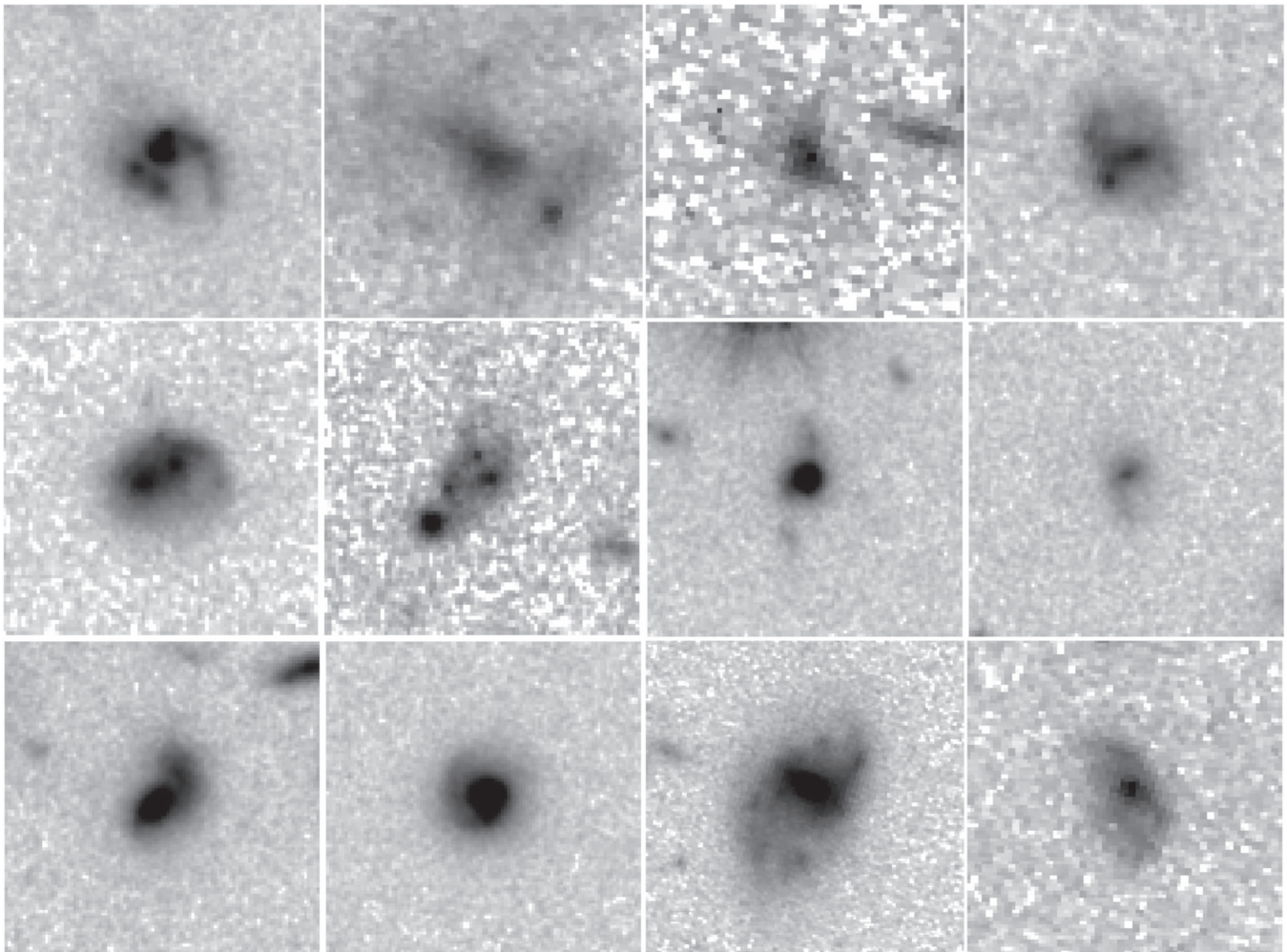
Figure 5. Sample *HST*/WFC3 postage stamps illustrating the different interaction classes. The sizes of the stamps follow the prescription described in Section 2.3.

outliers and correct any misunderstandings of the scheme. It also gives the classifiers a chance to get more comfortable with classifying—the classifiers agree that they approach things a bit differently for the first 20 galaxies in the set than they do for the last 20 in the set. As of the writing of this paper, we have had 65 different classifiers, all of whom have classified this calibration set. This has provided us with a unique sample of 200 galaxies with 65 classifications for comparison—we will describe these comparisons in detail in Section 4.

In the following subsections, we describe each of the three major components of the classification scheme in detail. This scheme is outlined in detail in Figure 3.

3.1. Main Morphology Class

The top level of the classification scheme is called the Main Morphology Class. These are based on the typical broad Hubble sequence types and there are five different options to choose from. More than one of these options can be selected for each galaxy, allowing the classifier to indicate intermediate cases. These classes are as follows.

(1) *Disk:* Disk galaxies have a disk structure that may or may not show clear spiral arms. Disks may also have a central bulge (spheroid component), in which case, the classifier may select both "Disk" and "Spheroid" classes and indicate whether the

disk is bulge- or disk-dominated in the structure flags described below.

(2) *Spheroid:* Spheroid galaxies appear centrally concentrated, smooth, and roughly round/ellipsoidal, regardless of their size, color, or apparent surface brightness. Selecting both "Disk" and "Spheroid" with a bulge-dominated Structure Flag indicates a more early-type galaxy with a modest disk component.

(3) *Irregular/Peculiar:* These are galaxies that do not easily fall into one of the other categories. This class is meant to indicate galaxies with irregular structure, regardless of surface brightness. This includes objects that are strongly disturbed, such as mergers (see Interaction Classes in the next section) but can also include disks or spheroids that have slightly disturbed morphologies. For example, an object that has a warped disk or asymmetric spiral arms should have both "Disk" and "Irregular/Peculiar" checked. Or, a spheroidal galaxy with strong asymmetries should have both "Spheroid" and "Irregular/Peculiar" checked.

(4) *Compact/Unresolved:* These objects are either clear point sources, unresolved compact galaxies, or are so small that the internal structure cannot be discerned. A small but clearly resolved spheroidal galaxy should be classified as a spheroid. This class is not meant to be used if the dominant galaxy is another class but has an embedded point source—there is a Point Source Contamination flag in the Structure flags below.

(5) *Unclassifiable:* These objects are problematic and cannot be classified in any of the other main morphology classes, either because of a problem with the image (e.g., satellite trail, too close to a bright star or galaxy, etc.), the object is not real and should be ignored (e.g., is part of a diffraction spike or is otherwise spurious), or because they are too faint for any structure to be seen. This class is not meant to be used in combination with any of the other classes. If the object can be classified in any of the bands, that band's classification should be used instead.

As noted above, these classes are not mutually exclusive because additional information can be gleaned by choosing more than one class. For example, choosing both disk and spheroid would identify galaxies with both a disk and bulge component. Choosing disk and irregular identifies objects where the disk is still visible but the morphology is slightly disturbed. As a result, there are a total of nine possible combinations of classifications: Disk, Spheroid, Irregular, Disk +Spheroid, Disk+Irregular, Disk+Spheroid+Irregular, Spheroid+Irregular, Compact/Unresolved, and Unclassifiable. Examples of galaxies in these different classes are shown in Figure 4, ordered by their *H*-band magnitude.

3.2. Interaction Class

In order to understand the role that galaxy mergers and interactions play in galaxy evolution, we adopted an interaction class that is separate from the main morphology class above. By keeping them separate, the classifier is not forced to choose between a disk and an interaction, for example. These classes are also intentionally kept separate from determining whether or not a galaxy has a clumpy morphology, which is done later on with the clumpiness flags (described in Section 3.3). There are four different options for the interaction class and only one of the four (or none) can be selected.

(1) *Merger:* These galaxies are single objects (including sources with double nuclei) that appear to have undergone a merger by evidence of tidal features/structures such as tails, loops, or highly irregular outer isophotes. All objects classified as mergers should have Irregular/Peculiar selected as one of their main classifications but not all galaxies classified as Irregular/Peculiar are necessarily mergers.

(2) *Interaction within SExtractor segmentation map:* The primary galaxy appears to be interacting with a companion galaxy within the same *H*-band segmentation map. Interactions have clear signatures of tidal interaction (e.g., tidal arms, bridges, dual asymmetries, off-center isophotes, or otherwise disturbed morphologically)—being apparent close pairs is not enough. To choose interaction over merger, two distinct galaxies must be visible. If more than one companion is present, the classification should be based on the one that appears to dominate the morphology—usually the larger/brighter one.

(3) *Interaction beyond SExtractor segmentation map:* The primary galaxy appears to be interacting with a companion galaxy that has its own distinct *H*-band segmentation map. By differentiating between interactions within and beyond the segmentation map we can identify galaxies with possible deblending problems. This information is also useful when comparing to various other galaxy properties measured by the CANDELS team, such as photometry, photometric redshifts, stellar masses, and automated classification measures since

these will all be based off of the initial source identification done by SExtractor.

(4) *Non-interacting companion:* These galaxies have a close (visible within the field of view of the large postage stamp) companion (in projection), yet no evidence of tidal interaction or disturbed morphology is apparent. The companion galaxy may be within or beyond the primary galaxy's segmentation map. If each galaxy resides in its own segmentation map, the companion galaxy's segmentation map must be separated from the primary galaxy's segmentation map by less than the diameter of whichever galaxy's segmentation map is larger. Additional information can be used later to determine if the two galaxies are at the same redshift, but either have not yet interacted or lack visible signatures, or if they are simply chance projections. One of the benefits of including this class is that it forces the classifier to consider whether or not they actually see merger/interaction signatures rather than being tempted to call everything with a companion an interaction.

Examples of galaxies in each of the different interaction classes are shown in Figure 5.

3.3. Flags

We include four different types of flags in our scheme in order to indicate various other structures and features that are not specified in the above morphology and interaction classes.

Quality flags: If there are any issues with the images that affect the galaxy or cause the classification to be marked as unclassifiable, then the classifier can choose from three different quality flags. The first of these is "Bad deblend" for cases where the *H*-band segmentation map has a problem and the galaxy is either over or under deblended. The second quality flag is "Image Quality Problem." This flag is meant for image problems such as a nearby bright object, the galaxy being too close to the edge of the mosaic, artifacts produced by diffraction spikes or cosmic rays, etc. And finally, there is a flag for "Uncertain" for cases where there are no image quality problems but the classifier is just unsure about their classification.

K-correction flags: These flags are for cases where the difference in morphological structure between the *H* band and any of the bluer bands is so severe that the classifier would select a different classification for that band. The classifier can check any band that meets this condition. This flag should be checked if the object is invisible or substantially fainter in the other bands as well. This flag should not be checked if the differences appear to be solely due to resolution differences.

Structure flags: There is a wide variety of structure flags that can be marked to indicate the presence of interesting/notable features. These are: tidal arms, double nuclei, asymmetric, spiral arms/ring, bar, point source contamination, edge-on disk, face-on disk, tadpole galaxy, chain galaxy, disk-dominated, and bulge-dominated.

Clumpiness/patchiness flags: Finally, there are a set of flags designed to denote how clumpy/patchy the light distribution of a galaxy is. These flags are set in a 3×3 grid, shown on the right side of Figure 2 along with some examples. Clumps are concentrated independent knots of light while patches are more diffuse structures. A central concentration of light is a bulge, not a clump. An object with a continuous surface brightness profile is not considered patchy. Clumps and patches are most clearly seen in the bluer bands so classifiers are asked to look at these for this set of flags.

4. RESULTS

Table 1 summarizes the various catalogs that are being released with the publication of this paper. There are two different flavors of catalogs—the "raw" and the "fractional." The raw catalogs are simply collections of all of the raw classifications. Each object therefore has multiple entries (as many entries as there are classifiers) and each classifier is identified by a unique number. The second set of catalogs, the "fractional" ones, are the ones that are most likely to be useful to the community. These catalogs contain one entry per object and each classification is marked by the fraction of people who

checked that box. So if one out of three classifiers classified an object as a disk, one as irregular, and two as a spheroid, then the disk column will have the value 0.33, the spheroid column 0.67, and the irregular column as 0.33. We have created separate catalogs for each depth. There is a 2-epoch depth catalog covering the entire GOODS-S field, one covering just the deep region, and one each covering the deep region at 4- and 10-epoch depths. The calibration set of 200 galaxies classified by all 65 classifiers is included in both 2-epoch depth catalogs. In future papers (D. McIntosh 2015, in preparation and M. Mozena 2015, in preparation), we will present catalogs that have combined these classifications into other scientifically useful metrics.

4.1. Calibration Set

The calibration set of 200 galaxies classified by all 65 of our classifiers provides a unique resource for examining our classification scheme. Figure 6 shows how often each morphological type is chosen by a given classifier. All nine of the different combinations of main morphology class (except disk+spheroid+irregular since there are very few times all three of these are selected) are shown along with each of the interaction classes. The median number of objects is shown as the dashed line and the standard deviation is shown as the dotted lines above and below the median. This illustrates how likely each classifier is to choose a particular type. For example, we were curious to see if some people were more likely than others to choose "merger." For the most part, the results are as expected. The disk category is chosen the most often and unclassifiable is rarely selected. One aspect that is immediately noticeable is that there are two people that chose

Table 1
CANDELS: GOODS-S Visual Classification Catalog Data Release

Depth	Type[a]	Area	# of Entries	# of Objects
2-epoch	raw	GOODS-S deep +wide+ERS	42695	7634
2-epoch	fractional	GOODS-S deep +wide+ERS	7634	7634
2-epoch	raw	GOODS-S deep	25059	2534
2-epoch	fractional	GOODS-S deep	2534	2534
4-epoch	raw	GOODS-S deep	12670	2534
4-epoch	fractional	GOODS-S deep	2534	2534
10-epoch	raw	GOODS-S deep	13477	2534
10-epoch	fractional	GOODS-S deep	2534	2534

Note.
[a] The raw classification catalogs follow the format of Table 2 while the fractional catalogs follow the format of Table 3.

Table 2
CANDELS: GOODS-S 2-epoch Depth Raw Visual Classification Catalog

ID	R.A.	Decl.	Spheroid[a]	Disk	Irregular	Compact	Unclassifiable	Interaction Class	Classifier	Comments
GDS_deep2_10000	53.054728	−27.769708	0	0	0	1	0	0	1	…
GDS_deep2_10000	53.054728	−27.769708	0	0	0	1	0	0	2	…
GDS_deep2_10000	53.054728	−27.769708	1	0	0	0	0	0	3	…
GDS_deep2_10000	53.054728	−27.769708	0	0	0	1	0	0	4	…
GDS_deep2_10000	53.054728	−27.769708	1	0	0	0	0	0	5	…

Note.
[a] All of the classifications discussed in this paper are presented in the online version of the paper. Here, only the main morphology class and interaction class columns are shown as an example.

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

Table 3
CANDELS: GOODS-S 2-epoch Depth Fractional Visual Classification Catalog

ID	R.A.	Decl	#[a]	Spheroid[b]	Disk	Irregular	Compact	Unclassifiable	Merger
GDS_deep2_4407	53.0746	−27.8473	5	0.6	0.6	0.6	0.0	0.0	0.0
GDS_deep2_4418	53.1035	−27.8473	5	0.6	1.0	0.2	0.0	0.0	0.0
GDS_deep2_4420	53.0902	−27.8479	5	0.2	1.0	0.6	0.0	0.0	0.0
GDS_deep2_4422	53.0758	−27.8466	5	1.0	0.4	0.0	0.2	0.0	0.0
GDS_deep2_4423	53.0915	−27.8468	5	0.4	0.2	0.6	0.0	0.0	0.0

Notes.
[a] Number of people that classified this object.
[b] All of the classifications discussed in this paper are presented in the online version of the paper. Here, only the main morphology class and interaction class columns are shown as an example.

(This table is available in its entirety in machine-readable form.)
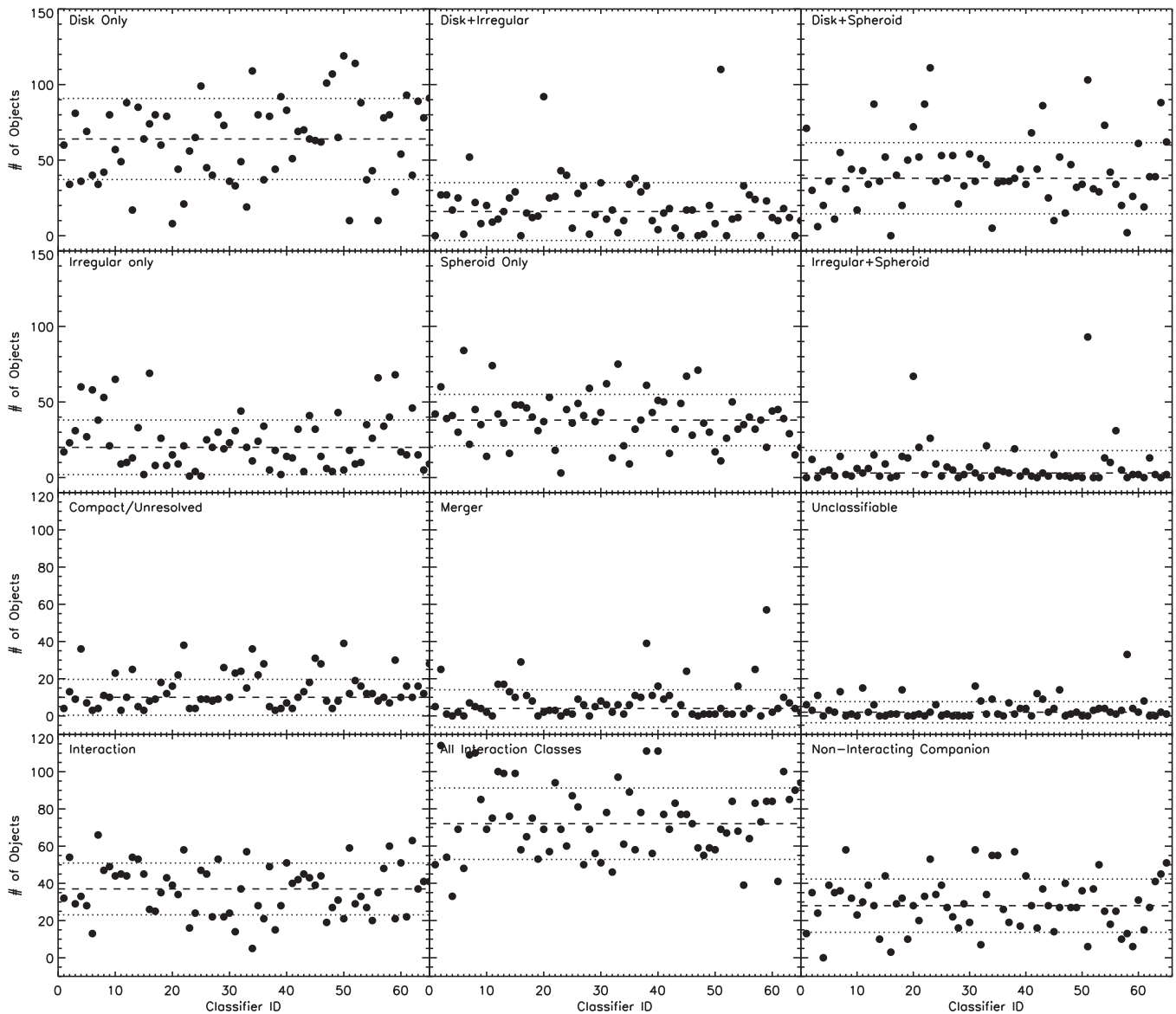
Figure 6. Number of objects that each person classified as a given morphological type. The dashed line represents the median number of objects and the standard deviation is shown as dotted lines above and below the median.

irregular in combination with disk and spheroid often, but those two people are no more likely than others to choose only irregular as a class. Interaction is chosen more often than merger and there are a couple of people more likely to identify an object as a merger than others.

For each of the 200 galaxies in the calibration set, we looked at the overall distribution of classifications. This was an interesting exercise, because while we were looking for evidence of agreement or disagreement, what we found was that the overall distribution of classifications contains useful information. The objects with the highest level of agreement are the simplest cases (for example, the top object in Figure 7 is a pure disk with high classifier agreement) while the objects with the lowest level of agreement are the ones with quite complex morphological structure (see object in bottom of Figure 7). Each classifier is seeing a slightly different aspect of the same galaxy when they classify and this extra information is lost if we were to only use one person's classification.

4.2. Internal Consistency

One of the first things to look at with these sets of classifications is internal consistency. How often do classifiers choose the same classifications? Figure 8 shows the fraction of objects with high agreement, where high agreement is defined as $>3/5$ of classifiers choosing a single main morphology class, as a function of H-band magnitude for the 2534 galaxies in the deep area of the field. This fraction is plotted separately for each of the three different depths (with independent classifiers). This plot indicates that the level of agreement is clearly a function of galaxy magnitude, as one would expect. For the brightest galaxies, the agreement is rather high, $\sim 96\%$ of objects have classifications that agree for $>3/5$ of classifiers. The classifications at all three depths agree with each other for these brightest galaxies. The level of agreement stays above $\sim 90\%$ until $H > 22$ and then starts to fall off at fainter magnitudes. The dispersion between the different depths also begins to increase. For the faintest galaxies, $H > 24$, the level
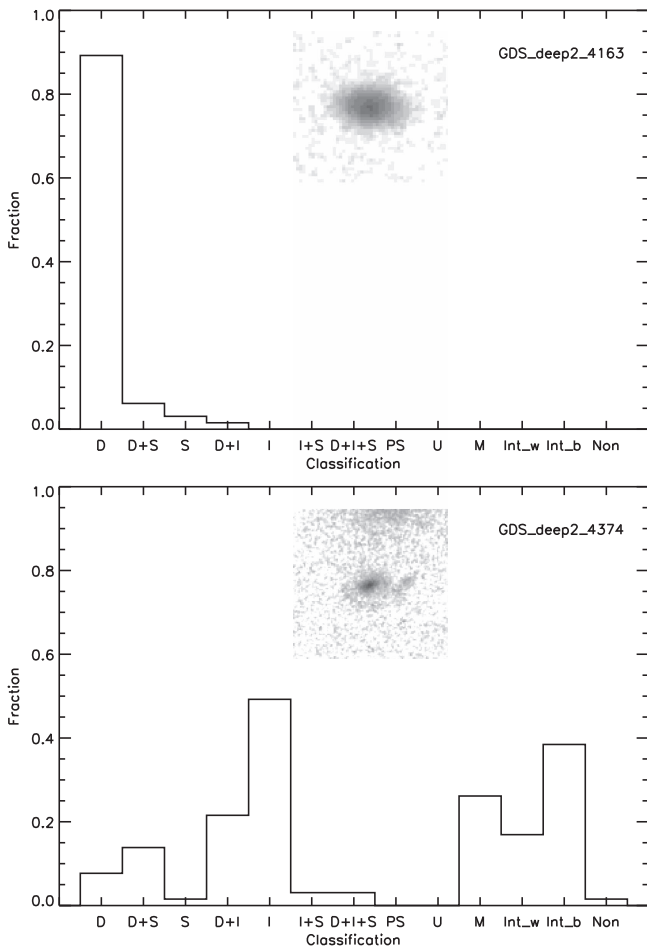
Figure 7. Distribution of morphological classifications for two sample galaxies in the calibration set of 200 galaxies classified by 65 people. The top galaxy is one with high agreement where almost everyone agrees it is a disk. A few people also note a bulge component and very few people call it irregular. The bottom galaxy is one with a low level of agreement. The irregular classification has the highest level of agreement, but some people call it a disk and a few see a spheroid component. There is also little agreement about the interaction class (though almost everyone thinks it is merging/interacting on some level).
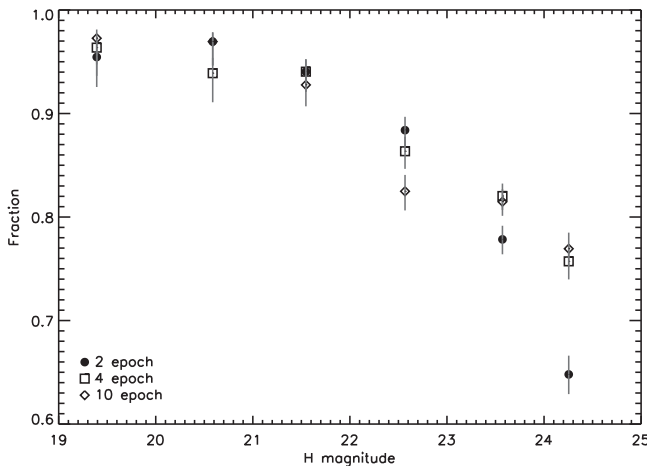
Figure 8. Fraction of objects for which $>3/5$ classifiers agree on the classification as a function of H-magnitude and separated by the depth of the images classified. The error bars on each point reflect the 1σ binomial confidence limits given the number of objects in each category, following the method of Cameron (2011). Even at the faintest magnitudes included in our sample, the fraction of objects with high agreement is still above $\sim70\%$.

of agreement is lowest for the 2-epoch depth images, but is still above 70% for the 4- and 10-epoch depth images.

Figure 9 shows the same information, but split by the individual main morphology class. Again, a threshold of $>3/5$ classifiers is shown for each type. Since the fractions show how often each class was chosen, the absolute fraction on the y-axis cannot be used to determine the level of agreement, but rather how often a given type was selected with a high confidence level. For example, the fraction of objects classified as disks is high while those classified as irregular are low. The fraction of objects classified as disks with high agreement is consistent across the different depths, until $H > 23$, and even then the difference is only slight. However, for spheroids, the difference between the different depths is more pronounced. The classifiers are less likely to agree that an object has a spheroid component in the deepest images, and the fraction of objects with a spheroid component decreases slightly with increasing magnitude. Perhaps this is due to the presence of a disk component becoming easier to see in the deepest images and for brighter galaxies. Overall, there is a low fraction of objects that are classified as irregular with high agreement, but this fraction increases for the faintest galaxies. Classifiers are also more likely to call something irregular for the 2-epoch images than for the deepest images. The consistency of the increasing fraction of irregular galaxies for the 4- and 10- epoch deaths suggests that this is a physical trend and not just the result image depth. The compact/unresolved fraction agrees for all of the depths and the fraction of objects decreases with increasing magnitude. This makes sense since a number of the brightest objects in our sample are point sources. And finally, there is a very low fraction of objects that is classified as "unclassifiable" and this increases only slightly for the faintest galaxies. With this small number of objects, there is no discernible difference between the different depths.

Figure 10 shows the fraction of objects for which the classifiers disagree (defined as only one or two of the classifiers choosing a given class) on the particular morphological type of interest. For each main morphology class, the fraction of objects that the classifiers disagree on increases with increasing magnitude. For disks, irregulars, and unclassifiable objects, there is a significant difference between the different depths for the faintest galaxies such that there is a higher fraction of galaxies with disagreement at the 2-epoch depth than for the deeper images.

Figure 11 illustrates the level of agreement in a different way. Plotted is the difference in the fraction of classifications for each galaxy between the wide (2-epoch) and deep (10-epoch) depths. For example, if for a given galaxy, $3/5$ of classifiers classified it as a disk at both depths, then its value on this plot would be zero; likewise if $2/5$ or $5/5$ classified it as a disk at both depths. However, if $3/5$ classified it as a disk at the 10 epoch depth and only $1/5$ classified it as a disk at the 2-epoch depth, then its value on this plot would be 0.4. This difference is shown on the top plot for disks, spheroids, and irregulars. All objects with a value of zero have complete agreement between the two different depths for completely independent sets of classifiers). This plot illustrates that the level of agreement is highest for disks and the lowest for irregulars. The asymmetry in the distributions for spheroids and irregulars (both have more objects with a negative difference) indicates that a higher fraction of people choose those classifications in the shallower images. The differences for
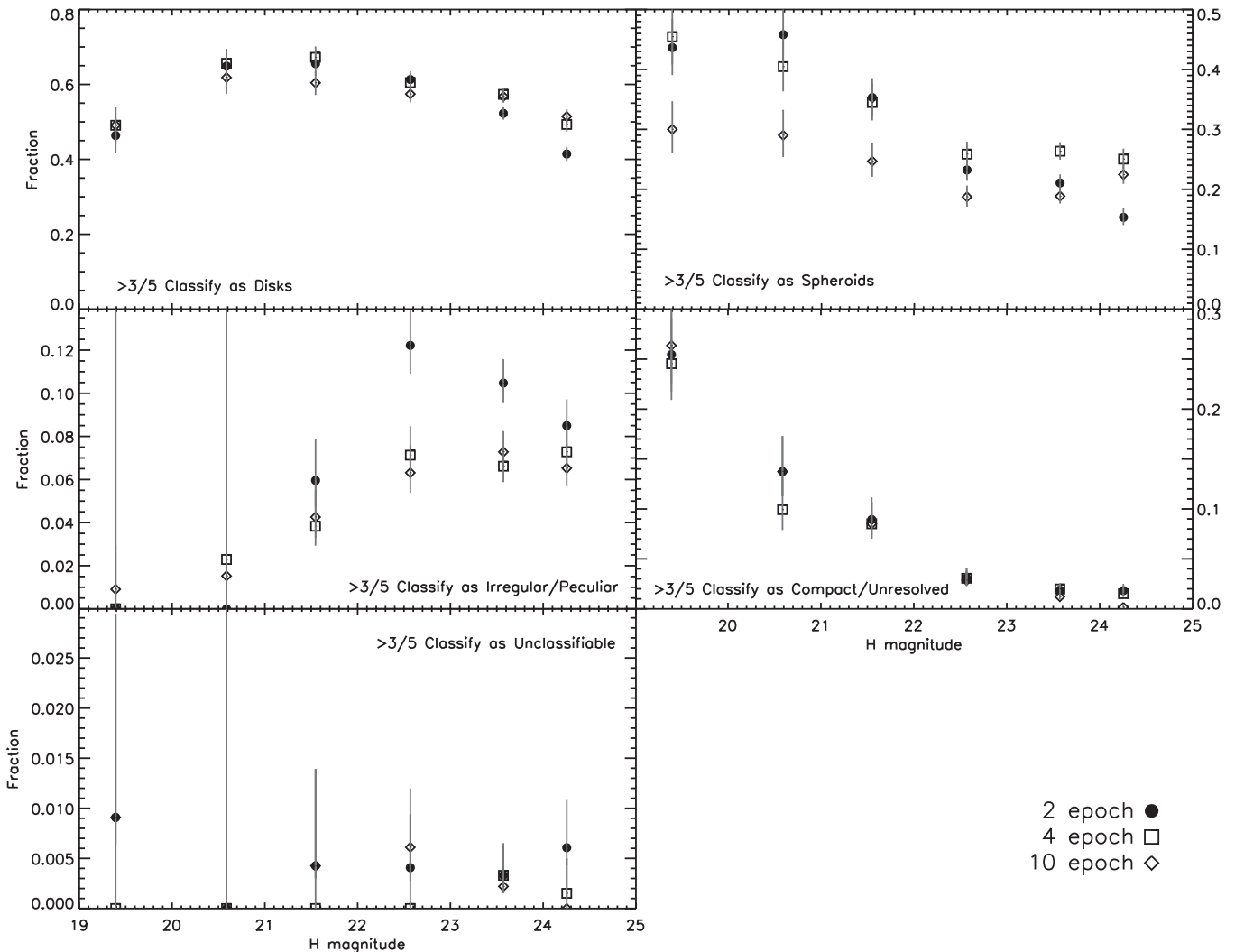
Figure 9. Fraction of objects for which >3/5 classifiers agree on the classification as a function of *H*-magnitude and separated by the depth of the images classified. Each panel represents a separate main morphology class. The error bars on each point reflect the 1σ binomial confidence limits given the number of objects in each category, following the method of Cameron (2011).

the interaction classes are shown in the bottom plot. Overall, the level of agreement is worse than for the main morphology plot (the distributions are broader). There is the highest agreement for the "any interaction" set, which includes mergers and both interactions within and beyond the segmentation map, as expected. There is a similar level of agreement for the non-interacting companion class.

4.3. Disk/Spheroid Separation

Ideally, one would like to apply various quantitative measures for galaxy morphology whenever possible. Several methods have been developed for automatically classifying galaxies using image statistics in the local and low redshift universe, such as CAS (Concentration, Asymmetry, and Clumpiness; Abraham et al. 1996; Conselice et al. 2000; Conselice 2003) and Gini/M_{20} (Abraham et al. 2003; Lotz et al. 2004). While some of these methods seem to work well for separating disks from spheroids at low redshifts, they typically do not work as well for identifying mergers and interactions (e.g., Kartaltepe et al. 2010) or are only sensitive to mergers at certain stages (Lotz et al. 2008) and they have yet to

be thoroughly tested and calibrated at high redshifts with large samples of visually classified galaxies. Our large sample of visually classified galaxies will enable detailed studies of these methods at high redshifts for the first time (e.g., J. Lotz et al. 2015, in preparation). The identification of mergers will be the subject of a future paper. Here, we investigate how well our scheme separates disks from spheroids.

First, we compare our visual classifications to the Sérsic index, typically used as a way to separate disks from bulges, using GALFIT (Peng et al. 2002) measurements for the CANDELS data in GOODS-S (van der Wel et al. 2012). GALFIT fits the two-dimensional galaxy light profile in an image using a χ^2 minimization routine to estimate the best-fit Sérsic profile of the galaxy. Reliable fits were obtained for 6225 of our visually classified galaxies in GOODS-S, after excluding objects in noisy areas of the mosaic and objects with unrealistic parameters. The distribution of Sérsic indices (n, where $n = 0.5$ corresponds to a Gaussian profile, $n = 1$ to an exponential profile, and $n = 4$ to a de Vaucouleurs profile) are shown in Figure 12, separated by their main morphology class in the top panel, and their interaction class in the bottom panel. In the top panel, galaxies are separated by their relative
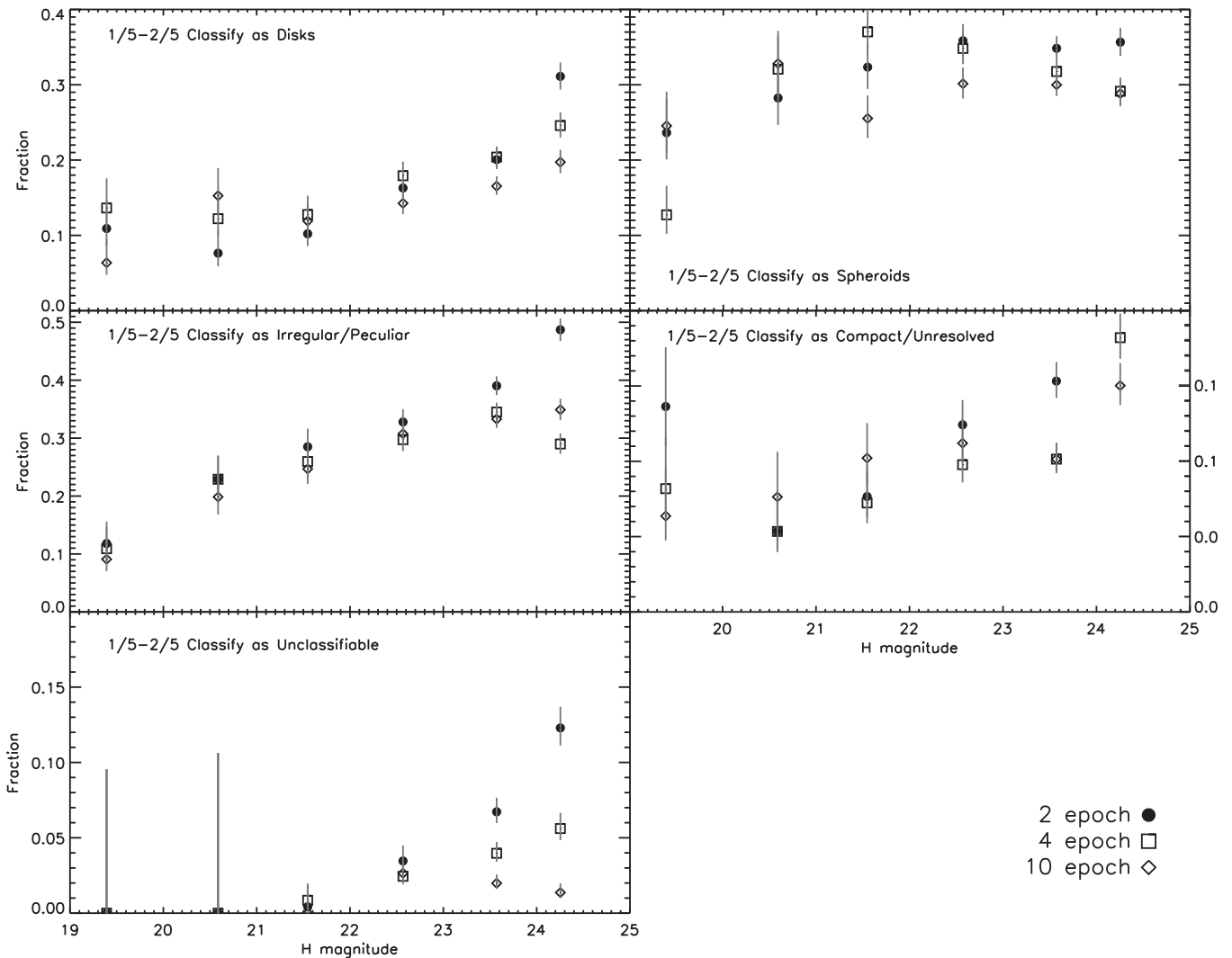
Figure 10. Fraction of objects for which 1/5–2/5 classifiers agree on the classification as a function of *H*-magnitude and separated by the depth of the images classified. Each panel represents a separate main morphology class. The error bars on each point reflect the 1σ binomial confidence limits given the number of objects in each category, following the method of Cameron (2011).

diskiness/bulginess, into "Mostly Disk" ($>2/3$ classifiers call it a disk and $<2/3$ classifiers call it a spheroid), "Mostly Spheroid" ($>2/3$ classifiers call it a spheroid and $<2/3$ classifiers call it a disk), "Disk + Spheroid" ($>2/3$ classifiers call it a disk and $>2/3$ classifiers call it a spheroid), and "Irregular" ($>2/3$ classifiers call it irregular, $<2/3$ classifiers call it a disk, and $<2/3$ classifiers call it a spheroid). This plot shows a significant difference in the distribution of Sérsic indices for each of these classes. The "Mostly Disk" group is narrowly peaked around a value of 1 ($\langle n \rangle = 1.01$). The "Mostly Spheroid" group has a broad distribution with $\langle n \rangle = 2.98$, and the "Disk + Spheroid" group is in between ($\langle n \rangle = 2.53$), as expected. The "Irregular" group, which excludes galaxies that are clearly disks or spheroids more closely resembles the disk distribution with $\langle n \rangle = 1.33$.

For the interaction classes in the bottom panel, we plot the distributions of Sérsic indices for mergers, interactions within, and interactions beyond the segmentation map (in all cases, we use a threshold of $\geqslant 2/3$ classifiers). There is a difference

between the distributions for these interaction classes, though slight. The mergers have the lowest mean Sérsic index with $\langle n \rangle = 1.15$. For the interactions beyond the segmentation map, the distribution is similar to the irregulars and disks in the top panel (with $\langle n \rangle = 1.45$), which could be explained by the companion being distant enough that it does not affect the light distribution of the main galaxy by much. Finally, the interaction within the segmentation map shows the broadest distribution as well as the largest mean value ($\langle n \rangle = 2.06$). Studies of the Sérsic indices of merging galaxies in the local universe have found a similar low mean Sérsic index, similar to that of disks (Kim et al. 2013). This illustrates that the low Sérsic index of mergers in our sample is not surprising and that the Sérsic index itself is not a good measure of whether or not a galaxy is a merger.

To further explore how well our classifications separate disks and spheroids, we plot the positions of the sample on a color–color diagram (*UVJ* diagram: $U - V$ versus $V - J$) in Figure 13. In this diagram, star forming galaxies follow a diagonal sequence with redder $V - J$ colors due to dust
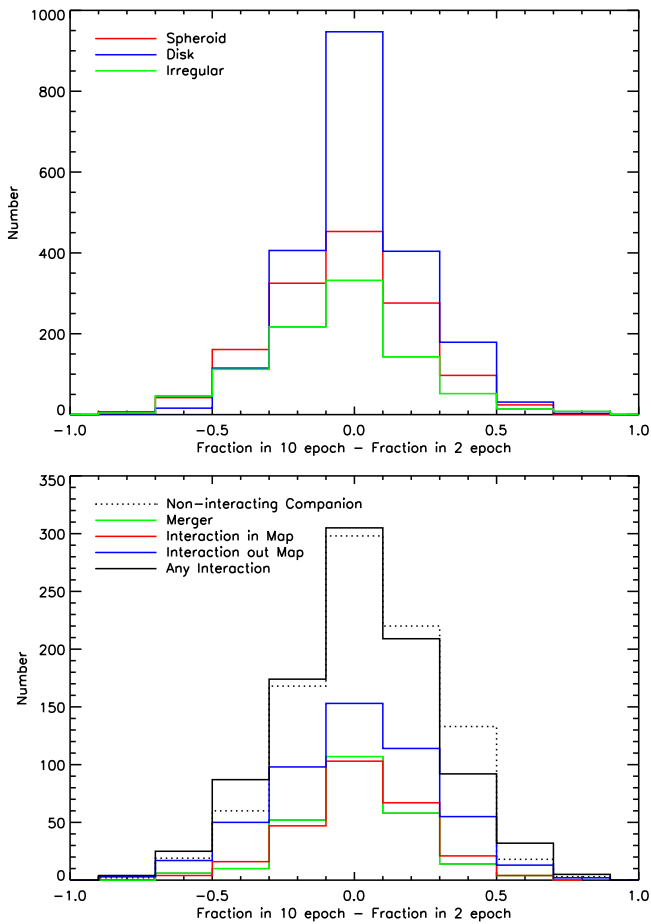
13

Figure 11. Distribution of the differences between fraction of classifiers choosing a given class for a particular object at two different CANDELS depths. Top: difference for the Disk, Spheroid, and Irregular main morphology classes. Bottom: difference for each interaction class.

Figure 12. Sérsic index for visually classified galaxies in GOODS-S color coded by their main morphology class (top) and the interaction class (bottom). The main morphology classes are further refined by their relative diskiness/ bulginess, into "Mostly Disk" ($>2/3$ classifiers call it a disk and $<2/3$ classifiers call it a spheroid), "Mostly Spheroid" ($>2/3$ classifiers call it a spheroid and $<2/3$ classifiers call it a disk), "Disk + Spheroid" ($>2/3$ classifiers call it a disk and $>2/3$ classifiers call it a spheroid) and "Irregular" ($>2/3$ classifiers call it irregular, $<2/3$ classifiers call it a disk, and $<2/3$ classifiers call it a spheroid).

reddening while quiescent galaxies lie in a clump above the sequence with redder $U - V$ colors but bluer $V - J$ colors (e.g., Wuyts et al. 2007, 2009; Williams et al. 2009). We split the sample into six different redshift bins and color-code the points by their main morphology class. Galaxies that are "Mostly Disks" separate cleanly from those that are "Mostly Spheroids," such that the disks fall onto the star forming sequence while the spheroids fall into the quiescent region (with some scatter). Galaxies classified as disk+spheroid mostly fall onto the star-forming sequence with some scatter up into the quiescent region while irregular galaxies mostly fall onto the star-forming sequence and are concentrated at blue $V - J$ colors.

4.4. Morphological K-corrections

Since the classifiers look at all of the bands at once, instead of individually, we cannot directly compare classifications between the different bands. However, we included a k-correction flag in the classification scheme so that the classifiers could mark cases that they think they would have classified differently in one or more of the other bands. This allows us to track and search for galaxies with large morphological k-
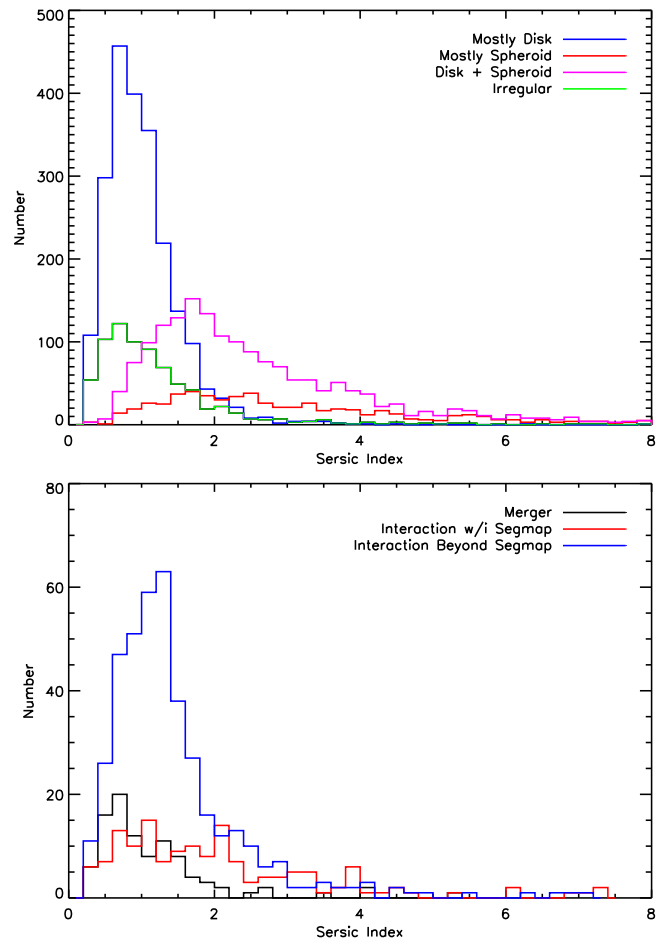
corrections. Figure 14 contains a sample of objects that more than half of the classifiers flagged as being different in the V-band. This represents a total of 84 objects out of the full sample with $\langle z \rangle = 0.9$. Both the V and H-band stamps are shown side by side for comparison. Many of the objects that people mark as different are simply faint or undetected in the optical bands. However, most of the galaxies in Figure 14 have clumpy irregular light distributions in the optical and are more regular and smooth and often have more prominent bulges in the near-infrared. For this reason, the classifiers are asked to look at the optical images when they select among the clumpiness/ patchiness flags and in the near-infrared when choosing between bulge and disk dominated flags. These galaxies illustrate some of the pitfalls of basing classifications solely on the rest-frame UV light of high redshift galaxies—some that seem highly irregular are actually normal looking disks in the rest-frame optical. Similarly, some galaxies that appear to be a
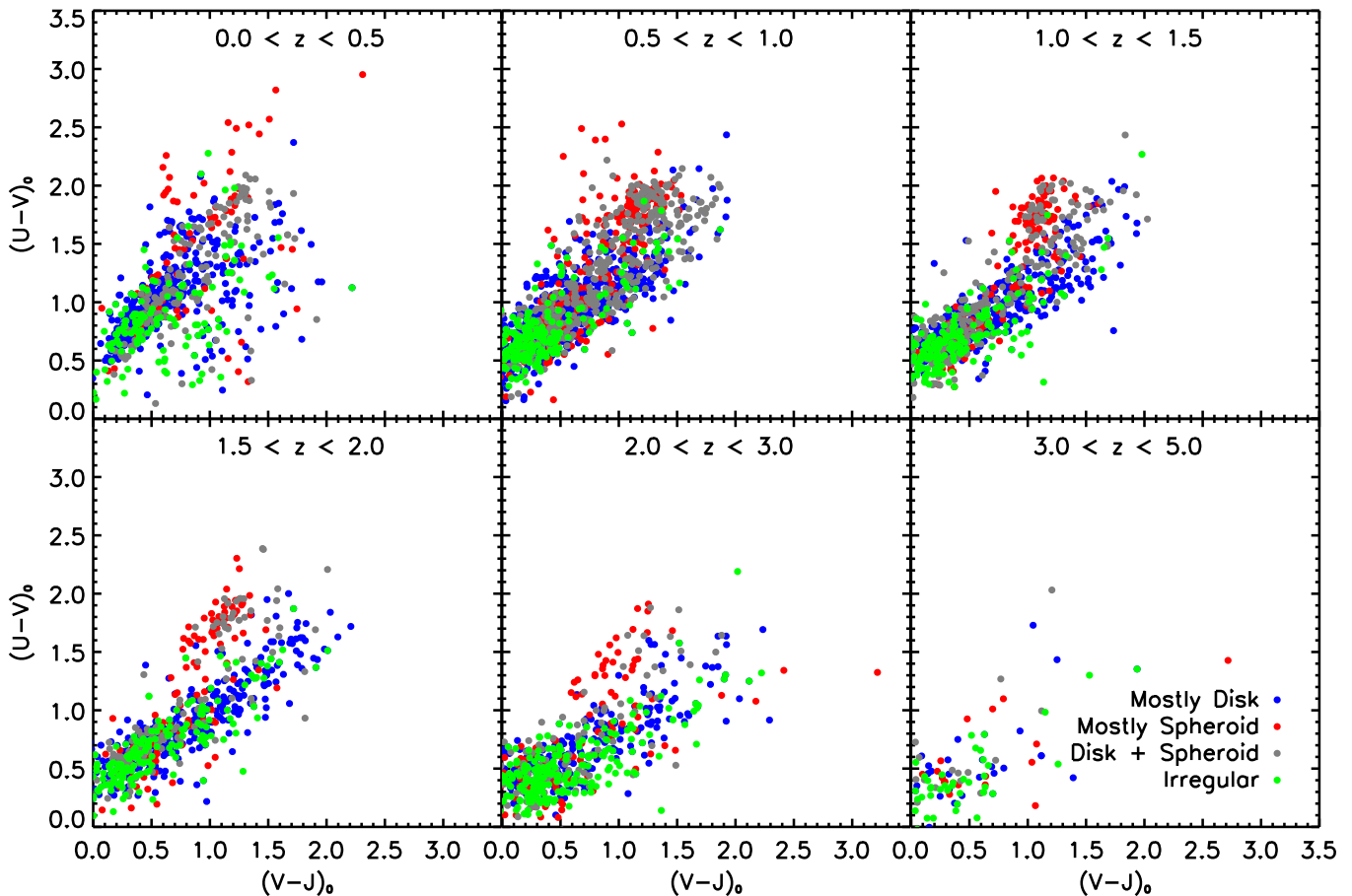
Figure 13. *UVJ* diagram for our sample divided into six redshift bins and color coded by their dominant main morphology class (mostly disk, mostly spheroid, disk +spheroid, and irregular). Note that the disks and spheroids separate from each other as expected, the irregular galaxies do not occupy the quiescent region of the diagram, and the irregular galaxies are concentrated at blue $(V - J)_0$ colors.

single object in the rest-frame UV show up as pairs in the rest-frame optical.

5. SUMMARY

We have presented an ambitious program to visually classify all galaxies in the five CANDELS fields down to $H < 24.5$. Once completed, we expect to have detailed morphological classifications for over 50,000 galaxies spanning $0 < z < 4$. Our visual classification scheme was designed to cover a wide range of science goals and we have described each of its components in detail. With the publication of this paper, we release to the public all of the visual classifications in GOODS-S, including separate catalogs for the full field and deep region, with the deep region classified by five different people at each of three different depths. The remainder of the field (the "wide" portion and the ERS) have been classified by at least 3 people. Included in this data release is the calibration set of 200 galaxies that have been classified by all 65 classifiers. We also release the software for the Perl/Tk GUI that we developed to classify galaxies.

We have made a number of comparisons among the classifiers to test for internal consistency and find that the level of agreement is dependent on both the brightness of the galaxy (H-band magnitude) and the type of galaxy itself. The

classifications are also consistent across the different depths for all but the faintest galaxies. Our detailed calibration set of galaxies illustrates the complexity of galaxy morphology for these faint objects and indicates the necessity of having multiple classifiers for each object. A comparison of our visual classifications with Sérsic indices measured by GALFIT shows broad agreement, where galaxies classified as disks tend to have lower Sérsic indices and galaxies classified as spheroids tend to have higher Sérsic indices. We also find that the colors of our morphologically selected galaxies are consistent with what we would expect for their galaxy types, such that disks are on the star forming sequence and spheroids are mostly quiescent. Irregular galaxies and galaxy mergers and interactions are the hardest to identify in our sample and tend to show the highest level of disagreement. A future paper (J. S. Kartaltepe et al. 2015, in preparation) will investigate galaxy mergers and interactions in these classifications in more detail.
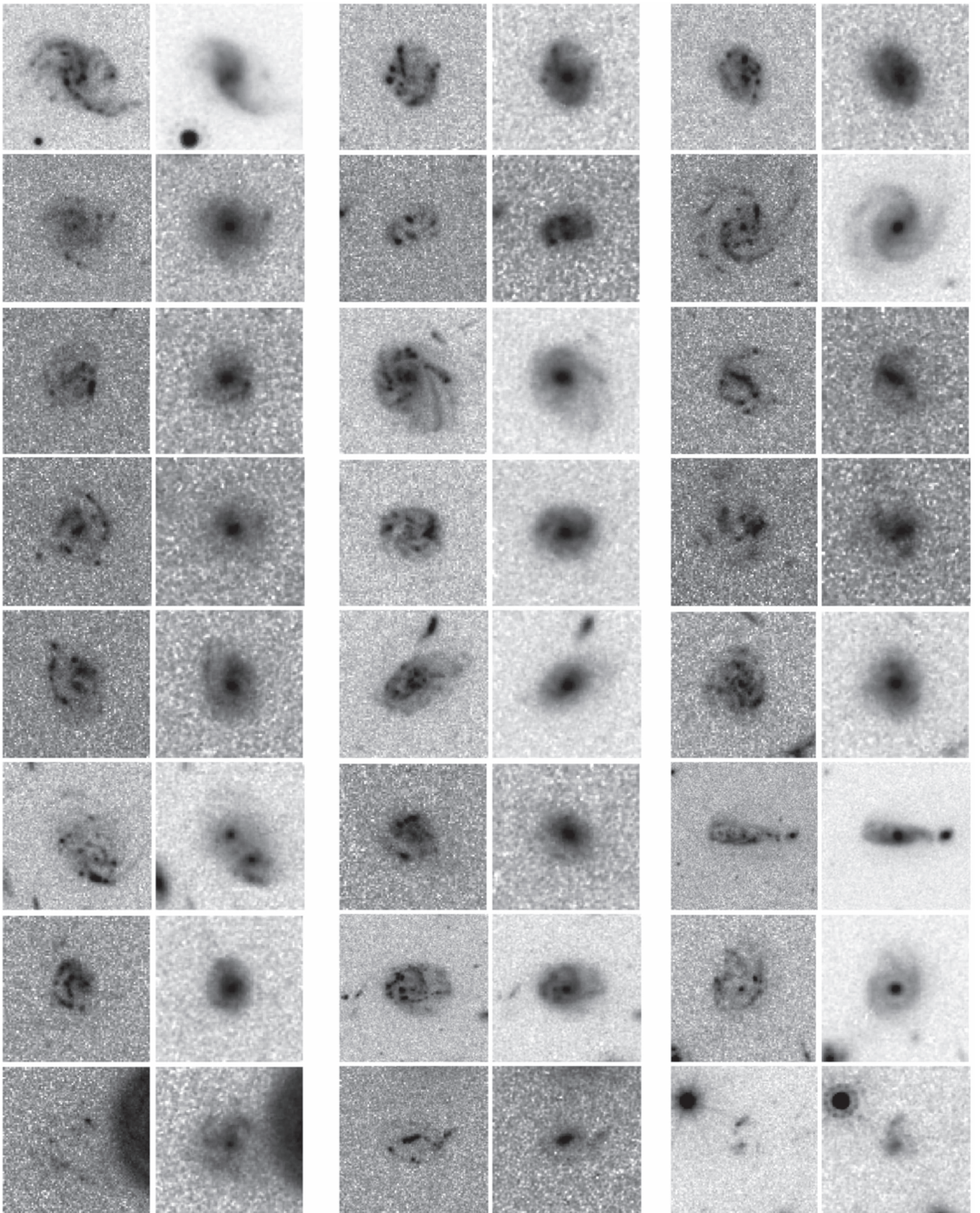
Figure 14. Postage stamps of galaxies with a large morphological *k*-correction between the *V* band (left, rest-frame UV) and *H* band (right, rest-frame optical). The sizes of the stamps follow the prescription described in Section 2.3.

Facility: HST (WFC3).

REFERENCES

Abazajian, K., Adelman-McCarthy, J. K., Agueros, M. A., et al. 2003, AJ, 126, 2081
Abraham, R. G., van den Bergh, S., Glazebrook, K., et al. 1996, ApJS, 107, 1
Abraham, R. G., van den Bergh, S., & Nair, P. 2003, ApJ, 588, 218
Atkinson, A. M., Abraham, R. G., & Ferguson, A. M. N. 2013, ApJ, 765, 28
Bell, E. F., Papovich, C., Wolf, C., et al. 2005, ApJ, 625, 23
Bell, E. F., van der Wel, A., Papovich, C., et al. 2012, ApJ, 753, 167
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Bridge, C. R., Carlberg, R. G., & Sullivan, M. 2010, ApJ, 709, 1067
Brinchmann, J., Abraham, R., Schade, D., et al. 1998, ApJ, 499, 112
Brinchmann, J., & Ellis, R. S. 2000, ApJL, 536, L77
Bruce, V. A., Dunlop, J. S., Cirasuolo, M., et al. 2012, MNRAS, 427, 1666
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Bundy, K., Ellis, R. S., & Conselice, C. J. 2005, ApJ, 625, 621
Caldwell, J. A. R., McIntosh, D. H., Rix, H., et al. 2008, ApJS, 174, 136
Cameron, E. 2011, PASA, 28, 128
Cameron, E., Carollo, C. M., Oesch, P. A., et al. 2011, ApJ, 743, 146
Cassata, P., Giavalisco, M., Guo, Y., et al. 2010, ApJL, 714, L79
Chabrier, G. 2003, PASP, 115, 763
Conselice, C. J. 2003, ApJS, 147, 1
Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, ApJ, 529, 886
Conselice, C. J., Bluck, A. F. L., Buitrago, F., et al. 2011b, MNRAS, 413, 80
Conselice, C. J., Bluck, A. F. L., Ravindranath, S., et al. 2011a, MNRAS, 417, 2770
Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, ApJ, 775, 93
Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010, MNRAS, 401, 1043
Davis, M., Guhathakurta, M., Konidaris, N. P., et al. 2007, ApJL, 660, L1
Giavalisco, M., Ferguson, H. C., Koekemoer, A., et al. 2004, ApJL, 600, L93
Glikman, E., Simmons, B., Mailly, M., et al. 2015, ApJ, 806, 218
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, ApJ, 800, 39
Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24

Häussler, B., McIntosh, D., Barden, M., et al. 2007, ApJS, 172, 615
Jogee, S., Miller, S., Penner, K., et al. 2009, ApJ, 697, 1971
Kartaltepe, J. S., Dickinson, M., Alexander, D. M., et al. 2012, ApJ, 757, 23
Kartaltepe, J. S., Sanders, D. B., Le Floc'h, E., et al. 2010, ApJ, 721, 98
Kim, D.-C., Evans, A. S., Vavilkin, T., et al. 2013, ApJ, 768, 102
Kocevski, D. D. 2015, MorphGUI, v1.0, Zenodo
Kocevski, D. D., Faber, S. M., Mozena, M., et al. 2012, ApJ, 744, 148
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kriek, M., van Dokkum, P. G., Franx, M., Illingworth, G. D., & Magee, D. K. 2009, ApJL, 705, L71
Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, PASP, 119, 1325
Law, D. R., Steidel, C. C., Shapley, A. E., et al. 2012, ApJ, 745, 85
Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599
Lee, B., Giavalisco, M., Williams, C., et al. 2013, ApJ, 774, 47
Lintott, C., Schawinski, K., Bamford, S., et al. 2011, MNRAS, 410, 166
Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179
Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2008, MNRAS, 391, 1137
Lotz, J. M., Primack, J., & Madau, P. 2004, AJ, 128, 163
McGrath, E. J., Stockton, A., Canalizo, G., Iye, M., & Maihara, T. 2008, ApJ, 682, 303
Morishita, T., Ichikawa, T., & Kajisawa, M. 2014, ApJ, 785, 18
Mortlock, A., Conselice, C. J., Hartley, W. G., et al. 2013, MNRAS, 433, 1185
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJS, 206, 8
Nair, P. B., & Abraham, R. G. 2010, ApJS, 186, 427
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
Roberts, M. S., & Haynes, M. P. 1994, ARA&A, 32, 115
Rosario, D. J., McIntosh, D. H., van der Wel, A., et al. 2015, A&A, 573, 85
Santini, P., Ferguson, H. C., Fontana, A., et al. 2015, ApJ, 801, 97
Schawinski, K., Thomas, D., Sarzi, M., et al. 2007, MNRAS, 382, 1415
Scoville, N., Abraham, R. G., Aussel, H., et al. 2007, ApJS, 172, 38
Simmons, B. D., Melvin, T., Lintott, C., et al. 2014, MNRAS, 445, 3466
Szomoru, D., Franx, M., Bouwens, R. J., et al. 2011, ApJL, 735, L22
Targett, T. A., Dunlop, J. S., Cirasuolo, M., et al. 2013, MNRAS, 432, 2012
Trump, J. R., Barro, G., Juneau, S., et al. 2014, ApJ, 793, 101
van der Wel, A., Bell, E. F., Haussler, B., et al. 2012, ApJS, 203, 24
Villforth, C., Hamann, F., Rosario, D. J., et al. 2014, MNRAS, 439, 3342
Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, MNRAS, 435, 2835
Willett, K. W., Schawinski, K., Simmons, B. D., et al. 2015, MNRAS, 449, 820
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879
Windhorst, R. A., Cohen, S. H., Hathi, N. P., et al. 2011, ApJS, 193, 27
Wuyts, S., Labbe, I., Franx, M., et al. 2007, ApJ, 655, 51
Wuyts, S., Franx, M., Cox, T. J., et al. 2009, ApJ, 700, 799
Wuyts, S., Förster Schreiber, N. M., Genzel, R., et al. 2012, ApJ, 753, 114